Disease-Causing Mutations in The Human Metal Ion Transporter ZIP14 Alter Its Structure, Ion Uptake Function and Trafficking

For the award of PhD

...

University College London

Charmian Nicole Dawson

December 2022

Declaration

I, Charmian Nicole Dawson, confirm that the work presented in my thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Abstract

ZIP14 is a member of the LIV-1 family, a sub-family of the Zrt-, Irt-like Protein (ZIP) family. These proteins import zinc into the cytosol, regulating zinc concentrations to control physiological processes. As a zinc transporter, ZIP14 regulates inflammation, the sensitivity of multiple signaling pathways, and in cancer, over-expression of ZIP14 in skeletal muscle causes cachexia. During iron overload, ZIP14 directs excess iron into the pancreas and liver. In the intestines, ZIP14 prevents excessive manganese absorption from the diet. Pathological mutations in human ZIP14 cause manganism or hyperostosis cranialis, due to effects on manganese and zinc transport, respectively. This is the first detailed study of ZIP14 structure and function, and the effects of the disease-causing mutations. Homology modelling was used to predict the structure of ZIP14, and further mutations were designed based on this structure. Epitope-tagged ZIP14 and ZIP14 mutants were expressed in HeLa cells and studied with a range of techniques, including western blotting to measure total cellular levels of the protein, immunofluorescence microscopy to assess intracellular localisation, flow cytometry to quantify total and cell surface protein levels, and Fe^{55} uptake to assess ion transport. Mutations inhibiting dimerisation of the extracellular domain and subsequent formation of an inter-molecular disulphide bridge decreased localisation at the cell surface in a concentration-dependent manner, as well as preventing iron transport. Effects of mutations in

the predicted metal transport pore differed: P379L primarily altered trafficking, G383R prevented iron uptake. A combination of the coevolution analysis of the transmembrane region, and the effects of the L441R and N469K mutations, showed that dimerisation of the transmembrane domain was also important for cell surface localisation, but not for metal transport. In summary, this research provides a convincing structure for the ZIP14 protein, and sheds light on how aspects of this structure relate to ZIP14 trafficking and function.

Impact statement

ZIP14 is a transporter protein capable of importing zinc, manganese, iron and cadmium into cells, and is widely expressed throughout the body. The transport of zinc by ZIP14 helps to control the response of many signalling pathways to their hormones and coordinates immune responses so that they are effective against pathogens without harming the host. In cancer patients, ZIP14 can be overactive and cause severe weight loss, increasing the risk of mortality. ZIP14 also regulates absorption of manganese from the diet. Most mutations in ZIP14 lead to parkinsonism symptoms due to the deposition of excess manganese in the brain, with one causing a disorder in bone overgrowth due to effects on intracellular concentrations of zinc.

In this study, the structure of ZIP14 was predicted using computational methods, and the effects of various ZIP14 mutations were tested for their effects on the location of ZIP14 in the cell, and ZIP14 transport function. The experimental results and computational analysis led to a greater understanding of ZIP14 structural features, which could contribute to research into the effects of, and treatments for pathological ZIP14 mutations.

These findings have the potential to impact the research of others; those working on ZIP14 or homologous transporters, or transporters that use similar mechanisms. In the long-term, these findings could contribute to research in immunology, management of cachexia, or numerous other conditions involving

hormones whose action is affected by ZIP14 function.

The skills developed during this doctoral training, including knowledge of experimental techniques, protocol optimisation, data analysis, and bioinformatics, has also informed my pedagogical development. This development will strongly influence my future design and delivery of educational activities for Life Sciences students, including more opportunities for experiences that will equip students with the transferable skills needed to succeed in research and other fields.

To disseminate research findings, I have presented at the 2022 Young Researcher Symposium (London Metallomics Consortium), and manuscripts are being prepared for publication in peer-reviewed journals. In addition, pedagogical projects influenced by the skills learnt during my doctoral degree are being delivered to, or developed for, UCL Life Sciences undergraduates, and I also have plans to engage in outreach activities for young people interested in pursuing careers in STEM.

Acknowledgements

To my family and friends: I will never be able to accurately describe the patience and support you have given me over all these years. This has meant all the more because I know that my absence was missed, and yet you have been my cheerleaders anyway. Thank you for being understanding, and for patiently listening to me talk in great detail about subjects that are niche, at best.

To my supervisors, Dr Katherine Bowers and Prof. Kaila Srai: Thank you so much for all your support, encouragement and advice. I especially appreciate the space I was given to develop my own ideas and make my own mistakes along the way.

To my lab-mates, lab managers, other PhD students, post-docs and other researchers: Thank you all for being so generous with your time, hard-won expertise, and friendships. There is just no substitute for shared misery...

To my office-mates and colleagues: Thank you all for being so understanding over all these years, and for being amazing people that I love working with. I have been really touched by all the encouragement and offers of advice. It has been so lovely to feel that everyone is rooting for me to succeed.

To everyone who discouraged me from doing a PhD: I can now see that you were merely looking out for my mental, emotional, physical and financial well-being; thank you for resisting the urge to say I told you so!

And to all these people who have supported me over these nine long years, I say: What have you done for me lately?

Dedication

My grandma never got to see this day, but she helped me get here.

Contents

Abbreviations

AEC	Automated Efficiency Control
BCA	Bicinchoninic acid
bp	basepair
BSA	Bovine serum albumin
CASP	Critical Assessment of Structure Prediction
CLP	Cecal ligation puncture
cpm	Counts per minute
DAPI	4’,6-diamidino-2-phenylindole
DEPC	Diethyl pyrocarbonate
DMEM	Dulbecco’s Modified Eagle Medium
DMP	DeepMetaPSICOV
DMT1	Divalent metal transporter 1
DNA	Deoxyribonucleic acid

dpm	Disintegrations per minute
ECD	Extracellular domain
EDTA	Ethylenediaminetetraacetic acid
FBS	Fetal bovine serum
FBSapE	FACS permeabilizing blocking buffer
GAPDH	Glyceraldehyde-3-phosphate dehydrogenase
GFP	Green Fluorescent Protein
HA	Haemagglutinin
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HFE	High Fe (iron)
HH	Hereditary haemochromatosis
HMM	Hidden Markov Model
HRD	Helix-rich domain (of ZIP4)
IKK	Inhibitory-k kinase
IL-1β	Interleukin-1β
IL-6	Interleukin-6
KeV	kilo-electronvolts
LPS	Lipopolysaccharide

MFI	Median fluorescence intensity
MSA	Multiple sequence alignment
NTBI	Non-transferrin bound iron
PBS	Phosphate buffered saline
PBSE	2.5 µM EDTA in PBS
PCR	Polymerase Chain Reaction
PDE	Phosphodiesterase
PFAE	2.5 µM EDTA, 0.5% PFA in PBS
PFA	Paraformaldehyde
PtdIns	Phosphatidyl inositol
PTP	Phosphotyrosine phosphatase
PTP1B	Phosphotyrosine phosphatase 1 B
rcf	relative centrifugal force
RMSD	Root Mean Square Deviation
RNA	Ribonucleic acid
rpm	Rotations per minute
RPMI	Roswell Park Memorial Institute
RT-qPCR	Real-time quantitative PCR

SNX	Sorting nexin
TBI	Transferrin bound iron
TE	Tris-EDTA
TEMED	Tetramethylethylenediamine
TfR1	Transferrin receptor 1
TM	Transmembrane helix
TMD	Transmembrane domain
tSIE	Transformed Spectral Index of the External standard
WBC	White blood cell
WT	Wild-type
ZIP	Zrt-, Irt-related protein

List of Figures

List of Tables

Chapter 1

Introduction

Zinc, iron and manganese are essential micronutrients, acting as cofactors and components of numerous proteins and enzymes (see Salgueiro et al. [1], Gurzau et al. [2] and Chen et al. [3] for reviews). Deficiencies in each of these metals lead to distinctive disorders, including anaemia, growth retardation and immune dysfunction (see Ackland and Michalczyk [4], Dev and Babitt [5], and Finley and Davis [6] for reviews). These metals are not only harmful when deficient, as excessive amounts lead to toxic and potentially fatal effects (see Plum et al. [7], Siah et al. [8] and Hauser et al. [9] for reviews). Transporters of these metals, such as ZIP14, are carefully regulated in order to differentially control the local concentrations of these metals according to the organ, tissue, subcellular location and metabolic status of the organism.

1.1 ZIP14 Homology and Family

ZIP14, also known as SLC39A14, is a member of the Zrt-, Irt-like Proteins (ZIP) family, named after its founding members Zrt1 and Irt1, a zinc transporter from

Saccharomyces cerevisiae and an iron transporter from *Arabidopsis thaliana*, respectively [10, 11]. This family includes transporters found in a wide range of organisms, including archeae and mammals [10]. Based on homology with other members, ZIP14 is placed within the LIV-1 subfamily, named after ZIP6 (figure 1.1) [11, 12]. The ZIP family are generally zinc transporters, transporting zinc into the cytoplasm from outside the cell, or from organelles [11].

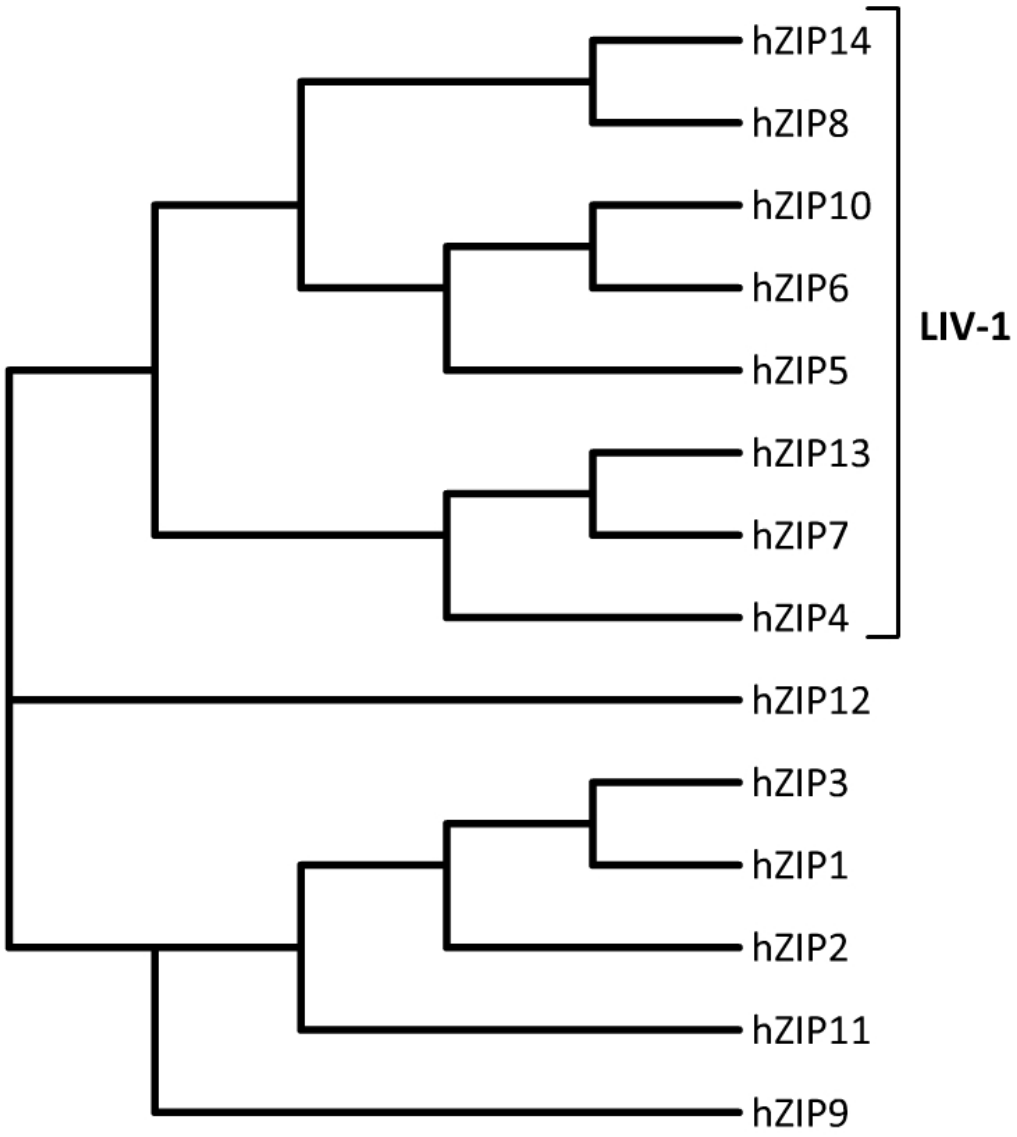


Figure 1.1: **Phylogenetic tree of human ZIP proteins.** LIV-1 family members are indicated. Adapted from Taylor et al. [12].

Based on the homology of ZIP14 with the LIV-1 subfamily and the larger ZIP family, ZIP14 has been predicted to have eight transmembrane helices (TMs), with both N- and C- extracellular termini, and a large, variable, intracellular loop between TMs 3 and 4 (figure 1.2) [11]. The LIV-1 family members have conserved residues on both TMs 4 and 5, HNF and HEXPHEXGD respectively, believed to be required for the transport of zinc, and for zinc selectivity [11]. ZIP14, and the closely related ZIP8, both differ somewhat from other members of the LIV-1 family, in that the first residue of the HEXPHEXGD motif is a glutamate, possibly altering the solute profile [11, 13].

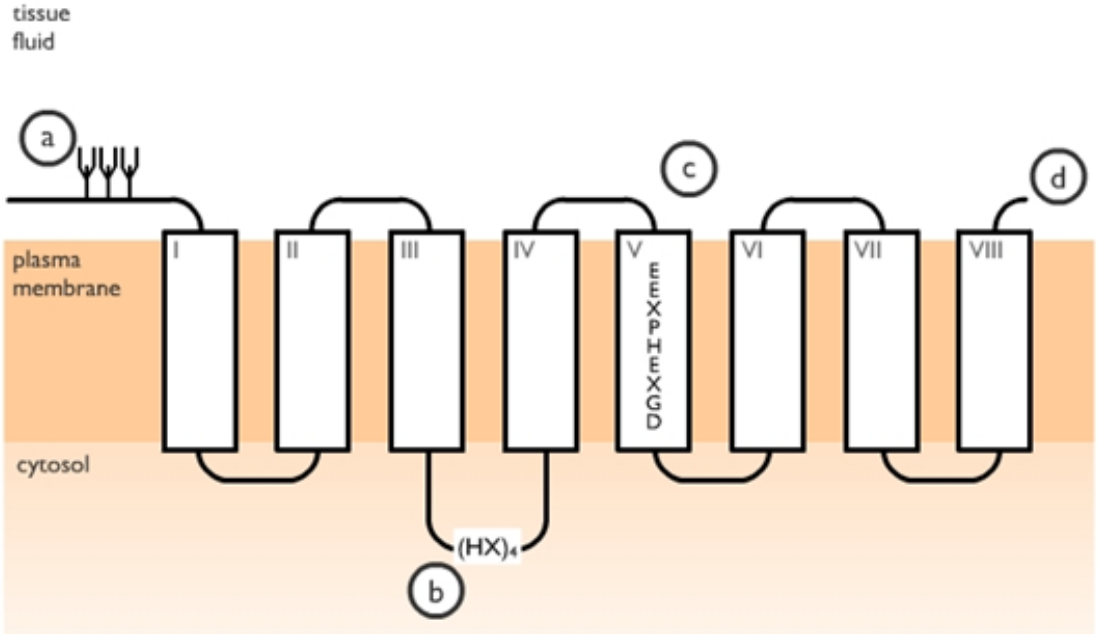


Figure 1.2: **Predicted topology of ZIP14.** a) The long extra-cellular N-terminus, with three N-glycosylation sites. b) The long intracellular loop with histidine-rich region. c) The modified HEXPHEXGD sequence, part of the predicted metal translocation pore. d) The short extracellular C-terminus. Adapted from Taylor et al. [13].

ZIP14 contains three N-glycosylation sites, at residues 77, 87, and 102 (numbering according to the human sequence). These sites are all in the predicted extracellular N-terminus [13, 14]. Immunofluorescent staining and co-localisation with protein markers locates ZIP14 in early endosomes, lysosomes,

and at the plasma membrane [13, 15]. In rat liver slices, ZIP14 is located primarily at the sinusoidal membranes, which are in contact with the blood supply [16].

Human ZIP14 has two alternative exons, 4A and 4B, though the resultant proteins are nearly identical, differing by only 22 amino acids out of ZIP14's 492, and there is no yet known physiological significance of the two isoforms [14], although they may be of use as cancer markers. Alternative splicing of ZIP14 is regulated by Wnt signalling, and the ratio of the two isoforms is altered in tumours where Wnt signalling is dysregulated [17]. In papers referencing the two isoforms, the naming is inconsistently used. For future reference, the isoform incorporating exon 4B (Q15043-1) will be referred to as ZIP14_A, and the isoform incorporating exon 4A (Q15043-3) as ZIP14_B. An alternative exon 9 has also been reported, giving rise to isoform C (Q15043-2), although there is no information on its function or its expression patterns.

1.2 Expression of ZIP14

The various ZIP family members have different patterns of tissue expression, subcellular location, and regulate different aspects of zinc transport [18]. The pattern of ZIP14 tissue expression has been examined in mice and humans. The results vary, but ZIP14 is expressed in many tissues, with increased levels of ZIP14 mRNA found in liver, small intestine, pancreas, and heart [13, 14]. Some differences in tissue expression were observed between isoforms A and B. In mice, ZIP14_A expression was fairly similar in all the tissues tested, while ZIP14_B showed higher levels of expression in the liver and duodenum [14]. In humans, ZIP14_A was again expressed in all the tissues tested. ZIP14_B expression was much lower than ZIP14_A expression in all tissues tested, with

the exception of the liver and the pancreas [19].

Regulation of ZIP14 expression is extremely responsive to inflammatory cytokines [20–25], and has also been associated with other phenomena, such as adipogenesis. In mouse 3T3-L1 cells (an adipocyte cell line), ZIP14 expression spiked during early differentiation, rapidly decreasing again within 24 hours [26].

1.2.1 Knockout Phenotype

ZIP14 knockout mice have altered zinc [21], manganese [27], cadmium [28] and iron [16] metabolism in the liver. They display abnormalities in growth, motor function, glucose homeostasis, and immune function. These abnormalities are typically related to manganese dyshomeostasis, and faulty regulation of signalling pathways by zinc.

Zip14 knockout mice have impaired growth, difficulty standing on their hind legs and moderately increased body fat [21, 29, 30]. There are higher concentrations of manganese in the blood and areas of the brain, with a low concentration in the liver [27]. Pancreatic islets are enlarged and plasma insulin is increased, which may explain the lower plasma glucose, decreased expression of gluconeogenic enzymes and increased expression of lipogenic enzymes observed in these mice [21, 31]. Knockout mice also have increased intestinal permeability leading to mild chronic endotoxemia [32] and insulin insensitivity of adipose tissue [33]. Regulation of zinc transport during inflammation is strongly affected, resulting in aggressive, but ultimately less effective immune responses [21].

1.3 Metal Transport by ZIP14

ZIP14 transports zinc, manganese, iron and cadmium [13–15, 19, 26, 34–36], and is believed to be a carrier, more specifically a bicarbonate symporter. During ZIP14 transport there is no net movement of charge, indicating that another ion is also transported [35]. Zn^{2+}, Mn^{2+}, Fe^{2+} and Cd^{2+} uptake by ZIP14 is stimulated by the presence of bicarbonate in the medium [14, 35], and Cd^{2+} uptake is inhibited by DIDS (4,4'-diisothiocyanostilbene-2,2'-disulfonic acid) [14], known to inhibit bicarbonate transporters [37].

Attempts to determine the V_{max} and K_m of mouse ZIP14_A and B for ZIP14 solutes have produced inconsistent results (table 1.1) [14, 36]. Despite the dependence of V_{max} values on transporter concentration, metal uptake was reported relative to total cellular protein in stably-transfected mouse fetal fibroblasts, or per oocyte for *Xenopus laevis* oocytes, as the protein content of oocytes "varied widely" [14, 35, 36]. More relevant than total ZIP14 expression, is the expression of ZIP14 at the cell surface, which was also not determined. Given these caveats, mouse ZIP14_B had a higher V_{max} than mouse ZIP14_A for transport of Mn and Cd, while mouse ZIP14_A had a higher V_{max} for transport of Zn and Fe (table 1.1) [14, 36]. In stably-transfected mouse fetal fibroblasts, both ZIP14_A and B had a low V_{max} for Cd relative to Mn. In *X. laevis* oocytes, ZIP14_A had a low V_{max} for transport of Cd relative to Zn and Fe, while ZIP14_B had a low V_{max} for Fe relative to Zn and Cd (table 1.1) [14, 36]. Like the results reported for mouse ZIP14, human ZIP14_B is reported to transport manganese more rapidly than ZIP14_A [14, 19].

Unlike V_{max}, K_m is unaffected by transporter concentration, though measured values will be elevated by the activity of other transporters in the experimental system. Endogenous expression of surface transporters is very low in *Xenopus*

Table 1.1: **Reported K_m and V_{max} values for mouse ZIP14 isoforms.** a) Metal uptake was measured in mouse fetal fibroblasts stably transfected with mouse ZIP14 isoforms [14]. b) Metal uptake was measured in *X. laevis* oocytes injected with capped RNA encoding mouse ZIP14 isoforms [35, 36].

	ZIP14_A		**ZIP14_B**		
a)	**K_m (mM)**	**V_{max} (pmol/min/mg protein)**	**K_m (mM)**	**V_{max} (pmol/min/mg protein)**	Refs.
Mn	4.4±0.5	330±29	18.2±2	1140±100	[14]
Cd	0.14±0.02	25±3	1.1±0.02	113±12	[14]
b)	**K_m (μM)**	**V_{max} (pmol/oocyte/hr)**	**K_m (μM)**	**V_{max} (pmol/oocyte/hr)**	Refs.
Cd	0.31±0.01	0.51±0.01	0.54±0.78	1.6±0.07	[36]
Zn	0.78±0.027	1.67±0.10	0.36±0.03	1.05±0.01	[36]
Zn			1.9±0.6		[35]
Fe	0.64±0.16	0.92±0.08	0.19±0.12	0.38±0.07	[36]
Fe			2.3±0.5		[35]

laevis oocytes, making them an ideal system for the study of transporters, and explaining the much lower K_m values (∼1000-fold lower) obtained for ZIP14 transport in oocytes, relative to those obtained in mouse fetal fibroblasts (table 1.1) [14, 35, 36, 38]. Since low K_m values correspond to high affinity for the solute, these results suggest that mouse ZIP14_A has a higher affinity than ZIP14_B for Mn and Cd, and a lower affinity for Zn and Fe (table 1.1) [14, 36]. The results of Girijashanker et al. [14] suggest a much higher affinity of both ZIP14_A and B for Cd relative to Mn. Results obtained in oocytes suggest that ZIP14_A shows the highest affinity for Cd relative to Zn and Fe, while ZIP14_B shows the highest affinity for Fe relative to Zn and Cd (table 1.1) [35, 36].

If the kinetic properties of mouse and human ZIP14 isoforms are similar, then a comparison of typical human serum concentrations of Mn (ranges from ∼7-16

nM in healthy individuals, higher in liver cirrhosis [39]), Zn (ranges from ∼9-15 µM in healthy individuals, is lower in liver cirrhosis [40]), Fe (NTBI should not be present in healthy individuals, but may reach ∼20 µM in iron-overload conditions [41]), and Cd (∼2 nM in healthy individuals, is typically higher in smokers [42]) with the relevant K_m values may be instructive (figure 1.3). Such a comparison shows that the K_m values for uptake of Cd by ZIP14_A and B are much higher than typical serum Cd concentrations, with the result that uptake of Cd should be highly responsive to changes in serum [Cd] (table 1.1). Even considering that the ZIP14 K_m values for Mn are likely to be artificially high, it remains possible that the the true K_m values for Mn are higher than serum [Mn] (K_m values obtained for Cd in *X. laevis* oocytes are three orders of magnitude lower than those obtained in mouse fetal fibroblasts for the same metal, while serum [Mn] is six orders of magnitude lower than the K_m values for Mn obtained in mouse fetal fibroblasts), meaning that uptake of Mn by ZIP14 would also be responsive to changes in serum [Mn].

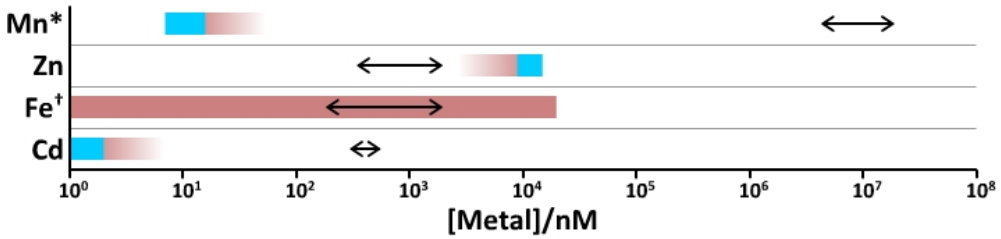


Figure 1.3: **Measured K_m values of mouse ZIP14 relative to typical concentrations of metals in human serum.** Blue bars represent approximate metal concentration ranges in serum of healthy individuals, red bars/gradients represent ranges/changes in the serum concentrations of metals in specific disease conditions (Mn - liver cirrhosis; Zn - liver cirrhosis; Fe - iron overload; Cd - smoking), double-headed arrows represent the range of K_m values measured in *X. laevis* oocytes for mouse ZIP14 A and B isoforms (table 1.1). *The measured K_m values for Mn were obtained in mouse fetal fibroblasts and may be relatively high. †In healthy individuals, plasma iron is bound to the protein transferrin. However, in iron-overload disorders, saturation of transferrin leads to free plasma iron (see section 1.3.1 for more information).

In contrast, the K_m values for uptake of Zn, and potentially Fe during iron overload, are lower than human serum concentrations for both ZIP14_A and B (table 1.1). ZIP14 would therefore be saturated relative to zinc, as well as iron in iron overload conditions. The implication is that while controlling ZIP14 uptake of Mn and Cd could involve regulation of cell surface ZIP14 expression, controlling ZIP14 uptake of Zn and Fe would necessitate regulation of cell surface ZIP14 expression.

Competition studies are conflicting [14, 35, 36, 43]. Carried out in different cell types, and using different concentrations of metals, they are not directly comparable and do not agree in generalities or specifics. Added variability may be due to regulatory effects of certain metals on ZIP14, or on other metal transporters. In one study, a 10-fold excess of unlabelled iron led to only a 60% decrease in uptake of ^{55}Fe, rather than the expected 90%, suggesting an unexpected complexity in the experimental system being tested [35].

The optimum pH for Cd^{2+} and Fe^{2+} uptake by mouse ZIP14 is $\sim$pH 7.5 [14, 35], while some iron transport can be observed at pH 6.5, and none at pH 5.5 [15, 35]. Mouse ZIP14 zinc uptake is functional over a much broader range of pH values, though the optimum is also pH 7.5 [35].

1.3.1 Iron

Iron, a redox-active transition metal, is capable of causing significant oxidative stress via Fenton chemistry [44]. In healthy persons, plasma iron is typically bound to transferrin in the ferric form (Fe^{3+}), referred to as transferrin bound iron (TBI) [8]. At high concentrations of iron in the plasma, increased transferrin saturation leads to the formation of non-transferrin bound iron (NTBI), the majority of which deposits in the liver, followed by the pancreatic acinar cells,

and then the heart [16, 45]. ZIP14 transports both TBI and NTBI [15, 34].

1.3.1.1 Non-Transferrin Bound Iron

Initially, ZIP14 was believed to only transport NTBI. Overexpression of ZIP14 increases NTBI uptake [34, 35], ZIP14 suppression significantly decreases it [34, 46], and in ZIP14 knockout mice, iron overload does not lead to the usual deposition of NTBI in the liver and pancreatic acinar cells [45].

Both NTBI and TBI are in the ferric form; however, ZIP14 transports ferrous iron (Fe^{2+}). Overexpression of the prion ferrireductase (ferrireductases convert ferric iron to ferrous iron) has been found to increase ZIP14 uptake of ferric, but not ferrous iron [47], and ferrous iron chelation prevents ZIP14 uptake of ferric iron, showing that ferric iron must be converted to the ferrous form before transport [34].

1.3.1.2 Transferrin Bound Iron

ZIP14 is also implicated in the transport of endosomal TBI into the cytosol. Holotransferrin binds to the transferrin receptor, causing its endocytosis. As the pH of the endosome decreases, the iron is released from transferrin, and must then be reduced to ferrous iron by ferrireductases in the endosome before being transported from the endosome to the cytosol [8, 48, 49]. This last transport function is certainly carried out by divalent metal transporter 1 (DMT1) [50], but there is evidence that in the liver, ZIP14 contributes to the process.

DMT1 is suited to the mildly acidic endosomes (pH 6.8-5.9) [51], its optimum pH has been reported to be pH 5.5 [52], or pH 6.75 [53]. However, uptake of TBI into the cytosol can occur independently of DMT1. In liver-specific DMT1 knockout mice, 60% of hepatic TBI uptake capacity remains [54]. In mice, hepatic DMT1

expression is low relative to ZIP14 [16], increasing in iron-deficient diets [16, 52]. Accordingly, the role of DMT1 in hepatic TBI uptake is proposed to be more important in iron-deficiency [54].

Overexpression of ZIP14 in HEK 293T cells increases TBI uptake without altering the rate of transferrin endocytosis [15]. The optimum pH of ZIP14 iron transport is pH 7.5, however, it retains activity at pH 6.5, and intracellular ZIP14 showed 56% co-localisation with transferrin receptor 1 (TfR1) [15]. It should be noted that although ZIP14 knockout mice have normal hepatic TBI uptake [45], liver DMT1 mRNA is increased relative to wild-type, which may compensate for the absence of ZIP14 [21].

1.3.1.3 Iron Homeostasis and ZIP14

Iron, and regulators of iron homeostasis appear to regulate ZIP14. Iron-deficient HepG2 cells (a liver-derived cell line) extract ZIP14 from the membrane for proteasomal degradation, while iron-loading protects ZIP14 and extends its half-life [55]. For reasons as yet unknown, ZIP14 iron-sensitivity depends on glycosylation at N102 [55]. The closely related ZIP8 is also post-transcriptionally regulated by iron. Iron-loading of cells increased both ZIP14 and ZIP8 protein levels, however ZIP8 increased at the plasma membrane [56], while ZIP14 accumulated intracellularly [55].

Increased ZIP14 protein in high iron diets has been observed in the liver and the pancreas, but not the heart [16], or primary pancreatic islets [46]. Loading of HepG2 cells with iron also increases ZIP14 expression [16].

Hereditary haemochromatosis protein (HFE, for High Fe) and hepcidin are both key regulators of iron homeostasis. Hereditary hemochromatosis (HH), a condition characterised by excessive iron absorption, is most commonly caused

by defective HFE protein, leading to underproduction of hepcidin [8]. Ordinarily, HFE and hepcidin decrease serum iron by preventing absorption of dietary iron and increasing iron retention in certain cells, such as the hepatocytes [8].

Increased HFE expression in HepG2 cells shortens ZIP14 half-life, decreasing NTBI uptake [57], while in HFE knockout mice, ZIP14 mRNA in the small intestine is increased [58]. In hepcidin knockout mice, pancreatic acinar cells express more ZIP14, and take up more iron [59]. It is not clear whether hepcidin and/or HFE are directly affecting ZIP14, or indirectly due to altered intracellular iron concentrations. Combined, it appears that in HH, increased ZIP14 in the liver and pancreatic acinar cells increases uptake of NTBI.

Overall, ZIP14 appears to contribute to normal iron homeostasis, and becomes more important in iron overload.

1.3.2 Manganese

Manganese is an essential cofactor for a number of enzymes, including glycosyltransferases, and superoxide dismutase [60], and is actually required by members of all enzyme classes [61]. Manganese is abundant in the diet, absorption of $\sim$2-3% is sufficient to meet manganese requirements, and although high iron diets depress manganese absorption, deficiency is rare [62, 63]. Excess manganese intake typically occurs through inhalation of manganese-containing dust [64], which bypasses normal regulatory processes, or liver failure, which impairs manganese removal [9]. Excess manganese accumulates in the globi pallidi of the brain [9], causing manganism, which shares symptoms with Parkinson's disease [60]. Like iron, manganese can be taken up into cells via transferrin and DMT1 [24, 65]. It is also transported by certain members of the ZnT (Zinc Transporter) family, and the ZIP family (ZIP8

and ZIP14) [14].

Manganese sequestration is becoming recognised as a component of inflammation [66], as a lack of manganese apparently renders pathogens more susceptible to host defence strategies [67, 68]. Upregulation of ZIP14 in response to interleukin-6 (IL-6) has been shown to increase manganese uptake in SH-SY5Y neuronal cells [24], but the potential significance of increased tissue manganese during inflammation has not been investigated.

1.3.2.1 ZIP14 Mutations Affecting Manganese Transport

The majority of research into ZIP14 function was initially focused on zinc, and then iron. However, in humans with mutated ZIP14, it is typically manganese homeostasis that is grossly disturbed [19, 69–73]. Some mutations result in severe truncations of the protein; with the loss of the majority of the predicted transmembrane region, loss of function is expected [19, 73]. One of the mis-sense mutations, G383R, disrupted the highly conserved EEXPHEXGD motif (to EEXPHEX*R*D) which is believed to be involved in metal transport [19]. The other mis-sense mutations impaired manganese transport for reasons unknown [19, 73]. One truncation mutation in particular only affected ZIP14_B, which is expressed in fewer tissues than ZIP14_A, and yet the symptoms were essentially the same [19]. The main symptoms were childhood-onset of hypermanganesemia and accumulation of manganese in certain areas of the brain, causing significant damage and resulting in parkinsonian symptoms, a number of which indicate faulty brain control of muscle function, such as tremors, muscle spasms and scoliosis; chelation of excess manganese by intravenous $Na_2CaEDTA$ can alleviate the parkinsonian symptoms, though it is not always effective [19, 73]. Similar symptoms were seen in ZIP14 knockout mice [27]. The initial explanation for these symptoms was that ZIP14 was

required for uptake of excess manganese into the liver, prior to excretion of that manganese in the bile secreted by the liver [19]. However, liver-specific ZIP14 knockout mice did not show manganese accumulation in the brain or other tissues when fed a normal diet, and only displayed greater manganese accumulation than controls, when given a manganese-rich diet, and even then, the increased concentration of manganese in the brain did not impair motor function [27]. Subsequent studies showed that knocking out ZIP14 expression in the intestine caused more severe symptoms, though preventing expression in both the liver and the intestines had a larger effect [74]. Based on studies in cell lines and mice, ZIP14 is believed to regulate intestinal absorption of manganese by reabsorbing excess manganese from the blood, and facilitating its export into the intestinal lumen [75, 76]. Hepatic ZIP14 may be unable to independently manage excess manganese, as exposure to excess manganese increases lysosomal degradation of ZIP14 [77].

While the hypermanganesemia and parkinsonian symptoms are clear defects in those with ZIP14 mutations, immune dysfunction may be an overlooked symptom. In the study by Tuschl et al. [19], although the nine identified patients were very young, three had died before the age of eight, two from infection, the third from unknown causes. This is far greater than the under-5 mortality rate (approximately 50 per thousand live births) [78], raising the possibility of immune system dysfunction.

1.3.3 Zinc

Zinc is essential for growth [79], a cofactor for hundreds of enzymes, and an essential component of some protein domains, such as the zinc finger domain commonly found in transcription factors [80].

1.3.3.1 Zinc Uptake

The role of zinc uptake has been primarily assigned to ZIP4, as ZIP4 mutations lead to acrodermatitis enteropathica (AE [MIM (Mendelian Inheritance in Man) 201100]), an inherited disorder of zinc absorption in the intestinal tract [81, 82]. Zinc homeostasis is maintained over a wide range of intake levels [83]. This can be partially explained by the regulation of ZIP4, both in terms of transcription and its subcellular location in enterocytes, in response to the level of dietary zinc [84, 85]. Other transporters may be involved, as the zinc deficiency resulting from faulty ZIP4 can be overcome with zinc supplementation [86], and a number of incidences of AE cannot be attributed to any defects in ZIP4 or its expression [4]. ZIP14 is also involved in zinc absorption; ZIP14 is found on the basolateral membranes, and within the cells, of enterocytes in mouse small intestine, with the highest expression in the duodenum, at the tips of the villi [32]. In ZIP14 knockout mice, absorption of orally administered zinc is decreased by $\sim$50% [21]. ZIP14 has been proposed to play a role in the absorption of zinc from enterocytes into the villi capillaries [32].

1.3.3.2 Zinc Deficiency

Zinc deficiency is a common problem worldwide, estimated to affect approximately 20% of the global population [7, 87]. Depending on the severity of the deficiency, symptoms can include skin lesions, alopecia, diarrhoea, impaired growth, and immune system dysfunction [7]. Zinc supplementation in zinc-deficient populations restores immune function and decreases mortality from infection [82, 88, 89].

1.3.3.3 Zinc Excess

Excess zinc has the potential to be toxic, but ingestion at high enough levels to be detrimental is unusual as the regulation of intestinal absorption is extremely effective, over the range of intakes that occur with ordinary food [83]. Excess intake of zinc does occasionally occur, for example, in the unintentional intake of zinc in dental adhesives [90], or kidney dialysis using water from zinc-coated tanks [91]. In such cases, copper deficiency, rather than excess zinc, explains the resulting symptoms. Excess zinc upregulates expression of metallothionein in enterocytes, which binds zinc and prevents its absorption into the body. The increased metallothionein also binds and prevents the absorption of copper, leading to zinc-induced copper deficiency [92]. The only known genetic disorders leading to excessive zinc absorption, are caused by overexpression of zinc-binding proteins [93, 94]. However, the excess zinc is sequestered, so patients are either asymptomatic, or their symptoms mimic zinc deficiency, due to the excessive binding of zinc by the overexpressed protein [94].

1.3.3.4 A Human ZIP14 Mutation Affecting Zinc Transport

Although a number of disease-causing mutations in ZIP14 have been reported in the literature, only one, L441R, has been connected to zinc transport [19, 69–73, 95]. With this specific mutation, ZIP14 fails to traffic to the cell surface. Despite the intracellular localisation of L441R ZIP14, overexpression in HEK293T cells increased intracellular levels of zinc, suggesting that L441R ZIP14 causes accumulation of zinc within cells by decreasing zinc export. This mutation affected the development of a subset of bone cells, which was suggested to be related to high intracellular zinc increasing the activity of the

cAMP-CREB signalling pathway [95].

1.3.4 Metallothionein

Metallothioneins are small, cysteine-rich proteins capable of binding both essential and toxic metals, such as copper, zinc and cadmium [96]. Metallothionein acts as an intracellular store of zinc, and plays a key role in the homeostasis of intracellular zinc [97, 98].

Metallothionein expression, like that of ZIP14, is responsive to inflammatory status. In mouse liver, lipopolysaccharide (LPS) administration led to a 32.7-fold increase in metallothionein mRNA when tested 16 hours later [20]. This increase in metallothionein expression appears to be partially dependent on increased zinc uptake by ZIP14, as in ZIP14 knockout mice, where inflammation fails to increase hepatic zinc uptake, there is an attenuated increase in metallothionein [21]. In addition, metallothionein expression is required for the increased uptake of zinc by the liver during inflammation [97].

1.4 ZIP8

Within the LIV-1 family, ZIP8 is the most closely related to ZIP14. In particular, it shares the alteration of the otherwise characteristic sequence, HEXPHEXGD, to EEXPHEXGD [14]. Approximately 50% of the ZIP8 amino acid sequence is identical to that of ZIP14 [99], and it also has three glycosylation sites in the extracellular N-terminus [56].

ZIP8 transports zinc [36, 56, 100], iron [36, 56], manganese [14] and cadmium [14, 36]. Although ZIP8 and ZIP14 both transport cadmium, studies attempting to explain genetic variation in sensitivity to cadmium toxicity identified ZIP8 as

the culprit, suggesting it plays the more important role in cadmium-induced testicular necrosis [101, 102]. This is most likely due to its higher expression in the testes, rather than a difference in cadmium affinity [14, 36].

The optimal pH of ZIP8 iron transport is $\sim$pH 7.5, and some activity is retained at pH 6.5, but not pH 5.5 [56]. Like ZIP14, post-transcriptional regulation leads to an increase in ZIP8 protein in iron-loaded cells, and a transient effect was also observed with zinc [56].

The pattern of ZIP8 expression in human tissues differs markedly from that of ZIP14, with the highest expression found in lungs [56, 102], testes [102], and kidneys [102], while expression in the liver and small intestine is very low [56]. ZIP8 has been detected at the plasma membrane [56], in early endosomes [56], lysosomes [103] and mitochondria [104].

Like ZIP14, ZIP8 is involved in inflammation [100]. It was first identified as a gene induced during activation of monocytes and dendritic cells by bacterial components [105], and was later shown to be induced by LPS or tumour necrosis factor alpha (TNFα) in a number of other cell types [25, 100]. In human septic and critically ill patients, monocyte ZIP8 mRNA was greatly increased relative to healthy people, and ZIP8 expression was correlated with decreased plasma zinc, a key component of inflammatory responses [104]. The importance of ZIP8 in the regulation of the immune response is illustrated by the fact that NF-kB activation increases ZIP8 expression, while ZIP8 expression negatively regulates NF-kB [100]. The NF-kB pathway is a key signal transduction pathway, regulating inflammation in numerous cell types [106]. ZIP8 and ZIP14 both seem to play roles in the regulation of inflammation, particularly in respect to negative feedback.

1.5 Regulation of Other ZIP proteins

ZIP proteins are regulated by a variety of methods, affecting transcription, degradation and trafficking. The regulation of ZIP4, the main transporter responsible for absorption of dietary zinc, has received the most attention, and studies have revealed layered modes of control. Expression of ZIP4 is upregulated in the enterocytes of mice on zinc-deficient diets [84]. In addition, in cells treated with a zinc chelation agent in order to mimic zinc deficiency, ZIP4 localisation to the plasma membrane was dramatically increased within 1 hour of treatment [85]. At high zinc concentrations, ZIP4 is ubiquitinated, targeting it for degradation [107]. Certain other ZIP proteins also appear to be either up- or down-regulated by zinc concentration, and in some cases, this has been attributed to the zinc-responsive transcription factors, metal regulatory transcription factor-1 and 2 [106, 108]. Various other ZIP proteins are regulated by hormones or cytokines [18], such as ZIP6 by oestrogen [109], and ZIP8 by TNFα [25, 100].

1.6 Trafficking of ZIP14

ZIP14 is found at the plasma membrane and on intracellular vesicles [13, 15]. The mechanism of ZIP14 endocytosis has not been determined, though intracellular ZIP14 has been shown to partially co-localise with transferrin, known to be internalised by clathrin-mediated endocytosis (CME) [15, 110]. While studying the effects of cellular iron on ZIP14, Zhao et al. [15] found deglycosylation of ZIP14 to be significantly decreased by dynasore, an inhibitor of dynamin, a GTP-ase involved in the scission of vesicles from the plasma membrane [111, 112]. This was interpreted as an indication that prior to

deglycosylation of ZIP14, the protein is endocytosed, though the effect of dynasore on the cellular location of ZIP14 was not observed [55]. Dynasore is known to decrease clathrin-mediated endocytosis through its inhibition of dynamin, however, dynamin is involved in additional, clathrin-independent forms of endocytosis, and many proteins are internalised by more than one mechanism [113]. Furthermore, dynasore has also been shown to alter cell behaviour through additional, dynamin-independent mechanisms [111, 114, 115]. Therefore, experimental results pertaining to ZIP14 trafficking have not unequivocally confirmed or ruled out any specific form of endocytosis.

As a transmembrane protein, degradation of ZIP14 would be expected to occur in lysosomes [116]. In HepaRG cells exposed to manganese, the resulting decrease in ZIP14 protein was prevented by treatment with bafilomycin A1, a drug that inhibits the vacuolar ATP-ases required for the low vesicular pH at which lysosomal proteases function [77, 117, 118]. Treatment with bafilomycin A1 also increased co-localisation of ZIP14 with vesicles containing the lysosomal marker Lamp1 [77]. Together, these results suggest that the degradation of ZIP14 promoted by manganese exposure occurred in lysosomes, although it should be noted that bafilomycin A1 has the off-target effect of decreasing cellular uptake of iron, which could be relevant to ZIP14 behaviour [119]. Surprisingly, the degradation of ZIP14 in response to iron-depletion of HepG2 cells proceeded in the presence of inhibitors of lysosomal degradation, but was prevented by two inhibitors of proteasomal degradation, MG132 and epoxomicin, though proteasomal degradation is known as the main pathway for degrading intracellular, soluble proteins [55, 120–123]. Although MG132 is a non-selective proteasome inhibitor that additionally inhibits the calcium-dependent protease m-calpain, epoxomicin is believed to specifically target proteasomes [124, 125].

1.6.1 ZIP14 Recycling

Recycling of internalized ZIP14 to the plasma membrane has not been specifically studied, however ZIP14 abundance at the cell surface has been shown to decrease in response to siRNA-knockdown of sorting nexins 27 and 17 (SNX27 and SNX17), components of two distinct protein complexes involved in recycling of proteins to the plasma membrane [126, 127]. SNX27 and 17 act as adaptor proteins for the multiprotein trafficking complexes retromer and retriever, respectively, enabling selection of specific cargo proteins for recycling [128–131]. SNX27 and 17 each have cargo-binding FERM-domains that recognize NPxY/NxxY motifs (where x is any amino acid), while SNX27 also possesses a PDZ domain that binds a C-terminal S/Txφ motif (where φ is any hydrophobic amino acid) [132, 133]. The ZIP14 amino acid sequence contains neither of these motifs, though it co-precipitates with SNX27; ZIP14 may possess a currently unknown motif mediating interaction with these adaptor proteins, or there may be an indirect interaction involved [126].

1.7 ZIP14 and Signalling

Zinc affects the activity of a number of enzymes related to signalling, over a range of concentrations [134–138]. Proteins with a relatively low IC_{50} (half maximal inhibitory concentration) for zinc have partially depressed activity, and zinc deficiency leads to the protein being overactive. This can be observed with phosphodiesterase (PDE) [29], an enzyme that breaks down cAMP, the second messenger that activates protein kinase A. In ZIP14 knockout mice, the lowered concentration of zinc in hepatocytes causes overactive PDE, resulting in decreased sensitivity to glucagon [29]. In the chondrocytes of ZIP14 knockout

mice, the overactive PDE decreases the sensitivity to parathyroid hormone-related peptide, decreasing bone growth [29].

Proteins with an IC_{50} for zinc close to the cytosolic zinc concentration will be regulated by any stimulus that alters zinc concentration in the cytosol. Phosphotyrosine phosphatase 1 B (PTP1B) is one such protein. PTP1B, like other PTPs, counteracts certain receptor tyrosine kinase-induced pathways, by removing the phosphates that are added during signal transduction [139]. In AML12 hepatocytes, increased zinc uptake by ZIP14 has been linked to an increased response of the c-Met receptor to its ligand, hepatocyte growth factor [136].

1.7.1 Zinc Signalling and Fuel Metabolism

Perhaps due to the high expression of ZIP14 in the liver and pancreas [13, 14], ZIP14 transport of zinc appears to be necessary for the normal regulation of fuel metabolism. As mentioned, ZIP14 contributes to glucagon sensitivity of hepatocytes through maintenance of the zinc concentration and zinc's effect on phosphodiesterase [29].

The relationship between ZIP14 and insulin is more complicated. In wild-type mice, hepatic glucose uptake upregulates ZIP14, and liver zinc content increases [31]. Zinc is known to have an insulinomimetic effect [140, 141], which is believed to be due to its inhibition of PTPs [134, 135, 141]. The overall effect is that during fasting, the liver will respond appropriately to glucagon, and after carbohydrate ingestion, hepatic insulin sensitivity will be increased, again leading to appropriate metabolic behaviour.

In ZIP14 knockout mice, hepatic insulin sensitivity, surprisingly, is increased, as the decreased zinc levels inhibit insulin-degrading enzymes, thereby increasing

insulin signalling [31]. This increased sensitivity occurs all the time, rather than in the post-absorptive state alone, leading to an attenuated shift in metabolic behaviour when transitioning between the fed and fasted state. Constitutively increased insulin sensitivity in the liver explains the low blood sugar, the decreased gluconeogenesis and the increased glycogenic and lipogenic activity of the liver in ZIP14 knockout mice [21, 31]. The transition from the fasting state to the fed is also affected. For example, the fasting plasma insulin concentration is higher in ZIP14 knockout mice than in WT, but after glucose intake, the increase in plasma insulin is far less than expected [31]. The result is that there is very little difference in the concentration of plasma insulin between the fed and fasting states in ZIP14 knockout mice, suggesting that pancreatic ZIP14 may be involved in the regulation of insulin secretion. The effect of ZIP14 on liver sensitivity to glucagon and insulin is an example of how zinc may be used to regulate the delicate balance between signalling pathways.

ZIP14 is also involved in adipocyte metabolism. ZIP14 expression correlates closely with expression of the adipogenic transcription factor, PPARγ [142, 143]. Adipocyte ZIP14 expression is decreased in the obese [142, 143], but increases after weight loss [142], and is believed to provide the zinc required for adipogenesis [142].

1.7.2 Zinc Signalling and Cachexia

Cachexia, the progressive wasting of muscle that occurs in some cancer patients, not only affects quality of life for those affected, but is a significant cause of mortality [144]. Cachexia decreases tolerance for cancer treatments, and the wasting of muscle can even lead to cardiac and respiratory failure [144]. Cachexia is described as a systemic condition, and is not the result of anorexia

or weight loss, as it occurs despite nutritional supplementation and exercise regimes [144].

On studying the transcriptome of cachectic muscles, Wang et al. [145] found that proteins related to zinc metabolism were significantly up-regulated, ZIP14 in particular. Through the study of mouse cell lines, various mouse models of cancer, and samples from human patients, ZIP14 has been identified as a central point in the development of cachexia [145, 146]. TNFα and TGF-β, both cytokines that may be released by metastatic tumours, upregulate muscle expression of ZIP14, both preventing muscle repair, and weakening existing muscle by increasing the degradation of a key myofibrillar protein [145]. Tumour-bearing ZIP14 knockout mice did not develop cachexia, even when supplemented with zinc, while zinc supplementation exacerbated cachexia in tumour-bearing wild-type mice [145]. It has been suggested that ZIP14 may therefore make a promising drug target for the treatment of cachexia [147].

1.7.3 Zinc Signalling and Immunity

During an infection, the initial stage of defence, known as the acute phase response, is initiated by the recognition of pathogen-associated molecular patterns, such as lipolysaccharide (LPS). This is a basic, in-born level of immunity which goes on to stimulate adaptive immunity, involving T-cells and B-cells, which carry out attacks more tailored to the infectious agent [148]. While inflammation has the obvious aim of removing and destroying pathogens, the immune response has to be precisely tailored for maximal effectiveness against the specific pathogen, while minimizing harm to the host [149]. Research indicates that the spatial and temporal regulation of trace metal concentrations is required for an effective immune response, as well as protection of the host.

1.7.3.1 Lipopolysaccharides

Lipopolysaccharides, also known as endotoxin, are a class of compounds found on the outer membrane of Gram-negative bacteria [150]. LPS is the archetypal pathogen-associated molecular pattern, and stimulates the production of numerous cytokines; the three cytokines interleukin-6, interleukin-1β (IL-1β) and TNFα are upstream of many LPS effects, and have been referred to as the 'classical pro-inflammatory triad' [150]. LPS is a powerful immunostimulator, and is commonly used in studies of inflammation, as are the members of the classical pro-inflammatory triad. LPS, and each member of the pro-inflammatory triad, has been found to alter ZIP14 expression, with tissue-specific effects [25, 151].

LPS increases ZIP14 mRNA in macrophages, the liver and other tissues [21–23] and causes a ZIP14-dependent increase in hepatic zinc content [21]. Compared to the ∼3-fold change in expression in the liver [20], adipose tissue expression of ZIP14 increases more than 30-fold, skeletal muscle expression nearly 20-fold [21]. IL-6 is largely responsible for LPS-induced increases in ZIP14 mRNA. Hepatic ZIP14 mRNA does not increase in IL-6 knockout mice [20], and IL-6 induces ZIP14 mRNA in a large number of cell types tested [24, 25]. IL-1β is capable of inducing ZIP14 expression, partly via the activation of inducible nitric oxide synthase [151]. TNFα also induces ZIP14 expression [25].

1.7.4 Nutritional Immunity

Nutritional immunity is a broadly used defence against bacteria [66]. Sequestering trace elements deprives invading pathogens of nutrients needed for their survival, and has been observed with iron, zinc, and manganese [66, 152, 153]. Nutritional immunity can also involve using high concentrations

of metals to kill pathogens. For example, in macrophages, phagocytosed bacteria may be attacked by the removal of trace metals such as zinc and manganese from the phagosome [154], or alternatively, by pumping metals, such as zinc or copper, into the phagosome to reach toxic concentrations [155].

During inflammation, plasma concentrations of iron, zinc and manganese are all decreased [66, 152, 153]. The hypozincemia of inflammation is greatly dependent on the induction of ZIP14, unlike hypoferremia [21]. IL-6 is believed to cause hypoferremia independently of ZIP14, via hepcidin [156], a hormone involved in inflammation and regulation of iron transport [157], and in ZIP14 knockout mice, hypoferremia occurs as usual [21]. In SH-SY5Y cells (neuroblastoma cells), IL-6 led to an increase in mRNA and protein of ZIP8 and ZIP14, and the down-regulation of Znt10, a manganese and zinc exporter. The accumulation of manganese was attributed to the combined regulation of these transporters [24]. It is possible that ZIP14 may play a part in establishing hypomanganesemia during inflammation. The body of research on ZIP14 and inflammation indicates that the regulation of trace metal transport is core to immune defence, and that ZIP14 is an important transporter in this response.

1.7.4.1 Manganese

Manganese is an essential element for bacteria [68], and mutations in the manganese transport proteins of pathogens decrease virulence [158]. Hypomanganesemia is a component of the acute phase response [153], and manganese depletion from bacterial abscesses has also been documented [159]. Calprotectin demonstrates the efficacy of manganese sequestration. Calprotectin is secreted by neutrophils [67], and binds both zinc and manganese [68]. Manganese binding is key to calprotectin's effectiveness against infections. When calprotectin is mutated, such that Zn-binding is intact, but Mn-binding is

ablated, its ability to inhibit the growth of *Staphylococcus aureus* is severely compromised. Mn-sequestration deprives *S. aureus* of the manganese required for superoxide dismutase, making it vulnerable to the oxidative stress created by attacking neutrophils [68]. The identity of the transporters responsible for hypomanganesemia during inflammation is unknown, but ZIP14 is a promising contender [24].

1.7.4.2 Zinc

As in eukaryotes, a significant proportion of bacterial proteins require zinc in order to function [160]. Depriving pathogens of the zinc they need for growth and survival is an effective strategy, and accordingly, pathogen proteins that enhance zinc uptake are powerful virulence determinants [152]. ZIP14, in concert with metallothionein [97], plays an important role in the removal of zinc from the plasma [20]. Much of the plasma zinc is redistributed to the liver [161], although other cell types also increase zinc uptake during inflammation.

Zinc and Sepsis

Sepsis, with its high mortality rate of $\sim$40% [162], exemplifies the risks of a dysregulated immune response, and its close ties to zinc status demonstrate the importance of zinc in moderating inflammatory signalling [163, 164]. In humans, septic patients exhibited lower serum zinc levels than their critically ill counterparts, and cytokine concentrations were inversely correlated with plasma zinc concentrations [165]. Animal studies have been used to elucidate the relationship between zinc and sepsis.

Experimentally, sepsis is most often induced with a procedure called cecal ligation puncture (CLP), which involves puncturing the cecum so that fecal

matter enters the peritoneal cavity, causing serious infection by gut bacteria [166]; the severity of the infection can be modified by altering the thickness of the puncturing needle [167]. Deaths during sepsis may occur during an initial, aggressive immune response, or later after a prolonged and unsuccessful immune response [168].

During studies of sepsis, the expected redistribution of zinc from the plasma to the liver is observed [164, 167]. Hepatic ZIP14 expression is more strongly affected than other ZIP proteins, and expression of both inflammatory and anti-inflammatory proteins increases [167]. Zinc deficient diets prior to CLP lead to a large decrease in survival [163], increased production of cytokines [163, 169], increased bacterial load [169] and increased tissue damage [163]. Zinc supplementation of zinc-deficient mice, prior to induction of sepsis, results in a decreased bacterial load [167, 169], increased survival and cytokine production similar to that seen with normal zinc intake [163].

In ZIP14 knockout mice, the progress of mild sepsis is greatly altered. An initial decrease in serum zinc returns to baseline levels within less than 24 hours, compared to the more than 72 hours seen in the wild-type mice. Additionally, liver zinc did not increase, though metallothionein, surprisingly was greatly increased. Indicative of the flawed immune response, the white blood cell (WBC) count initially decreased rather than increased, and apoptosis of liver cells was increased [167]. Liver production of key inflammatory and anti-inflammatory cytokines was decreased, and in WBCs, IL-10 (anti-inflammatory) production was decreased [167]. The need for zinc sufficiency and proper regulation of its distribution for a normal immune response, shows that the redistribution of zinc is not simply a matter of depriving pathogens of an essential nutrient, but also serves to regulate host defence and repair [152].

Zinc Signalling In Inflammation

In its capacity as a modulator of signalling pathways, zinc regulates inflammation. Microarray analysis of gene expression following CLP in mice fed zinc-deficient/adequate/supplemented diets identified the JAK/STAT and NF-kB pathways as the most affected by zinc [170]. These are both essential inflammatory pathways, and adequate zinc is needed for effective negative feedback modulation. NF-kB activation requires the phosphorylation of the inhibitory protein IkB by IK-kinase (IKK), which targets IkB for degradation, releasing NF-kB to translocate to the nucleus, where it alters transcription [171]. Zinc directly inhibits IKK [100, 170], thereby enhancing negative feedback on this pathway. During inflammation zinc deficiency leads to increased activation of NF-kB, while zinc supplementation suppresses NF-kB signalling [169]. Zinc enhances negative feedback of the JAK/STAT3 signaling pathway by increasing the activity of SHP-1, a phosphatase that dephosphorylates the STAT3 transcription factor, preventing its translocation to the nucleus [170]. Zinc also increases expression of SOCS-3, which binds and inhibits JAK kinase [172]. Zinc inhibition of the JAK/STAT3 pathway decreases production of acute phase proteins [170]. Zinc concentrations in tissues may still be altered hours, or even days after the initial stimulus [167], but uptake immediately after exposure to LPS has been demonstrated in monocytes and macrophages (peak concentration reached at 2 minutes), and this zinc was required for the subsequent induction of cytokine production [138]. Zinc is now increasingly recognised as a second messenger, like calcium, whose concentration changes are able to effect changes within minutes to hours (referred to as a ‘zinc wave’) or even within seconds (referred to as ‘zinc flux’) [173]. In monocytes, the increased zinc enhanced LPS-induced responses, by inhibiting tyrosine phosphatases, thereby protecting LPS-induced tyrosine phosphorylation in the

MAP kinase pathway [138]. Zinc also has effects on adaptive immunity. In activated T-cells, the increase in intracellular zinc is concentrated at the location of the T-cell receptor synapse, a signalling complex that forms between T-cells and the dendritic cells that activate them [174]. Zinc also alters polarization of T-cells, which in turn affects the development and proliferation of other immune cells during an immune response [152]. As well as regulating inflammation, zinc is required to protect the host from harm during infection. Inflammatory strategies targeted at pathogens also increase the likelihood of apoptosis in host cells [175]. Supplemental zinc in the medium of LPS-treated sheep pulmonary endothelial cells significantly decreases apoptosis, an effect that is dependent on the expression of ZIP14 and partially dependent on metallothionein [22]. This was attributed to direct inhibition of caspase-3 by zinc, a key enzyme in apoptosis [22]. During an inflammatory response, zinc-deficient mice have increased caspase-3 activity in lung tissue, relative to controls and zinc supplemented mice [169]. In fact, zinc is an inhibitor of the caspase family, at physiologically relevant concentrations [176]. ZIP14 expression increases during inflammation in a number of tissues, including the liver. The liver plays an important role in both the defence and repair components of an immune response [162]. 80% of macrophages are found in the liver (Kupffer cells) [177], and it secretes the acute phase proteins crucial for the initial inflammatory response [178]. The ZIP14-mediated increase in hepatic zinc during inflammation regulates liver production of acute phase proteins [170], regulates the secretory profile of other tissues [33], and in the absence of ZIP14, the immune response is altered [21, 167].

1.7.5 Other Aspects of Inflammation

Inflammation is involved in more phenomena than simply defence against pathogens. Inflammation also occurs in wound healing, allergic reactions and different types of stress. After experimental abdominal surgery or partial hepatectomy in mice, the resultant inflammation includes hypozincemia, increased liver zinc and increased expression of ZIP14 [136, 179]. In ZIP14 knockout mice that underwent partial hepatectomy, liver zinc content did not increase, and liver regeneration was significantly impaired. In the wild-type mice, the additional zinc increased the response to hepatocyte growth factor [136]. In a mouse model of allergic inflammation, zinc supplementation significantly decreased the number of eosinophils and lymphocytes in bronchoalveolar lavage fluid (a method of sampling lung tissue) [180]. During the allergic inflammation, ZIP14 expression was the most affected of the ZIP proteins, increasing 5.3-fold [180]. In rats experiencing psychological stress over two weeks, serum corticosterone, a stress hormone, increased as expected. The changes in zinc metabolism typically seen during an immune response were observed, albeit over a longer time-scale; a redistribution of zinc from the plasma to the liver, and an increase in hepatic ZIP14 mRNA [181]. Endoplasmic stress in mice can be induced using tunicamycin, or a high fat diet. In both cases, hepatic ZIP14 expression increases, and the subsequent zinc uptake inhibits apoptosis in the liver [137]. In ZIP14 knockout mice, the increased zinc uptake does not occur, there is increased accumulation of fat in the liver, relative to wild-type mice, and increased apoptosis [137]. These variations on inflammation serve to confirm the importance of ZIP14 during inflammation, and highlight some examples of how its regulation benefits the organism, in moderating inflammation.

1.8 Aims

The ZIPs are the only known family of zinc importers, and there is growing recognition of the role of zinc in both short and long term regulation of key cellular processes, such as signalling and gene expression. ZIP14 is an unusual member of the ZIP family, as it is able to transport manganese and iron, as well as zinc, and has different physiological roles in the regulation of each of these three metals. A number of disease-causing mutations have been identified in ZIP14, and little is known about the mechanisms through which they prevent ZIP14 from functioning. While most ZIP14 mutations affect manganese metabolism, one has been shown to alter ZIP14 trafficking, and impacts zinc metabolism instead. The aims of this thesis were as follows:

1. To study the effects of a selection of disease-causing point mutations on cellular localisation and iron-uptake function of ZIP14.

2. To use computational biology to model a structure of ZIP14, and make predictions about the structure-function relationship of ZIP14.

3. To experimentally test predictions resulting from the computational model of ZIP14, and draw conclusions on the structure-function relationship of ZIP14.

1.8.1 Overview

In this thesis, the known disease-causing point mutations of human ZIP14 were studied in order to understand key aspects of ZIP14 function and behaviour. Mutant ZIP14 was expressed exogenously and compared to wild-type, in terms of metal uptake and cellular localisation. The mutants varied from wild-type

ZIP14 in a number of ways, and these differences were then studied further in order to identify the key reasons for dysfunction. The structure of human ZIP14 was predicted with homology modelling, and new mutations were designed to test the model, and predictions based on the results of the disease-causing mutations. The experimental results, in combination with evolutionary analysis of the homology model, led to a number of novel conclusions about the structure-function relationship of ZIP14; some of these conclusions may be specific to ZIP14, while others will be relevant for the wider ZIP family.

Chapter 2

Materials and Methods

2.1 Materials and Methods

2.1.1 Plasmids and Plasmid Construction

The full-length cDNA of the genes encoding ZIP14 (SLC39A14) isoforms (A: Q15043-1, B: Q15043-3 and C: Q15043-2) had been previously sub-cloned into the mammalian expression vector pIRESneo2 (Takara Bio Europe/ Clontech), with a hemagglutinin (HA) tag at the 3' end of each gene. The plasmid encoding ZIP14_A had also been used to make the F98V and N469K mutants by Samuel Du (Dr Bowers, University College London). The full-length cDNA of ZIP14_A had also been sub-cloned into a pEGFP-N1 plasmid (Takara Bio Europe/ Clontech), such that the code for GFP was positioned at the 3' end of ZIP14 (Dr Bowers, University College London).

2.1.1.1 Site-directed Mutagenesis

Site-directed mutagenesis was used to insert mutations into the sequences of ZIP14 isoforms A and B with primers, (see tables 2.1 and 2.2, for disease-causing and designed mutations, respectively). Primers were ordered from Sigma-Aldrich, USA, at a concentration of 100 µM, desalted, in TE buffer.

Table 2.1: **Disease-causing mutation primers.** Forward and reverse primers for site-directed mutagenesis of the pIRESneo2 plasmids encoding hZIP14-HA.

Mutation	Primer sequences (forward, reverse)
F98V	5'-GTTCTGGAGACCTCGTCACTGCCCACAATTTC-3' 5'-GAAATTGTGGGCAGTGACGAGGTGTCCAGAAC-3'
S104I	5'-CTGCCCACAATTTCATCGAGCAGTCGCGGATTG-3' 5'-CAATCCGCGACTGCTCGATGAAATTGTGGGCAG-3'
R128W	5'-GCAGCTGGATTCCTGGGCCTGCACC-3' 5'-GGTGCAGGCCCAGGAATCCAGCTGC-3
P379L	5'-CTCTGTGAGGAGTTCCTACATGAGCTAGGAGAC-3' 5'-GTCTCCTAGCTCATGTAGGAACTCCTCACAGAG-3
G383R	5'-CCCACATGAGCTAAGAGACTTTGTCAT-3' 5'-ATGACAAAGTCTCTTAGCTCATGTGGG-3'
L441R	5'-CTTGTATATTTCTCGGGCTGATATGTTCCCTGAG-3' 5'-CTCAGGGAACATATCAGCCCGAGAAATATACAAG-3
N469K	5'-CCATTTATCCAGAAACTGGGCCTCCTGAC-3' 5'-GTCAGGAGGCCCAGTTTCTGGATAAATGG-3'

Site-directed mutagenesis was carried out using either Pfu Turbo DNA polymerase (Agilent Technologies, Germany) or Herculase II Fusion DNA polymerase (Agilent Technologies, Germany); the reagents and concentrations used are described in table 2.3.

The PCR conditions were carried out in a Bio-Rad DNA Engine® Thermal Cycler (Bio-Rad Laboratories, California, US), programmed as described in table 2.4. PCR products were examined using agarose gel electrophoresis to confirm

Table 2.2: **Designed mutation primers.** Forward and reverse primers for site-directed mutagenesis of the pIRESneo2 plasmids encoding hZIP14-HA.

Mutation	Primer sequences (forward, reverse)
L62A	5'-CTCTGCAGCAGGCGAAGGCCCTGCTCAAC-3' 5'-GTTGAGCAGGGCCTTCGCCTGCTGCAGAG-3'
L69D	5'-CTGCTCAACCACGACGATGTGGGAGTGGG-3' 5'-CCCACTCCCACATCGTCGTGGTTGAGCAG-3'
C118D	5'-CTCCAGGAGTTCGACCCCACCATCCTCCAGCA-3' 5'-TGCTGGAGGATGGTGGGGTCGAACTCCTGGAG-3
L125D	5'-ATCCTCCAGCAGGACGATTCCCGGG-3' 5'-CCCGGGAATCGTCCTGCTGGAGGAT-3
L206A	5'-CTCTTCCAGGCCATCCCGGAGGCATTTG-3' 5'-CAAATGCCTCCGGGATGGCCTGGAAGAG-3
E209Q	5'-CCAGCTCATCCCGCAGGCATTTGGTTTCAAC-3' 5'-GTTGAAACCAAATGCCTGCGGGATGAGCTGG-3'
Q365E	5'-GTGTCAGTTTTCGAAGGCATCAGCAC-3' 5'-GTGCTGATGCCTTCGAAAACTGACAC-3
E376H	5'-GCCATCCTCTGTCACGAGTTCCCAC-3' 5'-GTGGGAACTCGTGACAGAGGATGGC-3'

expected size.

PCR reaction products were treated with 20 units of DpnI (New England Bioloabs, UK) at 37 ℃ for ∼1-2 hours to digest the template plasmid. Competent DH5α *E.coli* (100 µL) were incubated on ice with 5-10 µL of DpnI-treated PCR product for 15 minutes, subjected to heat shock at 42 ℃ for 1 minute, and returned to ice for 1 minute. 750 µL of Luria broth [182] was added and the mixture incubated at 37 ℃ for 1 hour. The bacteria were spun down (15700 rcf for 1 minute) and plated on Luria broth agar [182] plates with ampicillin (100 µg/mL) (Sigma-Aldrich, USA); the plates were incubated at 37 ℃ for ∼18 hours.

The resultant colonies were used to inoculate 5 mL Luria broth, including

Table 2.3: **Reagent concentrations for site-directed mutagenesis.**

Reagents	Concentration (Pfu Turbo)	Concentration (Herculase II)
Deoxynucleotides (A, C, G and T) (Thermo Fisher Scientific, USA)	0.2 mM each	0.25 mM each
Forward, reverse primers	0.2 µM each	0.25 µM each
DNA polymerase	2.5 U per 50 µL	1 µL per 50 µL (concentration not provided)
Template plasmid	∼1 ng/µL	∼1 ng/µL
Enzyme buffer	1 x stock buffer	1 x stock buffer
Dimethyl sulphoxide (Sigma-Aldrich, USA)	0-8% final conc.	0-8% final conc.
Distilled water	To final volume	To final volume

Table 2.4: **PCR conditions for site-directed mutagenesis.** *Temperatures were adjusted according to the melting temperatures (T_m) of the primers; at step 7 the temperature was typically 3-5 ℃ below the lowest primer T_m, the step 3 temperature was 3 ℃ below this.

Step	Temperature (℃)	Duration/Instruction
1.	95	30 sec
2.	95	30 sec
3.	50*	1 min
4.	68	1 min per kb
5.		Go to step 2, 5 times
6.	95	30 sec
7.	53*	1 min
8.	68	1 min per kb
9.		Go to step 6, 13 times
10.	4	Forever

ampicillin (100 µg/mL), and incubated in a New Brunswick Innova 4200 Incubator-Shaker (National Institute of Standards and Technology, U.S Department of Commerce, Maryland, U.S.A) (250 rpm) at 37 ℃ for ∼18 hours. 3 mL of the resulting bacteria were used to purify the plasmid using the QIAprep Spin Miniprep Kit (Qiagen, Germany), according to the manufacturer's instructions.

Samples of the plasmid were sent for sequencing to Source Bioscience Ltd, with appropriate primers (provided by Dr Bowers). Concentrations of DNA and primers were as specified by Source Bioscience Ltd.

2.1.1.2 Creation of Deletion Mutants

In order to make the △40-130 mutant, PCR amplification of the pIRESneo2 hZIP14-HA plasmids was carried out using a forward primer whose 5' end annealed to the codon for T131, and a reverse primer whose 5' end annealed to the codon for S39 (see the LinAmp primer pair in table 2.5 and the PCR conditions for amplification in table 2.6). This step amplified the entire plasmid, excluding the codons for residues 40-130 of hZIP14-HA.

This PCR product was ligated to circularise the linear DNA, and the ligation products were used to transform bacteria. As described in section 2.1.1.1, plasmids were isolated from colonies and sent for sequencing. Site-directed mutagenesis, as described in section 2.1.1.1, was then used to correct errors in the ligation site (see the RepairLig primer pair in table 2.5).

Table 2.5: **△40-130 deletion** Forward and reverse primers for deletion of the △40-130 residues from the pIRESneo2 plasmids encoding hZIP14-HA.

Primer pair	Primer sequences (forward, reverse)
LinAmp	5'-ACCTCGGAGAACCAGGAAAACGAGGAGAATGAGC-3' 5'-GCTGATAGCTGGTGCACCCGGGGATGAAG-3'
RepairLig	5'-ATCAGCACCTCGGAGAACCAGGAAAACGAGGAG-3' 5'-CGAGGTGCTGATAGCTGGTGCACCCGGGGA-3'

Table 2.6: **PCR conditions for plasmid amplification.**

Step	Temperature (℃)	Duration/Instruction
1.	95	2 mins
2.	95	20 sec
4.	72	4 mins
5.		Go to step 2, 29 times
10.	4	Forever

2.1.2 Cell Culture

HeLa and HEK293 cells were provided by Dr. Bowers, HepG2 cells by Prof. Srai. Cells were cultured in Dulbecco's Modified Eagle Medium (DMEM), supplemented with 10% foetal bovine serum (FBS), 2 mM L-glutamine, penicillin and streptomycin (10,000 U and 10 mg/mL, respectively). Growth medium for HeLa cells stably expressing ZIP14 or variants was supplemented with 0.2 mg/mL G418. Information regarding the reagents used for cell line maintenance is detailed in table 2.7.

Cells were passaged every 3-4 days (up to 30 passages from thawing) by washing with phosphate buffered saline (PBS) (Sigma-Aldrich, USA), incubation with trypsin-EDTA solution (Sigma-Aldrich, USA) for approximately 5 minutes in

Table 2.7: **Reagents used in maintenance of cell lines.**

Reagents	Information
Dulbecco's Modified Eagle Medium (DMEM)	Sigma-Aldrich, USA
Foetal bovine serum (FBS)	Sigma-Aldrich, USA
L-glutamine	Sigma-Aldrich, USA
Penicillin/Streptomycin	Sigma-Aldrich, USA
Geneticin® (G418)	Thermo Fisher Scientific, USA
Phosphate-buffered saline (PBS)	Sigma-Aldrich, USA
Trypsin-EDTA solution	Sigma-Aldrich, USA

order to detach and separate the cells, and a proportion of the detached cells were transferred to new dishes with fresh growth medium.

Between passages, HeLa cells were incubated at 37 ℃ in 5% CO_2.

2.1.2.1 Transfection Protocols

HeLa cells were transfected using the TransIT-HeLaMONSTER® Transfection Kit (Mirus Bio LLC), according to the manufacturer's instructions. Transfected cells were used for experiments after 48 hours.

Monoclonal Stable Transfectants

48 hours after transfection, the growth medium of transiently transfected cells was replaced with growth medium supplemented with 0.2 mg/mL G418. Mock-transfected cells (cells transfected with the elution buffer of the QIAprep Spin Miniprep Kit) were used as a negative control. Upon the death of the cells in the negative control, the remaining cells constituted a polyclonal collection of stable transfectants, with a range of expression levels.

The surviving transfected cells were detached using trypsin-EDTA and added to 25 mL of growth medium at a concentration of 5 cells/mL. The cells were transferred to a 96-well plate, 200 µL per well, and allowed to grow until the colonies of cells were visible to the naked eye. Colonies in wells containing single colonies were then cultured and checked for expression by immunofluorescence microscopy or flow cytometry.

2.1.3 Reverse Transcription PCR

2.1.3.1 RNA Extraction

Cells were lysed in Trizol (Ambion, USA), and incubated in chloroform (Sigma-Aldrich, USA) for 5 minutes at room temperature. The samples were centrifuged at 4 ℃ for 15 minutes, and the top layer extracted and mixed gently with isopropanol. Samples were incubated at room temperature for 10 minutes, and then centrifuged for 30 minutes at 4 ℃. The pellets were washed in 1 mL of ethanol (Sigma-Aldrich, USA), then centrifuged at 4 ℃ for 5 minutes. The ethanol was removed and the pellets left to air-dry before being dissolved in DEPC-treated water (Ambion, USA) at 60 ℃ for 10 minutes. RNA concentrations were quantified by a Nanodrop 2000 (Thermo Fisher Scientific, USA).

2.1.3.2 cDNA Synthesis

RNA (1 µg) was treated with 1 unit DNase (New England Biolabs, UK) at 37 ℃ for 10 minutes, followed by a 10 minute incubation at 70 ℃. The DNase-treated RNA was used for cDNA synthesis with the Verso^{TM} cDNA Synthesis Kit (Thermo Fisher Scientific, USA); the reagents are detailed in table 2.8. Reaction mixtures were incubated at 42 ℃ for 1 hour, then at 95 ℃ for 2 minutes.

Table 2.8: **Reagents for cDNA synthesis.**

Reagents	Volume (µL)
5 x cDNA Synthesis Buffer	4
dNTP Mix	2
Verso RT-Enzyme Mix	1
RT Enhancer	1
Anchored Oligo dT	1

2.1.3.3 Use of cDNA to Detect Gene Expression

From the cDNA produced as described in section 2.1.3.2, 1 µL was used for PCR, with primers appropriate for the detection of particular genes; the reagents are detailed in table 2.9. The PCR conditions were carried out in a Bio-Rad DNA Engine® Thermal Cycler (Bio-Rad Laboratories, California, US), programmed as described in table 2.10.

Table 2.9: **Reagents for PCR of cDNA in order to detect gene expression.**

Reagents	Volume (µL)
5 x Herculase Buffer	5
100 mM dNTPs	0.4
10 µM forward primer	0.75
10 µM reverse primer	0.75
Distilled H_2O	19.5
DMSO	1
Herculase II polymerase	0.5
cDNA	0.5

Primers (detailed in table 2.11) were designed to amplify specific regions of the hZIP14 isoforms, and to span introns in order to prevent amplification of genomic DNA. Primers were ordered from Sigma-Aldrich, USA, at a

Table 2.10: **Conditions for PCR of cDNA in order to detect gene expression.**

Step	Temperature (℃)	Duration/Instruction
1.	95	30 sec
2.	95	30 sec
3.	65	1 min
4.	68	20 seconds
5.		Go to step 2, 29 times
6.	68	5 minutes
7.	4	Forever

concentration of 100 µM, desalted, in TE buffer.

Table 2.11: **Primers used to amplify specific regions of cDNA.** The primer numbers correspond to those seen in figure 3.3.

Primer	Primer sequence
GAPDH, forward	5'-CCTCCTGCACCACCAACTGC-3'
GAPDH, reverse	5'-ATGTTCTGGAGAGCCCCGCG-3
Primer 1	5'-CTGGATGTGGGAGTGGGCCG-3'
Primer 2	5'-GCTGGAGGATGGTGGGGCAG-3'
Primer 3	5'-CTGCCCCACCATCCTCCAGC-3'
Primer 4	5'-GCAGGGCAGGACGAGGACTC-3
Primer 5	5'-GGGAGCAGAGGGAGATGACGG-3
Primer 6	5'-GCTGCTACCTGGGTCTGGCC-3'
Primer 7	5'-TGCTGCCCTTCCTTTCATCCTCT-3'
Primer 8	5'-CGGCAGAGGTTGCAGTGAGC-3'

2.1.4 Immunofluorescence Microscopy

Information regarding the reagents used for immunofluorescence microscopy is detailed in table 2.12.

Table 2.12: **Reagents used in immunofluorescence microscopy.**

Reagents	Information
Phosphate-buffered saline (PBS)	Sigma-Aldrich, USA
Paraformaldehyde	Dissolved in PBS
Quenching solution	50 mM ammonium chloride dissolved in PBS
Staining solution	PBS, 5 mg/mL bovine serum albumin, 0.05% w/v saponin
4',6-diamidino-2-phenylindole (DAPI)	Thermo Fisher Scientific, USA
Möwiol	Heimer and Taylor [183]
ProLong™ Diamond Antifade Mountant	Thermo Fisher Scientific, USA
Binding buffer	RPMI medium with 10 mM HEPES and 0.2% w/v BSA, pH 7.35-7.42
Trypsin-EDTA solution	Sigma-Aldrich, USA

2.1.4.1 Staining of Permeabilized Cells

Cells were seeded onto 13 mm glass coverslips and allowed to attach overnight. Coverslips were washed with PBS, then fixed in 3% paraformaldehyde for 15 minutes, washed with PBS and quenched with quenching solution. Coverslips were then washed with PBS, permeabilized with staining solution and then stained with primary antibodies in staining solution for 1 hour. Coverslips were then washed with staining solution and stained with secondary antibodies dissolved in staining solution for 45 minutes. See table 2.13 for antibody concentrations. Coverslips were washed with PBS, the nuclei were stained with 300 nM DAPI and coverslips were mounted onto slides using Möwiol. Alternatively, coverslips were mounted with ProLong™ Diamond Antifade Mountant.

Table 2.13: **Antibodies used for immunofluorescence microscopy, with corresponding concentrations.**

Antibody	Concentration	Information
Mouse Anti-Hemagglutinin	1:500	Clone 16B12 (Covance, USA)
Rat Anti-Hemagglutinin	1:250	Clone 3F10 (Roche, Switzerland)
Rabbit Anti-Calreticulin	1:500	PA3-900 (Affinity BioReagents, USA)
Alexa Fluor 488 Goat Anti-Rat	1:500	(Abcam, UK)
Alexa Fluor 546 Goat Anti-Mouse	1:500	(Abcam, UK)
Alexa Fluor 546 Goat Anti-Rabbit	1:500	(Abcam, UK)

2.1.4.2 Staining of Non-permeabilized Cells

Cells were seeded onto 13 mm glass coverslips, as described in section 2.1.4.1. Prior to fixation, coverslips were kept on ice, washed with chilled binding buffer, then stained with primary antibodies in chilled binding buffer for 45 minutes. Coverslips were then washed with binding buffer and stained with secondary antibodies dissolved in chilled binding buffer for 30 minutes. Coverslips were then washed with chilled PBS, fixed, quenched, stained with DAPI and mounted onto slides, as described in section 2.1.4.1.

2.1.4.3 Surface/Intracellular 2-stage Staining

Cells were seeded, stained and fixed as described in section 2.1.4.2. Coverslips were then washed with PBS, permeabilized and stained again, as described in section 2.1.4.1.

2.1.4.4 Cell Imaging

For conventional fluorescence microscopy, coverslips were viewed using a Zeiss Axioskop 2 Plus microscope and images were taken with a Hamamatsu ORCA-ER digital camera. For confocal fluorescence microscopy, coverslips were viewed and imaged using a Leica TCS SPE confocal microscope (Leica Camera).

Colocalization Analysis

Image-processing was carried out using FIJI ImageJ [184] and standardized within experiments. Images were produced from the median signal intensity of 2x 0.173 µm-thick slices, followed by use of the ImageJ 'Subtract Background' tool (set to between 20 and 50 px, sliding paraboloid) and converted into composites using ImageJ. Comparably sized areas were selected from the images (an area within the cell, excluding the nucleus) and used for colocalization analysis with the ImageJ Coloc2 plugin, in order to obtain the Pearson's correlation coefficient.

2.1.5 Western Blotting

2.1.5.1 Protein Extraction

Cells grown on 6 cm dishes were chilled on ice, washed with chilled PBS and then lysed with chilled lysis buffer (0.5% NP-40, 1.25 mM EDTA, 0.15 M NaCl, 20 mM TrisHCl pH 8, EDTA-free protease inhibitor cocktail (Sigma-Aldrich, USA)) on ice for 5 minutes. The lysate was scraped and transferred to chilled 1.5 mL Eppendorf tubes. An equal volume of lysis buffer was again added to the dishes; the lysate was scraped and added to the previous lysate. Lysate

was centrifuged at 4 ℃ for 13 minutes at 15700 rcf, and the supernatant was retained. The protein concentration was quantified using the BCA Protein Assay kit (Thermo Fisher Scientific, USA), according to the manufacturer's instructions.

2.1.5.2 Gel Preparation and Loading

SDS-polyacrylamide gels were handcast at 1.5 mm thick, typically 10% acrylamide with ten lanes, according to the recipes provided by Kielkopf et al. [185]. Protein samples to be loaded were diluted to equal concentrations of protein in lysis buffer and mixed with 3x Laemmli buffer [186] (5% β-mercaptoethanol) at a 2:1 ratio. Unused lanes were loaded with lysis buffer mixed with Laemmli buffer at a 2:1 ratio. Gels were run at 80 V for 30 minutes, followed by 180 V for 1 hour.

2.1.5.3 Protein Transfer and Visualization

Protein was transferred onto a Biotrace NT Nitrocellulose transfer membrane (Thermo Fisher Scientific, USA) at 300 mA for 1 hour. Membranes were blocked in blocking solution (5% dried skimmed milk powder (Marvel, Premier Foods, Ireland), 0.1% Tween 20 (Sigma-Aldrich, USA) in Tris-buffered saline) at 4 ℃ for 12-18 hours. Membranes were incubated in primary antibodies for 2 hours, and then washed in 0.1% Tween 20 (Sigma-Aldrich, USA) in Tris-buffered saline (3x). Membranes were then incubated in secondary antibodies for 1 hour, and then washed in 0.1% Tween 20 in Tris-buffered saline (4x 5 minutes), and Tris-buffered saline (2x 5 minutes). Antibodies were diluted in blocking solution, see table 2.14 for concentrations. Membranes were visualized using a LI-COR Odyssey CLx Near-Infrared Fluorescence Imaging System, with Image Studio 5.0 software.

Table 2.14: **Antibodies used for western blotting, with corresponding concentrations.**

Antibody	Concentration
Mouse Anti-Hemagglutinin (Clone 16B12, Covance, USA)	1:1000
Rabbit Anti-Calreticulin (PA3-900, Affinity BioReagents, USA)	1:1000
IRDye 800CW Goat Anti-Mouse (LI-COR, USA)	1:5000
IRDye 680RD Goat Anti-Rabbit (LI-COR, USA)	1:5000

2.1.6 Flow Cytometry

2.1.6.1 Staining of Permeabilized Cells

The compositions and abbreviations of the reagents used are detailed in table 2.15. Cells were grown in 6-well plates to a confluence of approximately 100%. Cells were detached and separated by washing with PBSE, followed by incubation with trypsin-EDTA solution for 30 minutes at 37 ℃, 5% CO_2 to produce a single cell suspension. A maximum of 6 x 10^5 cells in trypsin-EDTA were transferred to 1.5 mL Eppendorf tubes and washed in PBSE, then fixed in PFAE for 15 minutes. Cells were washed in PBSE, then permeabilized in FPBB for 5 minutes. Cells were stained with primary antibody in 100 µL of FPBB for 1 hour, then washed in FPBB (2x). Cells were then stained with secondary antibody in 100 µL of FPBB for 45 minutes, then washed in FPBB (2x). Cells were then prepared for scanning - they were washed with PBSE, centrifuged and resuspended in 150 µL PBSE, then filtered into a new Eppendorf tube with 40 µm Falcon cell strainers (Thermo Fisher Scientific, USA).

Washes were carried out by centrifuging the samples at 4 ℃ for 5 minutes, at 400 rcf prior to fixation and 500 rcf after fixation. Antibody concentrations are detailed in table 2.16.

Table 2.15: **Reagents used to prepare cells for flow cytometry.**

Reagents	Information
PBS and ethylenediaminetetraacetic acid (EDTA) (PBSE)	2.5 µM EDTA
Paraformaldehyde and EDTA (PFAE)	3% paraformaldehyde, 2.5 µM EDTA in PBS
Flow cytometry permeabilizing blocking buffer (FPBB)	0.05% saponin, 1 mg/mL bovine serum albumin (BSA), 2.5 µM EDTA in PBS
Flow cytometry blocking buffer (FBB)	1 mg/mL BSA, 2.5 µM EDTA in PBS

Table 2.16: **Antibodies used for flow cytometry staining, with corresponding concentrations.**

Antibody	Concentration
Mouse Anti-Hemagglutinin (Clone 16B12, Covance, USA)	1:250
Rat Anti-Hemagglutinin (Clone 3F10, Roche, Switzerland)	1:250
Alexa Fluor 488 Goat Anti-Rat (Abcam, UK)	1:250
Alexa Fluor 647 Goat Anti-Mouse (Invitrogen)	1:500

2.1.6.2 Surface Staining of Non-Permeabilized Cells

A maximum of 6 x 10^5 cells in trypsin-EDTA was prepared, as described in section 2.1.6.1. Cells were stained prior to fixation, on ice, with primary and secondary antibodies, as described in section 2.1.6.1, with antibodies dissolved in FBB (table 2.15). Cells were then fixed and prepared for scanning, as described in section 2.1.6.1.

2.1.6.3 Surface/Intracellular 2-stage Staining Protocol

A maximum of 6 x 10^5 cells in trypsin-EDTA was prepared, as described in section 2.1.6.1. Cells were stained for surface antigen (using mouse anti-HA and anti-mouse Alexa Fluor 647) and then fixed, as described in section 2.1.6.2. Cells were then washed in PBSE, permeabilized and stained for intracellular antigen (using rat anti-HA and anti-rat Alexa Fluor 488), as described in section 2.1.6.1. Cells were then prepared for scanning, as described in section 2.1.6.1.

2.1.6.4 Flow Cytometry of Stained Samples

Stained samples were scanned using a BD Accuri™ C6 flow cytometer. Alexa Fluor 488 was detected in the FL1 channel (laser excitation: 488 nm, emission detection: 533/30 nm (detects wavelengths from 518 to 548 nm) [187]) and Alexa Fluor 647 in the FL4 channel (laser excitation: 640 nm, emission detection: 675/25 nm (detects wavelengths from 662.5nm to 687.5 nm) [187]). The flow cytometer was set to collect 10,000 events within the appropriate forward-scatter/side-scatter gate. For transiently transfected cells, an additional gate was used to exclude events with the same fluorescence as mock-transfected cells.

2.1.6.5 Analysis of Flow Cytometry Data

The data collected by the BD Accuri™ C6 flow cytometer were exported as .fcs files, and the files were analysed using the FlowCal 1.3.0 Python library [188].

FlowCal's 'FlowCal.gate.density2d' function was used in the forward-scatter and side-scatter channels in order to exclude debris. In order to select transiently transfected cells, mock-transfected cells were scanned to set a minimum

fluorescence threshold in the FL1 channel.

2.1.6.6 Titration of Antibodies for Flow Cytometry

In order to determine the best concentration of antibody to use, HeLa cells and HeLa cells stably expressing wild-type ZIP14_A-HA were permeabilized and stained with a range of antibody concentrations. The staining index (the MFI of the ZIP14-expressing cells divided by the MFI of the HeLa cells) was determined for each concentration, and the antibody concentration with the highest staining index was selected for use in future experiments.

2.1.7 Iron Uptake

Cells growing in 6-well plates between 80 and 100% confluency were placed on ice, washed in 1 mL chilled PBS three times, and then incubated in iron uptake buffer (see table 2.17 for composition) for 30 minutes at 37 ℃. Each sample's uptake was measured in triplicate.

Table 2.17: **Composition of iron uptake buffer.**

Reagent	Concentration
10x RPMI	1x
1 mM HEPES	10 mM
5 M NaOH	Add until pH is between 7.35 and 7.42
$FeSO_4.H_2O$	10 µM
$^{55}FeCl_3$ (PerkinElmer, USA)	0.8 kBq/mL
L-Ascorbic acid (Sigma-Aldrich, USA)	100 µM

After incubation, the cells were placed on ice and the iron uptake buffer was removed. Cells were washed in 1 mL chilled PBS three times, and then

incubated on ice with 750 µL lysis buffer (see section 2.1.5.1 for composition) for a minimum of 5 minutes. The lysates were scraped and transferred to chilled Eppendorf tubes, then centrifuged at 4 ℃ for 13 minutes at 15700 rcf. The supernatant was retained, and the 30 µL was used to quantify the protein concentration, using the BCA Protein Assay kit (Thermo Fisher Scientific, USA), according to the manufacturer's instructions.

2.1.7.1 Iron-55 Quench Curve

Scintillation vials were filled with 5 mL scintillation fluid (Sigma-Fluor LSC Cocktail for Aqueous Medium, Sigma-Aldrich, USA), 0.5 mL lysis buffer and 30,000 disintegrations per minute (dpm) of $^{55}FeCl_3$. Each tube was scanned in a TriCarb 2900 LTR PerkinElmer Scintillation Analyzer to determine the counts per minute (cpm), and the scanning was terminated when the cpm reached 0.5%2S statistical significance, meaning that the true cpm is expected to be within a range of plus or minus 0.5% of two standard deviations of the measured cpm. Vials with counts that differed from the mean by more than 2% were discarded. Acetone was added to the remaining tubes over a range of volumes from 0 to 450 µL, and then scanned for 10 minutes using the Quanta LLC Quench programme, between the energies of 0 and 10 KeV; the parameter used to determine quench was tSIE/AEC. The resulting quench curve was used by the Quanta LLC software to calculate dpm.

2.1.7.2 Scintillation Counting

For each sample, 0.5 mL of the lysate was added to 5 mL of scintillation fluid. The vials were vortexed for approximately 5 seconds, and then wiped with a damp tissue to remove static. The vials were scanned in a TriCarb 2900 LTR

PerkinElmer Scintillation Analyzer; cpm were collected for 10 minutes, between the energies of 0 and 10 KeV. Cpm were converted to dpm using a quench curve, described in section 2.1.7.1. Results were presented as dpm/µg protein.

2.1.8 Statistics

Where appropriate, one-way ANOVA was performed, followed by Tukey-Kramer post-hoc analysis to identify significant differences between means ($P < 0.05$). Statistical analysis was performed with JMP ® Pro 17.0.0 (SAS Institute Inc, NC, U.S.A).

2.1.9 Sequence Alignment and Analysis

2.1.9.1 Pairwise Alignment

Pairwise alignments were carried out using the EMBL-EBI EMBOSS Needle Pairwise Sequence Alignment Tool [189], with default settings.

2.1.9.2 Multiple Sequence Alignment

Multiple sequence alignments were carried out with a maximum of 500 sequences using the EMBL-EBI T-COFFEE online interface (https://www.ebi.ac.uk/Tools/msa/tcoffee/) [189].

2.1.9.3 Profile-Profile Alignment

Profile-profile alignments were performed with pairs of multiple sequence alignments using the MUSCLE program [190].

2.1.9.4 Conservation Scoring of Multiple Sequence Alignments

Mulitple sequence alignments in the ClustalW format were submitted to the AL2CO web server (http://prodata.swmed.edu/al2co/al2co.php) [191]; scores were normalised relative to the maximum score at 1, and the minimum score at 0.

2.1.9.5 Calculation of Percentage Identity and Similarity

A python script was written to calculate percentage identity and similarity between short alignments, to match the scores given by the EMBL-EBI EMBOSS Needle Pairwise Sequence Alignment Tool (section 2.1.9.1).

Similar amino acids are detailed in table 2.18 (for the python script, see appendix B.1).

Table 2.18: **Similar amino acids.** Amino acids considered as similar in order to calculate percentage similarity for short pairwise alignments.

A	C	D	E	F	G	H	I	K	L	M	N	P	Q	R	S	T	V	W	Y
S		E	D	W		N	L	E	I	I	D		E	K	A	S	I	F	F
		N	K	Y		Y	M	Q	M	L	H		K	Q	N		L	Y	H
			Q				V	R	V	V	S		R		T		M		W

2.1.10 Homology Modelling

Modeller: After installation of Modeller 9.21, the target-template (PIR format) alignment was used for homology modelling according to the user documentation [192].

Phyre2: The protein sequence was submitted to the Phyre2 Protein Fold Recognition server, using the Normal mode (recommended if a homologous

protein is known) [193].

SWISS-MODEL: The target-template alignment (FASTA format) was submitted to the SWISS-MODEL Expasy web server, in the alignment mode [194].

Side-chain positions were optimised by SCWRL4 [195].

2.1.10.1 Multimer Modelling

Sequences were entered as both 'sequence_1' and 'sequence_2' into section 3 of the AlphaFold Google CoLab site, with 'run_relax' disabled [196]. Numbering of residues in the resultant PDB files was corrected with the 'resrenum' command in UCSF Chimera [197].

2.1.11 Structure Quality Assessment and Analysis

2.1.11.1 Root Mean Square Deviation Calculation

The UCSF Chimera program was used to align two protein structures, using the Matchmaker command with the Needleman-Wunsch alignment algorithm and the BLOSUM-62 matrix [197]. The 'rmsd' command was used for relevant pairs of α-carbons, and an average taken of the results.

2.1.11.2 Identification of Residue Contacts

Structure .pdb files were primarily imaged and analysed in the UCSF Chimera program [197]. Residue contacts were identified using the 'Find Clashes/Contacts' command, with Van Der Waals overlap of $\geq$ -0.4 Å.

2.1.11.3 Identification of Mutation Sites

The UCSF Chimera program [197] was used to identify residue contacts (section 2.1.11.2). A python script was used to extract contacts that only involved R-groups (for this script, see appendix C.2.1).

2.1.11.4 Quality Assessment Programs

ProSA-Web: PDB files were submitted to the ProSA-Web server at prosa.services.came.sbg.ac.at/prosa.php. Chains were specified for assessment as necessary, according to instructions [198]. ANOLEA: PDB files were submitted to ANOLEA at melolab.org/anolea. Chains were specified for assessment as necessary, according to instructions [199]. QMEAN: PDB files were submitted to the QMEAN website at swissmodel.expasy.org, with the QMEANDisCo (N-terminal extracellular domain) or QMEANBrane (transmembrane domain) options selected [200].

2.1.12 Co-evolution Analysis

2.1.12.1 Human ZIP14 Distance Matrix

Distances between residues in the homology model were extracted from the .pdb file using the Biopython library [201, 202], and distances of 10 Å or less were used to produce a scatter plot (appendix D.3).

The sequence of human ZIP14, isoform A (Q15043-1), residues 154-257;321-486 was submitted to the GREMLIN web-server (http://gremlin.bakerlab.org/submit.php), with options detailed in table 2.19.

Textual output of the co-evolution analysis was downloaded, and residue

Table 2.19: **Options used for co-evolution analysis of hZIP14 on the GREMLIN web-server.**

Option	Selection
Generate MSA using	HHblits
E-value	1E-10
Iterations	4
Filter coverage	75
Remove gaps	75
GREMLIN Prior	Vanilla

numbering corrected; strength of coevolution was indicated by the 'r_sco' column (appendix D.1). Relative to a maximum co-evolution score of 0.4244, scores over 0.15 were considered a strong signal.

The sequence of human ZIP14, isoform A (Q15043-1), was submitted to the PSIPRED Workbench (http://bioinf.cs.ucl.ac.uk/psipred/); the DeepMetaPSICOV 1.0 option was selected. Textual output of the co-evolution analysis was downloaded; strength of co-evolution was shown in the fifth column (appendix D.2). Results for residues less than 4 amino acids away from each other were not included in the analysis. Relative to a maximum co-evolution score of 1, less than 0.2 was not include in the analysis, greater than 0.4 was considered a strong signal.

Co-evolution data was parsed by a python script to produce a scatter plot of co-evolving amino acid positions; colour and size of points was altered to indicate strength of co-evolution signal (appendix D.3).

Chapter 3

Effects of Disease-Causing Mutations on Iron Uptake and Cellular Localisation of ZIP14

3.1 Introduction

In this chapter, ZIP14 mutations documented to cause human disease will be studied in order to establish functional consequences of each mutation, leading to a study of the structural implications of these mutations in the following chapters. As with proteins in general, the primary sequence of human ZIP14 (hZIP14) is the determinant of all aspects of its function, either due to the features resulting directly from its own three-dimensional structure, or due to the resulting interactions with other proteins and components in its environment. There are variations in protein sequences between healthy individuals, as proteins can tolerate amino acid substitutions without deleterious effects, so disease-causing mutations can highlight regions of the protein that

are functionally important [203].

Figure 3.1 shows the number of documented non-synonymous single nucleotide polymorphisms (SNPs) at each position of the hZIP14 protein sequence, only a handful of which are known to be harmful [204]. The deleterious hZIP14 mutations known at the time of experimental design, and which will be studied in this chapter, are also indicated in figure 3.1, and are shown to be in areas with low variation in mammalian ZIP14 sequences (figure 3.1), suggesting these regions are important for ZIP14 function. In addition, if these pathogenic mutations negatively affect ZIP14 function in different ways, this information may also lead to a greater understanding of the structure-function relationship of ZIP14.

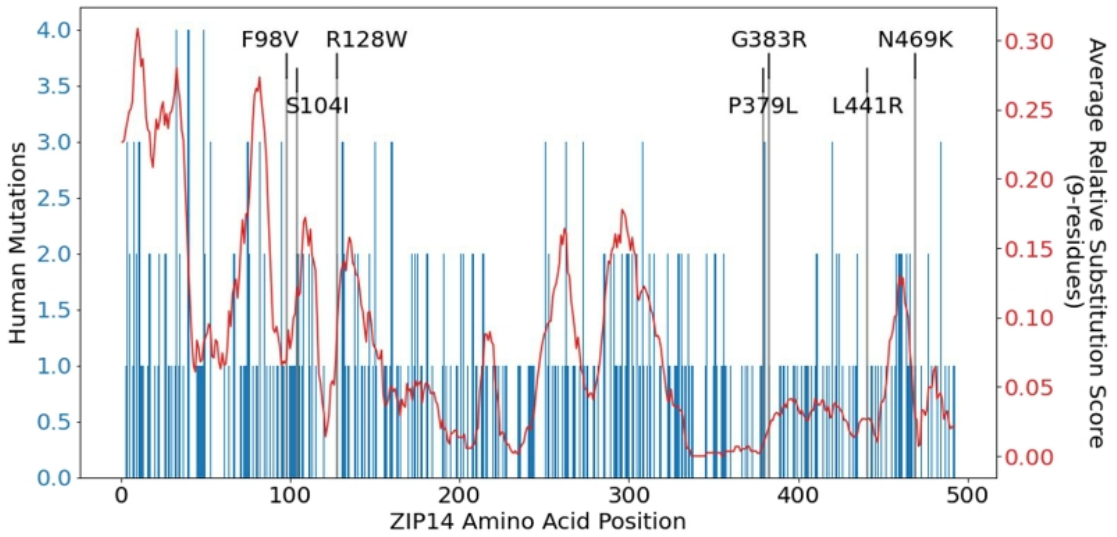


Figure 3.1: **Variation across the human ZIP14 sequence.** Graph depicting, at each position in the human ZIP14 sequence, the number of non-synonymous SNPs (blue) /hfl(for both isoforms) and the sequence variation within mammalian ZIP14 proteins (red, as 'average relative substitution score', averaged over 9 residues) [204, 205]. Positions of disease-causing mutations to be studied in this chapter are depicted in grey.

3.1.1 Exogenous Expression of Wild-Type and Mutant Human ZIP14

The reported disease-causing point mutations of hZIP14 are listed in table 3.1, the mutations studied in this thesis are indicated in bold, and are those reported in 2018 or earlier. As shown in figure 3.2, the mutations are found in both the predicted extracellular N-terminal region, and in three of the predicted transmembrane helices.

Table 3.1: **Reported disease-causing point mutations in ZIP14.** Mutations are listed with reported symptoms. Mutations used in this study are emphasised in bold.

Mutation	Symptoms	Published
F98V, G383R, N469K	Hypermanganesemia with dystonia	2016 [19]
R128W	Hypermanganesemia with dystonia	2017 [69]
S104I	Hypermanganesemia with dystonia	2018 [72]
P379L	Hypermanganesemia with dystonia	2018 [71]
L441R	Hyperostosis cranialis interna	2018 [95]
L59Q, N102, R108W, A432P, Y438C	Hypermanganesemia with dystonia	2022 [73]

The majority of the mutations to be studied were reported to be recessive, suggesting that expression alongside wild-type (WT) hZIP14 could mask the effects of the faulty copy. For this reason, I chose to identify a cell line where endogenous ZIP14 expression was low or absent, in which to exogenously express the wild-type and mutant ZIP14 proteins. A useful aspect of exogenous expression is that the focus is on the behaviour of the proteins, as regulation of expression need not be considered.

Transient transfections were used in the study of all cell lines, and were especially useful when studying changes in ZIP14 behaviour at different levels

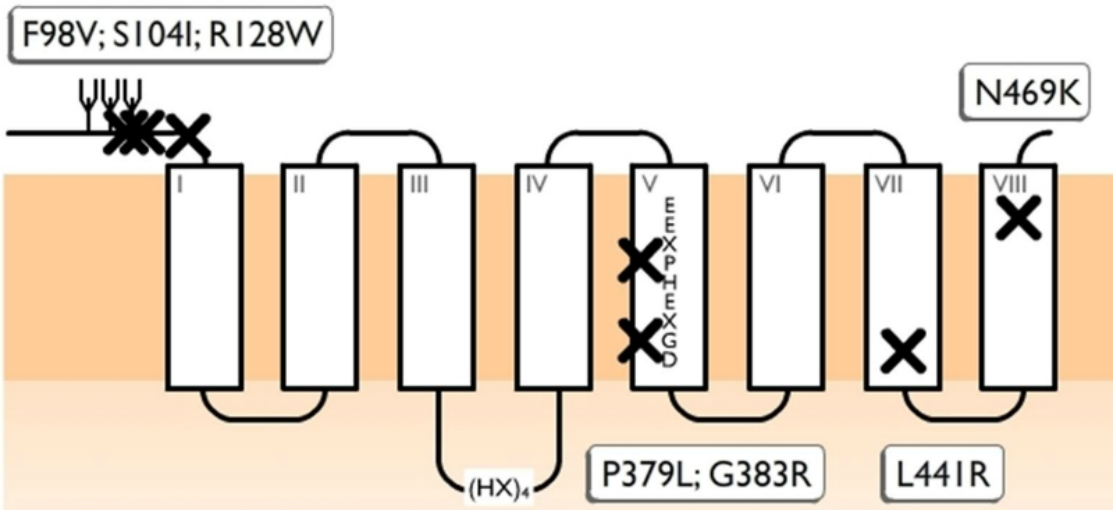


Figure 3.2: **Positions of disease-causing mutations relative to predicted topology of hZIP14.** The positions of the mutations studied in this thesis are indicated on the predicted topology with black crosses. Predicted locations are based on analysis by Taylor et al. [13].

of expression. Monoclonal stable cell lines were used where behaviour of all cells in the sample needed to be consistent, particularly where normalisation relative to hZIP14 expression or hZIP14 cell surface levels within experiments was not possible.

3.1.2 Experiments on hZIP14 Variants

While mutations of transporters can affect a range of characteristics, such as solute preference, or appropriate responses to regulation, I focused on the two most important features of any transporter: the ability to transport solute, and correct localisation within the cell. Uptake of ^{55}Fe by stably transfected cells was used to measure uptake function of the hZIP14 variants. Overexpression of transporters is considered an acceptable technique for functional assays, as long as the parental cell line is used as a negative control [206, 207]. Staining of hZIP14 was achieved with anti-HA primary antibodies, and fluorescently labelled secondary antibodies. The exact staining protocols were adapted for

various purposes, in order to prepare cells for immunofluorescence microscopy or flow cytometry, and where necessary, to differentiate between surface and internal hZIP14. Immunofluorescence microscopy allowed me to identify the subcellular localisation of hZIP14 within transfected cells; flow cytometry allowed me to study the behaviour of hZIP14 over a range of expression levels in a quantitative manner. The overall outcome was a more detailed understanding of the various effects of hZIP14 disease-causing mutants, most of which have not previously been characterised [19, 69, 71, 72].

3.2 Results

Little is known about the mechanisms through which hZIP14 mutations cause disease. In this chapter, disease-causing mutations of hZIP14 are assessed for their effects on iron-uptake function, cellular localisation and trafficking to the plasma membrane.

3.2.1 Selection of the HeLa Cell Line for Transfections

Documented disease-causing mutations in hZIP14, with the exception of L441R [95], only cause symptoms in those who are homozygous for the mutated gene [19, 69–71]. For this reason it was decided to use reverse-transcriptase PCR (RT-PCR) to identify a cell line expressing little or no ZIP14, for transfection with plasmids encoding hZIP14 with a C-terminal hemagglutinin (HA) tag, as there was no suitable antibody for detection of hZIP14.

It was also of interest to know which hZIP14 isoforms were being expressed, so primers were designed that could test for hZIP14 expression, and distinguish between the hZIP14_A, B and C isoforms (Q15043-1, Q15043-2 and Q15043-

3, respectively [208]). Primers 1 and 2 (table 2.11) span the first exon junction, and were designed to amplify all isoforms of hZIP14 present, as this region of the mRNA is shared by all isoforms (figure 3.3).

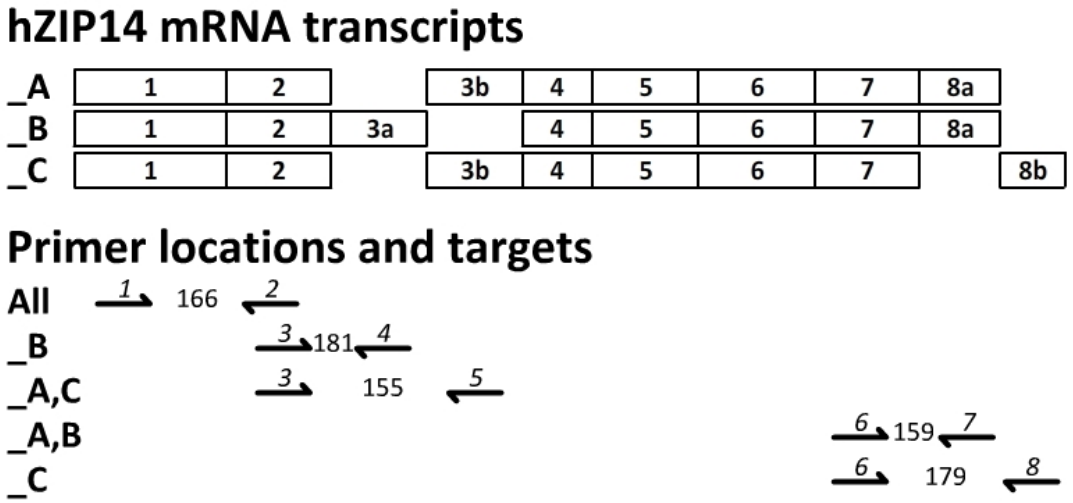


Figure 3.3: **Positions of primers used for PCR on hZIP14 cDNA transcripts.** White numbered boxes represent protein-coding exons (exons not coding for protein are excluded from this diagram). The numbers above the primers correspond to those in table 2.11. The numbers between primer pairs are the sizes of the PCR products, in base pairs.

The arrangement of exons in hZIP14_A mRNA is such that it contains no regions unique to this isoform, with a 5' end identical to hZIP14_C and a 3' end identical to hZIP14_B (figure 3.3). Primer 3 binds to protein-coding exon 2, present in all hZIP14 isoforms. When paired with primer 4, which binds exon 3a, an exon that is only found in hZIP14_B, only hZIP14_B mRNA is amplified. When paired with primer 5, both hZIP14_A and C mRNA are amplified, but not hZIP14_B mRNA.

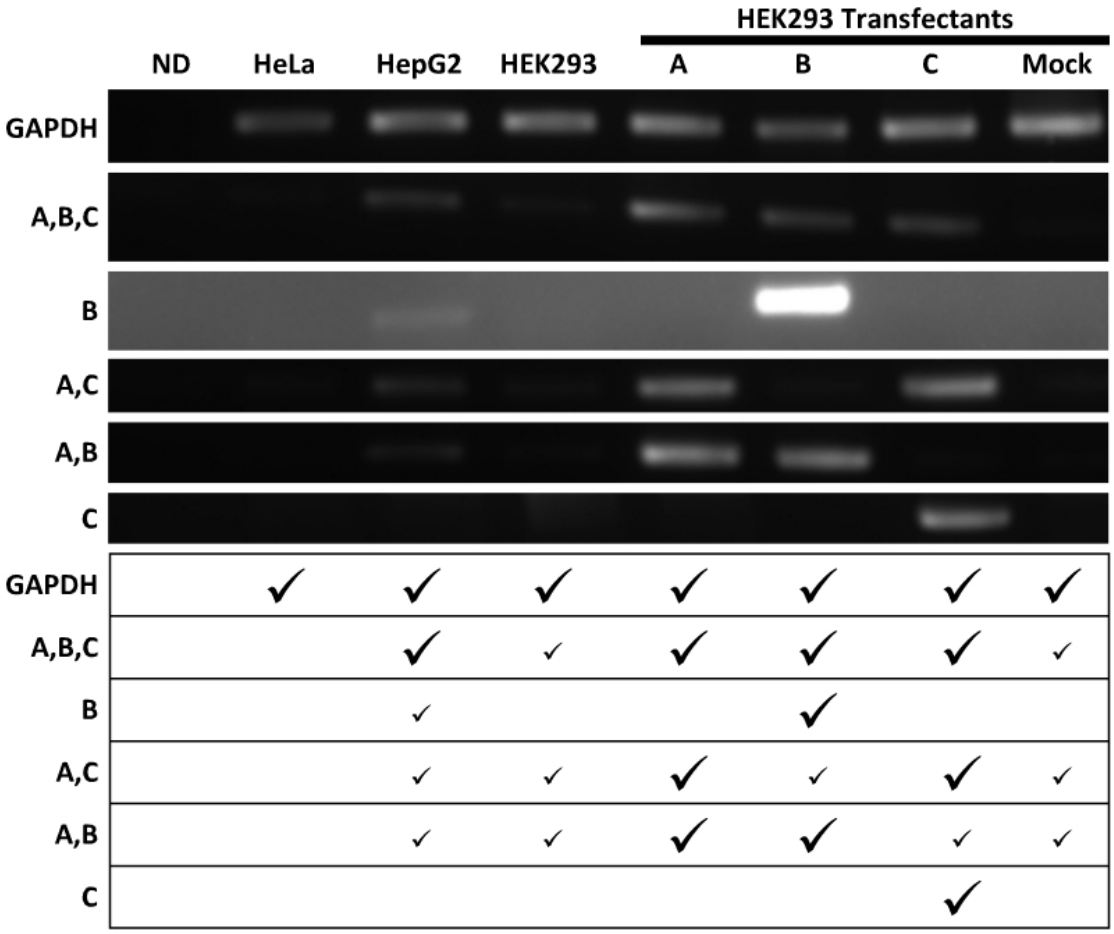
Likewise, primer 6 binds to all hZIP14 isoforms, and can amplify hZIP14_A and B mRNA when paired with primer 7, but only hZIP14_C mRNA when paired with primer 8. Testing cDNA against all the primer pairs depicted in figure 3.3 makes it possible to determine whether cells are expressing hZIP14_A alone, or hZIP14_A with one other isoform. However, if cells express both hZIP14_B and C, then a primer pair amplifying a region from protein-coding exon 3b to protein-

coding exon 8a would be needed to confirm the expression of hZIP14_A.

Glyceraldehyde-3-phosphate dehydrogenase (GAPDH) was chosen as a house-keeping gene, expressed in all three cell lines tested. Although the levels of GAPDH expression may vary between the three cell lines, positive results would confirm the successful synthesis of cDNA in each sample. HEK293 cells transfected with pIRESneo2 plasmids encoding hZIP14_A, B or C (section 3.2.2) were used as positive controls, and to test the ability of the hZIP14 primers to distinguish between the hZIP14 isoforms.

Agarose gel electrophoresis of the HepG2 RT-PCR products gave positive results for all the hZIP14 primer pairs except the primer pair specific for hZIP14_C, showing that the HepG2 cell line expresses hZIP14_A and B. For the HEK293 RT-PCR products, very faint bands were observed for all primer pairs designed to amplify hZIP14_A, but not the primers specifically targeting hZIP14_B or C, and therefore expresses hZIP14_A. HeLa RT-PCR products showed no visible bands for any of the hZIP14 primer pairs, with the positive result for the GAPDH primers showing that HeLa cDNA had been successfully synthesised (figure 3.4).

Real-time quantitative PCR (RT-qPCR) would provide a more sensitive analysis of hZIP14 expression in these cell lines, but requires primers with particular characteristics in order to be effective. The high similarity between the DNA sequences of the two exons by which hZIP14_A and B differ (protein-coding exons 3b and 3a, respectively) (figure 3.5) greatly restricts the ability of primers to distinguish between these two isoforms, and it was not possible to design primer pairs suitable for RT-qPCR for all isoforms. However, the primary concern was identifying levels of endogenous hZIP14 expression that could interfere with characterisation of the mutants. Only HepG2 hZIP14 cDNA was detected at levels similar to that of the transiently transfected controls. On the basis of



	ND	HeLa	HepG2	HEK293	A	B	C	Mock
GAPDH		✓	✓	✓	✓	✓	✓	✓
A,B,C			✓	✓	✓	✓	✓	✓
B			✓			✓		
A,C			✓	✓	✓	✓	✓	✓
A,B			✓	✓	✓	✓	✓	✓
C							✓	

Figure 3.4: **Reverse-transcriptase PCR for endogenous ZIP14 expression.** RNA was extracted from cells, the mRNA used to synthesize cDNA, and regions specific to GAPDH and hZIP14 isoforms were amplified by PCR. ND refers to a no-cDNA negative control; Mock refers to mock-tranfected HEK293 cells. Agarose gel electrophoresis of PCR products are shown, exposure was adjusted for detection of faint bands. The table indicates presence or absence of bands, small ticks indicate faint bands detected by ImageJ analysis [184] of the images or high exposure of the gel.

these results, the HeLa cell line was selected for the further study of hZIP14. As hZIP14_A is more widely expressed in humans than hZIP14_B [19], and there is no experimental evidence of hZIP14_C expression, the majority of the following research was conducted on hZIP14_A.

Figure 3.5: **Alignment of hZIP14 protein-coding exons 3a and 3b.** hZIP14 protein-coding exons 3b and 3a contribute to hZIP14_A and B, respectively. The nucleotides highlighted in blue mark differences between the two exons; the underlined sequences show the positions of the primers used to distinguish between them.

3.2.2 Creation of Plasmids Encoding Wild-Type and Disease-Causing Mutations of hZIP14-HA

pIRESneo2 plasmids encoding wild-type hZIP14-HA, isoforms A, B and C, were supplied by Dr Bowers, as well as the F98V and N469K mutations in the hZIP14_A isoform. Site-directed mutagenesis was used to introduce the F98V and N469K mutations into the hZIP14_B isoform, and the S104I, R128W, G383R and L441R mutations into both A and B isoforms. Plasmids were sequenced to confirm the desired mutations, and to discard plasmids with unwanted mutations.

3.2.3 hZIP14 and Disease-Causing Mutants Can Be Expressed in HeLa Cells

As point mutations have the potential to disrupt folding to the extent that mutants are rapidly targeted for degradation, SDS-PAGE analysis was used to check for successful expression of the hZIP14 mutants [209, 210]. HeLa cells were transiently transfected with plasmids encoding hZIP14_A with a 3' hemagglutinin

tag (hZIP14_A-HA), wild-type and the disease-causing mutants. Protein was extracted, subjected to SDS-PAGE and analysed by western blot (figures 3.6 and 3.7).

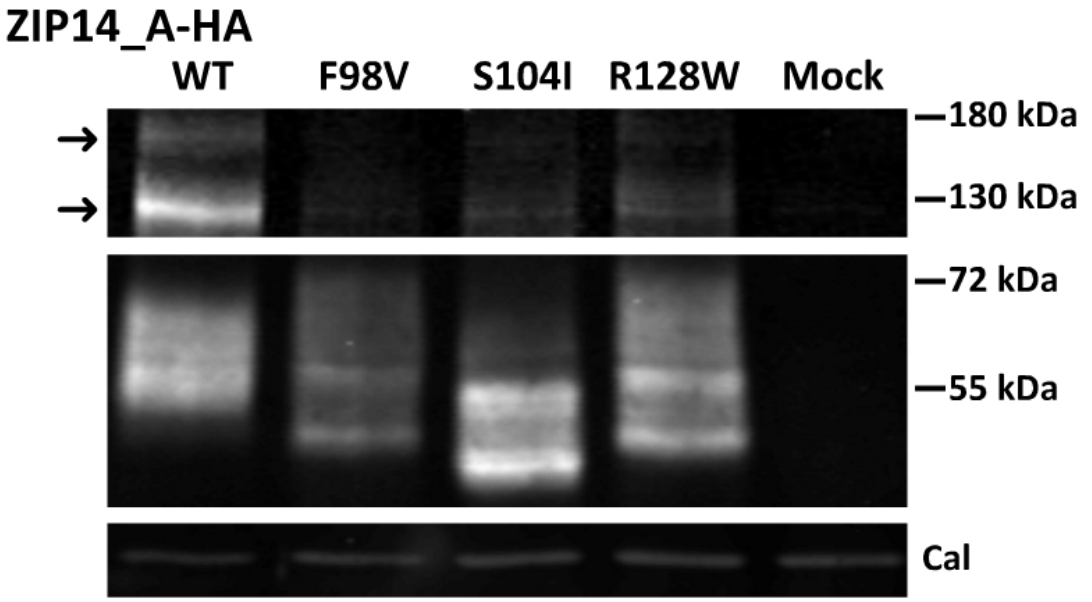


Figure 3.6: **N-terminal disease-causing hZIP14_A-HA mutants were expressed in HeLa cells.** HeLa cells were transiently transfected with plasmids encoding hZIP14_A-HA variants, or mock-transfected. Transfected cells were lysed after 48 hours and the lysates (25 µg of protein) were analysed by SDS-PAGE (8% polyacrylamide) and western blotting, for hZIP14_A-HA (top two panels) and calreticulin (bottom panel). Arrows indicate the positions of higher molecular mass bands seen.

During optimization, the heat denaturation step typically used in western blotting procedures (95 ℃ for 5 minutes) was removed, as it led to most of the hZIP14_A-HA protein appearing in high molecular mass aggregates, as has been described for some membrane proteins [211, 212]. Instead, lysates were thawed on ice, and samples in loading buffer were allowed to reach room temperature before loading onto the acrylamide gel, preventing the formation of aggregates.

After the removal of the 30-nucleotide signal sequence, the expected size of

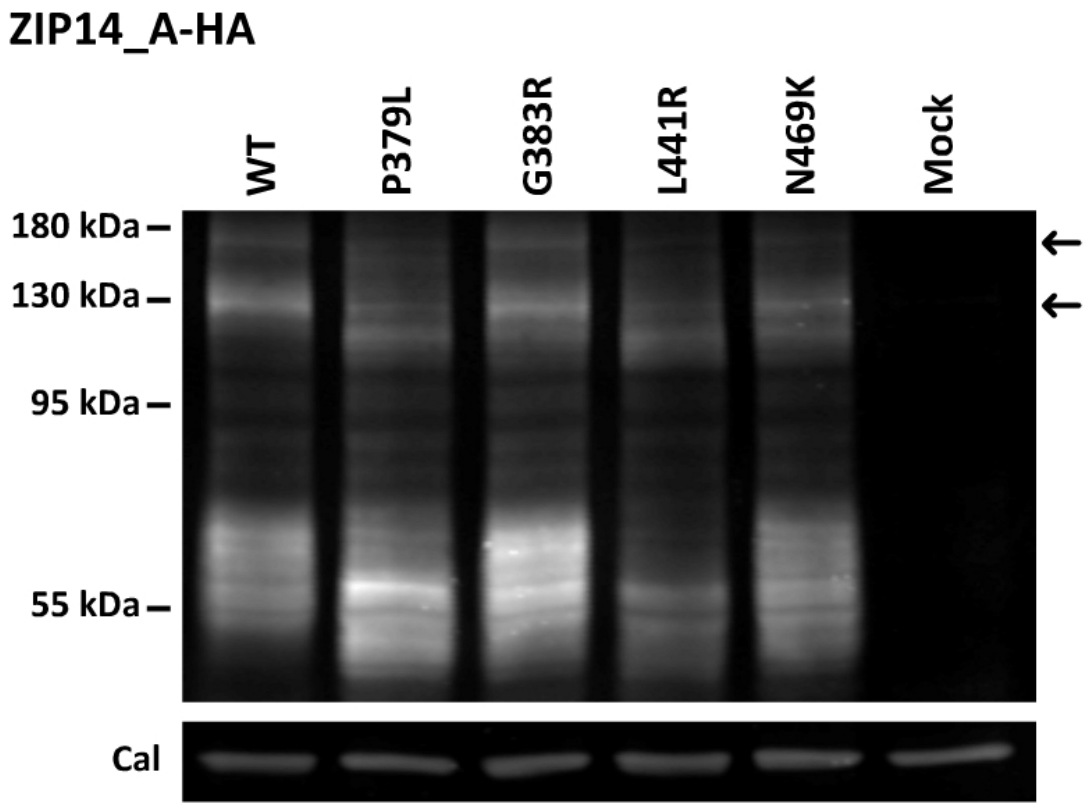


Figure 3.7: **C-terminal disease-causing hZIP14_A-HA mutants were expressed in HeLa cells.** HeLa cells were transiently transfected with plasmids encoding hZIP14_A-HA variants, or mock-transfected. Transfected cells were lysed after 48 hours and the lysates (25 µg of protein) were analysed by SDS-PAGE (8% polyacrylamide) and western blotting, for hZIP14_A-HA (top two panels) and calreticulin (bottom panel). Arrows indicate the positions of higher molecular mass bands seen.

hZIP14-HA is approximately 52 kDa; glycosylation of the N-terminus should also result in spreading of this band into higher molecular masses. Figure 3.6 shows the western blot of the N-terminal disease-causing mutants.

Multiple bands of around 55 kDa could be observed for each mutant, likely reflecting hZIP14_A-HA in different stages of glycosylation, though the apparent size ranges of the bands differed. The S104I bands had the lowest range of molecular masses, probably due to the disruption of the N-linked glycosylation site, NFS (102-104) , causing a decrease in size with the loss of glycosylation

at N102 [55]. The size-ranges of the ∼55 kDa bands for the mutants all differed from that of the wild-type. 'Gel-shifts' caused by amino acid substitutions in transmembrane proteins have been documented [213, 214], and are not easily interpreted.

Two higher molecular mass bands were typically observed for wild-type hZIP14_A-HA, at ∼130 and ∼180 kDa (figure 3.6), which could represent hZIP14 which is not fully denatured, possibly in oligomeric forms or interacting with other proteins. The upper of these two bands was not seen for the N-terminal mutants, while the lower was fainter or absent. Western blotting confirmed that the N-terminal disease-causing mutants of hZIP14_A-HA could be expressed in HeLa cells.

Figure 3.7 shows the western blot of the C-terminal disease-causing mutants. Bands were observed for all mutants at approximately 55 kDa, and the higher molecular mass bands at ∼130 and 180 kDa could also typically be seen, though they did slightly differ from those seen in the wild-type hZIP14_A-HA protein (figure 3.7). The upper band was fainter in the mutants (excepting G383R) and an additional band appears just below the ∼130 kDa band, for each of the C-terminal mutants.

3.2.4 hZIP14_A-HA Disease-Causing Mutations Impact Iron Uptake Function

3.2.4.1 Creation of Stable Monoclonal Cell Lines Expressing Wild-type hZIP14_A-HA and Disease-causing Mutants

In order to measure the iron uptake abilities of hZIP14 disease-causing mutants, stable monoclonal cell lines expressing hZIP14_A-HA were developed

as described in section 2.1.2.1, for wild-type hZIP14_A-HA, and the R128W, P379L, G383R and N469K mutants.

Regarding the N-terminus of hZIP14, little was known about its potential function, and all documented N-terminal mutations were associated with manganism. The N-terminal mutant most similar to wild-type by western blotting, R128W, was chosen as an example of the group (figure 3.6).

Regarding the C-terminal hZIP14 mutations, symptoms varied - although most had been linked to manganism, P379L had also been associated with iron deficiency anaemia [71], while L441R caused hyperostosis cranialis interna [95], and some were predicted to be in the metal translocation pore [13]. Each C-terminal mutant was of interest, though attempts to stably transfect HeLa cells with L441R failed, as the protein seemed to be toxic.

Total hZIP14_A-HA expression, as assessed by flow cytometry, found expression of the wild-type hZIP14_A-HA protein to be at slightly higher levels than that of the mutants, but all cell lines expressed their hZIP14_A-HA variant well enough to be clearly distinguished from the background fluorescence of the negative control (HeLa) (figure 3.8).

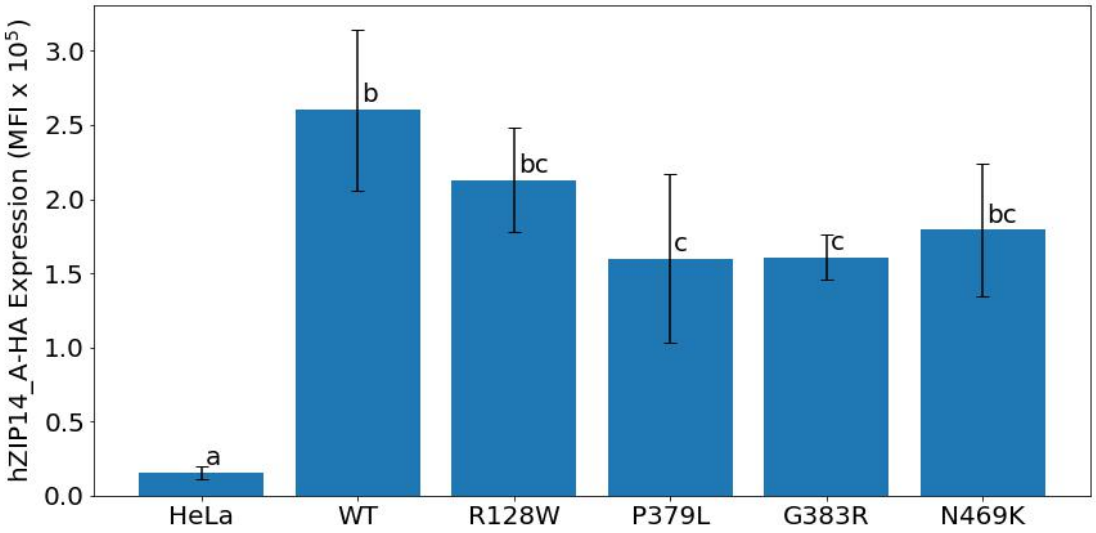


Figure 3.8: **Expression of hZIP14_A-HA by monoclonal stable cell lines.** HeLa cells were transfected with pIRESneo2 plasmids encoding hZIP14_A-HA variants, or mock transfected. Cells were treated with G418 to select for cells that had incorporated neomycin resistance, and monoclonal cell lines were cultivated from the survivors. Samples from cell lines were permeabilised, fixed, and then stained for the HA tag for analysis by flow cytometry. For HeLa and WT, n=5; for mutants, n=3. Error bars represent the standard deviation; different letters represent means that are significantly different from each other (one-way ANOVA with Tukey-Kramer post-hoc analysis, $P < 0.05$).

3.2.4.2 Iron Uptake by HeLa Cells Expressing hZIP14_A-HA or Mutants

To measure iron uptake, cells were incubated in a growth medium containing the radioactive isotope, ^{55}Fe, for a fixed amount of time. Cells were then washed to remove excess medium containing ^{55}Fe, and lysed in a fixed amount of lysis buffer. This lysate was analysed for protein concentration and radioactivity.

HeLa cells likely express transporters capable of transporting iron, with their own characteristics, such as optimum pH and K_m values. Over-expression of exogenous hZIP14 should alter the characteristics of iron uptake by the stably transfected HeLa cells to more closely match the characteristics of hZIP14 function. Optimisation of this experiment involved identifying the conditions that

best captured the differences in iron uptake between the HeLa and hZIP14_A-HA-expressing cell lines.

Figure 3.9 shows the effect of varying concentrations of cold iron ($FeSO_4$) on uptake of ^{55}Fe. For both cell lines, as the concentration of $FeSO_4$ increased, there was an initial increase in ^{55}Fe uptake, followed by a gradual decrease. The $FeSO_4$ concentration at which peak ^{55}Fe uptake occurred differed between the two cell lines, most likely due to the differing properties of the iron transporters in the two cell lines. The higher rate of ^{55}Fe uptake in the hZIP14_A-HA-expressing cell line reflects the increased number of iron transporters present in that cell line. The gradual decrease in ^{55}Fe uptake at higher $FeSO_4$ concentrations is caused by dilution of the ^{55}Fe with cold iron. Based on these results, 10 µM was selected as the optimal $FeSO_4$ concentration for measuring hZIP14_A-HA-mediated iron uptake.

The ability of the cell lines expressing disease-causing hZIP14_A-HA mutants to transport iron was decreased relative to the wild-type protein (figure 3.10), though the decrease did not reach statistical significance for the N469K mutant. The R128W and G383R mutants appeared to be non-functional. The P379L cell line had very little iron-uptake ability; though not statistically significant, the P379L iron uptake was always slightly higher than that of the HeLa cells and the R128W and G383R cell lines. The cell line expressing the N469K mutant was able to take up iron, albeit slightly less than the wild-type hZIP14_A-HA cell line. With the exception of N469K, the differences in iron uptake were not mirrored by differences in expression of the hZIP14_A-HA variants (figure 3.8); P379L was expressed at $\sim$60% of the level of wild-type hZIP14_A-HA, but average iron uptake by that cell line was $\sim$25% of that of the wild-type hZIP14_A-HA cell line.

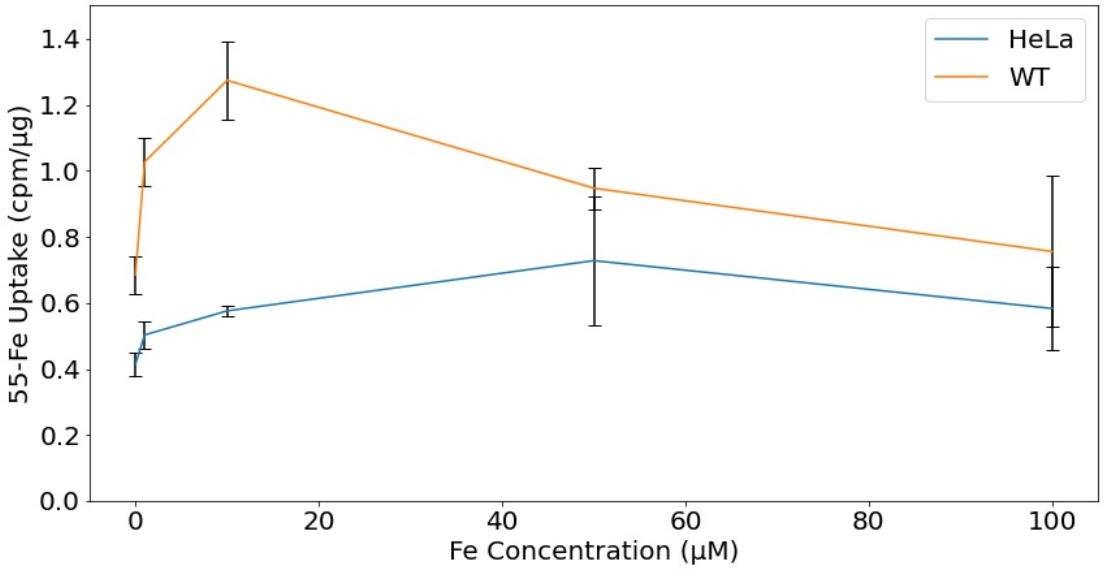


Figure 3.9: **Optimum concentration of $FeSO_4$ for measuring ^{55}Fe uptake.** HeLa and a HeLa cell line expressing wild-type (WT) hZIP14_A-HA were incubated in varying concentrations of $FeSO_4$ labelled with a fixed amount of ^{55}Fe. ^{55}Fe uptake was measured by liquid scintillation counting, and protein concentration by BCA assay. Data points represent the average of triplicates, the error bars represent the standard deviation. The optimum $FeSO_4$ concentration chosen for the iron uptake assay was 10 µM $FeSO_4$.

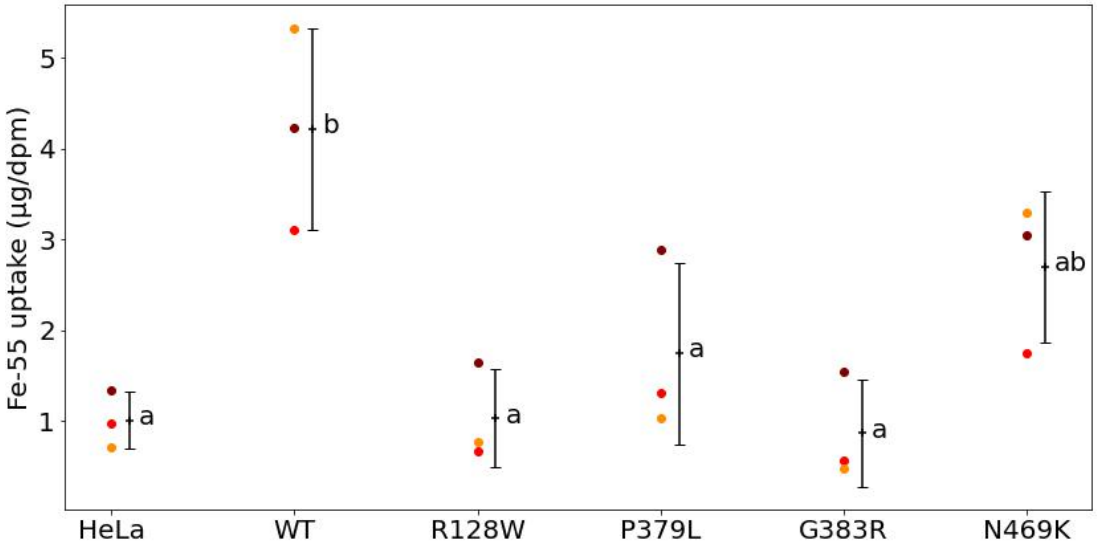


Figure 3.10: **Disease-causing mutations decrease ability to transport iron.** HeLa and HeLa cell lines expressing hZIP14-HA, wild-type and mutants, were incubated in 10 µM $FeSO_4$ labelled with ^{55}Fe. ^{55}Fe uptake was measured by liquid scintillation counting, and protein concentration by BCA assay. Data points represent averages of triplicates, coloured points represent three separate experiments.Error bars represent the standard deviation; different letters represent means that are significantly different from each other (one-way ANOVA with Tukey-Kramer post-hoc analysis, $P < 0.05$).

Differences in Iron Uptake Were Not Caused By Differences in Surface Levels of Mutants

Alterations in uptake of iron by the disease-causing mutations could be caused by impairments in iron uptake function, or impairments in trafficking of the mutants to the cell surface.

In order to determine whether decreased uptake of iron by the hZIP14_A-HA mutants was caused by decreased amounts of protein on the plasma membrane, cell surface levels of the hZIP14_A-HA variants were measured by flow cytometry. Relative iron uptake and relative surface levels of hZIP14_A-HA variants (relative to the HeLa and wild-type hZIP14_A-HA-cell lines) are shown in figure 3.11.

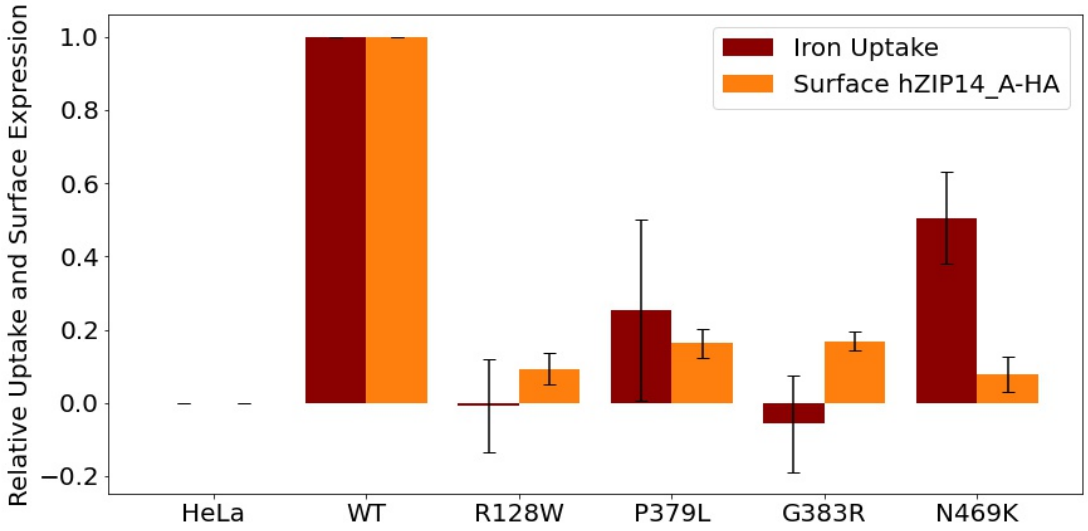


Figure 3.11: **Relative iron uptake compared to relative surface levels of hZIP14_A-HA disease-causing mutants.** HeLa cells and monoclonal stable cell lines were stained for the HA tag and analysed by flow cytometry. Average uptake of ^{55}Fe and surface levels of hZIP14_A-HA variants were normalized relative to HeLa at 0 and WT at 1, for each experiment.

Surface levels of all mutants were at approximately 10-20% of the wild-type hZIP14_A-HA, but iron uptake varied greatly between mutants. The R128W and G383R mutants failed to take up iron despite their presence at the surface of

the cell. The P379L mutant took up more iron than the G383R mutant, despite similar levels of cell surface expression. Relative to the wild-type hZIP14_A-HA cell line, the decreased iron uptake by P379L could be a result of lower amounts of protein at the cell surface. Surprisingly, the N469K mutant took up more iron than would be expected from the relatively low amounts detected at the cell surface. Comparing the relative iron uptake function to the cell surface levels of the mutants suggests that the R128W and G383R mutations render hZIP14 non-functional, while the N469K mutation enhances the iron-uptake function of hZIP14. The P379L mutation may not have any effect on iron-uptake function, for the proportion of the protein that has reached the plasma membrane.

A comparison of total levels of expression to surface levels shows that relative surface levels are also greatly decreased compared to wild-type (figure 3.12). This effect is most pronounced for R128W and N469K, whose cell lines have the highest expression of their respective mutants, and the lowest levels of the mutants at the surface.

Taken together, it is clear that the disease-causing mutations tested affect hZIP14_A-HA in different ways. Effects on iron uptake, both positive and negative, were observed, while trafficking to the surface was negatively affected for all mutations tested.

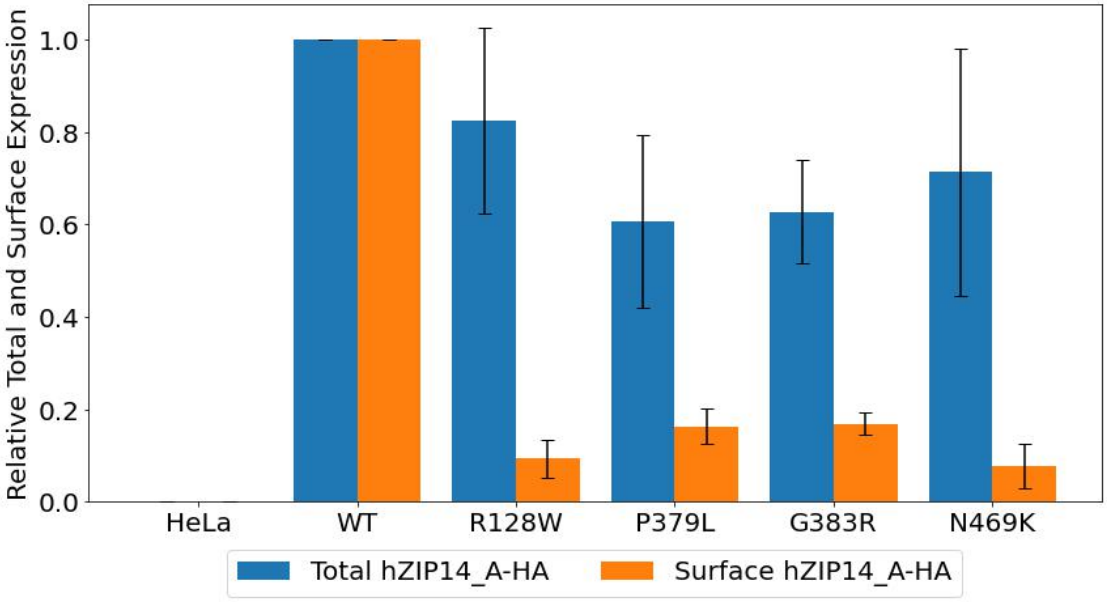


Figure 3.12: **Comparison of relative total and surface levels of hZIP14_A-HA disease-causing mutants.** Surface and total levels of hZIP14_A-HA variants were determined by staining for the HA tag in non-permeabilised and permeabilised cells, respectively. Expression levels were normalized relative to HeLa at 0 and WT at 1, for each experiment.

3.2.5 Cellular Localisation of hZIP14-HA is Affected by Disease-Causing Mutations

Measurements of protein expression and surface levels of the hZIP14_A-HA disease-causing mutants in stable cell lines illustrated alteration in trafficking for each of the mutants tested (figure 3.12). This raised the question of how the cellular location of hZIP14-HA was affected by the disease-causing mutations.

Immunofluorescence microscopy was used to study the effects of all the disease-causing mutations on the cellular localisation of hZIP14-HA. Transiently transfected HeLa cells were permeabilised and stained for hZIP14-HA. Wild-type hZIP14_A-HA could be seen at the surface of the cell and in intracellular puncta distributed throughout the cytoplasm (figure 3.13). N-terminal mutations had a dramatic effect on the distribution of hZIP14_A-HA. The N-terminal

mutants did not appear to be at the cell surface, the intracellular pattern of distribution also differed from that of the wild-type, resembling staining of the endoplasmic reticulum (ER), with a bright ring surrounding the cell nucleus.

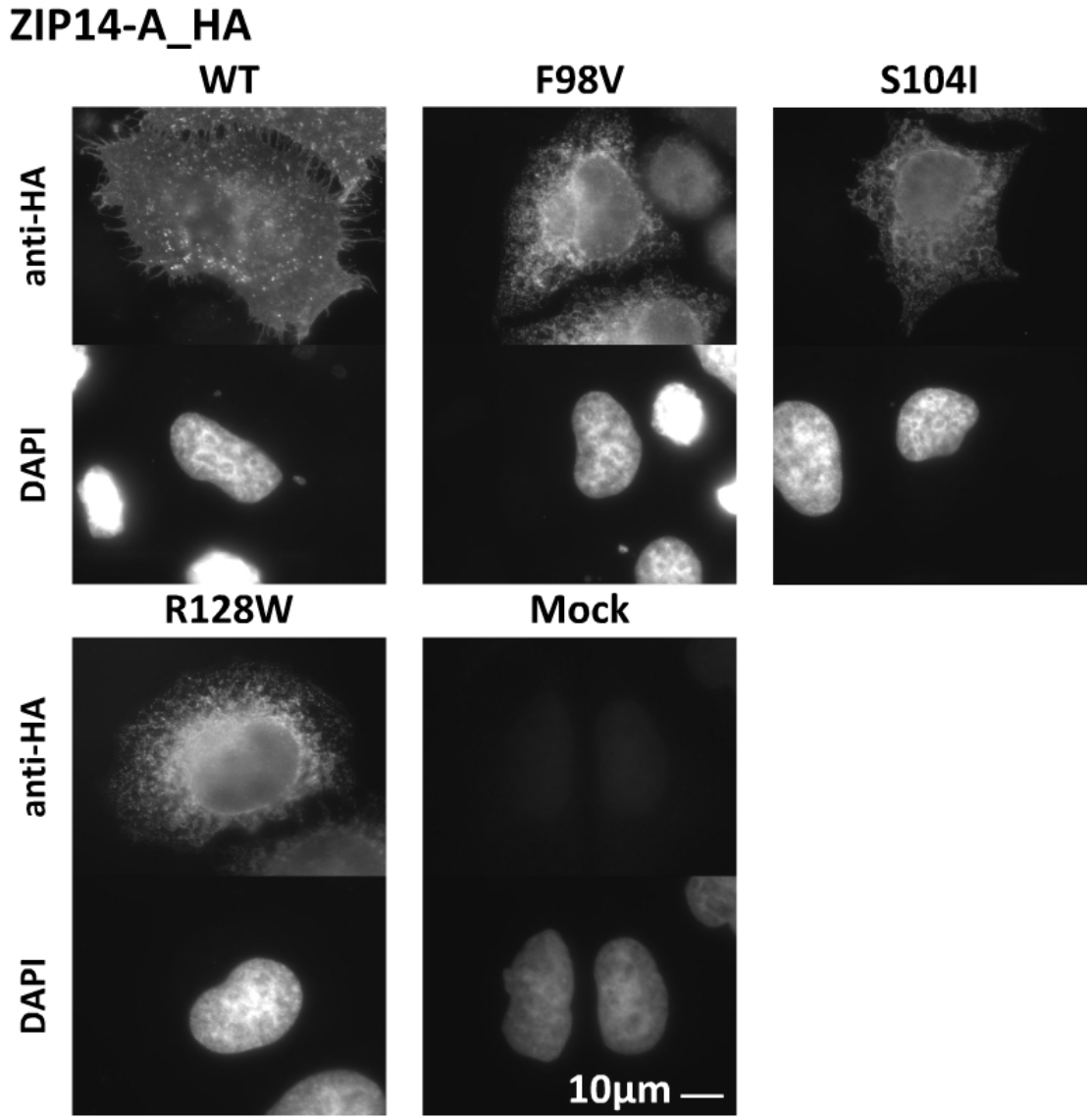


Figure 3.13: **N-terminal hZIP14_A-HA disease-causing mutations alter cellular distribution.** HeLa cells were transiently transfected with plasmids encoding hZIP14_A-HA variants, or mock-transfected. Transfected cells were seeded onto coverslips and stained with mouse anti-HA antibody, followed by anti-mouse Alexa Fluor 488. Coverslips were viewed by immunofluorescence microscopy.

The C-terminal disease-causing mutants P379L and G383R more closely

resembled the wild-type distribution (figure 3.14). They could also be seen at the plasma membrane and in bright intracellular puncta. N469K hZIP14_A-HA typically behaved like the wild-type protein, though there was some variation, further discussed in section 3.2.5.2, while the L441R mutant greatly differed from wild-type hZIP14_A-HA; it was not observed at the plasma membrane, and was distributed throughout the cytoplasm, with a bright ring around the nucleus.

These results illustrate that the disease-causing mutations differ in their impacts on cellular distribution of hZIP14_A-HA. Intracellular locations and cell surface levels are both affected by N-terminal disease-causing mutations, as well as the L441R mutation, and to some extent, the N469K mutation, while the P379L and G383R mutants did not appear to impact cellular distribution of hZIP14_A-HA.

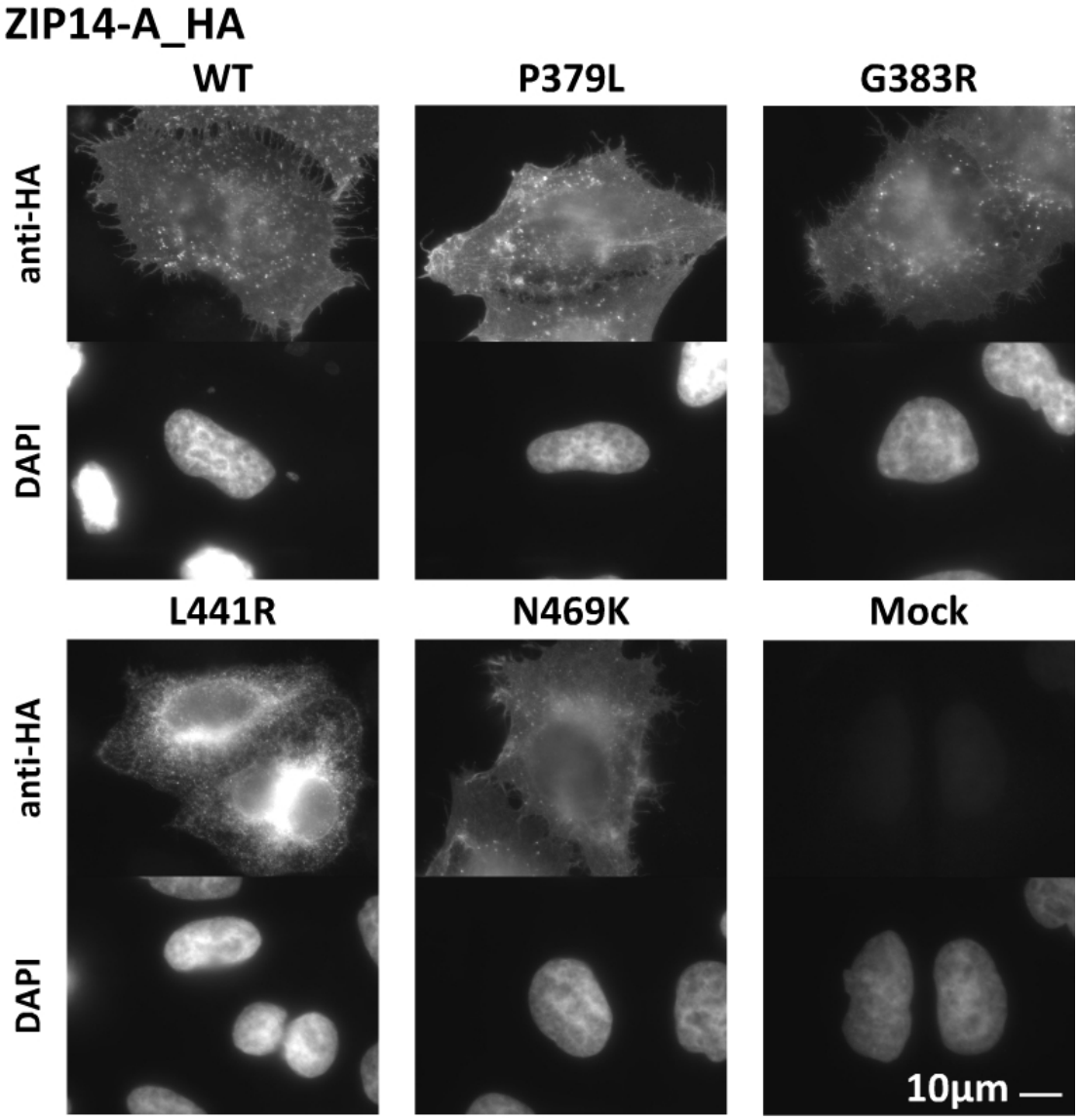


Figure 3.14: **Most C-terminal disease-causing mutations do not affect cellular distribution.** HeLa cells were transiently transfected with plasmids encoding hZIP14_A-HA variants, or mock-transfected. Transfected cells were seeded onto coverslips and stained with mouse anti-HA antibody, followed by anti-mouse Alexa Fluor 488. Coverslips were viewed by immunofluorescence microscopy.

3.2.5.1 Further Study of Cellular Distribution of hZIP14-HA Variants

In order to understand the impacts of the disease-causing mutations on hZIP14_A-HA intracellular localisation in more detail, several approaches were

tested: co-transfection and co-localisation with wild-type hZIP14_A-GFP, and co-localisation with an ER marker protein. The disease-causing mutations in hZIP14_B-HA were also tested to see if there was any difference in effects between isoforms. Wild-type hZIP14_C-HA was also included, of interest because it has an alternative transmembrane helix 8, which is where the N469K mutation is predicted to be located [13].

Co-transfection with hZIP14_A-GFP

HeLa cells were co-transfected with a plasmid encoding wild-type hZIP14_A-GFP, and a plasmid encoding an hZIP14_A-HA variant. Figure 3.15 A shows images of cells transfected with both, neither, or one type of plasmid. These results show that the signal from hZIP14_A-GFP was not detectable in the channel used to detect fluorescently labelled hZIP14_A-HA (stained with Alexa Fluor 546 antibody) and vice versa.

Co-transfection of HeLa cells with hZIP14_A-GFP and variants of hZIP14_A-HA was used to determine whether disease-causing mutants co-localised with the wild-type protein. This approach was tested on the C-terminal disease-causing mutants, as their intracellular distributions have been reported as being identical to that of wild-type hZIP14, including the N469K and L441R mutants [19, 95]. Co-localisation with a tagged wild-type hZIP14 could confirm whether or not wild-type and mutant hZIP14 traffic to identical intracellular locations.

Figure 3.15 B shows that hZIP14_A-GFP and wild-type hZIP14_A-HA had the same distribution pattern, both being observed at the plasma membrane, and occurring at the same intracellular locations. Likewise, the P379L and G383R hZIP14_A-HA mutants co-localised with hZIP14_HA-GFP, confirming a wild-type like distribution. The L441R mutant did not co-localise with hZIP14_A-GFP, and the N469K mutant partially co-localised with hZIP14_A-GFP. In the example

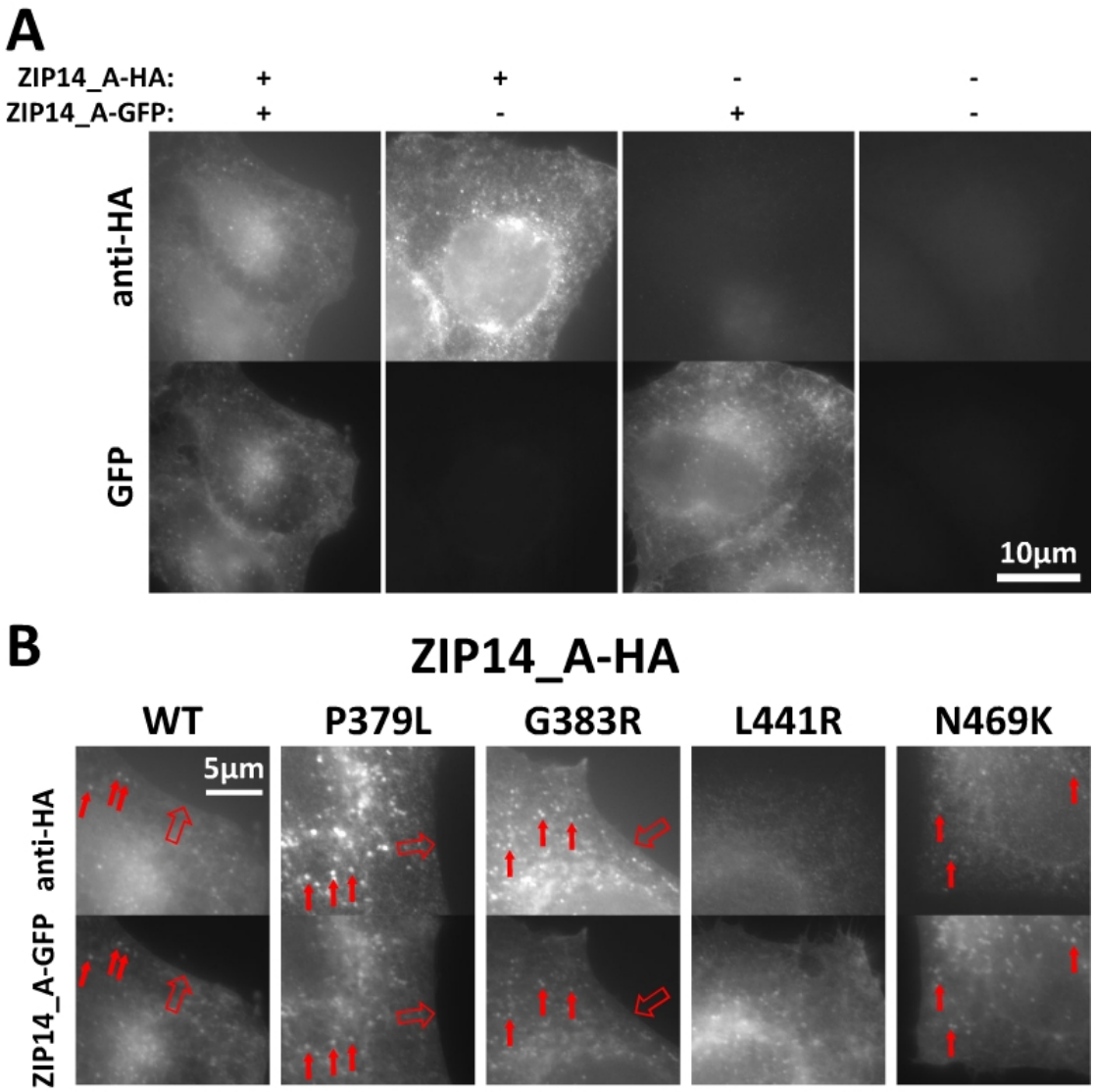


Figure 3.15: **Cellular distribution of hZIP14_A-HA variants and wild-type hZIP14_A-GFP.** A: HeLa cells were transiently transfected with plasmids encoding hZIP14_A-HA, hZIP14_A-GFP, both, or mock-transfected. Transfectants were stained with mouse anti-HA antibody, then anti-mouse Alexa Fluor 546. B: HeLa cells were transiently transfected with plasmids encoding hZIP14_A-HA variants and hZIP14_A-GFP. Transfectants were stained with mouse anti-HA antibody, then anti-mouse Alexa Fluor 546. Coverslips were viewed by immunofluorescence microscopy. Thin red arrows identify intracellular puncta seen in both channels; hollow red arrows identify protein at the plasma membrane seen in both channels.

pictured, intracellular N469K co-localised with hZIP14_A-GFP, but could not be seen at the plasma membrane. The variability of N469K appearance will be

discussed further in section 3.2.5.2.

One problem with this approach was in identifying cells that expressed both GFP and HA-tagged proteins at levels that could be clearly imaged. Cells typically expressed high levels of one plasmid or the other, rarely both. A second problem was the potential for hZIP14_A-GFP to alter the behaviour of the HA-tagged proteins. It is currently unknown why most of the mutants require homozygosity in order for symptoms to be experienced. While it is possible that the activity of one copy of hZIP14 is sufficient to avoid negative effects, another possibility is that wild-type hZIP14 alters the behaviour of mutant hZIP14, perhaps restoring its function. For these reasons, this approach was not pursued further.

Co-localisation with Calreticulin

Calreticulin is a protein that localises to the ER [215], and as such, acts as a marker for this location. A selection of the disease-causing mutants were transiently transfected into HeLa cells, which were stained for both hZIP14-HA and calreticulin, and then viewed by immunofluorescence microscopy (figures 3.16 and 3.17).

This approach was tested on one N-terminal mutant (F98V), one C-terminal mutant whose location appeared identical to wild-type (G383R) and N469K, a C-terminal mutant whose intracellular location seemed to differ from that of wild-type hZIP14-HA. An additional point of interest is that these three mutations have been previously reported to have an identical intracellular distribution to wild-type hZIP14-HA [19]. Mutants were examined in both isoforms A and B, to look for any differences in effect, and wild-type hZIP14_C-HA was also studied, to compare it to wild-type A and B isoforms.

Neither the A nor B isoforms of wild-type hZIP14-HA co-localised with

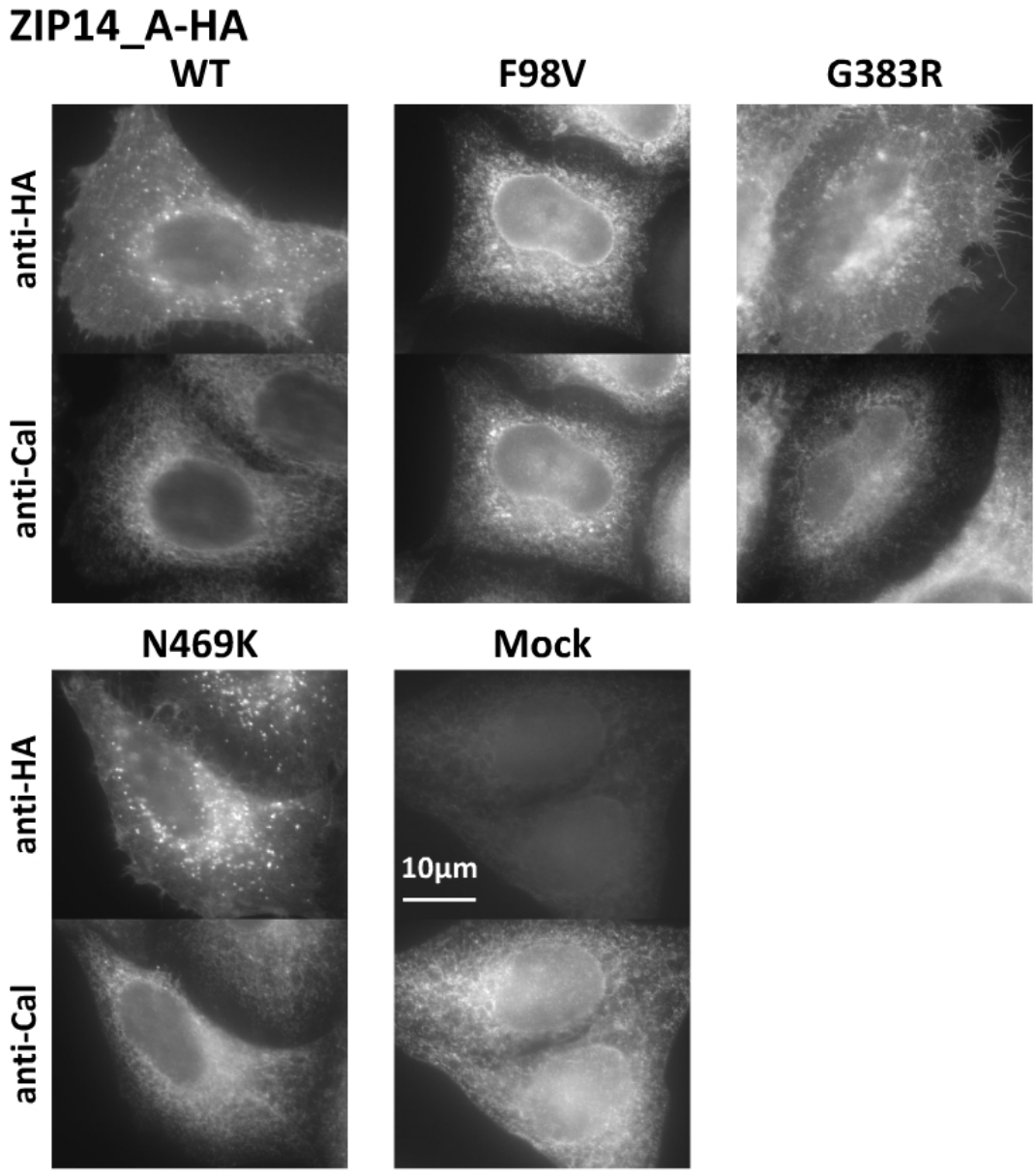


Figure 3.16: **hZIP14_A-HA F98V co-localises with calreticulin.** HeLa cells were transiently transfected with plasmids encoding hZIP14_A-HA variants, or mock-transfected. Transfectants were seeded onto coverslips and stained with mouse anti-HA and rabbit anti-calreticulin antibody, followed by anti-mouse Alexa Fluor 488 and anti-rabbit Alexa Fluor 546. Coverslips were viewed by immunofluorescence microscopy.

calreticulin (figures 3.16 and 3.17), while the C isoform strongly co-localised with calreticulin (figure 3.17). The F98V mutants also showed strong co-localisation with calreticulin (see figure 3.18 for a close-up comparison), and

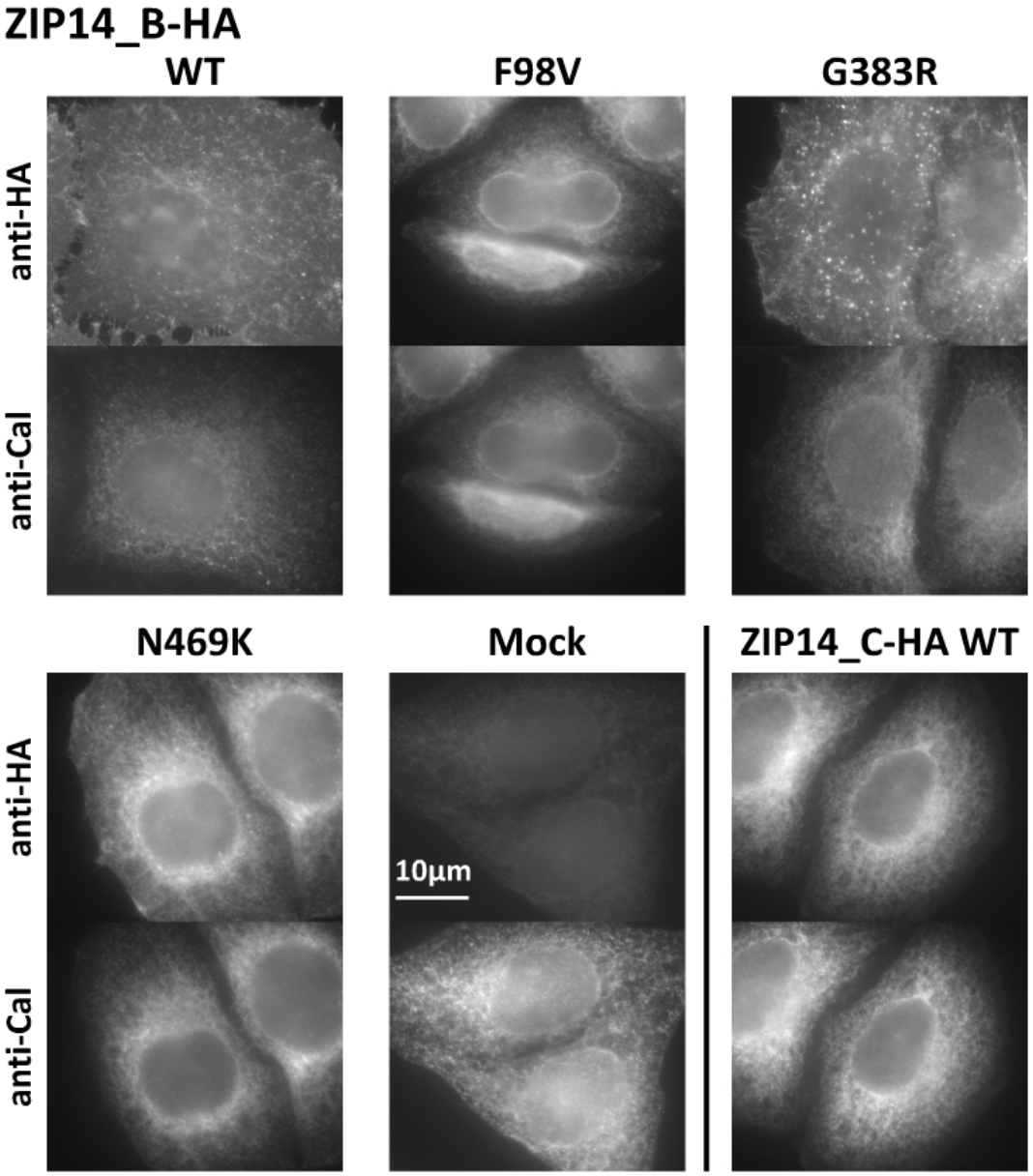


Figure 3.17: **hZIP14_B-HA F98V and hZIP14_C-HA co-localise with calreticulin.** HeLa cells were transiently transfected with plasmids encoding hZIP14_A-HA variants, or mock-transfected. Transfectants were seeded onto coverslips and stained with mouse anti-HA and rabbit anti-calreticulin antibody, followed by anti-mouse Alexa Fluor 488 and anti-rabbit Alexa Fluor 546. Coverslips were viewed by immunofluorescence microscopy.

the G383R mutants, like wild-type hZIP14-HA A and B, did not co-localise with calreticulin. The N469K mutants varied; in some cells, N469K behaved like wild-type hZIP14-HA, in others, it co-localised with calreticulin, to varying

degrees. This variability will be discussed further in section 3.2.5.2. There was no discernible difference in behaviour between the A and B isoforms of hZIP14-HA, for the disease-causing mutants or the wild-type protein.

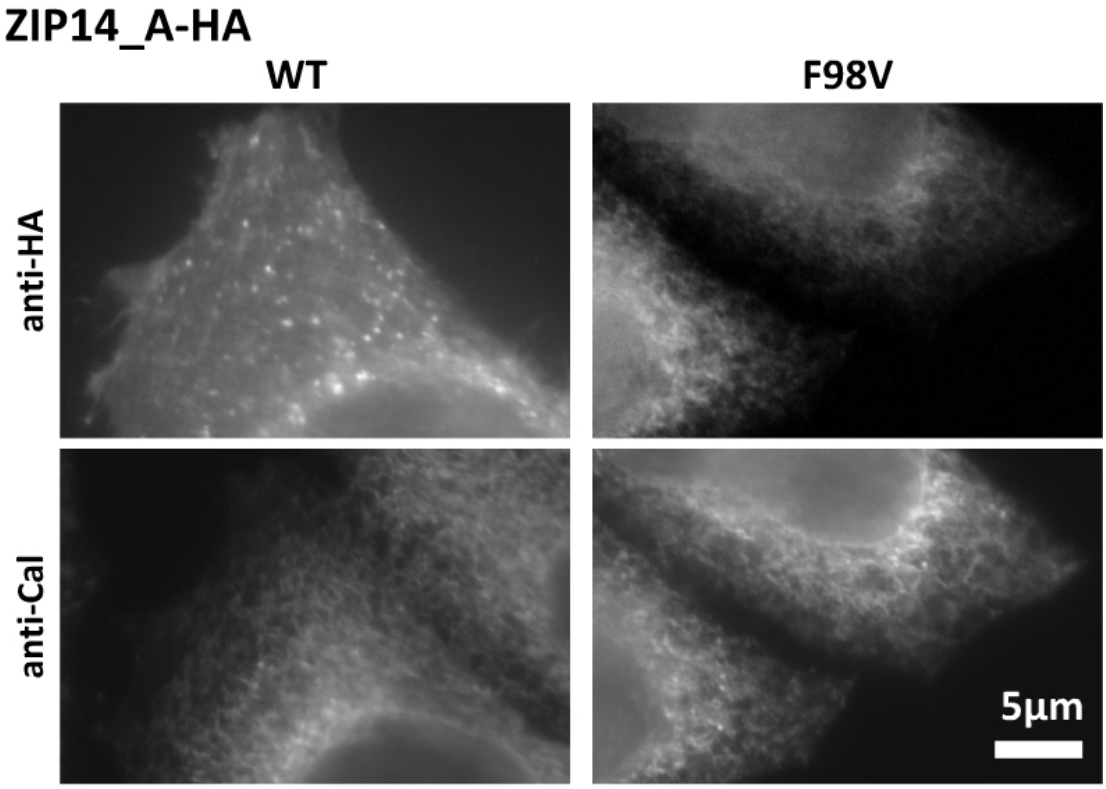


Figure 3.18: **Close-up comparison of calreticulin distribution with wild-type hZIP14_A-HA and the F98V mutant.** HeLa cells were transiently transfected with plasmids encoding wild-type hZIP14_A-HA and the F98V mutant. Transfectants were seeded onto coverslips and stained with mouse anti-HA and rabbit anti-calreticulin antibody, followed by anti-mouse Alexa Fluor 488 and anti-rabbit Alexa Fluor 546. Coverslips were viewed by immunofluorescence microscopy.

Confocal Microscopy

Co-localisation with calreticulin by standard immunofluorescence microscopy showed potential for clarifying one key difference between the wild-type hZIP14-HA and some of the mutants. Confocal microscopy is the preferred method for studying co-localisation, as the use of cross-sections (rather than imaging

the cell in its entirety) decreases the likelihood of proteins appearing to co-localise due to identical positions in the vertical plane. HeLa cells were transiently transfected, co-stained for calreticulin and hZIP14-HA, and studied using confocal microscopy. Images of separate channels and their composites are shown in figures 3.19, 3.20 and 3.21.

The composite images were in agreement with the results from standard immunofluorescence microscopy. Overlap between the F98V mutants and calreticulin was very strong, likewise for hZIP14_C-HA and calreticulin. The N469K hZIP14_A-HA example chosen did co-localise with calreticulin, while the N469K hZIP14_B-HA example did not. The variation in behaviour of cells transfected with N469K will be discussed further in section 3.2.5.2.

The example images shown in figures 3.19, 3.20 and 3.21 were used to quantify the co-localisation observed by calculating the Pearson's correlation coefficient (figure 3.22). The 2D histograms in figure 3.22 are density plots showing the correlation between the signals of each channel at each pixel of the image. A positive gradient indicates a positive correlation, namely, that the two proteins co-localise. A negative gradient indicates the opposite, that the two proteins are found in different locations. The steepness of the gradient is not useful, as a steep or shallow gradient relates to the relative brightness of the two channels. The Pearson's correlation coefficient indicates the strength of the correlation; a value of 1 means that the brightness, at any given point, in one channel perfectly predicts the brightness in the other channel, at that point, meaning that the two proteins are perfectly co-localised. To distinguish between chance overlap caused by high expression in one, or both, channels, it is useful to rotate one of the images and re-calculate co-localisation between the two channels. Quantification of the co-localisation captured in the example images of figures 3.19, 3.20 and 3.21 resulted in large differences between the

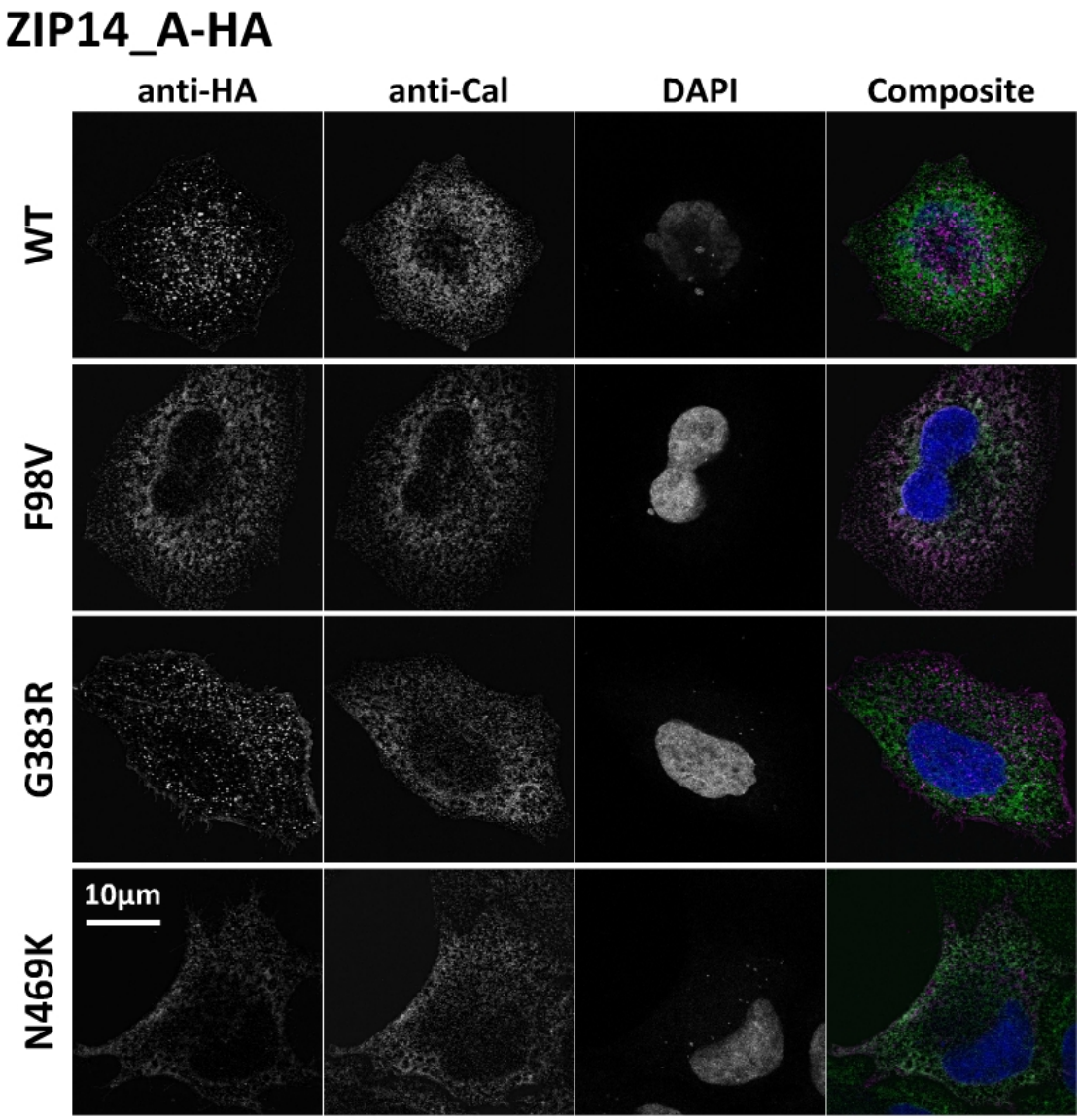


Figure 3.19: **hZIP14_A-HA F98V co-localises with calreticulin.** HeLa cells were transiently transfected with plasmids encoding hZIP14_A-HA variants, or mock-transfected. Transfectants were seeded onto coverslips and stained with mouse anti-HA and rabbit anti-calreticulin antibody, followed by anti-mouse Alexa Fluor 488 and anti-rabbit Alexa Fluor 546. Cells were viewed using confocal microscopy, and composite images produced following processing and analysis with ImageJ [184]. (The mock-transfected sample ('Mock') is pictured in figure 3.21.)

correlation coefficients of hZIP14-HA variants that did, or did not co-localise with calreticulin. For proteins that were not observed to co-localise with calreticulin, they typically had correlation coefficients that were very low, or even negative.

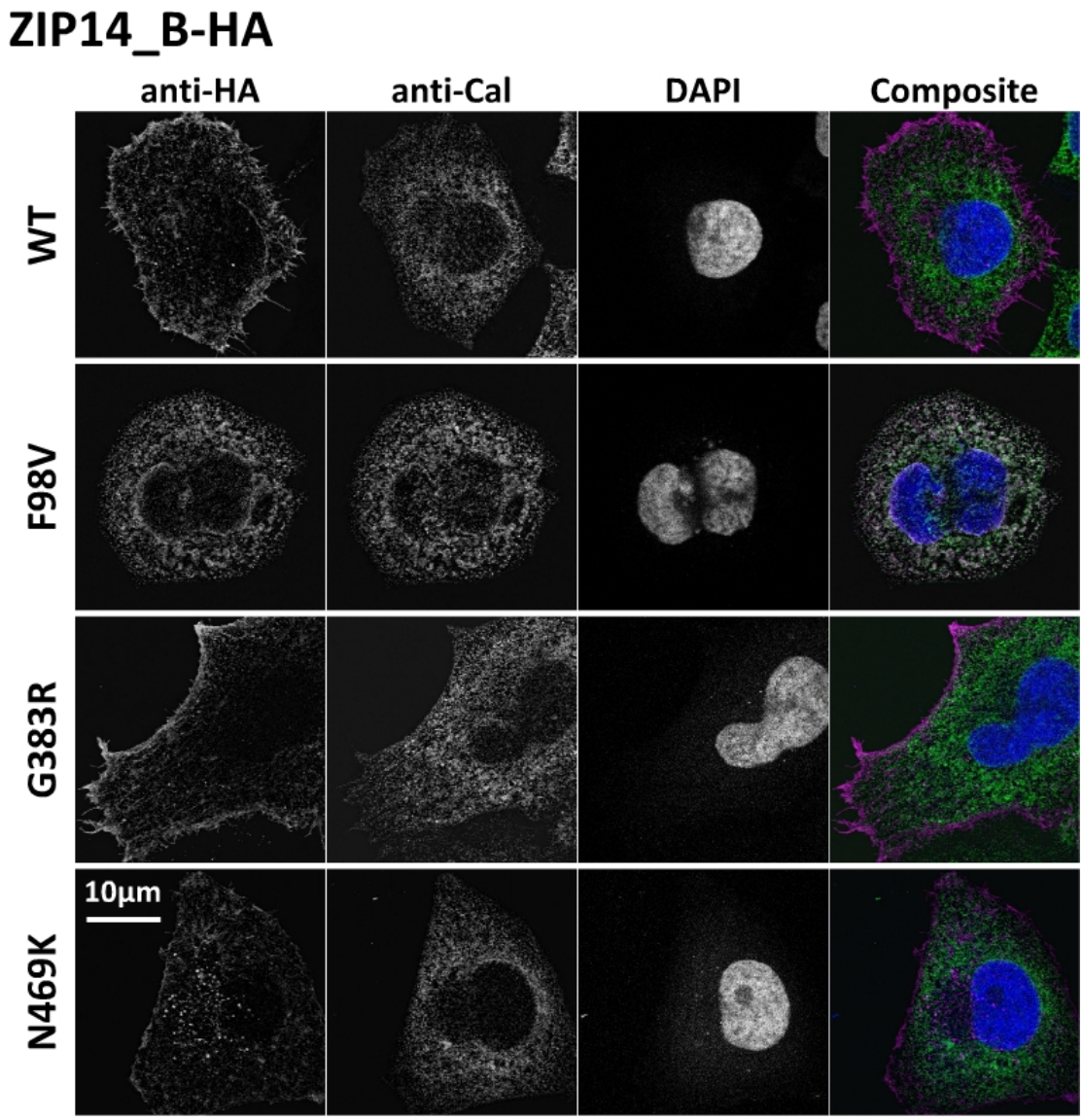


Figure 3.20: **hZIP14_B-HA F98V co-localises with calreticulin.** HeLa cells were transiently transfected with plasmids encoding hZIP14_A-HA variants, or mock-transfected. Transfectants were seeded onto coverslips and stained with mouse anti-HA and rabbit anti-calreticulin antibody, followed by anti-mouse Alexa Fluor 488 and anti-rabbit Alexa Fluor 546. Cells were viewed using confocal microscopy, and composite images produced following processing and analysis with ImageJ [184]. (The mock-transfected sample ('Mock') is pictured in figure 3.21.)

The correlation coefficients of the F98V mutants and wild-type hZIP14_C-HA were very high, ranging from $\sim$0.5 to 0.7. The correlation coefficients also captured the difference in co-localisation observed between the examples of

N469K hZIP14-HA A and B isoforms.

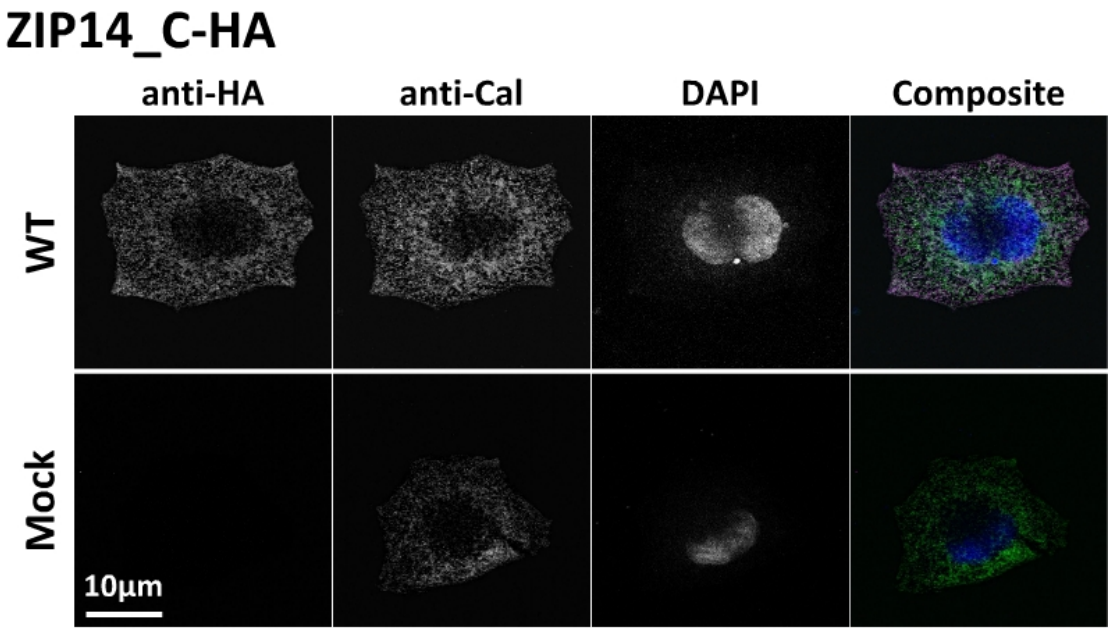


Figure 3.21: **hZIP14_C-HA co-localises with calreticulin.** HeLa cells were transiently transfected with plasmid encoding wild-type hZIP14_C-HA, or mock-transfected. Transfectants were seeded onto coverslips and stained with mouse anti-HA and rabbit anti-calreticulin antibody, followed by anti-mouse Alexa Fluor 488 and anti-rabbit Alexa Fluor 546. Cells were viewed using confocal microscopy, and composite images produced following processing and analysis with ImageJ [184].

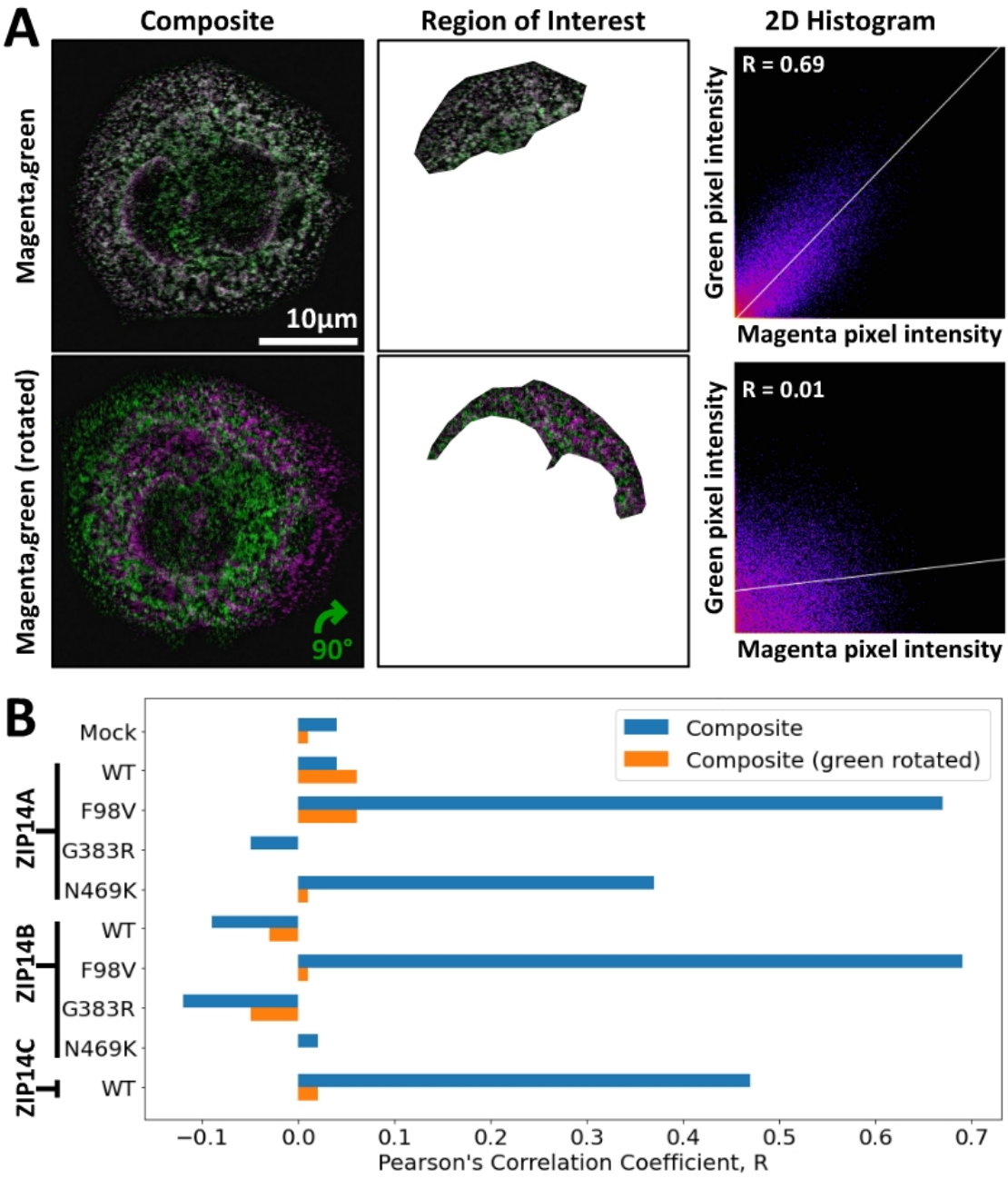


Figure 3.22: **Quantification of hZIP14-HA co-localisation with calreticulin.** A: The hZIP14-HA (magenta) and calreticulin (green) channels of the confocal microscopy images were converted into a composite image. A region of interest - an area of overlap within the cell, excluding the nucleus - was analyzed by the ImageJ Coloc2 plugin, providing a 2D histogram of the pixel intensities, and the Pearson's correlation coefficient [184]. The negative control was a region of interest with one channel rotated through 90°. B: Pearson's correlation coefficients for the co-localisation of hZIP14-HA variants with calreticulin.

3.2.5.2 Variability of Cellular Distribution in Some hZIP14-HA Mutants

Both within and between transfections, the N469K mutant cellular distribution showed some variation. While it typically appeared similar to wild-type, variations were common (figure 3.23). In some cells, the N469K mutant behaved like the N-terminal mutants - failing to traffic to the plasma membrane and displaying a reticular distribution. In other cells, the N469K mutant showed a mix of wild-type and N-terminal mutant characteristics, for example, both trafficking to the cell membrane (like wild-type) and showing increased amounts of protein surrounding the nucleus (like N-terminal mutants). In addition, the N-terminal mutants also displayed variation; images of the N-terminal mutants occasionally showed distributions indistinguishable from that of wild-type hZIP14-HA, including expression at the plasma membrane.

Fluorescence microscopy allowed imaging of individual cells, and showed clear differences between wild-type hZIP14-HA and some of the mutants, namely the N-terminal and L441R mutants. However, neither co-transfection with hZIP14_A-GFP nor studying co-localisation with calreticulin, could fully address the differences in surface levels of the hZIP14_A-HA variants observed previously (figure 3.12), or the variability in the effects of some of the disease-causing mutations (figure 3.23).

Fluorescence microscopy allowed a detailed look at the intracellular location of wild-type hZIP14-HA, isoforms A, B and C, as well as the disease-causing mutations, for both the A and B isoforms. These images revealed no differences between the A and B isoforms of hZIP14-HA (wild-type, or their respective mutants), while wild-type hZIP14_C-HA was retained on the ER and did not traffic to the cell surface. Co-transfection with wild-type hZIP14-GFP indicated that the P379L and G383R mutants localised to the

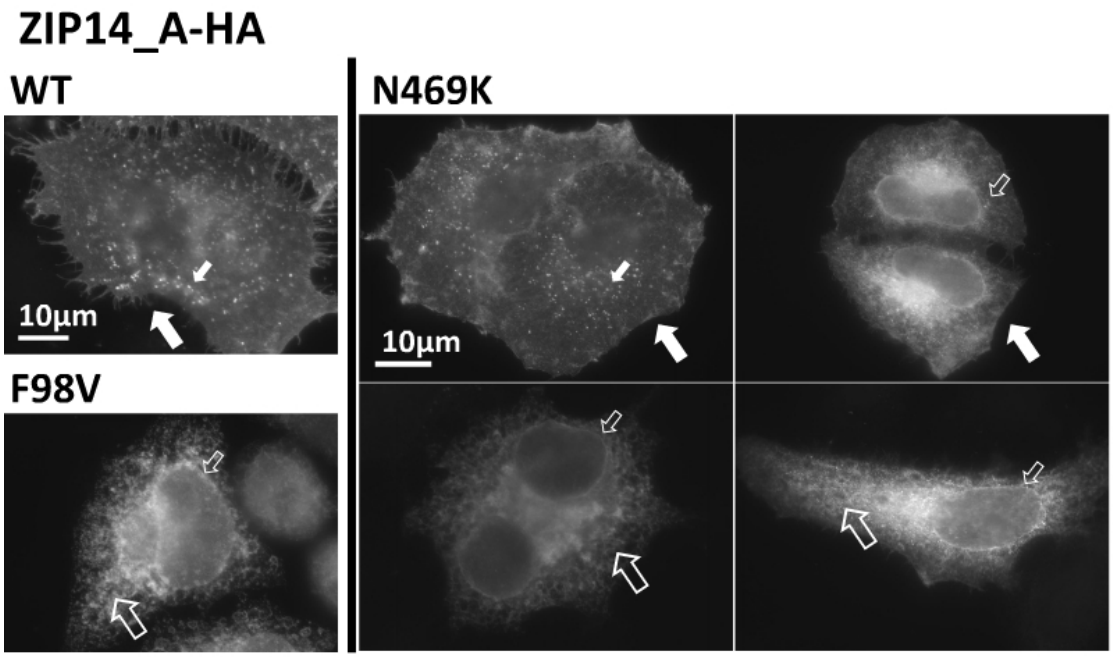


Figure 3.23: **The cellular distribution of the N469K mutant varies between that of wild-type hZIP14_A-HA and the N-terminal mutants.** HeLa cells were transiently transfected with plasmids encoding wild-type, F98V and N469K hZIP14_A-HA. Transfectants were seeded onto coverslips and stained with mouse anti-HA antibody, followed by anti-mouse Alexa Fluor 488. Coverslips were viewed by immunofluorescence microscopy. Solid arrows indicate similarity to wild-type hZIP14-HA localisation - large arrow: protein at the cell membrane, small arrow: bright intracellular puncta. Hollow arrows indicate similarity to N-terminal mutant hZIP14-HA localisation - large arrow: protein in the ER, small arrow: bright ring around the nucleus.

same compartments as wild-type hZIP14-HA, while L441R did not, and co-localisation with calreticulin showed that the N-terminal mutants were retained in the ER. These experiments also exposed a surprising variability in the behaviour of some of the mutants, most notably, N469K. The fluorescence microscopy methods employed successfully revealed effects of the disease-causing mutations that were not apparent from the previous experiments on stable monoclonal cell lines.

3.2.6 Surface Levels of hZIP14-HA in Transiently Transfected Cells

In section 3.2.4.2, it was shown that in the stably-transfected monoclonal cell lines, the R128W, P379L, G383R and N469K hZIP14_A-HA disease-causing mutants were present at the cell surface, and that their cell surface levels were low relative to total expression, when compared to wild-type hZIP14_A-HA (figure 3.12). The immunofluorescence microscopy results of transiently-transfected cells did not fully agree with these observations, as R128W was typically not seen at the cell surface, while both P379L and G383R appeared to traffic effectively to the plasma membrane (figures 3.13 and 3.14). It is unclear why mutants that appeared to be so similar by flow cytometry, might seem so different under the microscope. One potential difference was the wider range of expression levels that occur in transiently transfected cells, in comparison to monoclonal cell lines.

In order to view cell surface expression in transiently transfected cells, non-permeabilising immunofluorescence microscopy was initially used. To confirm that the non-permeabilising conditions were effective at preventing staining of intracellular antigens, HeLa cells were stained for calreticulin (a marker for the ER, calreticulin is not found on the cell surface), in permeabilising and non-permeabilising conditions (figure 3.24 A). In non-permeabilising conditions, cells were stained on ice, in order to prevent uptake of bound antibodies, and were stained prior to fixation, in order to prevent inadvertent permeabilisation of the plasma membrane. The lack of signal in the non-permeabilising conditions shows that antigens inside the cell were not detected in these conditions.

HeLa cells were then transiently transfected with wild-type, F98V and N469K hZIP14-HA, both A and B isoforms, and stained for hZIP14-HA in non-

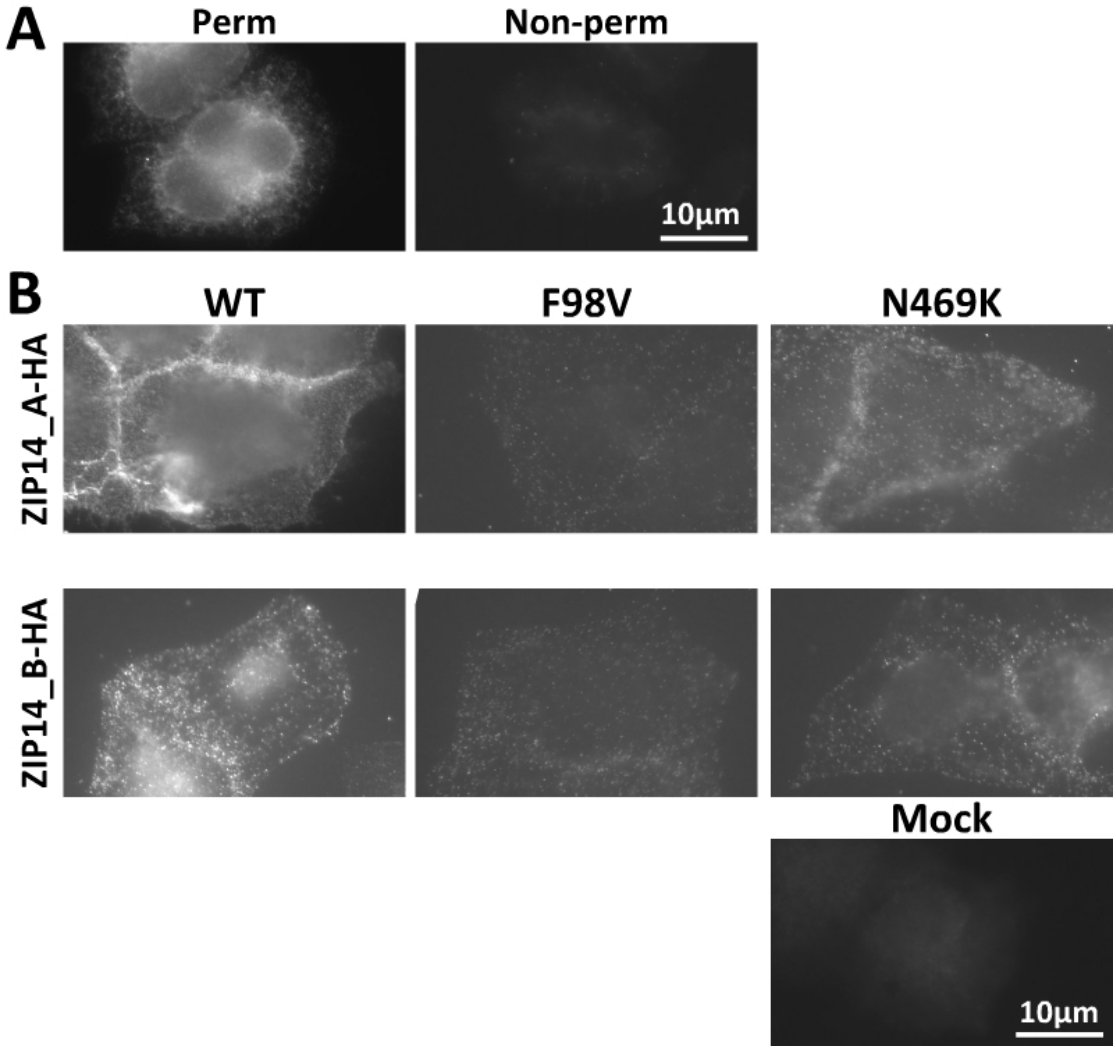


Figure 3.24: **hZIP14-HA disease-causing mutants are present at the plasma membrane.** A: Mock-transfected HeLa cells were stained with rabbit anti-calreticulin antibody followed by anti-rabbit Alexa Fluor 488. Cells were stained in permeabilising (Perm) or non-permeabilising (Non-perm) conditions. B: HeLa cells were transiently transfected with plasmids encoding hZIP14-HA variants or mock-transfected, and stained in non-permeabilising conditions, first with mouse anti-HA antibody, then anti-mouse Alexa Fluor 488. Coverslips were viewed by immunofluorescence microscopy.

permeabilising conditions. These mutants were chosen, as they had both exhibited variability in trafficking to the cell surface, and yet were reported to localise to the plasma membrane as effectively as wild-type hZIP14_A-HA [19].

Following staining for the HA tag in non-permeabilising conditions in order to

detect hZIP14_HA variants at the cell surface, positively staining cells were identified for the each of the transfections, both A and B isoforms, and none for the negative control, mock-transfected cells (figure 3.24). Cells expressing wild-type hZIP14-HA gave the strongest signals, followed by N469K, and cell surface F98V was also detected, but at very low levels. In addition to these differences, the numbers of positively staining cells varied greatly between hZIP14-HA variants; in the wild-type hZIP14-HA transfections, the proportion of positively staining cells was greater than in the N469K transfections, which was again greater than in the F98V transfections. Transfection efficiency typically varied between experiments, but the difference was rarely so pronounced, to the extent that very few cells expressing F98V at the surface could be identified.

A possible explanation for the apparent decrease in transfection efficiency of the hZIP14-HA mutants, is that there were cells that had been successfully transfected, but were nonetheless not detectable due to the absence of the hZIP14-HA protein at the cell surface.

3.2.6.1 Differentiating Between Internal and Surface hZIP14-HA

The variability displayed by mutant hZIP14-HA under immunofluorescence microscopy appeared to be related to the level of expression. The total and surface levels of hZIP14_A-HA had been measured by flow cytometry in stably expressing cell lines, but being monoclonal, the range in total expression levels was limited. Those experiments compared between mutants, but shed no light on the impact of expression on cell surface levels for each mutant. Using immunofluorescence microscopy on transiently transfected cells, which show a wider range of expression levels, both surface and total hZIP14-HA have been imaged, but never simultaneously on individual cells.

Figure 3.25 shows two alternative protocols designed to achieve this aim. In figure 3.25 A, the cells are incubated with anti-HA antibodies from two different species, before and after fixation and permeabilisation, then simultaneously incubated with two different secondary antibodies - one for each species, and each labelled with a different fluorophore. This protocol requires the secondary antibodies to distinguish between the two primary antibodies, each binding their own target.

In order to identify any relationship between expression and surface levels of hZIP14-HA, and to determine whether the strength of these relationships differs between hZIP14-HA variants, it was necessary to stain surface and intracellular hZIP14-HA with different fluorophores. This could be achieved by staining cells with different antibodies before and after permeabilisation. Transiently transfected cells were used in order to study these relationships over a wide range of expression levels.

In figure 3.25 B, the cells are stained with one set of primary and secondary antibodies in non-permeabilising conditions, the cells are fixed and permeabilised, and then stained with a second set of primary and secondary antibodies. Both of these protocols were tested to ensure that secondary antibodies were not binding to the incorrect primary antibody. The results in figure 3.26 illustrate how the protocol depicted in figure 3.25 A failed. When the anti-rat secondary antibody was incubated with cells that had only been stained with mouse anti-HA in the non-permeabilising stage, it could not be detected under the microscope, showing that the anti-rat secondary antibody did not bind non-specifically to mouse anti-HA. On the other hand, when anti-mouse secondary antibody was incubated with cells that had only been stained with rat anti-HA, it did bind to the incorrect antibody and could be detected above the background levels of the negative control (mock-transfected cells).

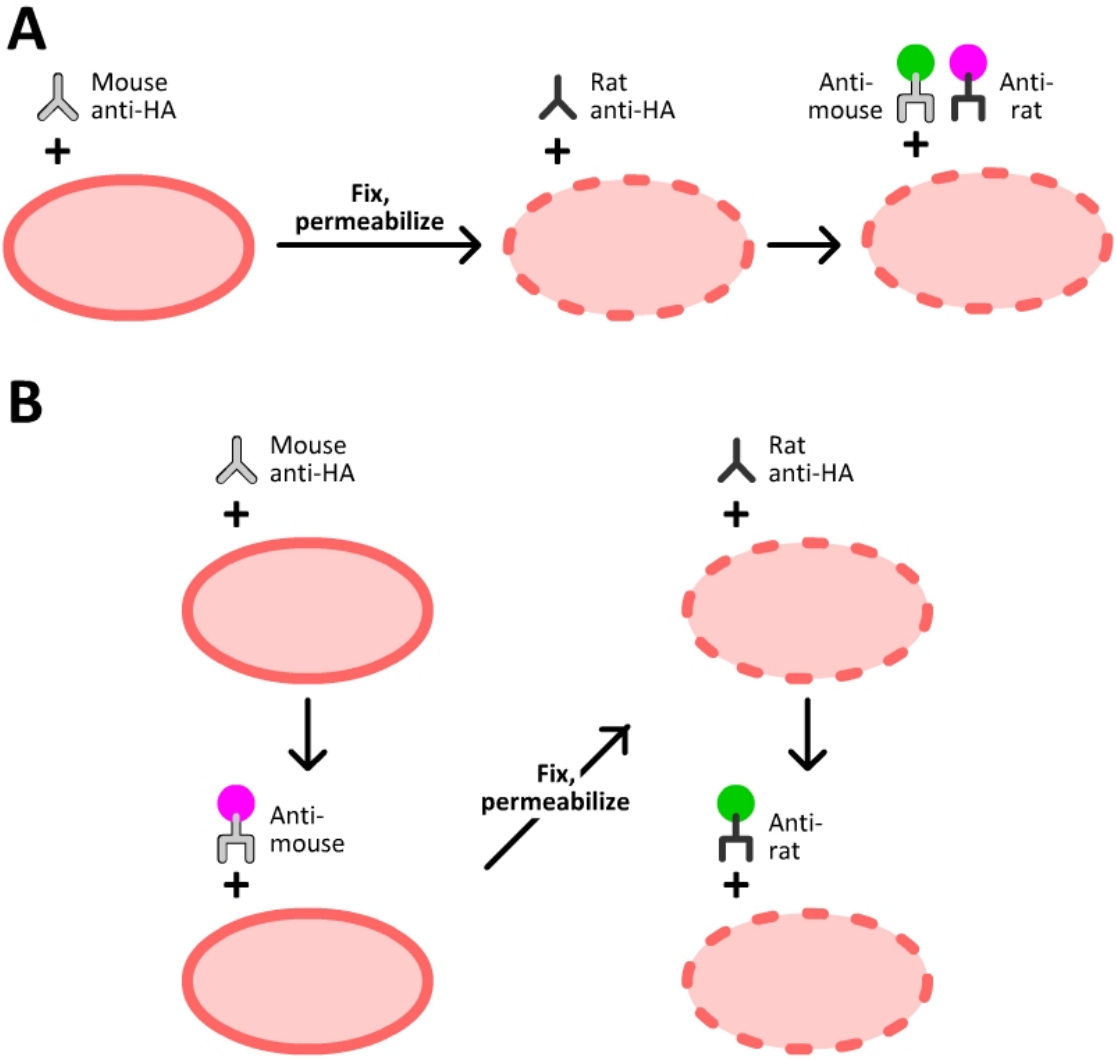


Figure 3.25: **Alternative staining protocols for distinguishing between surface and intracellular protein.** A: Cells are stained with mouse anti-HA antibody in non-permeabilising conditions, allowing access to surface antigen only, then fixed and permeabilised. Cells are then stained with rat anti-HA antibody in permeabilising conditions, providing access to intracellular antigen. Cells are finally incubated with both anti-mouse (Alexa Fluor 488) and anti-rat (Alexa Fluor 546) antibody, in permeabilising conditions. B: Cells are stained with mouse anti-HA antibody, followed by anti-mouse (Alexa Fluor 546) antibody in non-permeabilising conditions, allowing access to surface antigen only, then fixed and permeabilised. Cells are then stained with rat anti-HA antibody, followed by anti-rat (Alexa Fluor 488) antibody, in permeabilising conditions.

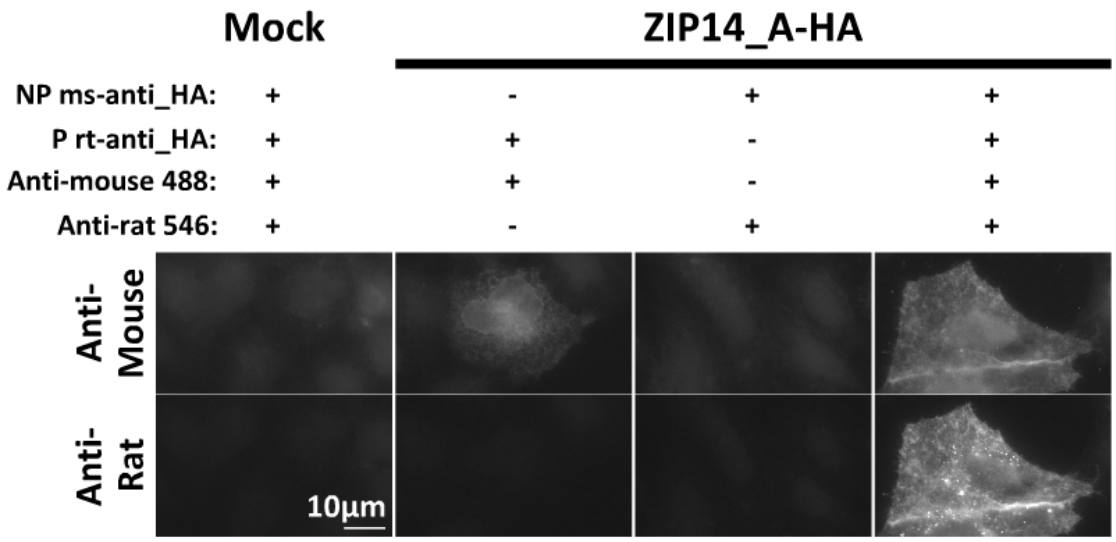


Figure 3.26: **Non-specific binding of anti-mouse antibody to rat anti-HA antibody.** HeLa cells were transiently transfected with wild-type hZIP14_A-HA variants, or mock-transfected. Transfectants were seeded onto coverslips and stained with mouse anti-HA antibody in non-permeabilising (NP) conditions, then fixed and permeabilised. Cells were then stained with rat anti-HA antibody, then anti-mouse Alexa Fluor 488 and anti-rat Alexa Fluor 546, all in permeabilising (P) conditions, as described in figure 3.25 A. Primary or secondary antibodies were excluded for controls, as indicated. Coverslips were viewed by immunofluorescence microscopy.

The protocol illustrated in figure 3.25 B prevents non-specific binding of the anti-mouse secondary antibody to rat anti-HA, as it is used prior to fixation and permeabilisation, and is therefore never exposed to rat anti-HA. With this protocol, the fluorescence of anti-mouse secondary antibody reflects the presence of hZIP14-HA at the plasma membrane, not hZIP14-HA within the cell.

Internal and Surface Levels of hZIP14-HA Mutants by Immunofluorescence Microscopy

HeLa cells were transiently transfected with hZIP14-HA variants and differentially stained for surface and internal hZIP14-HA, allowing the identification of cells that expressed hZIP14-HA without expressing it at the cell surface. Figures 3.27 and 3.28 show typical examples of the cells observed. Cells expressing wild-type hZIP14_A-HA or B always showed strong expression of the protein at the cell surface, while wild-type hZIP14_C-HA was never seen on the cell surface. G383R behaved like wild-type hZIP14-HA, and was always present at the cell membrane. Cells expressing F98V hZIP14_A-HA or B typically did not have detectable amounts of protein at the cell surface, though there were exceptions, and the N469K mutants showed a great deal of variation. Figure 3.29 shows an example of a cell expressing N469K hZIP14_B-HA with none detectable at the cell surface, and cells expressing F98V hZIP14_A-HA, with some of those cells expressing the protein at the cell surface.

Observation by eye suggested that for the hZIP14-HA variants whose expression at the membrane varied, the strength of expression could be affecting cell surface levels. Transient transfections lead to a wide range of expression levels within each transfected sample, but for wild-type and G383R hZIP14-HA, surface protein could be detected even with weak expression, while moderately high levels of expression appeared necessary in order for N469K to reach the surface, and F98V was only detected at the surface of cells with very high expression levels.

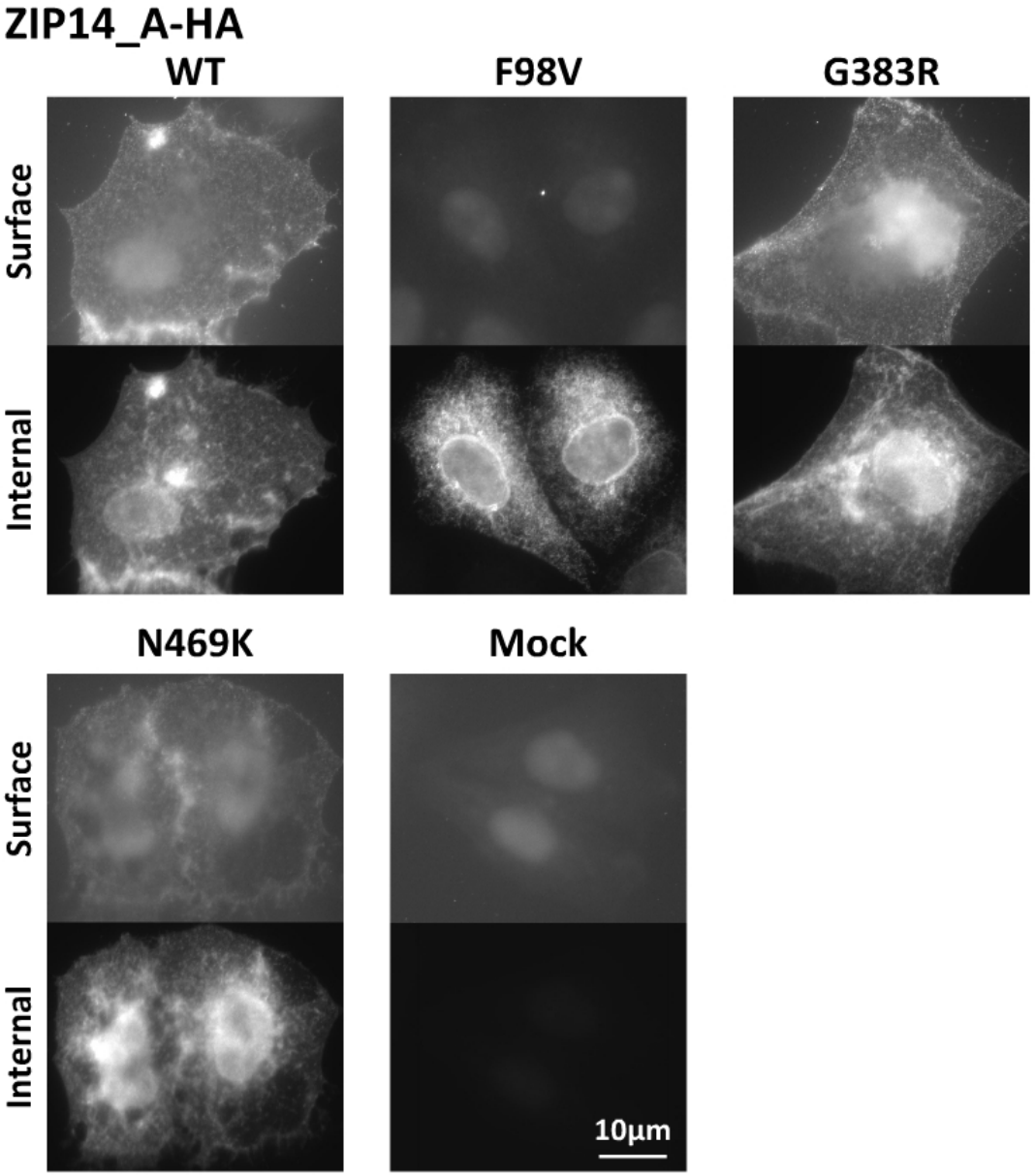


Figure 3.27: **The hZIP14_A-HA F98V mutant has decreased expression at the plasma membrane.** HeLa cells were transiently transfected with hZIP14_A-HA variants, or mock-transfected. Transfectants were seeded onto coverslips and stained with mouse anti-HA antibody, then anti-mouse Alexa Fluor 546, in non-permeabilising conditions, then fixed and permeabilised. Cells were then stained with rat anti-HA antibody, then anti-rat Alexa Fluor 488, in permeabilising conditions, as described in figure 3.25 B. Coverslips were viewed by immunofluorescence microscopy.

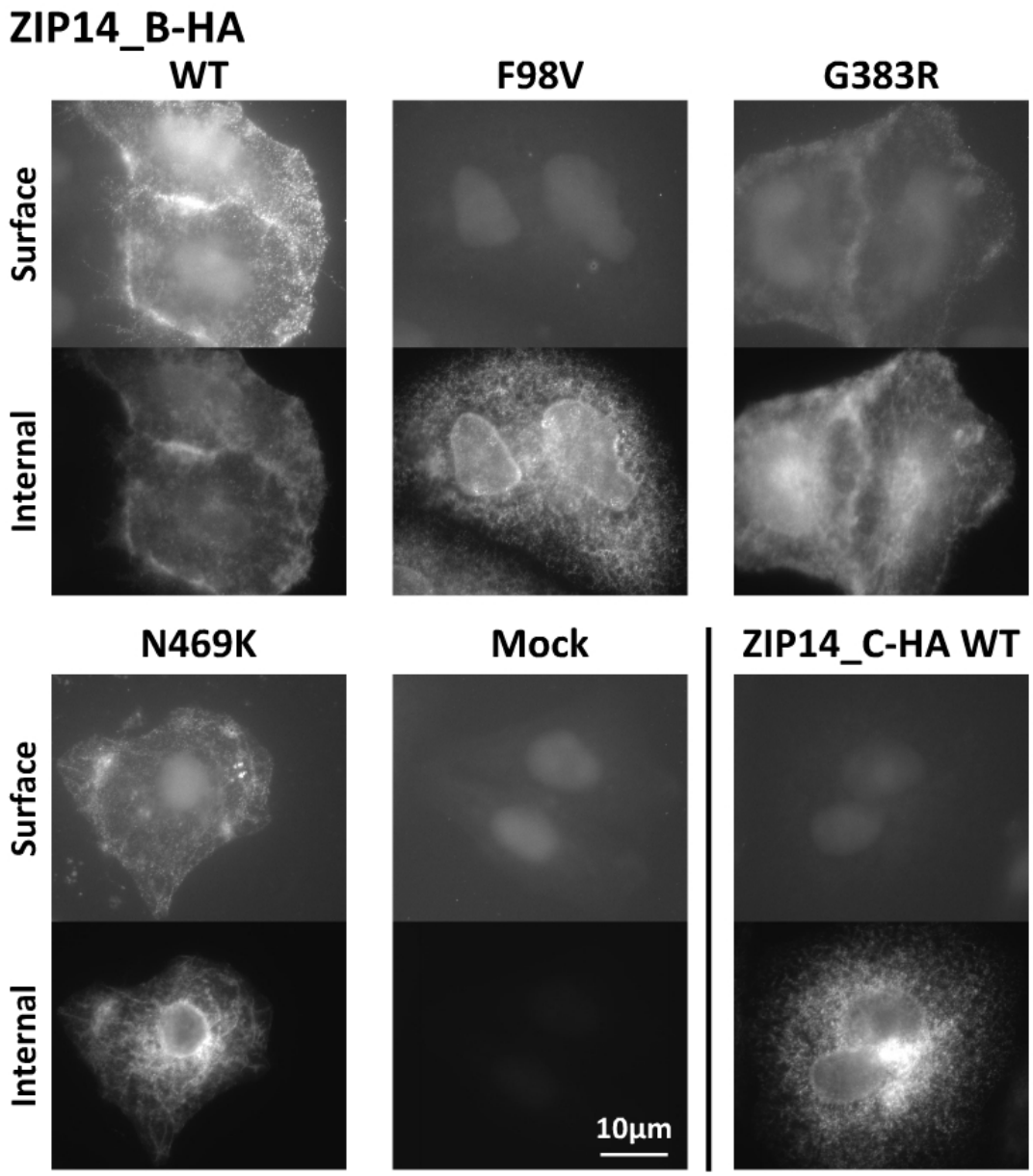


Figure 3.28: **The hZIP14_B-HA F98V mutant and wild-type hZIP14_C-HA have decreased expression at the plasma membrane.** HeLa cells were transiently transfected with hZIP14_B-HA variants, wild-type hZIP14_C-HA, or mock-transfected. Transfectants were seeded onto coverslips and stained with mouse anti-HA antibody, then anti-mouse Alexa Fluor 546, in non-permeabilising conditions, then fixed and permeabilised. Cells were then stained with rat anti-HA antibody, then anti-rat Alexa Fluor 488, in permeabilising conditions, as described in figure 3.25 B. Coverslips were viewed by immunofluorescence microscopy.

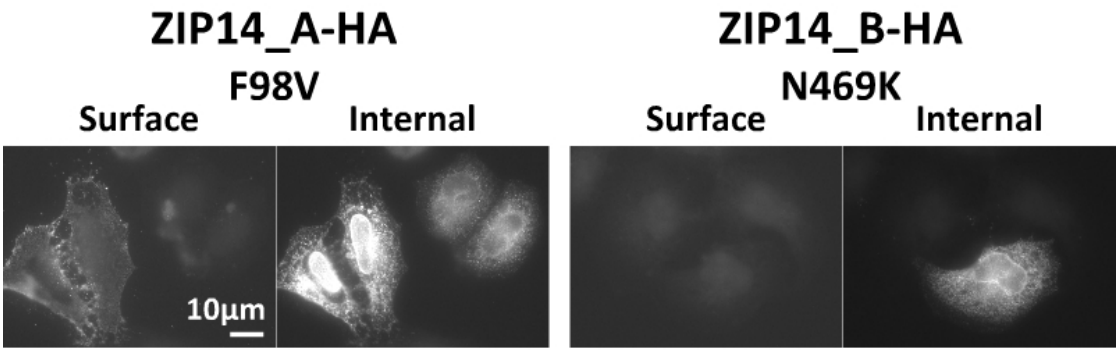


Figure 3.29: **Expression of hZIP14-HA F98V and N469K at the plasma membrane varies.** HeLa cells were transiently transfected with plasmids encoding hZIP14-HA variants. Transfectants were seeded onto coverslips and stained with mouse anti-HA antibody, then anti-mouse Alexa Fluor 546, in non-permeabilising conditions, then fixed and permeabilised. Cells were then stained with rat anti-HA antibody, then anti-rat Alexa Fluor 488, in permeabilising conditions. Coverslips were viewed by immunofluorescence microscopy.

3.2.7 Comparing Expression and Surface Levels By Flow Cytometry

Immunofluorescence microscopy, with differential staining of interior and surface hZIP14-HA, suggested that for some of the disease-causing mutants, expression impacted levels at the cell surface. In order to determine whether the relationship between the expression of hZIP14-HA and the amount on the surface differed between the wild-type protein and the mutants, a more quantitative approach was required.

Like microscopy, flow cytometry can also discriminate between surface and intracellular protein. However, flow cytometry is a more quantitative approach for assessing protein levels than microscopy, and can be used to rapidly scan thousands of cells. In order to better understand trafficking of hZIP14_A-HA variants to the cell surface, the protocol depicted in figure 3.25 B was adapted for flow cytometry, and used to look at transiently transfected cells.

3.2.7.1 Compensation for Alexa Fluor 488 and Alexa Fluor 647

A quantitative study of the relationship between the internal and surface levels of hZIP14-HA, requires that fluorescence spillover into the channels used to detect the two fluorophores is compensated for. Figure 3.30 shows the emission spectra of the Alexa Fluor 488 and 647 fluorophores, and the range of wavelengths detected by channels FL-1 and FL-4. The emission peak of Alexa Fluor 488 fits within the FL-1 channel, and the emission peak of Alexa Fluor 647 fits within the FL-4 channel. The emission spectrum of Alexa Fluor 647 is unlikely to be detected within the FL-1 channel, though there is a slight possibility that Alexa Fluor 488 could spillover into the FL-4 channel.

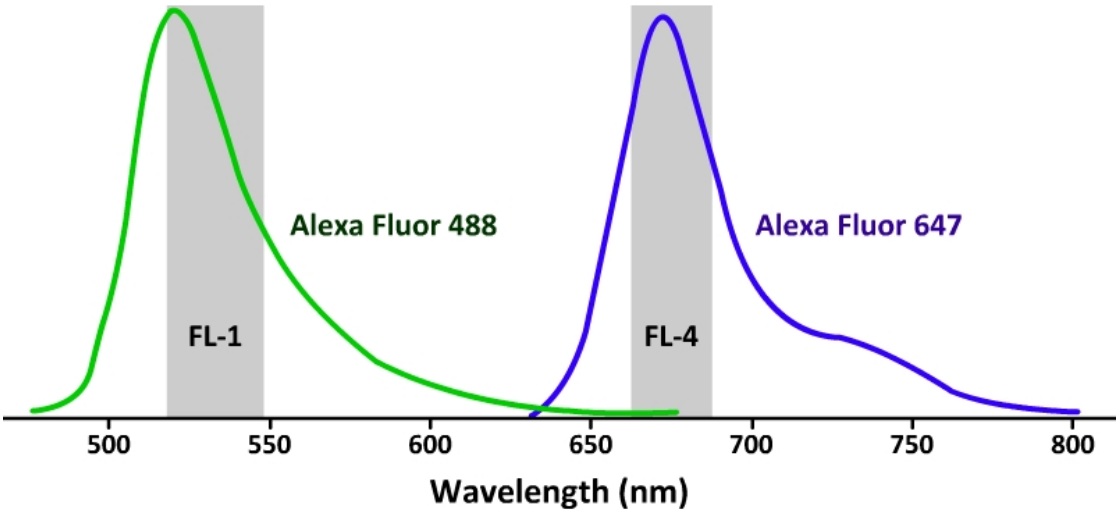


Figure 3.30: **Emission spectra of the Alexa Fluor 488 and 647 fluorophores.** The green spectrum represents the emission spectrum of Alexa Fluor 488, the blue spectrum, Alexa Fluor 647. The grey boxes represent the ranges of wavelengths detected within the FL-1 and FL-4 channels of the BD Accuri^{TM} C6 flow cytometer [187].

To test this, cells were stained for antigens using a range of antibody concentrations, with secondary antibodies conjugated to either Alexa Fluor 488 or 647, and the median fluorescence intensity of each sample was recorded in channels FL-1 and FL-4 (figure 3.31).

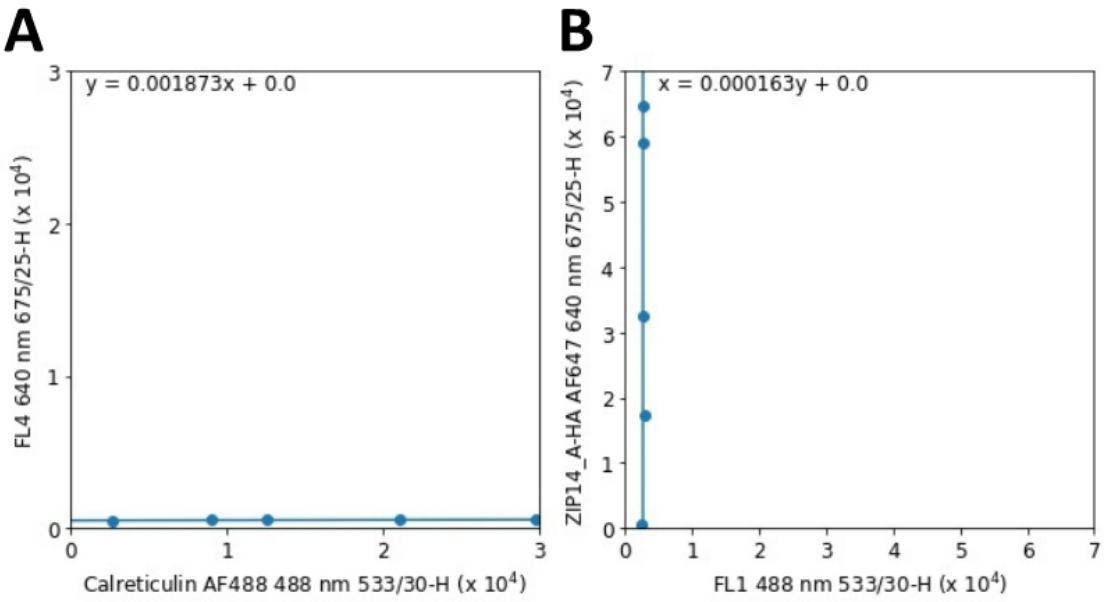


Figure 3.31: **Alexa Fluor 488 and 647 do not spill over into the FL-4 and FL-1 channels.** HeLa cells in suspension were stained with fluorescently-labelled antibodies and median fluorescence intensity was measured in the FL-1 and FL-4 channels of a BD Accuri^{TM} C6 flow cytometer. A: Permeabilised HeLa cells were stained with a range of concentrations of rabbit anti-calreticulin antibody, then anti-rabbit Alexa Fluor 488 antibody. B: Permeabilised hZIP14_A-HA-expressing cells were stained with mouse anti-HA antibody, then a range of concentrations of anti-mouse Alexa Fluor 647 antibody.

Figure 3.31 A shows the results obtained when HeLa cells were permeabilised and stained for calreticulin using an Alexa Fluor 488 secondary antibody. As the concentration of primary antibody used increased, the median fluorescence intensity of the sample detected in the FL-1 channel increased, while the reading in the FL-4 channel was unaffected. The line of best fit for the relationship between the FL-1 and FL-4 readings had a gradient close to zero, showing that there is effectively no spillover of Alexa Fluor 488 into the FL-4 channel.

Figure 3.31 B shows the results obtained when HeLa cells, stably expressing hZIP14_A-HA, are stained for hZIP14_A-HA using an Alexa Fluor 647 secondary antibody. In this case, the concentration of secondary antibody was varied, altering the fluorescence intensity of samples in the FL-4 channel, while the

reading in the FL-1 channel was unaffected. The line of best fit for the relationship between the FL-4 and FL-1 readings also had a gradient close to zero, showing that there was no spillover of Alexa Fluor 647 into the FL-1 channel. Checking for fluorescence spillover was important, because in order to distinguish between surface and interior antigens, the signals from the fluorophores used to stain these antigens need to be distinct.

3.2.7.2 Permeabilising and Non-permeabilising Conditions Differentiate Between Internal and Surface Proteins

Flow cytometry is a more sensitive method of detecting protein than immunofluorescence microscopy, and as such, the flow cytometry protocol for differentially staining interior and surface hZIP14-HA needs to be checked to ensure that the non-permeabilising conditions do not permeabilise cells, and conversely, that the permeabilising conditions do permeabilise cells.

To this end, HeLa cells were stained for calreticulin in permeabilising and non-permeabilising conditions; permeabilised HeLa cells that had not been incubated with the anti-calreticulin antibody were used as a negative control (figure 3.32). HeLa cells stained for calreticulin in non-permeabilising conditions had a much lower fluorescence intensity than cells stained in permeabilising conditions, and an identical fluorescence intensity to the negative control. These results confirm that the non-permeabilising conditions used did not allow intracellular antigens to be detected, and that the permeabilising conditions used did make intracellular antigens accessible to antibodies.

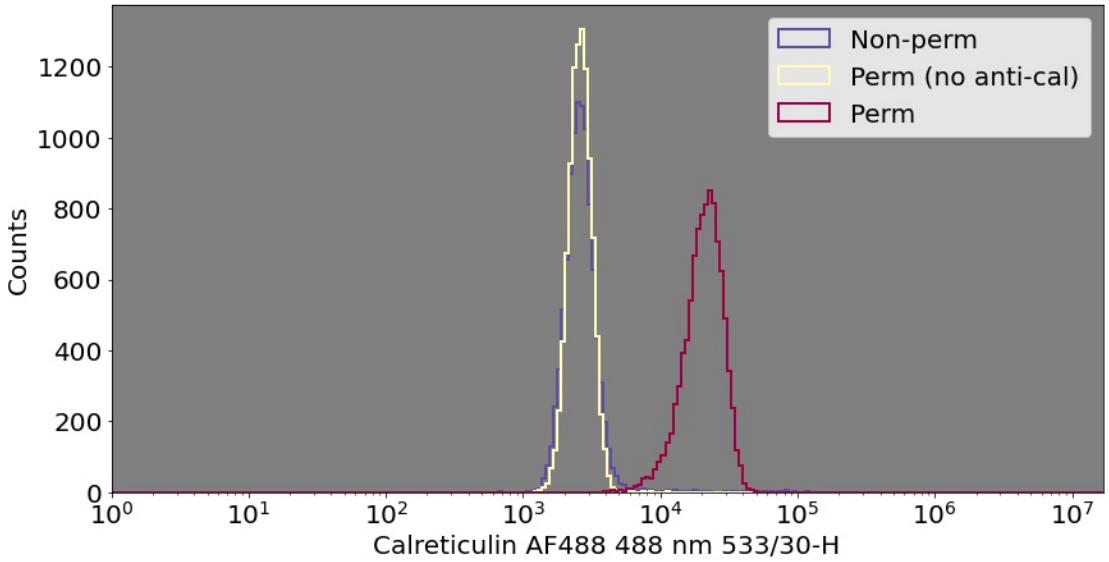


Figure 3.32: **Non-permeabilising conditions prevent staining of intracellular epitopes.** HeLa cells in suspension were incubated with or without ('no anti-cal') rabbit anti-calreticulin, then anti-rabbit Alexa Fluor 488 antibody. Cell readings were taken in the FL-1 channel of a BD Accuri™ C6 flow cytometer.

3.2.7.3 Secondary Antibodies Specifically Bind Their Target Antibodies

In order to be certain that the secondary antibodies to be used in the two stages of staining are specific for their primary antibodies, HeLa and the stable hZIP14_A-HA cell line were stained with or without the secondary antibodies in the non-permeabilising and permeabilising stages (figure 3.33).

Regarding surface staining, the fluorescence intensity of HeLa cells was the same whether the cells were fully stained (both secondary antibodies used in their appropriate stages) or if the secondary antibody was missing in the non-permeabilising stage (staining surface antigen). Results for HeLa cells in the permeabilising stage (staining intracellular antigen) showed a slight increase in fluorescence intensity when the secondary antibody for the permeabilising stage was present. This indicates some non-specific binding of the secondary antibody at this stage, which is likely caused by the fact that the cells are

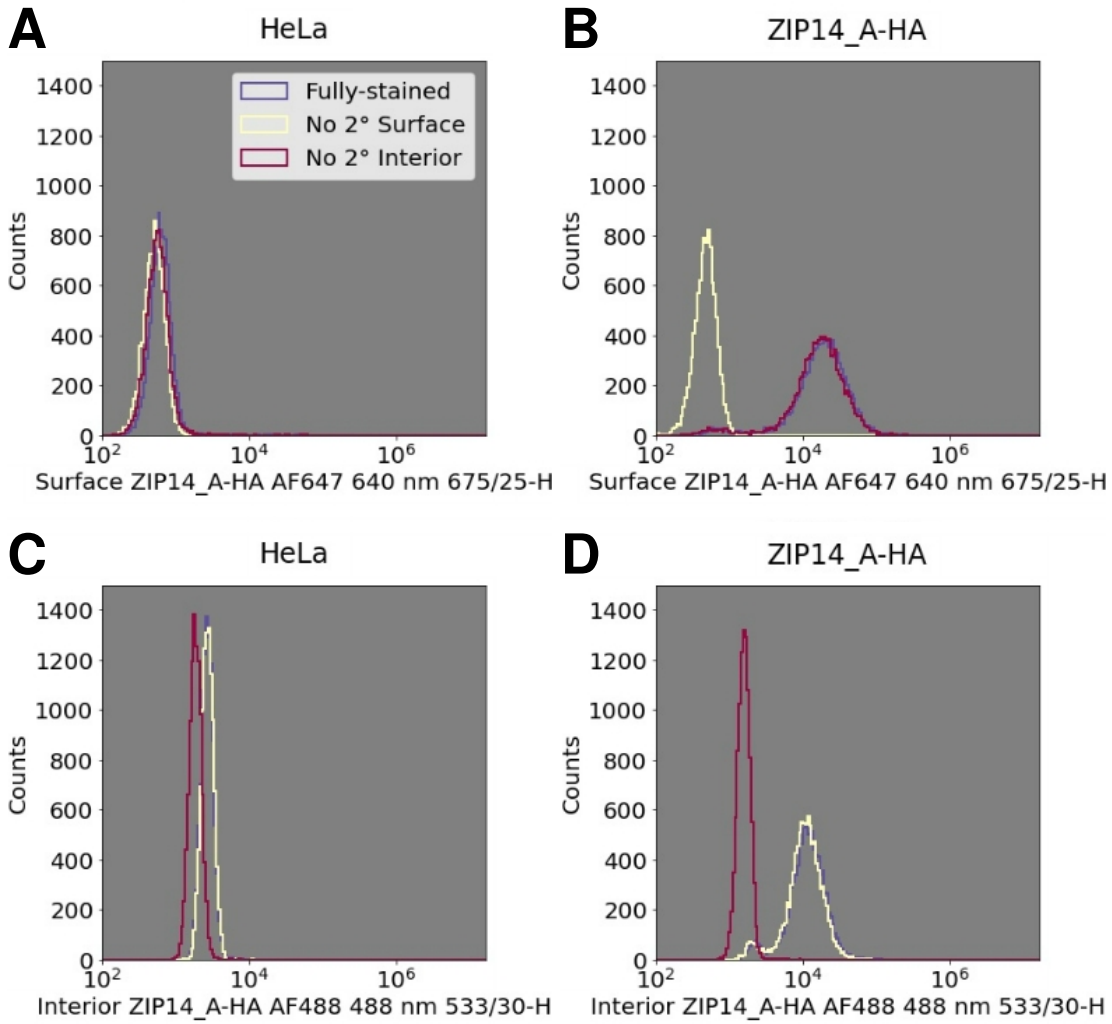


Figure 3.33: **Secondary antibodies are specific for target primary antibodies.** HeLa cells (A, C) and HeLa cells stably expressing hZIP14_A-HA (B, D) in suspension were incubated with mouse anti-HA, then with/without anti-mouse Alexa Fluor 647 antibody, in non-permeabilising conditions. Cells were then fixed and permeabilised, then incubated with rat anti-HA, then with/without anti-rat Alexa Fluor 488 antibody. Cell readings were taken in the FL-1 and FL-4 channels of a BD Accuri^{TM} C6 flow cytometer. A, B: Detection of Alexa Fluor 647, representing surface hZIP14_A-HA. C, D: Detection of Alexa Fluor 488, representing internal hZIP14_A-HA.

permeabilised so there is more opportunity for non-specific binding than when antibody is only exposed to the surface of the cells.

For cells expressing hZIP14_A-HA, non-specific binding of the secondary antibody would manifest as an increased fluorescence in the absence of the

other secondary antibody, due to binding both primary antibodies, compared to binding only the target. For surface staining, no increased fluorescence was expected, as the secondary antibody used in the non-permeabilising stage is never exposed to the incorrect primary antibody, and none was observed. For staining of intracellular antigen, there was no increase in fluorescence intensity when the secondary antibody for the surface-staining stage was absent, relative to fully stained cells. This shows that in the permeabilising stage, the anti-rat secondary antibody was only binding its target antibody, even though mouse primary antibody was present at the surface, available for off-target binding.

Figure 3.34 shows density plots for fully stained HeLa and hZIP14_A-HA expressing cells, with quadrants separating cells that are negative for surface and internal hZIP14_A-HA (lower-left quadrant), from cells that are positive for surface and internal hZIP14_A-HA (upper-right quadrant). The small number of cells in the upper-left quadrant of the HeLa density plot reflect non-specific binding of secondary antibody in the surface staining stage, most likely attached during the fixation step. The small number of cells in the lower-left quadrant of the ZIP14_A-HA density plot may represent cells that have lost expression of hZIP14_A-HA.

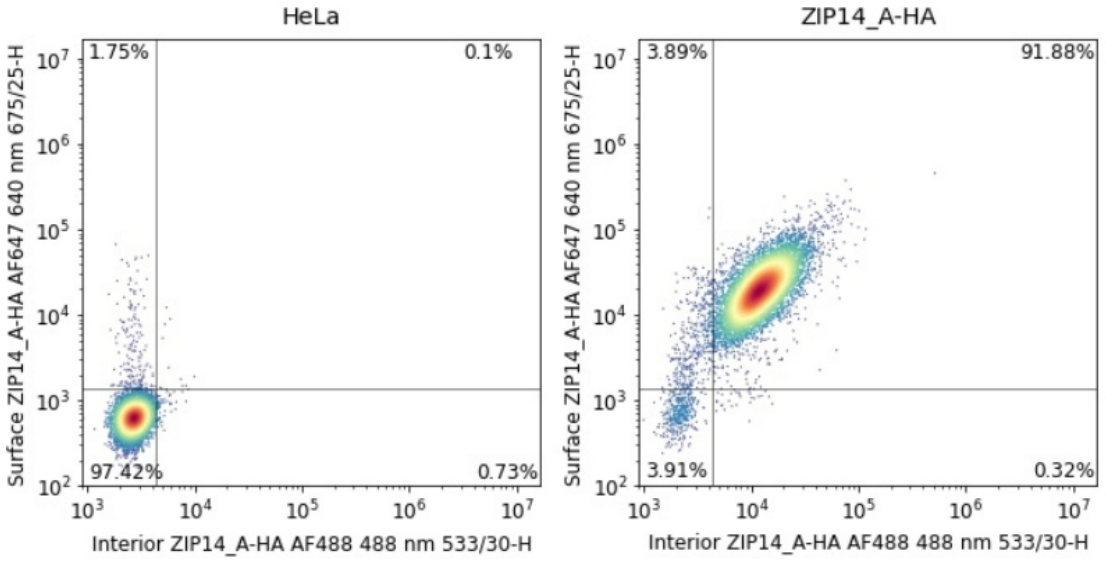


Figure 3.34: **Density plots for positive and negative controls.** HeLa cells and HeLa cells stably expressing hZIP14_A-HA in suspension were incubated with mouse anti-HA, then anti-mouse Alexa Fluor 647 antibody, in non-permeabilising conditions. Cells were then fixed and permeabilised, then incubated with rat anti-HA, then anti-rat Alexa Fluor 488 antibody. Cell readings were taken in the FL-1 and FL-4 channels of a BD Accuri^{TM} C6 flow cytometer.

3.2.8 Disease-Causing Mutations Affect Trafficking to the Cell Surface

HeLa cells were transiently transfected with hZIP14_A-HA variants and differentially stained for surface and intracellular hZIP14_A-HA.

An important point to consider in interpretation of these density plots, is that the x-axis is not an independent measure of expression (figure 3.35). A horizontal gradient indicates that as expression levels increase, no additional protein is being directed to the surface. A vertical gradient indicates that as expression levels increase, all the additional protein is being directed to the surface. The steepness of the gradient therefore reflects the proportion of additional protein being targeted to the cell surface, as opposed to remaining inside the cell. Due to the potential differences in affinity of the mouse and rat anti-HA antibodies,

as well as the secondary antibodies, it is not possible to use these density plots to calculate the total expression of hZIP14_A-HA by simply adding together the fluorescence intensities of the separate channels. However, these density plots do highlight quantitative differences in the relative proportions of protein that are retained intracellularly, or situated on the plasma membrane, and thus, shed light on which mutations alter trafficking to the cell surface.

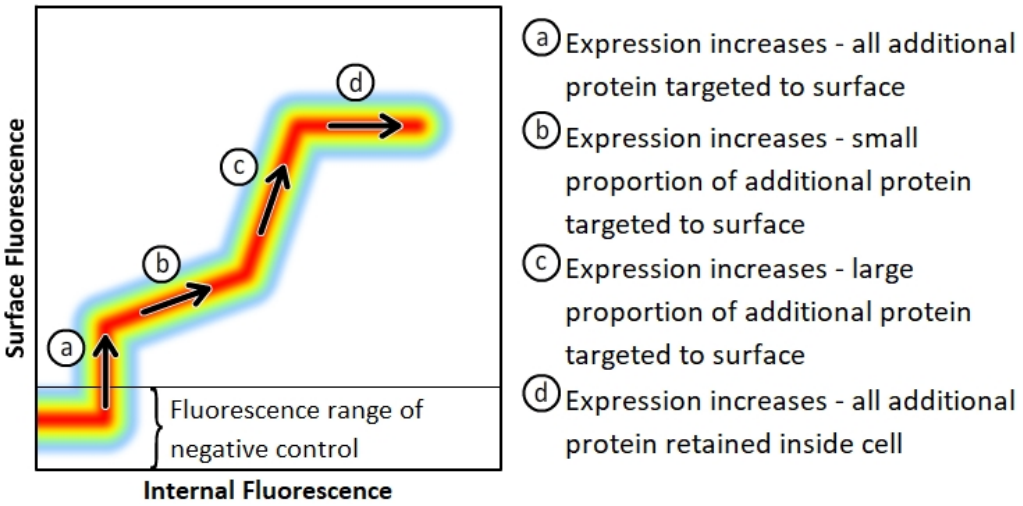


Figure 3.35: **Interpretation of density plots for differentially-stained samples.** Density plots of differentially-stained cells, where internal and surface proteins have been stained with different fluorophores, compare the brightness of the cell's internal fluorescence to that of its surface fluorescence. a) A vertical gradient indicates that as the protein's expression increases, all additional protein is being directed to the cell's surface, such that internal fluorescence is unchanged. b) A shallow gradient indicates that as protein expression increases, a relatively small proportion of the additional protein is being directed to the cell's surface, and the increase in expression is mostly detected by the increasing internal fluorescence. c) A steep gradient indicates that as protein expression increases, a relatively large proportion of the additional protein is being directed to the cell's surface, and the increase in expression is mostly detected by the increasing surface fluorescence. d) A horizontal gradient indicates that as the protein's expression increases, none of that additional protein is directed to the surface, instead remaining inside the cell.

The fluorescence intensity of mock transfected cells, in the channel used to

detect intracellular hZIP14_A-HA, was used to gate for transfected cells in the other samples, and representative density plots (each hZIP14_A-HA variant was tested a minimum of three times) for these cells are shown in figures 3.36 and 3.37.

Internal wild-type hZIP14_A-HA correlated with the amount of hZIP14_A-HA protein detected at the surface with a near linear relationship, potentially beginning to plateau at very high levels of expression (figure 3.36). At all levels of expression, wild-type hZIP14_A-HA was detectable at the cell surface.

The relationship between internal and surface levels of the N-terminal disease-causing mutants (F98V, S104I and R128W) resembled a sigmoidal curve (figure 3.36). At low expression, the median fluorescence intensities for surface hZIP14_A-HA were lower than wild-type. As expression of these mutants increased, the gradient of this relationship changed, such that further increases in expression resulted in much greater increases in the amount of the protein detected at the cell surface, slightly surpassing wild-type levels. To a lesser extent than wild-type hZIP14_A-HA, at very high expression of the N-terminal mutants, there appeared to be a slight tendency towards a plateau, where the effect of increasing expression on surface levels was diminished. The N-terminal mutants consistently displayed this sigmoidal relationship, with no clear differences between any of the N-terminal mutations. These results suggest that at most levels of expression, N-terminal mutants have less expression at the cell surface than wild-type hZIP14_A-HA, and that at very high levels of expression, these mutants can reach high levels at the cell surface, though the density plots show that only a small proportion of cells displayed such high expression of the N-terminal mutants.

The relationship between internal and surface P379L hZIP14_A-HA had a similar gradient to that of wild-type hZIP14_A-HA at low to moderate levels of

expression, though P379L surface levels were higher than wild-type hZIP14_A-HA for a given amount of interior hZIP14_A-HA. In cells transfected with the P379L mutant, there is a distinct population of cells where internal P379L hZIP14_A-HA is very high, while surface levels are moderately high, represented by a circular region towards the vertical centre of the density plot. This is possibly due to down-regulation of P379L protein from the surface of the cell to its interior, though as total protein levels are unknown, it is unclear whether the expression of P379L in these cells is greater, or equivalent, to any other transfectants on this plot. It is clear however, that for the P379L mutant, it is possible to have high levels of protein at the cell surface, with two very different levels of total expression.

Like P379L, the relationship between internal and surface levels of G383R was similar to that of wild-type, with slightly higher amounts of protein at the surface level, for a given amount of internal protein. The ball-like distribution described for the P379L mutant was present for G383R, though to a much lesser degree, potentially because much higher surface levels of G383R were reached before internalisation of the protein occurred.

The density plot for L441R appeared to show a sigmoidal curve, though with a distinct plateau occurring at moderate levels of internal L441R hZIP14_A-HA, and very low levels of protein at the cell surface over all expression levels, relative to wild-type hZIP14_A-HA. In addition, the density plot shows that the majority of transfectants have no detectable L441R hZIP14_A-HA at the cell surface; the distribution closely follows the line demarcating background fluorescence.

The N469K hZIP14_A-HA density plot also showed that the majority of cells were clustered at low expression levels, with no detectable N469K at the surface. However, N469K hZIP14_A-HA was detectable at the surface across

the range of interior expression levels, albeit at lower levels than wild-type, for a given level of interior expression. A plot of the relationship between interior and surface levels of N469K, using the median fluorescence intensities, would result in a sigmoidal curve. However, it would be more accurate to say that at low interior levels of N469K, the range of surface expressions observed was high, becoming progressively narrower as expression increased. As the range of surface levels observed relative to interior N469K decreased, it was the higher surface levels that were favoured, so that the overall trend was that the proportion of N469K at the surface increased, as expression increased, and a tendency towards a plateau was observed at the very highest expression levels.

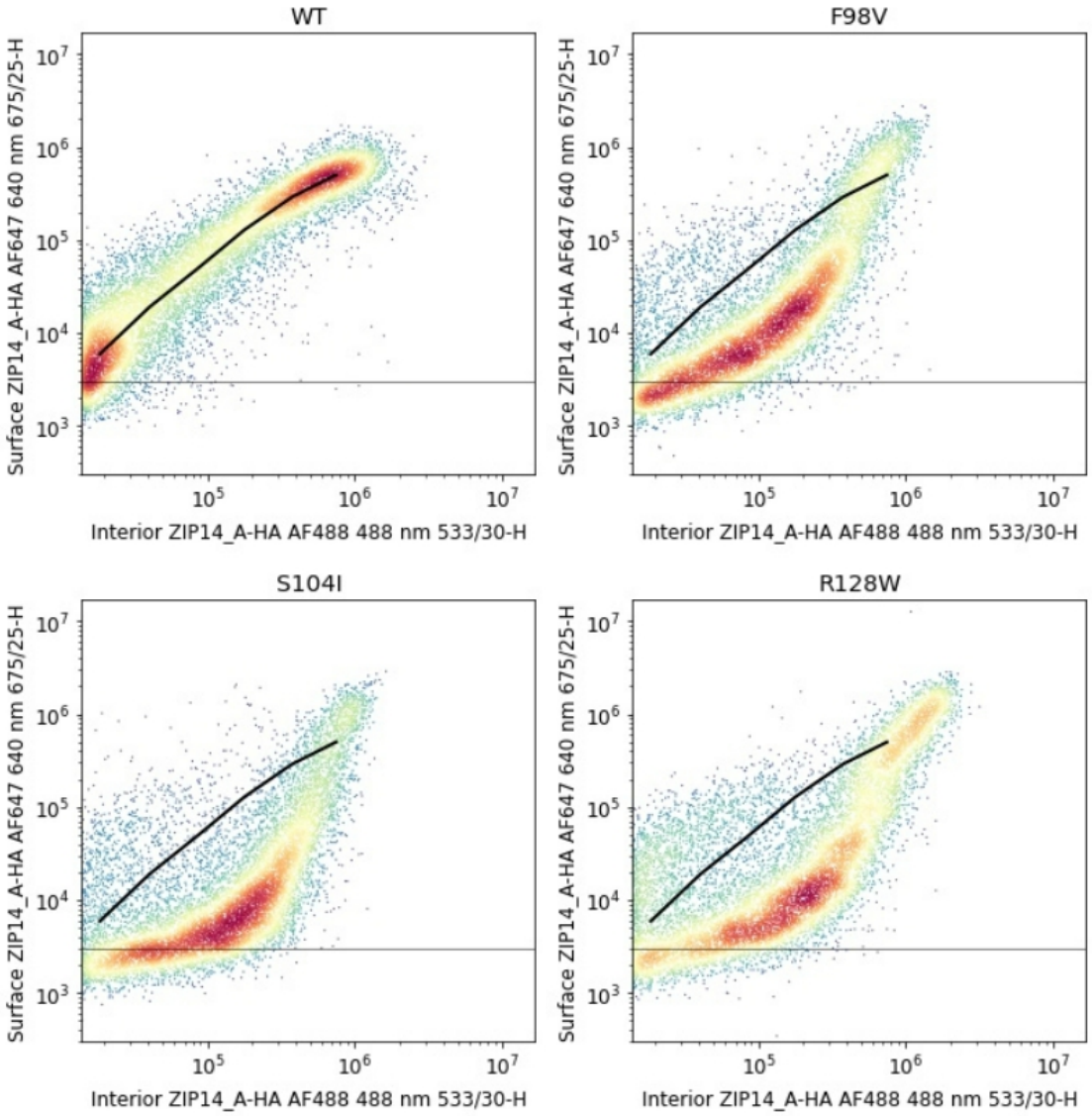


Figure 3.36: **Effect of N-terminal disease-causing mutations on cell surface levels of hZIP14_A-HA relative to internal levels.** HeLa cells were transiently transfected with plasmids encoding hZIP14_A-HA variants, or mock-transfected. Transfectants in suspension were incubated with mouse anti-HA, then anti-mouse Alexa Fluor 647 antibody, in non-permeabilising conditions. Cells were then fixed and permeabilised, then incubated with rat anti-HA, then anti-rat Alexa Fluor 488 antibody. Cell readings were taken in the FL-1 and FL-4 channels of a BD Accuri^{TM} C6 flow cytometer. Density plots show results for transfected cells, and are representative of three independent experiments. The horizontal line demarcates background surface fluorescence observed in mock-transfected cells; the solid black line represents wild-type hZIP14_A-HA MFIs.

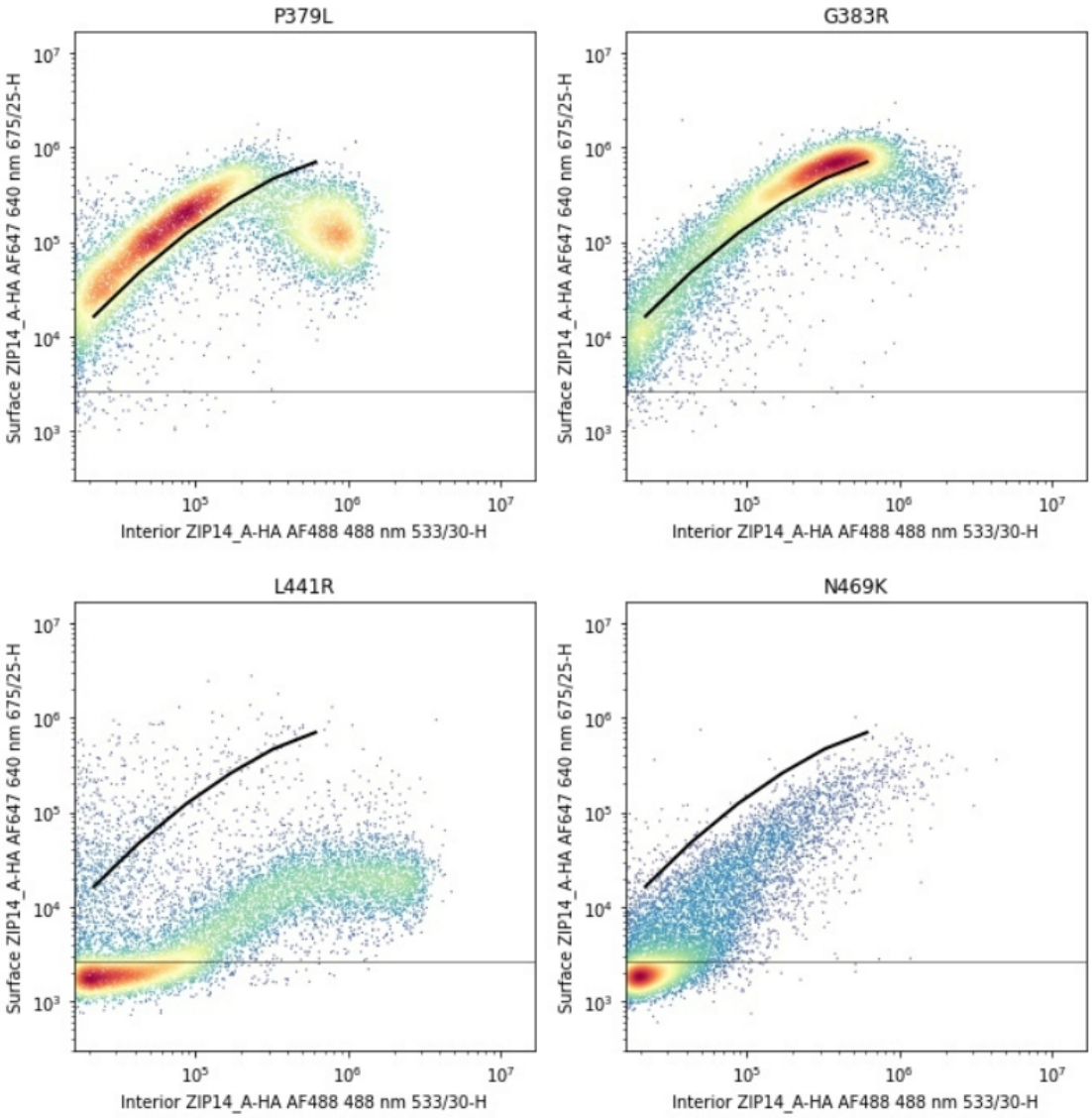


Figure 3.37: **Effect of C-terminal disease-causing mutations on cell surface levels of hZIP14_A-HA relative to internal levels.** HeLa cells were transiently transfected with plasmids encoding hZIP14_A-HA variants, or mock-transfected. Transfectants in suspension were incubated with mouse anti-HA, then anti-mouse Alexa Fluor 647 antibody, in non-permeabilising conditions. Cells were then fixed and permeabilised, then incubated with rat anti-HA, then anti-rat Alexa Fluor 488 antibody. Cell readings were taken in the FL1 and FL4 channels of a BD Accuri^{TM} C6 flow cytometer. Density plots show results for transfected cells, and are representative of three independent experiments. The horizontal line demarcates background surface fluorescence observed in mock-transfected cells; the solid black line represents wild-type hZIP14_A-HA MFIs.

3.3 Discussion

In this work, disease-causing mutations of hZIP14 were characterised in detail. Despite the similarities in the symptoms caused by the majority of these mutations, these results show that the disease-causing mutations have a range of impacts on iron uptake function and cellular localisation.

Wild-type hZIP14-HA localised to the cell surface over a wide range of concentrations, with bright intracellular puncta throughout the cytoplasm, and transported iron into cells at $\sim$pH 7.5, consistent with previous literature findings on ZIP14 [11, 13, 15, 35]. The N-terminal mutants (F98V, S104I and R128W) were largely retained in the ER (figure 3.13). In transiently transfected cells, the proportions of these mutant proteins localised to the cell surface was lower than wild-type, at low levels of expression (figure 3.36). In stably-transfected cells, the proportion of R128W protein localised at the cell surface was very low compared to wild-type hZIP14, and the R128W protein at the cell surface did not transport iron (figure 3.11).

The C-terminal disease-causing mutations (P379L, G383R, L441R and N469K) affected hZIP14 in a range of ways. The P379L mutation showed effects on cellular localisation - in transiently-transfected cells, cell surface levels were higher than wild-type at lower levels of expression, and lower than wild-type at higher levels of expression (figure 3.37). In stably-transfected cells, the proportions of the P379L mutant at the cell surface were lower than wild-type (figure 3.12). For protein at the cell surface, the P379L mutation showed no impact on iron uptake function (figure 3.11). The G383R mutation had little impact on cell surface levels in transiently transfected cells (figure 3.37), but in stable cell lines, proportions of G383R hZIP14 were low relative to wild-type (figure 3.12); G383R hZIP14 at the cell surface did not transport iron (figure

3.11). In stably-transfected cells, the N469K mutant surprisingly increased iron uptake (relative to cell surface levels), but decreased cell surface levels (figures 3.11 and 3.12). In transiently transfected cells, both the L441R and N469K mutations decreased the proportion of hZIP14 localised at the cell surface, especially at low levels of expression (figure 3.37).

Overall, the N-terminal mutations appeared to have very similar effects, decreasing cell surface localisation, and preventing iron uptake. In contrast, the C-terminal mutations decreased, increased, or had no effect on both cell-surface localisation and iron uptake.

3.3.1 Expression of hZIP14 Isoforms in Cell Lines

ZIP14 mRNA has been detected in a wide range of tissues, in both mice and humans [13, 14, 19]. The majority of that expression is due to isoform A, as mRNA of isoform B in adult humans is absent from key tissues including brain, heart and skeletal muscle, and where both are expressed, is typically found at lower levels than isoform A [19]. hZIP14 isoform C is found in databases such as Uniprot and Ensembl on the basis of computational predictions [208, 216], but there is no experimental confirmation of its expression. The HeLa, HEK293 and HepG2 cell lines originate from uterine, kidney and liver tissues, respectively [217]. In a multi-tissue expression array, Taylor et al. [13] detected ZIP14 cDNA in samples from human uterus, kidney and liver, as well as HeLa S3, a sub-clone of HeLa [217]. My detection of hZIP14_A expression in HEK293 cells, but not hZIP14_B, and of both A and B isoforms in HepG2, with hZIP14_B at relatively lower levels (figure 3.4), is consistent with the results for human kidney and liver reported by Tuschl et al. [19]. The failure to detect hZIP14 mRNA in HeLa cells may have been due to a lack of sensitivity of the method used, in which case,

expression of hZIP14 by HeLa cells is still much lower than in the other two cell lines tested, and therefore remains a more suitable cell line in which to study hZIP14 without interference from endogenous hZIP14 expression.

The predicted hZIP14_C differs from isoforms A and B from residue 444 to the end of the C-terminus, meaning that the predicted transmembrane helix 8 would have a different sequence, and would likely be shorter (hZIP14_C has 481 residues rather than 492). As shown in figures 3.17 and 3.21, hZIP14_C-HA was retained in the ER. With no experimental evidence of endogenous ZIP14_C expression in the literature, in any species, it is possible that the hZIP14_C-HA expressed in transfected cells was retained in the ER by ER quality control mechanisms because it could not be properly folded [215], and may not actually be expressed in human cells.

3.3.2 Metal Uptake by hZIP14 Variants

Iron uptake experiments were used to examine metal uptake function in stably-transfected cells, for the N-terminal mutant R128W, and the C-terminal mutants, P379L, G383R and N469K, relative to total expression, and cell-surface levels. Relative to cell surface levels, iron uptake was found to be unaffected by the P379L mutation, increased by the N469K mutation, and prevented by the R128W and G383R mutations (figure 3.11).

Iron uptake function of ZIP14 disease-causing mutants has not been measured previously, though zinc uptake by L441R hZIP14 has been studied [95], as well as manganese uptake by F98V, G383R and N469K hZIP14 [19].

Tuschl et al. [19] used stably transfected HEK293 cells to assay manganese uptake relative to total expression of the mutants. Although basal manganese uptake by HEK293 cells was not reported, limiting the comparisons that can be

made, these iron uptake results are in general agreement. Tuschl et al. [19] also found manganese uptake by G383R to be greatly decreased relative to wild-type. Uptake of Mn by N469K was lower than wild-type by ∼20% at similar expression levels [19], while iron uptake was decreased by ∼30%, relative to total expression levels (figures 3.8 and 3.10). By comparing iron uptake results to cell surface levels, I was able to determine that the decrease in iron uptake by the N469K mutant was due to decreased cell surface expression; cell surface N469K hZIP14_A-HA actually took up proportionately more iron than wild-type hZIP14_A-HA. Tuschl et al. [19] did not measure the surface levels of their hZIP14 mutants, but given these findings that cell surface levels of the N469K mutant are depressed relative to wild-type, in stably and transiently transfected cells, across a wide range of expression levels, it is likely that the effect of the N469K mutation on hZIP14 manganese uptake was similar to its effect on iron uptake.

Prevention of iron uptake by the G383R mutation was expected, since this position is part of the conserved metal transport motif (376-EEXPHEX**G**D). However, the results for the P379L mutant were surprising, as P379 is also part of the conserved metal transport motif (376-EEF**P**HELGD), and would therefore be expected to impact uptake function rather than trafficking. Nonetheless, these results showed that average iron uptake by hZIP14 P379L was decreased relative to wild-type in proportion with the decrease in cell surface levels of the P379L mutant, suggesting that iron uptake function was not affected, and that the mutation altered trafficking to the cell surface instead. Later flow cytometry experiments confirmed that the P379L mutant did alter trafficking to the surface in transiently transfected cells (figure 3.37).

Like G383R, the N-terminal R128W mutation abolished iron uptake function (figure 3.10). Tuschl et al. [19] reported very low levels of manganese uptake

(slightly higher than with G383R) for cells expressing hZIP14 with the F98V N-terminal mutation, though expression of this mutant was approximately twice that of wild-type hZIP14. These results both suggest that the N-terminal extracellular domain has a role in metal uptake by hZIP14.

3.3.2.1 Measurement of Manganese and Zinc Uptake Function May Provide Additional Insights

Iron uptake was used as a proxy for metal uptake function, as for logistical reasons, it was not possible to measure uptake of Zn and Mn. Selected point mutations in a founding member of the ZIP family, *Arabidopsis* IRT1, found some mutations that decreased Fe and Mn uptake without affecting uptake of Zn, and vice versa [218]. Therefore it is possible that some of the disease-causing mutations could alter iron uptake without affecting uptake of Mn or Zn to the same extent, or that iron uptake could be unaffected by a mutation that does alter uptake of Mn or Zn. It would therefore be ideal to measure manganese and zinc uptake function of hZIP14 and the pathological mutants; for example, by the use of inductively coupled plasma mass spectrometry, which would allow sensitive quantification of metal concentrations in cells incubated with different metals.

3.3.3 SDS-PAGE Analysis of hZIP14-HA Protein Extracts Suggests hZIP14-HA May Oligomerise

SDS-PAGE analysis of cell lysates was used to determine whether each of the hZIP14_A-HA variants could be successfully expressed in HeLa cells. Unexpectedly, the disease-causing mutations altered the patterns of bands in the western blots. As well as small shifts in the expected $\sim$52 kDa bands,

which were likely the result of alterations in detergent binding [213, 214], two additional bands at ∼130 and 180 kDa were consistently observed in wild-type hZIP14, and the C-terminal disease-causing mutants (figures 3.6 and 3.7).

Exogenous ZIP14 has been studied through western blotting previously, and the existence of higher molecular mass bands for this protein has been documented [13, 35]. Pinilla-Tenas et al. [35] suggested that the largest of the bands observed could represent an oligomeric form of murine ZIP14. Taylor et al. [13] described a 'trimer' band, and a 'higher molecular weight' band (above the range of their molecular marker), in addition to the bands in the expected size range. In non-reducing conditions, the lowest (∼50 kDa) band became fainter, while the other two darkened; after treatment with PNGase-F to remove N-linked oligosaccharides, all three bands decreased in size. These results suggested that the protein in these bands may have disulphide bridges, and that the higher molecular mass bands did not simply represent heavily glycosylated monomeric ZIP14.

Membrane proteins, particularly those with multiple membrane-spanning regions, are more hydrophobic than typical globular proteins, and as such, are not as easily denatured by sodium dodecyl-sulphate (SDS), making it possible for oligomeric forms of hZIP14 to persist in the conditions used for these experiments. Size determination of proteins in SDS-PAGE analysis relies on proportionate binding of negative SDS ions to unfolded protein, so if these higher molecular mass bands represent incompletely denatured oligomers of hZIP14, estimates of their size and therefore composition may be unreliable [213, 214].

The extracellular domain of *Pteropus alecto* ZIP4, a closely related protein, was crystallised in the dimeric form [219], and other ZIPs are also believed to form homo- or heterodimers [219–222]. Given that the N-terminal region of ZIP14

shares the highly conserved CPALLY motif at the core of the ZIP4 dimerisation interface [11, 219], it is highly probable that ZIP14 also forms a dimer, though this may only account for one of the two higher molecular mass bands.

These higher molecular mass bands were decreased or absent in the lysates of cells transfected with the N-terminal disease-causing mutants, only the $\sim$130 kDa band was faintly seen for the S104I and R128W mutants, and neither for the F98V mutant (figure 3.6). This is consistent with the hypothesis that the N-terminal region of hZIP14 is involved in its dimerisation, and that the N-terminal disease-causing mutations decrease that dimerisation.

The purpose of these western blots was primarily to confirm successful transfection and expression of the hZIP14 variants, rather than to detect the presence of oligomers. Other methods exist that are more suited to determine the quaternary structure of ZIP14, such as fluorescence correlation spectroscopy, which was recently used to determine that cell surface ZIP4 is dimeric [223], or analytical ultracentrifugation [224]. Further study of the oligomeric state of hZIP14 and its mutants may shed light on the role of oligomerisation in hZIP14 function.

3.3.4 Disease-Causing Mutations Affected Cellular Localisation and Cell Surface Levels of hZIP14

Immunofluorescence microscopy of cells transiently transfected with hZIP14 variants showed a strong impact on cellular localisation by many of the mutations, in particular those towards the N-terminus, which were largely retained in the ER (section 3.2.5). This finding has not been previously reported, and is in fact contradicted in the study of Tuschl et al. [19], who reported that F98V had no impact on cell localisation or cell surface levels. However,

these immunofluorescence microscopy results were in agreement with those of Hendrickx et al. [95], who found that the L441R mutation decreased cell surface levels of hZIP14.

Repeated immunofluorescent imaging of transiently transfected cells confirmed that hZIP14 mutants had different patterns of cellular localisation, and some time was spent testing various experimental methods that could best capture these differences (section 3.2.5.1). Imaging individual cells in order to gain quantitative data was time consuming (computational methods to achieve this were not available to me at that time), though seemed promising in terms of quantifying retention on the ER and the extent of co-localisation with wild-type hZIP14-GFP (figures 3.22 and 3.15). Regardless of the time constraints, in terms of understanding the effects of the mutations on the most salient aspects hZIP14 function, I decided to focus on the levels of the mutant proteins successfully reaching the plasma membrane. The differential staining protocol was successful in exposing a relationship between expression and cell surface levels that differed between mutants and wild-type hZIP14 - crudely by immunofluorescent imaging, and more sensitively by flow cytometry (figures 3.27, 3.28, 3.36 and 3.37). The effect of expression on cell surface levels of the N-terminal mutants suggests an explanation for the aforementioned contradiction with the F98V results of Tuschl et al. [19]. In those experiments, the cells imaged were stable transfectants, where the expression of F98V was approximately double that of the wild-type hZIP14. These flow cytometry results show that at high levels of expression, F98V can effectively localise at the cell surface (figure 3.36). Due to the use by Tuschl et al. [19] of monoclonal cell lines, in which only a narrow range of expression levels can be observed, the dependence of cell surface levels on expression by the F98V hZIP14 protein could not have been observed.

3.3.4.1 The Impact of Disease-Causing Mutations on Cell Surface Localisation at Different Levels of Expression

The flow cytometry experiments yielded several surprising results. The highly similar sigmoidal relationships between internal and surface levels of the N-terminal mutants were not anticipated - the relationship was hyperbolic for wild-type hZIP14 (figure 3.36). The very different, but nonetheless sigmoidal shape seen for the L441R mutant was also unexpected (figure 3.37). These mutants were all conspicuous in their failures to effectively localise to the cell surface, when observed under immunofluorescence microscopy.

As previously mentioned, the N-terminal extracellular domain may contribute to dimerisation of hZIP14, and mutations in this domain may therefore inhibit or prevent dimerisation. A possible explanation for the sigmoidal relationship between internal and surface hZIP14 for the N-terminal mutants, is that dimerisation of the extracellular domain is required for effective localisation at the cell surface, and the sigmoidal shape of the density plots reflects a lack of dimerisation at lower levels of hZIP14 expression, due to defects in the extracellular domain decreasing the affinity of the monomers.

Unlike F98V, S104I and R128W, L441R is *not* in the predicted N-terminal extracellular domain, which may explain why the sigmoidal relationship of L441R differs from the N-terminal mutants, but not why they all have this similar effect. The N469K mutation also resulted in a near-sigmoidal relationship between internal and surface hZIP14 (figure 3.37). Without structural information, it is difficult to hypothesize as to the cause of this effect by the N469K mutation.

Differential staining and flow cytometry also showed an impact of the P379L mutation on trafficking - at lower levels of expression, P379L was at higher levels at the cell surface than wild-type hZIP14_A-HA, while at high levels of

expression, there was a distinct decrease in cell surface levels of P379L (figure 3.37). As mentioned previously, the position of P379L in the predicted metal transport pore led to the reasonable prediction that its impact would be primarily on metal uptake rather than trafficking, yet it appears not to have affected hZIP14_A-HA iron uptake ability (figure 3.11). Additional experiments on uptake of other metals, and the regulation of hZIP14 trafficking by metals, could shed light on the unexpected behaviour of the P379L hZIP14_A-HA mutant.

Recent preliminary results from our lab [E. Pryke and Dr Bowers, unpublished data] indicate that the addition of zinc to the extracellular medium causes a rapid decrease in cell surface ZIP14, and it is unknown if this effect requires zinc uptake by ZIP14. Such results suggest additional complications, and further experiments would be needed to determine if short or long-term effects on trafficking were impacting cell surface levels of the human ZIP14 disease-causing mutants. It is notable that despite the increase in cell-surface localisation of hZIP14 P379L (over low and moderate expression levels) and G383R (at high expression levels) in transiently transfected cells (figure 3.37), in the stable cell lines all the mutants tested showed very low proportions of protein at the surface relative to wild-type (figure 3.12). This suggests that there may also be long-term regulation of hZIP14 cell surface levels that is affected by the disease-causing mutations. An important point to note when considering the effects of these mutations on cell surface levels of hZIP14, is that there are several points at which protein trafficking may be affected - decreased trafficking to the surface, increased endocytosis, decreased recycling or increased degradation could all decrease cell surface localisation.

A potential criticism of the flow-cytometry experiments is the use of transiently transfected cells, as they result in a wide range of expression levels, whose

physiological relevance is uncertain. ZIP14 expression is highly responsive to inflammation [20–25] - mRNA levels in mouse spleen and lung cells have been observed to increase as much as 90-fold in response to LPS (lipopolysaccharide) [23]. Since the resultant changes in ZIP14 protein expression have not been quantified, to what extent that range overlaps with these experimental conditions is not possible to say, however, a wide range of expression levels may well be physiologically relevant.

3.3.5 Summary of Effects of Disease-Causing Mutations

Based on the experiments in this chapter, the N-terminal mutations all appear to decrease trafficking to the cell surface, and are to a large extent retained on the ER (figures 3.13 and 3.36). The R128W mutant is also unable to transport iron (figure 3.10), the other N-terminal mutants may be similarly affected. Since the R128W mutation is not part of the transmembrane region, it is unlikely that this inability to transport metal is specific to iron. The N-terminal mutations appear to disrupt appropriate both cellular localisation, and metal uptake function.

The L441R and N469K mutations also impacted cell surface levels in an expression-dependent manner (figure 3.37). Even at high levels of expression, the proportions of these mutants at the cell surface did not approach that of wild-type hZIP14. Unlike the N-terminal mutations, the ability of these mutants to transport metal may not be negatively affected. L441R is reported to alter zinc transport within the cell, suggesting at least residual function [95]. N469K enhanced iron uptake function; though effects on manganese and zinc uptake may differ, the decrease in cell surface expression of N469K is sufficient to explain overall decreased function (figure 3.11).

The G383R mutant seemed to be largely unaffected in terms of trafficking over

the short-term (figure 3.14), though in stably-transfected cells, cell surface levels were much lower than would be expected from total expression (figure 3.12). In any case, cell surface levels were likely irrelevant; it did not transport iron (figure 3.10), and the position of the mutation in the key metal transport motif (376-EEFPHEL**G**D) makes it probable that the protein was non-functional.

The same was expected for the P379L mutation, found in the same conserved motif (376-EEF**P**HELGD). However, iron uptake was unaffected, while cell surface levels were decreased relative to wild-type, both in stably transfected cells and in transiently transfected cells expressing high levels of P379L hZIP14 (figure 3.11, 3.12 and 3.37). The lack of impact on iron uptake was not expected, though it remains possible that manganese and/or zinc transport were affected. However, like N469K, the decrease in cell surface levels is sufficient to explain the symptoms caused by the P379L mutation.

The experiments detailed in this chapter revealed effects of disease-causing mutations that were previously unreported in the literature, and suggested interesting avenues for further study. In contrast to previously held beliefs, these mutations have varied effects on trafficking and iron uptake function. Both increases and decreases in trafficking to the cell surface, relative to expression levels, were observed. Likewise, increases and decreases in iron uptake function were also found. At low levels of expression, N-terminal mutations appear to have similar impacts on trafficking, decreasing the levels at the cell surface and increasing ER retention, impacts that are diminished at higher expression levels, and C-terminal disease-causing mutations vary in their effects. These results suggest that the N-terminal ECD is involved in both trafficking and metal uptake of hZIP14, and raise questions about the mechanisms explaining some of the effects of the C-terminal mutations, which could be addressed with information on the locations of these mutations in the

three-dimensional structure of hZIP14.

Chapter 4

Homology Modelling of Human ZIP14

4.1 Introduction

The disease-causing mutations of human ZIP14 (hZIP14) had a range of impacts, affecting localisation and/or iron uptake function, in different combinations. The regions of hZIP14 in which these mutations are located may play different roles in the function of the protein, and so it was decided to develop a homology model of hZIP14. Such a model would allow me to hypothesize as to the mechanisms explaining the effects of the disease-causing mutations, and a second round of hZIP14 mutations could be designed to test these hypotheses.

4.1.1 Homology Modelling

Solving protein structures experimentally is a time-consuming and technically challenging process [225], particularly for membrane proteins, though there have been advances in this regard [226, 227]. For the purposes of understanding the effects of mutations, and designing new mutations, advances in computational modelling alongside increased availability of user-friendly interfaces, mean that structures of suitable accuracy can be produced in a much shorter time-frame [228]. The main strategies involved in protein structure prediction have been described as homology modelling, fold-recognition and *ab initio* folding [229–231]. Continuous developments mean that combinations of these strategies, as well as genetic information and the latest developments in technology, are increasingly being used to approach experimental levels of accuracy [232, 233].

Computational modelling accuracy is rigorously assessed in the biennial event organised by Critical Assessment of Structure Prediction (CASP) [234]; proteins whose structures have been experimentally solved but not published, are used as modelling targets. Participants use their modelling systems to predict the structure and the results are compared to the native structures. The most accurate models have previously been made by homology modelling [234], though this can only be done when a suitable homologous protein structure has been determined. Programs typically had similar levels of accuracy, as measured by deviation from the experimental structure [229, 234]. It should be noted however, that since the application of deep-learning strategies to protein structure prediction, the accuracy of models, including those that are considered 'difficult' (based on the existence/absence of similar experimental structures) has dramatically increased. In the latest round, CASP14, while

the quality of models generally increased, many AlphaFold2 models were considered to be more accurate than the experimentally determined structures [233, 234]. As these deep-learning methods were not publicly available when I needed a model of hZIP14, three different programs were selected for homology modelling of hZIP14: Phyre2, SWISS-MODEL and Modeller [192–194]. While Phyre2 accepts a target sequence (in this case, hZIP14), identifies a suitable homologous protein and develops its own alignment of the target sequence to the template sequence (the sequence of the protein structure to be used as a template), SWISS-MODEL and Modeller allow the user to submit their own alignment of template and target.

4.1.1.1 Developing the Structural Alignment

The accuracy of the alignment between the target and template sequences is a limiting factor in developing a high-quality homology model [235]. Pairwise alignment methods, which align two individual sequences, consider the similarity of amino acids, and produce an alignment that maximises the proportion of amino acids aligned with similar amino acids [236]. The main parameters used to optimise alignments are the choice of scoring matrix, and the gap penalty. Scoring matrices use different considerations, such as physico-chemical properties, or the observed likelihood of amino acid substitutions in homologues, to decide how similar two amino acids are [236]. Gap penalties are used to determine to which extent uninterrupted stretches of aligned amino acids should be prioritised over amino acid similarity within those aligned stretches [236]. For closely related proteins, a pairwise alignment may be sufficient for homology modelling [229].

For more distantly related homologues, pairwise alignments are unlikely to be accurate, in which case profile-profile alignments are more appropriate [229].

Profiles are multiple sequence alignments of homologous sequences which, with sufficient diversity in the sequences chosen, provide information on the most conserved regions of the target or template protein. The alignment algorithms are typically similar to those used in pairwise alignments, and employ strategies to manage the increased computational demands necessitated by the increased number of sequences. A common strategy is to use progressive alignments, where multiple pairs of sequences are aligned, and then those alignments are aligned, and so on, until all the sequences have been aligned to each other [237, 238]. A downside of this method is that misalignments can occur in the initial stages, which may then lead to an accumulation of errors [237]. Consistency-based algorithms, such as T-COFFEE, allow earlier alignments to be corrected when inconsistencies are detected, resulting in high levels of accuracy, though the number of sequences must be limited to manage computational requirements [229, 237]. Once profiles of the template and target proteins have been made, the most conserved residues are apparent, and an alignment between the two profiles makes use of this information to produce a more accurate alignment [229, 237]. Where two proteins are genuine homologues that share a protein structure, an accurate alignment will not contain gaps that interrupt regions of secondary structure; the final alignment between the template and target may require manual adjustment to meet this requirement [239]. For the modelling of the extracellular and transmembrane domains of hZIP14, multiple methods were used to produce alignments of the targets to the template sequences. Where there were disagreements, profile alignments of the templates were used to identify the most important residues in the template sequence, and the alignments that best preserved the physico-chemical properties of the conserved template residues were chosen for the final structural alignments. These alignments were then used for homology

modelling of hZIP14 in SWISS-MODEL and Modeller [192, 194]; only the target sequence was submitted to Phyre2 [193].

4.1.2 Quality Assessment

Assessment of protein models essentially aims to quantify the plausibility of the structure using molecular mechanics-based, or knowledge-based approaches [229]. Molecular mechanics treats proteins as detailed atomic models and uses equations to represent the laws of physics, while knowledge-based approaches use frequency distributions of features from experimentally-determined protein structures to assess the plausibility of those same features in the model being assessed [229]. While molecular dynamics approaches are computationally expensive, the huge number of high quality structures for analysis has improved the results of knowledge-based approaches [229]; in practice, these two approaches can be combined, with care to avoid over-emphasizing features by scoring them multiple times. Since I was using three different homology modelling programs, I used quality assessment programs to select the highest quality homology model produced by the three programs. The SWISS-MODEL server offers a quality assessment program (QMeanDISCO) that produces a detailed analysis of submitted structures, considering multiple aspects of the structure, and has even assessed the importance of those features in their contribution to the overall quality score [240]; additionally, there is a separate assessment program specifically trained on membrane proteins, which have some very different features to soluble proteins, due to their existence in the hydrophobic bilayer, rather than in an aqueous solution [200]. In case the Q-MEAN programs showed bias towards SWISS-MODEL structures as a result of having potentially used the same criteria in modelling as assessment [229], I also assessed the model with two additional programs, ANOLEA and ProSA-

Web, knowledge-based programs that make use of the frequency distribution of long-range residue interactions, and Cα distances and solvent accessibilities, respectively [198, 199]. The most highly ranked models were chosen, and used to predict the functional effects of the disease-causing mutations, and to choose new sites for mutagenesis.

4.2 Results

The aim of this work was to generate structural models of the extracellular and transmembrane domains of human ZIP14 (hZIP14), in order to use them as the basis for further experiments. The solved crystal structure of the *Pteropus alecto* (the 'black flying fox') ZIP4 N-terminal extracellular domain (P_ECD, PDB id: 4X82) [219] was used to model the N-terminal extracellular domain (ECD) of hZIP14 (section 4.2.2). The transmembrane domain (TMD) was modelled using the crystal structure of *Bordetella bronchiseptica* (bacterial) ZIP (BbZIP, PDB id: 5TSA) [241] (section 4.2.10). Three models were produced for each domain, using different software, which were then analysed by protein structure quality assessment programs (sections 4.2.3 and 4.2.11). The chosen structures were then used to design mutations to test hypotheses about the function of hZIP14 (sections 5.1.1 and 5.1.2).

4.2.1 Alignment of the N-Terminal Extracellular Domain

The crystal structure to be used as a template for the modelling of the hZIP14 ECD is from the species *Pteropus alecto* [219]. Zhang et al. [219] proposed a partial alignment of the ECDs of various members of the LIV-1 protein family with the C-terminal end of the *P. alecto* ZIP4 ECD. This alignment focused on the key regions, neglecting some parts of the ZIP4 and ZIP14 sequences.

As a first step towards a complete alignment, a simple pairwise alignment of *P. alecto* ZIP4 and human ZIP14 was carried out. This pairwise alignment differed from the partial alignment of [219] in ECD alpha-helix 10 (H 10), and the inter-helical sequence between Hs 10 and 11, but was otherwise identical (figure 4.1) (for the full pairwise alignment of the N-terminal regions, see appendix A.1).

```
                  H 10                H 11
                  ----------          ----
P_ECD        194  TPQYFVDFVFQQSHGNTPNISVAEL  218
hZIP14 Pw.    40  AASFLQDLIHRYGEGD--SLTLQQL  62
hZIP14 Zh.    39  SAASFLQDLIHRYGEGDS-LTLQQL  62
```

Figure 4.1: **Alignments covering helix 10 of P_ECD.** Pairwise alignment (Pw.) and alignment from Zhang et al. [219] (Zh.) of hZIP14 with the *P. alecto* ZIP4 ECD, over H 10 and the N-terminal end of H11; the black line indicates residues of Hs 10 and 11. The pairwise alignment was similar to the partial alignment provided by Zhang et al. [219] (Zh.), but differed over H 10 of the P_ECD.

Multiple sequence alignments ($\sim$50 sequences) of mammalian proteins are available at aminode.com [205]. These multiple sequence alignments provide information on which residues are the most conserved, and can aid in producing an accurate alignment. The ZIP4 (SLC39A4) and ZIP14 (SLC39A14) alignments were downloaded and used for a profile-profile alignment, which was in full agreement with Zhang et al. [219] and included the residues not shown in that partial alignment (for the full profile alignment of the N-terminal regions, see appendix A.2). This alignment was used as the structural alignment for the homology modelling of the extracellular domain of hZIP14.

4.2.2 Homology Modelling of the ECD

The ECD structural alignment was submitted to the SWISS-MODEL web interface [194] and the Modeller software [192] to generate structural models

of hZIP14. The Phyre2 web interface [193], a web server that accepts only the target sequence as input, was also used. For each model, the SCWRL software program [195] was then used to optimise the orientation of the side-chains. The resulting structures are shown in figure 4.2, with the template ZIP4 ECD (P_ECD) for comparison.

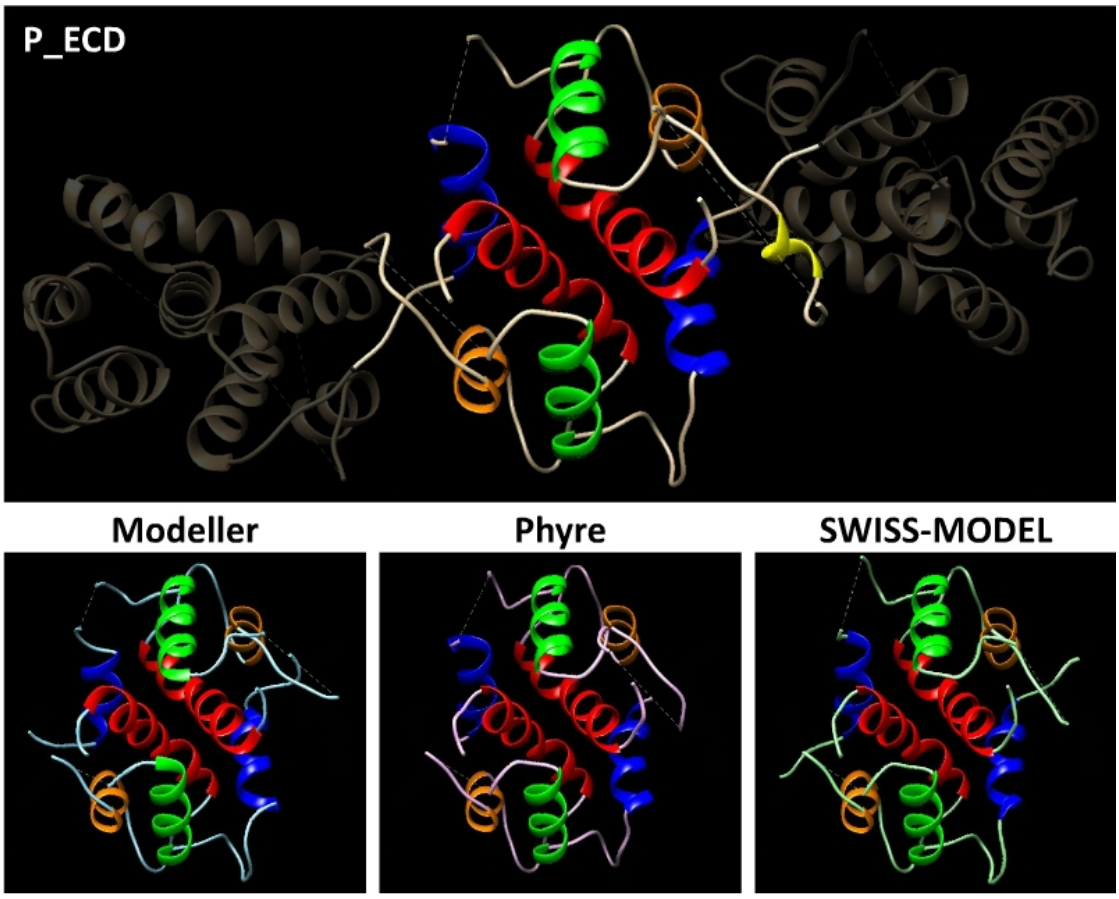


Figure 4.2: **Homology models of the N-terminal extracellular domain.** Upper panel: The N-terminal extracellular domain of *P. alecto* ZIP4, with helices 10-14 depicted in colour (blue to red). Lower panel: Homology models of the N-terminal extracellular domain of human ZIP14; helices 1-4 are coloured from blue to red.

The *P. alecto* ZIP4 ECD structure was solved in the dimeric form, with each monomer forming similar, though not identical structures. Chain B has 14 alpha helices, while chain A is missing the second and twelfth helices. The human ZIP14 ECD aligns with helices 10, 11, 13 and 14 of the *P. alecto* ZIP4 ECD, and

according to the SCWRL software [195], the C118 residues of each monomer form an inter-molecular disulphide bridge.

4.2.3 Quality Assessment of the N-Terminal ECD Homology Models

Quality assessment software use a variety of scoring functions to predict the reliability of protein structures. Three software programs that use different methods were used to score the quality of the homology models - ProSA-web [198], ANOLEA [199] and QMEANDisCo [240].

Based on a database of experimentally determined globular protein structures, the ProSA web-server assesses the plausibility of protein folds by looking at the distributions of Cα distances and solvent accessibilities of residues, and assigns structures an overall Z-score (a lower Z-score indicates a higher quality) [198, 242].

The ANOLEA software compares the non-local atomic interactions (more than 11 residues apart) of submitted structures to those in a database of structures of non-redundant proteins, returning an energy plot smoothed over 5 residues, and an overall Z-score for structures.

The QMEANDisCo software draws on a number of resources for quality assessment, such as programs that predict secondary structure or solvent accessibility and datasets of high-resolution structures, in order to assess several aspects of submitted structures. These assessments are combined to provide an overall quality score, ranging from 0 to 1 (low to high quality). The scores from each assessment program are shown in table 4.1.

Each assessment tool ranks the template structure as higher quality than the

Table 4.1: **The Phyre2 structure is ranked as the best quality homology model.** The ProSA-web and ANOLEA tools each assess structure quality using different parameters, producing a Z-score that summarises overall quality, and the QMEANDisCo (QMEAN) tool provides an overall quality score. Higher quality is indicated by a lower Z-score and a higher global score. ProSA-web and ANOLEA assessed each monomer (chains A and B) of the ECD homology model separately; QMEANDisCo assessed the dimeric structure.

	Z-Score							
Tool	**P_ECD**		**Modeller**		**Phyre2**		**SWISS-M**	
	A	B	A	B	A	B	A	B
ProSA-web	-7.84	-5.47	-3.85	-2.82	-3.89	-3.21	-3.65	-2.53
ANOLEA	0.95	1.72	6.59	9.36	4.36	4.89	6.18	7.06
	Global Quality Score							
QMEAN	0.89±0.05		0.58±0.07		0.63±0.07		0.62±0.07	

homology models, as expected. The Phyre2 structure was ranked as the best of the homology models, for both chains A and B. A trend observed for all the assessments was a better quality assessment for chain A than B, including for the template structure. The QMEANDisCo local quality scores also ranked chain A as higher quality than B for each of the structures. In the template 4X82 structure, multiple regions of chain B were unresolved; residues flanking these regions had the worst per-residue ratings, possibly explaining the lower ratings of chain B. For more detailed quality assessment results, see appendix C.1.

The error margin of the QMEANDisCo global quality scores shows that the scores were not significantly different for the three homology models. However, as all three assessment programs rated the Phyre2 model as the best quality, this was chosen as the basis for further experiments.

4.2.4 Pairwise Alignment Indicates Low Sequence Identity Between BbZIP and hZIP14

The recently solved crystal structure of *Bordetella bronchiseptica* ZIP (BbZIP) (resolution 2.40 Å) has been used to create a computational homology model of hZIP4 [241]. Since BbZIP is a bacterial protein, the evolutionary distance between BbZIP and hZIP14 is much greater than that between P_ECD (a mammalian protein, from the black flying fox, a fruit bat) and hZIP14. As an initial step in assessing the suitability of BbZIP for homology modelling of hZIP14, the sequences of BbZIP, hZIP4 and hZIP14 were compared by pairwise alignment.

The pairwise sequence alignments suggested slightly greater sequence identity and similarity between BbZIP and hZIP14, than between BbZIP and hZIP4 (table 4.2), supporting the use of the BbZIP crystal structure for modelling of hZIP14. A sequence identity of 18.8% is within the so-called 'twilight zone', where the accuracy of homology modelling is limited [235]. However, the pairwise alignments cover whole sequences, including loops between transmembrane helices which are not expected to be conserved. Alignments for conserved regions, such as transmembrane helices, are likely to have higher sequence identities, and modelling for these regions should be more accurate.

Closer examination of these pairwise alignments showed significant gaps and insertions in the transmembrane helices (TMs) of the BbZIP structure, (figure 4.3), making the pairwise alignment unsuitable for homology modelling; the alignment used by Zhang et al. [241] for homology modelling of hZIP4 was substantially adjusted to provide a more plausible structural alignment.

A further indication that the pairwise alignment was not suitable for homology modelling was the lack of alignment between the transmembrane helices of the

Table 4.2: **Sequence identity and similarity between BbZIP, hZIP4 and hZIP14.** Sequence identity and similarity by pairwise alignment. Alignments were produced using the Needleman-Wunsch algorithm [189]. The sequence identity and similarity between BbZIP and hZIP14 was sufficient to consider homology modelling.

	Sequence Identity		Sequence Similarity	
Protein	**hZIP4**	**hZIP14**	**hZIP4**	**hZIP14**
BbZIP	13.1%	18.8%	21.1%	30.2%
hZIP4	-	22.2%	-	34.0%

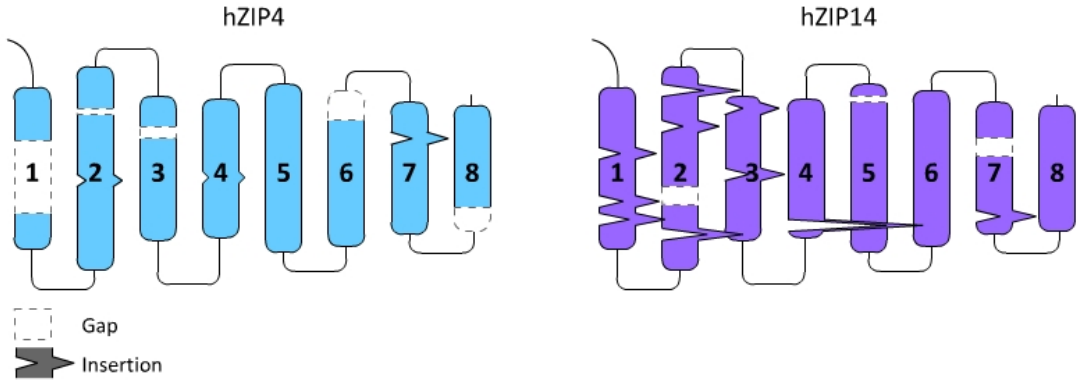


Figure 4.3: **Insertions and gaps in the pairwise alignments.** Diagram representing insertions and gaps in the pairwise alignments of hZIP4 and hZIP14, relative to the secondary structure of BbZIP; the lengths of the gaps and the widths of the insertions are proportional to the number of residues missing or inserted.

BbZIP crystal structure (PDB id: 5TSA) and the hZIP14 TM helices predicted by Taylor et al. [13] (figure 4.4). BbZIP TMs 1-3 corresponded particularly poorly with the predicted regions of TMs 1-3 in hZIP14, while TMs 4-8 of BbZIP were aligned relatively well with the predicted TMs 4-8 of hZIP14, albeit with both gaps and insertions (see appendix A.3 for pairwise alignments).

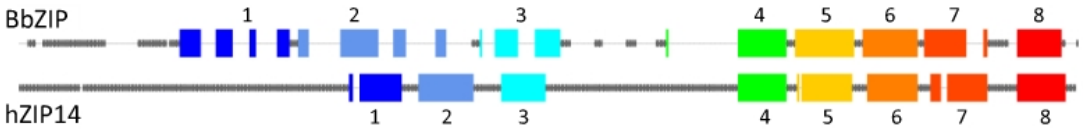


Figure 4.4: **Predicted TMs of hZIP14 are not all aligned with BbZIP TMs in the pairwise alignment.** Diagram illustrating the pairwise alignment of BbZIP TMs (based on the 5TSA crystal structure [241]) with the predicted TMs of hZIP14 (from Taylor et al. [13]).

4.2.5 Pfam-derived Alignment Covers TMs 4-8

In order to obtain a more suitable structural alignment of hZIP14 and BbZIP, hZIP14 was searched against Pfam, a protein family database of more than 17,000 families [243]. The Pfam database assigned both BbZIP and hZIP14 to the ZIP zinc transporter family (PF02535), and had a multiple sequence alignment (MSA) for this family of 17,347 sequences [243]. The MSA includes a truncated hZIP14 sequence (residues 151-483) and a truncated ZIP protein closely related to BbZIP (residues 145-305), from *Bordetella pertussis* (sequence identity 98.7%), corresponding to TMs 4-8 of the BbZIP sequence.

The near complete sequence identity between *B. pertussis* ZIP and *B. bronchiseptica* ZIP (BbZIP) meant that a pairwise alignment between these two proteins would be accurate. The BbZIP-*Bordetella pertussis* ZIP pairwise alignment was combined with the hZIP14-*Bordetella pertussis* ZIP alignment from Pfam's large MSA, to create a direct alignment between hZIP14 and BbZIP, covering TM helices 4-8 of the BbZIP structure, here referred to as the Pfam-derived alignment (see appendix A.4 for the Pfam-derived alignment).

A comparison of the BbZIP TM helices with the predicted TM helices of hZIP14 according to the Pfam-derived alignment shows smaller, and fewer gaps and insertions than those seen in the pairwise alignment for TMs 4-8 (figures 4.4, 4.5 and 4.6).

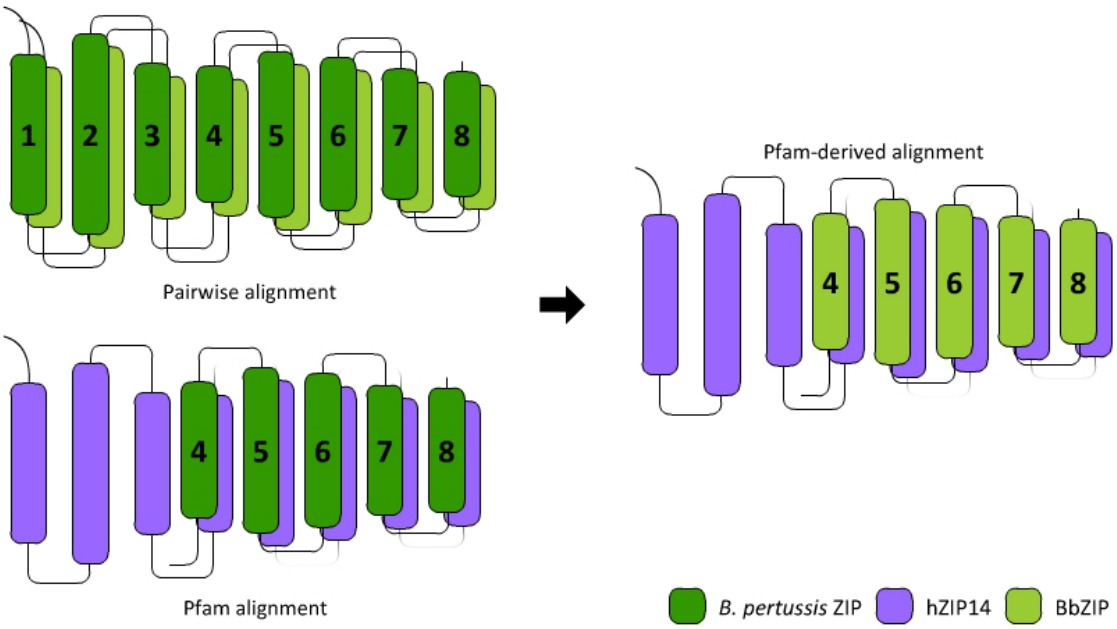


Figure 4.5: **The Pfam-derived alignment covers TMs 4-8.** The Pfam alignment of hZIP14 with *B. pertussis* ZIP was used to create a partial alignment of BbZIP with hZIP14, the Pfam-derived alignment.

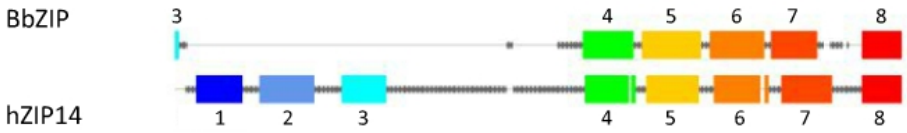


Figure 4.6: **Alignment of predicted TMs in hZIP14 with BbZIP TMs in the Pfam-derived alignment.** Diagram illustrating the Pfam-derived alignment of BbZIP TMs (based on the 5TSA crystal structure [241]) with the predicted TMs of hZIP14 (from Taylor et al. [13]). The Pfam-derived alignment aligned BbZIP TMs 4-8 with TM regions predicted for hZIP14.

4.2.6 Profile Alignment

An alternative alignment strategy that can be used to address the entire transmembrane region of BbZIP, and is superior to pairwise alignments, is to use profile-profile alignment - an alignment of two multiple sequence alignments. The T-COFFEE multiple alignment software [189] accepts a maximum 500 sequences. The BbZIP and hZIP14 (isoform A) protein sequences were each used to identify similar proteins with Blast [244], and the results were

filtered to exclude synthetic, predicted and partial sequences. Sequences from the Bordetella genus or the *Homo sapiens* species, respectively, were also removed, to increase the diversity of sequences in each alignment. This resulted in 398 homologous sequences for the hZIP14 MSA, and the remaining BbZIP homologous sequences were randomly sampled to obtain 499 homologous sequences.

With the addition of BbZIP and hZIP14 (isoforms A and B) to their respective sets of sequences, the two sets were each aligned using T-COFFEE [189], and then these multiple sequence alignments were aligned to each other using MUSCLE's 'profile-profile' option [190], resulting in an alignment between hZIP14 and BbZIP, here referred to as the profile alignment (see appendix A.5 for the profile alignment).

This alignment differed from the previous alignments in that it showed a strong overlap between the BbZIP TMs and the corresponding predicted TMs of hZIP14, for TMs 1-3 (figure 4.7). Gaps and insertions still occurred, particularly in TM 1, so this alignment still required analysis and adjustment. The results of the profile alignment for TMs 4-8 closely matched those of the Pfam-derived alignment.

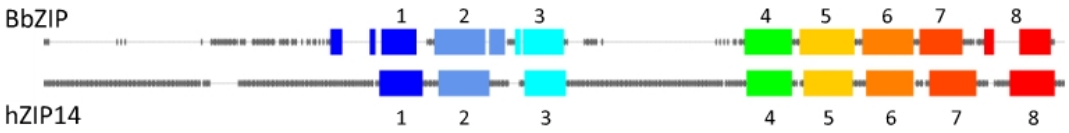


Figure 4.7: **Alignment of predicted TMs in hZIP14 with BbZIP TMs in the profile alignment.** Diagram illustrating the profile alignment of BbZIP TMs (based on the 5TSA crystal structure [241]) with the predicted TMs of hZIP14 (from Taylor et al. [13]).

4.2.7 Comparison of Alignments for Individual Transmembrane Helices

For several of the TM helices, the three alignments were largely in agreement and required only small adjustments to address gaps or insertions. The pairwise and profile alignments for TMs 1, 2 and 3 differed greatly and required more analysis to produce a structural alignment suitable for homology modelling.

4.2.7.1 Transmembrane Helices 4, 5, 6 and 8

For TM helices 4-8, there were small differences between the three alignments, occurring at the extreme ends of the helices (figures 4.8, 4.9, 4.10 and 4.11). For each of these TM helices, the alignment that covered the full helix without gaps or insertions was chosen for the final structural alignment (figure 4.22).

```
BbZIP       166  R--------------------------------VWLFVL  172
hZIP14 Pw.  303  EKVIVGSLSVQDLQASQSACYWLKGVRYSDIGTLAWMITL  342
hZIP14 Pf.  336  L-------------------------------AWMITL  342
hZIP14 Pr.  336  L-------------------------------AWMITL  342

BbZIP       173  TIILHNLPEGMAIGVSF                       189
hZIP14 Pw.  343  SDGLHNFIDGLAIGASF                       359
hZIP14 Pf.  343  SDGLHNFIDGLAIGA-S                       358
hZIP14 Pr.  343  SDGLHNFIDGLAIGASF                       359
```

Figure 4.8: **Alignments covering TM 4 of BbZIP.** Alignments of BbZIP and hZIP14; pairwise (Pw.), Pfam-derived (Pf.) and profile (Pr.) alignments. The alignments are very similar, with small differences at the N- and C-termini.

```
BbZIP       194  LRIGLPLTSAIAIQDVPEGLAVALALRA  221
hZIP14 Pw.  364  FQ-GISTSVAILCEEFPHELGDFVILLN  390
hZIP14 Pf.  363  VFQGISTSVAILCEEFPHELGDFVILLN  390
hZIP14 Pr.  363  VFQGISTSVAILCEEFPHELGDFVILLN  390
```

Figure 4.9: **Alignments covering TM 5 of BbZIP.** Alignments of BbZIP and hZIP14; pairwise (Pw.), Pfam-derived (Pf.) and profile (Pr.) alignments. The alignments are near identical.

```
BbZIP       226  IGRAVLVAVASGLMEPLGALVGVGIS  251
hZIP14 Pw.  395  IQQALFFNFLSACCCYLGLAFGILAG  420
hZIP14 Pf.  395  IQQALFFNFLSACCCYLGLAFGIL--  418
hZIP14 Pr.  395  IQQALFFNFLSACCCYLGLAFGILAG  420
```

Figure 4.10: **Alignments covering TM 6 of BbZIP.** Alignments of BbZIP and hZIP14; pairwise (Pw.), Pfam-derived (Pf.) and profile (Pr.) alignments. The alignments are near identical.

```
BbZIP       287  ETTAT-------------VGLMAGFALMMFLDTA  307
hZIP14 Pw.  465  FIIQN-------------LGLLTGFTIMVVLTMY  485
hZIP14 Pf.  465  FIIQN-------------LGLLTGFTIMVVLT--  483
hZIP14 Pr.  455  DE---RKGSILIPFIIQNLGLLTGFTIMVVLTMY  485
```

Figure 4.11: **Alignments covering TM 8 of BbZIP.** Alignments of BbZIP and hZIP14; pairwise (Pw.), Pfam-derived (Pf.) and profile (Pr.) alignments. The alignments are very similar, with differences at the N- and C-termini.

4.2.7.2 Transmembrane Helix 7

The Pfam-derived and profile alignments for TM 7 were in near total agreement, while the pairwise alignment introduced a gap and an insertion (figure 4.12). The Pfam-derived alignment was used for the final alignment, as it contained no gaps or insertions and covered the entire TM 7 region.

```
BbZIP        255  ALAYPISMGLAAGAMIFVVS--------HE  276
hZIP14 Pw.   424  SANWIFAL---AGGMFLYISLADMFPEMNE  450
hZIP14 Pf.   422  HFSANWIFALAGGMFLYISL--------AD  443
hZIP14 Pr.   423  FS-ANWIFALAGGMFLYISL--------AD  443
```

Figure 4.12: **Alignments covering TM 7 of BbZIP.** Alignments of BbZIP and hZIP14; pairwise (Pw.), Pfam-derived (Pf.) and profile (Pr.) alignments. The pairwise, Pfam-derived and profile alignments all share a common region; the Pfam-derived and profile alignments are the most similar.

4.2.7.3 Transmembrane Helices 1, 2 and 3

The pairwise and profile alignments for TM helices 1, 2 and 3 produced conflicting results (figures 4.13, 4.14 and 4.15). The profile alignment is based on multiple sequences, which provide information on conserved residues and improve the quality of the alignment, however the large evolutionary distance between *Bordetella bronchiseptica* and *Homo sapiens* increases the likelihood of misalignment. For this reason, the BbZIP-hZIP4 structural alignment by Zhang et al. [241] was consulted to aid in the creation of the BbZIP-hZIP14 structural alignment.

```
BbZIP         54  VHVGYA--------------VLG---G-------AAGFAA  69
hZIP14 Pw.    76  GNVTQH--------------VQG---HRNLSTCFSSGDLF  100
hZIP14 Pr.   132  SENQENEENEQTEEGRPSAVEV-WGYG-------LLCVTV  163

BbZIP         70  TA--------LGA----------LMALGL  80
hZIP14 Pw.   101  TAHNFSEQSRIGSSELQEFCPTILQQLDS  127
hZIP14 Pr.   164  IS--------LCS----------LLGASV  174
```

Figure 4.13: **Alignments covering TM 1 of BbZIP.** Alignments of BbZIP and hZIP14; pairwise (Pw.) and profile (Pr.) alignments. The pairwise and profile alignments differ substantially, with no similar regions.

```
BbZIP         85  ARTQD---------------AMLGFAAGMMLAASAFSL--  107
hZIP14 Pw.   132  SENQENEENEQTEEGRPSAVEVWGYG---LLCVTVISLCS  168
hZIP14 Pr.   184  KRLLL---------------YFIALAIGTLYSNALFQL--  206

BbZIP        108  -----ILP--GLD--------------AAGTI  118
hZIP14 Pw.   169  LLGASVVP--FMKKTFYKRLLLYFIALAIGTL  198
hZIP14 Pr.   207  -----IPEAFGFN--------------PLEDY  219
```

Figure 4.14: **Alignments covering TM 2 of BbZIP.** Alignments of BbZIP and hZIP14; pairwise (Pw.) and profile (Pr.) alignments. The pairwise and profile alignments differ substantially, with no similar regions.

```
BbZIP        123  P------AA-AAVVALGLG--------LGVLLMLGLDYF  146
hZIP14 Pw.   215  PLEDYYVSK-SAVVFGGFYLFFFTEKILKILLKQKNEHH  252
hZIP14 Pr.   222  --------SKSAVVFGGFY--------LFFFTEKILKIL  244
```

Figure 4.15: **Alignments covering TM 3 of BbZIP.** Alignments of BbZIP and hZIP14; pairwise (Pw.) and profile (Pr.) alignments. The pairwise and profile alignments differ substantially, with no similar regions.

4.2.8 hZIP4-Derived Alignment

The BbZIP crystal structure has been previously used to create a homology model of the transmembrane domain of hZIP4 [241], a protein that, like hZIP14, is a member of the LIV-1 family [11]. A comparison between the structural alignment used for that model and the alignment results for BbZIP and hZIP14 produced here could be instructive.

A profile-profile alignment between hZIP14 and hZIP4 was made in order to create a BbZIP-hZIP14 alignment based on the BbZIP-hZIP4 alignment used by Zhang et al. [241], here referred to as the hZIP4-derived alignment. For TMs 4, 5, 6, 7 and 8, the hZIP4-derived alignment was in agreement with the alignment constructed thus far (see appendix A.6 for the hZIP4-derived alignment). The hZIP4-derived alignment for TM2 was most similar to the profile alignment, but

had no gaps or insertions (figure 4.16); this was used for the final structural alignment. The hZIP4-derived alignment differed from the profile alignment for both TMs 1 and 3.

```
BbZIP         85  ARTQD---------------AMLGFAAGMMLAASAFSL--  107
hZIP14 Pw.   132  SENQENEENEQTEEGRPSAVEVWGYG---LLCVTVISLCS  168
hZIP14 Pr.   184  KRLLL---------------YFIALAIGTLYSNALFQL--  206
hZIP14 Z4.   184  KRLLL---------------YFIALAIGTLYSNALFQL--  206

BbZIP        108  -----ILP--GLD--------------AAGTI  118
hZIP14 Pw.   169  LLGASVVP--FMKKTFYKRLLLYFIALAIGTL  198
hZIP14 Pr.   207  -----IPEAFGFN--------------PLEDY  219
hZIP14 Z4.   207  -----IPE--AFG--------------FNPLE  217
```

Figure 4.16: **Alignments covering TM 2 of BbZIP.** Alignments of BbZIP and hZIP14; pairwise (Pw.), profile (Pr.) and hZIP4-derived (Z4.) alignments. The hZIP4-derived alignment is most similar to the profile alignment, but contains no gaps or insertions.

4.2.8.1 Transmembrane Helix 1

The profile alignment for TM 1 contained a gap and two insertions; figure 4.17 shows how this alignment was adjusted to preserve the majority of the alignment while closing the gap and removing the insertion. Despite these changes, the adjusted profile alignment and the hZIP4-derived alignment were offset by 4 residues.

The most conserved residues in the BbZIP TM 1 are T70 and G73 (figure 4.18); in the BbZIP MSA (section 4.2.6) these positions were invariant. In the BbZIP 5TSA crystal structure, the conserved residues of TM 1 line the side of the helix facing TMs 5 and 6 (figure 4.19 A), with T70 interacting with D208 on TM 5 (figure 4.19 B). According to the hZIP4-derived alignment, BbZIP residues T70 and D208 correspond to hZIP14 residues S168 and E377. Serine and threonine have similar chemical properties, as both side-chains have a hydroxyl

```
BbZIP         54  VHVGYA--------------VLG---GAAGFAATALGALM  76
hZIP14 Pr.   132  SENQENEENEQTEEGRPSAVEV-WGYGLLCVTVISLCSLL  170
hZIP14 A.Pr  132  PSAVEV--------------WGY---GLLCVTVISLCSLL  170
hZIP14 Z4.   152  --WGYG--------------LLC---VTVISLCSLLGASV  174

BbZIP         77  ALGL  80
hZIP14 Pr.   171  GASV  174
hZIP14 A.Pr  171  GASV  174
hZIP14 Z4.   175  VPFM  178
```

Figure 4.17: **Alignments covering TM 1 of BbZIP.** Alignments of BbZIP and hZIP14; profile (Pr.), adjusted profile (A.Pr) and hZIP4-derived (Z4.) alignments.

group; aspartate and glutamate also share similar chemical properties, each side-chain terminating in a carboxylic acid group. The other highly conserved BbZIP TM 1 residue, G73, is closely packed against TM 5 (figure 4.19 C); a larger residue would likely clash, potentially disrupting the structure of the protein. The hZIP4-derived alignment also has a glycine residue (G171) in the corresponding position, a highly conserved residue according to the hZIP14 MSA developed for the profile alignment (section 4.2.6) (see appendix B.2.2 for raw conservation scores).

Both the sequence identity and similarity of the TM 1 hZIP4-derived alignment were higher than the adjusted profile alignment (the sequence identity and similarity for the hZIP4-derived alignment are 21.4 and 39.3%, respectively, compared to 7.1 and 21.4% for the adjusted profile alignment). The hZIP4-derived alignment also preserved the properties of the most conserved BbZIP TM 1 residues, which may serve key functions in the structure of the protein (see appendix B.2.1 for raw conservation scores).

In contrast, the adjusted profile alignment had an isoleucine in place of the highly conserved T70, and a cysteine in place of the highly conserved G73. For TM 1, the adjusted profile alignment resulted in lower sequence identity

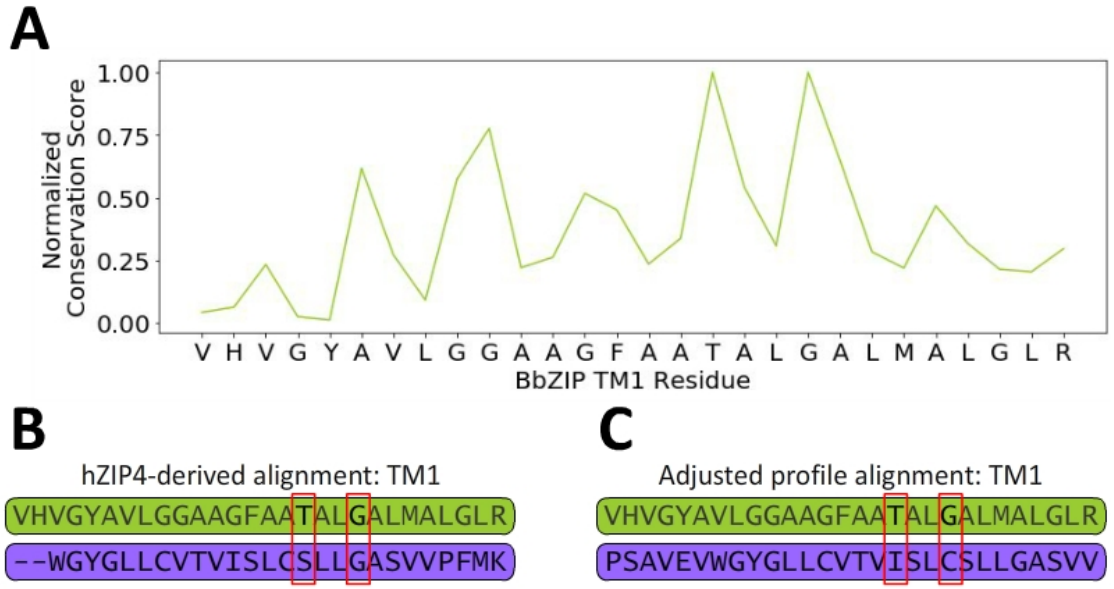


Figure 4.18: **Properties of conserved BbZIP TM 1 residues are retained in the hZIP4-derived alignment.** A: Graph of normalised conservation scores across the BbZIP TM 1. B: BbZIP-hZIP14 TM 1 hZIP4-derived alignment. C: BbZIP-hZIP14 TM 1 adjusted profile alignment. B,C: Red boxes indicate the most highly conserved residues of BbZIP TM 1. The physico-chemical properties of the most highly conserved BbZIP TM 1 residues are most closely matched in the hZIP4-derived alignment.

and similarity, and replaced conserved residues with dissimilar amino acids. Due to the higher sequence identity and similarity of the hZIP4-derived TM 1 alignment, as well as the agreement in the chemical properties of the most highly conserved BbZIP TM 1 residues, the hZIP4-derived TM 1 alignment was used for the final structural alignment.

4.2.8.2 Transmembrane Helix 3

The profile alignment for TM 3 contained one insertion; figure 4.20 shows how this alignment was adjusted to preserve the majority of the alignment while removing the insertion. Despite these changes, the adjusted profile and the hZIP4-derived alignment are offset by 4 residues.

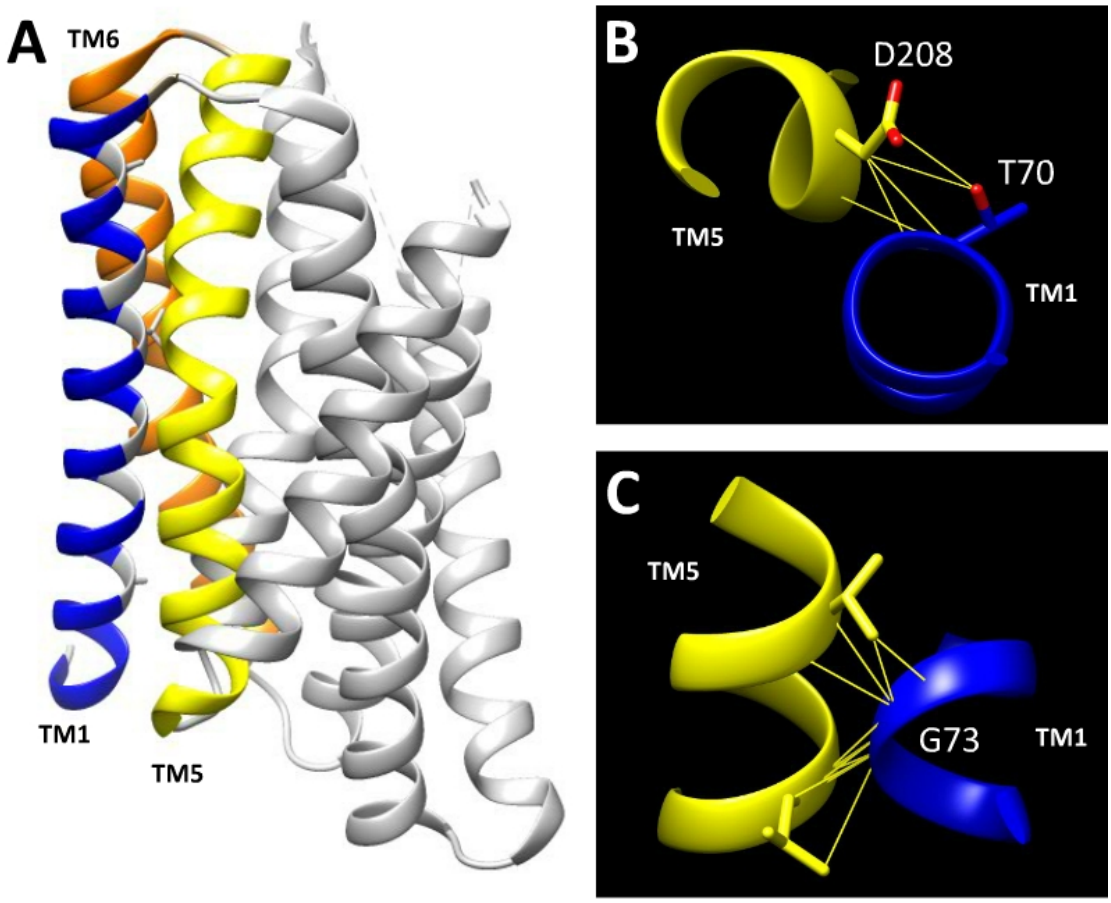


Figure 4.19: **TM 1 conserved residue interactions.** Images of conserved TM 1 residues and interactions. A: BbZIP ribbon representation, conserved TM 1 residues in white. B: Interaction between T70 and D208. C: Interactions between G73 and TM 5 residues. B,C: Thin lines link atoms close enough to interact (Van Der Waals overlaps of $\geq$ -0.4 Å, as identified by UCSF Chimera [197]). The conserved BbZIP TM 1 residues line the face of the alpha-helix facing TMs 5 and 6. T70 interacts with a TM 5 aspartate residue; G73 allows close packing between TM 1 and TM 5.

```
BbZIP       123  PAA-AAVVALGLGLGVLLMLGLDYF  146
hZIP14 Pr.  222  --SKSAVVFGGFYLFFFTEKILKIL  244
hZIP14 A.Pr 221  VSK-SAVVFGGFYLFFFTEKILKIL  244
hZIP14 Z4.  222  ------SKSAVVFGGFYLFFFTEKI  240
```

Figure 4.20: **Alignments covering TM 3 of BbZIP.** Alignments of BbZIP and hZIP14; profile (Pr.), adjusted profile (A.Pr) and hZIP4-derived (Z4.) alignments.

The hZIP4-derived alignment was not adjusted to cover the N-terminal region of the TM 3 alignment as this would lead to overlap with C-terminal hZIP14 residues from the TM 2 alignment, as a result, in this alignment the TM 3 helix is significantly shortened.

The adjusted profile alignment for TM3 had a higher sequence identity and similarity (25% and 33.3%) than the hZIP4-derived TM 3 alignment (8.3% and 20.8%). The adjusted profile alignment also shared more similar amino acid residues at the highly conserved positions of BbZIP TM 3 (figure 4.21); the invariant BbZIP A216 is similar in size to serine (if the hZIP4-derived sequence was adjusted, this position would be filled by a tyrosine residue), and the invariant BbZIP V129 is matched by an identical valine in the adjusted profile alignment, rather than the highly dissimilar lysine of the hZIP4-derived alignment (figure 4.21 B and C). Due to the increased sequence identity and similarity of the TM 3 adjusted profile alignment, and the relatively preserved chemical characteristics of the most conserved BbZIP TM 3 residues, the adjusted profile alignment of TM 3 was selected for the final structural alignment.

4.2.9 Final Structural Alignment

The full structural alignment can be found in appendix A.7; figure 4.22 provides an overview of how the final structural alignment was constructed.

A range of methods was used to align the BbZIP and hZIP14 sequences. For each BbZIP TM helix where the alignment methods were largely in agreement, the alignment was chosen that minimized gaps and insertions. Where alignment methods produced contradictory results, the alignment that best preserved the properties of the most conserved BbZIP residues was chosen for the final structural alignment. Table 4.3 shows the sequence identity and similarity of

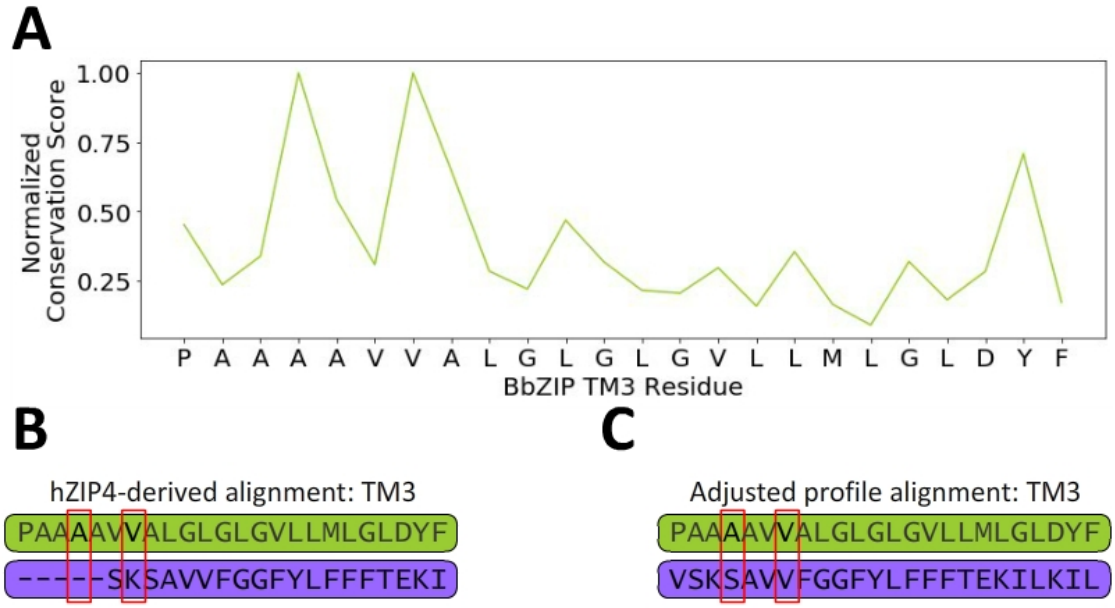


Figure 4.21: **Properties of conserved BbZIP TM 3 residues are partially retained in the adjusted profile alignment.** A: Graph of normalised conservation scores across the BbZIP TM 3. B: BbZIP-hZIP14 TM 3 hZIP4-derived alignment. C: BbZIP-hZIP14 TM 3 adjusted profile alignment. B,C: Red boxes indicate the most highly conserved residues of BbZIP TM 3.

each transmembrane helix, relative to BbZIP, in the final structural alignment.

The sequence identity and similarity obtained from the initial pairwise alignment of the full-length BbZIP and hZIP14 sequences was 18.8% and 30.2%, respectively. In the final alignment, the sequence identity and similarity across the TM helices varied, with the highest values for TM 4, of 45.8% and 62.5%. The higher similarity of the final structural alignment should lead to a higher quality homology model.

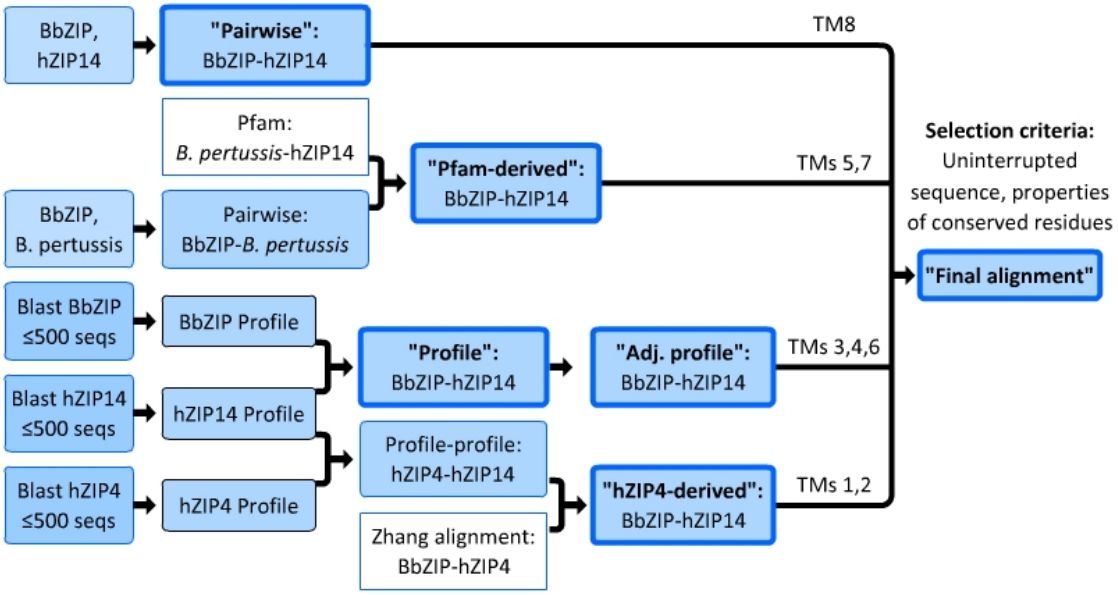


Figure 4.22: **Creation of structural alignment for modelling of hZIP14 transmembrane domain.** Diagram depicting the alignment methods used to create the final structural alignment. Various strategies were used to align the transmembrane helices of the BbZIP and hZIP14 sequences, sections of the alignments were chosen in order to arrive at a final structural alignment.

Table 4.3: **Sequence identity and similarity for each TM helix.** The final structural alignment was assembled from different alignments, with the aims of minimising gaps and insertions, maximising sequence identity and similarity, and preserving the physico-chemical properties of conserved BbZIP residues.

TM Helix	Identity	Similarity
1	21.4%	39.3%
2	17.6%	29.4%
3	25.0%	33.3%
4	45.8%	62.5%
5	21.4%	42.9%
6	23.1%	34.6%
7	13.6%	36.4%
8	28.6%	47.6%

4.2.10 Homology Modelling

The final structural alignment was submitted to the SWISS-MODEL web interface [194] and the Modeller software [192] to generate structural models of hZIP14. The Phyre2 web interface [193], a web server that accepts only the target sequence as input, was also used. For each model, the SCWRL software was then used to optimise the orientation of the side-chains [195].

The Phyre2 web interface automatically generated a structural alignment to guide the homology modelling and identified the BbZIP 5TSA crystal structure as the most suitable template for modelling. However, the Phyre2 structural alignment differed from the final structural alignment regarding TMs 3 and 8 (figures 4.23 and 4.24).

```
BbZIP     P----AAAAVVALGLGLGVLLMLGLDYF
Final     V----SKSAVVFGGFYLFFFTEKILKIL   Id: 25%  Sim: 33.3%
Phyre2    SKSAVVFGGFYLFFFTEKILKILLKQKI   Id: 8.3% Sim: 16.7%
```

Figure 4.23: **Comparison of final and Phyre2 structural alignments for BbZIP TM 3.** Sequence identity (Id) and similarity (Sim) are relative to the BbZIP TM 3. The Phyre2 alignment contains a 4-residue insertion at the N-terminus of TM3, and has lower sequence identity and similarity than the final structural alignment.

```
BbZIP     ET-------------------TATVGLMAGFALMMFLDTA
Final     FI-------------------IQNLGLLTGFTIMVVLTMY
          Id: 28.6% Sim: 47.6%
Phyre2    PEmnevcqederkgsilipfiIQNLGLLTGFTIMVVLTMY
          Id: 9.5%  Sim: 23.8%
```

Figure 4.24: **Comparison of final and Phyre2 structural alignments for BbZIP TM 8.** Sequence identity (Id) and similarity (Sim) are relative to the BbZIP TM 8; lower-case residues are not found in the Phyre2 model. The Phyre2 alignment contains a 19-residue deletion at the N-terminus of TM8; this is not a feasible alignment.

The BbZIP structure and the homology models are shown in figure 4.25 A; as expected, the positions of the alpha-helices resemble those of the BbZIP template. The hZIP14 residues in the TM helices correspond approximately to regions of increased conservation in the hZIP14 sequence (figure 4.25 B). However, the lengths and residues of the alpha helices vary between models. The Phyre2 model, in particular, diverges from the two other models, most noticeably for TMs 3, 7 and 8. Only for TM 3 does this result in a fundamental difference in the position of amino acids in the model (discussed further in section 4.2.11.3).

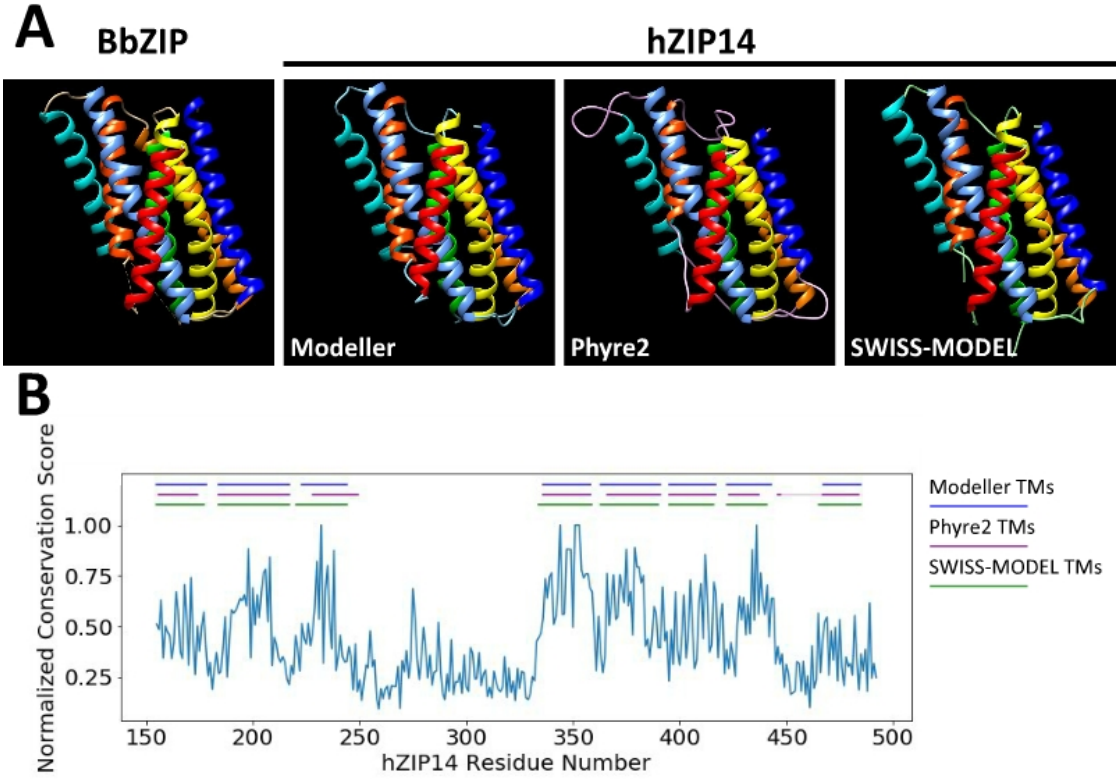


Figure 4.25: **TM helices of protein structures.** A: Ribbon structures of protein models - BbZIP and hZIP14 made using Modeller, Phyre2 or SWISS-MODEL; alpha helices are coloured from blue (N-terminus) to red (C-terminus). B: Graph of normalized conservation scores of hZIP14 residues, with positions of TM helices according to the homology models. The regions of the TM helices correspond to areas of increased sequence conservation.

4.2.11 Model Quality Assessment

As carried out for the hZIP14 ECD, the TM homology models were ranked by quality assessment software programs, in order to choose which model to use in designing future experiments. ProSA-web [198] and ANOLEA [199] were used as before. QMEANBrane [200] was used rather than QMEANDisCo, as QMEANBrane has been trained on membrane proteins specifically, and is more suitable for the TM domain. The Phyre2 structural alignment, which was not plausible due to the insertion in TM 3 and the missing residues in TM 8, was included in this assessment stage for the purposes of comparison.

4.2.11.1 ProSA-web Assessment Ranks the SWISS-MODEL Homology Model Most Highly

The ProSA-web server primarily assesses the plausibility of protein folds for submitted structures [198, 242], assigning submissions an overall Z-score, and providing an averaged energy profile of the sequence, where high energy regions indicate potentially low-quality regions of the model.

As shown in figure 4.26, the Z-scores of the BbZIP crystal structure and the homology models all fell within the range of Z-scores observed for experimentally determined NMR and crystal structures. The BbZIP crystal structure had a lower Z-score than the homology models (table 4.4), as expected for an experimentally determined structure. Of the homology models, the Phyre2 model was ranked as the lowest quality, and the SWISS-MODEL structure as the highest.

The ProSA-web energy plots of the protein structures show that the BbZIP TM helices tend to be areas of low energy, indicating that these regions of the model are of relatively high quality (figure 4.27), and the majority of the energy plot is

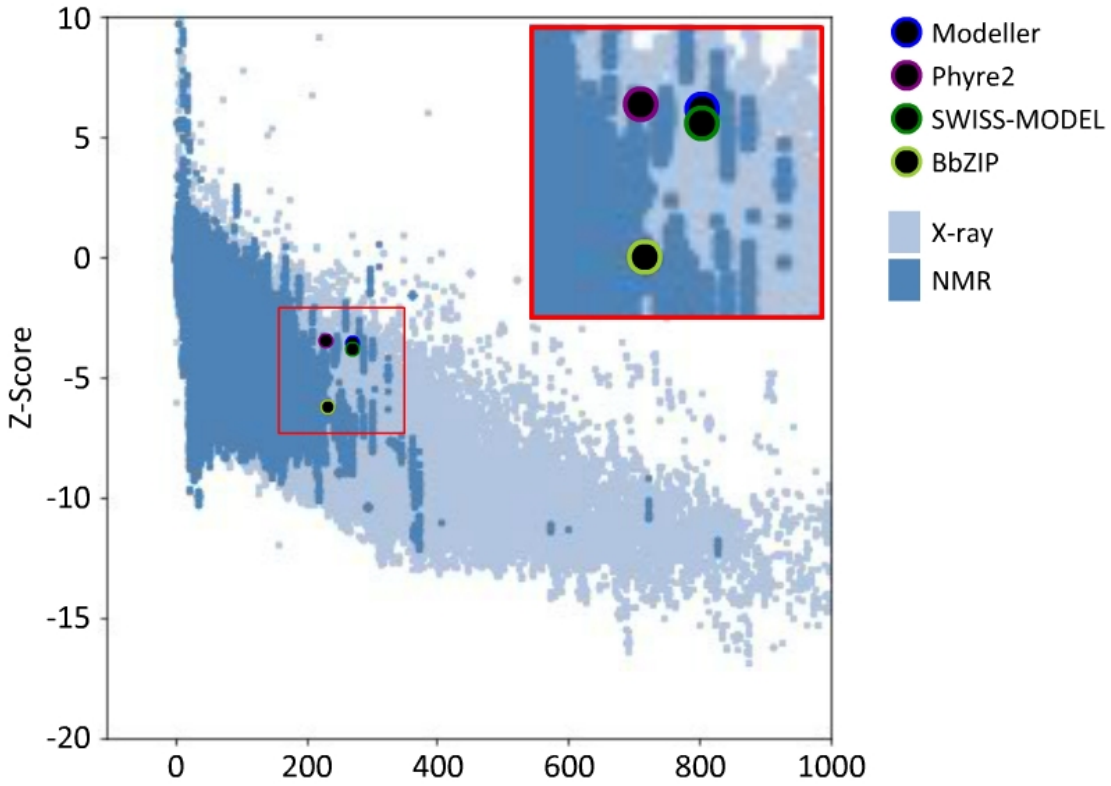


Figure 4.26: **ProSA-web Z-scores of overall model quality for BbZIP and the homology models relative to those of experimentally determined structures.** The Z-scores of BbZIP and the homology models fall within the range of those observed for experimentally determined structures, indicating the models are of good quality.

Table 4.4: **The SWISS-MODEL structure is ranked as the best quality homology model.** The ProSA-web and ANOLEA tools each assess structure quality using different parameters, producing a Z-score that summarises overall quality - a lower Z-score indicates higher quality.

	Z-Score			
Quality Assessment Tool	**BbZIP**	**Modeller**	**Phyre2**	**SWISS-MODEL**
ProSA-web	-6.16	-3.49	-3.42	-3.75
ANOLEA	3.19	11.30	14.57	9.00

below zero. The energy plots of the homology models fluctuate around zero, indicative of their decreased quality relative to the BbZIP template. The Phyre2 energy plot shows some clear differences from those of the other two models;

an energy spike between TMs 1 and 2 and an energy trough at TM 3 (window size 10). This suggests that the relatively shorter Phyre2 TM 1 is less plausible than the longer TM 1 of the other two models (figure 4.25). Interestingly, the ProSA-web software also finds the Phyre2 TM 3, whose alignment with BbZIP differs from the final structural alignment, to be more favourable than TM 3 in the other two models. However, the Z-score takes the energy level across all residues into account, and still ranks the SWISS-model structure as being of a higher quality than the Phyre2 and Modeller structures.

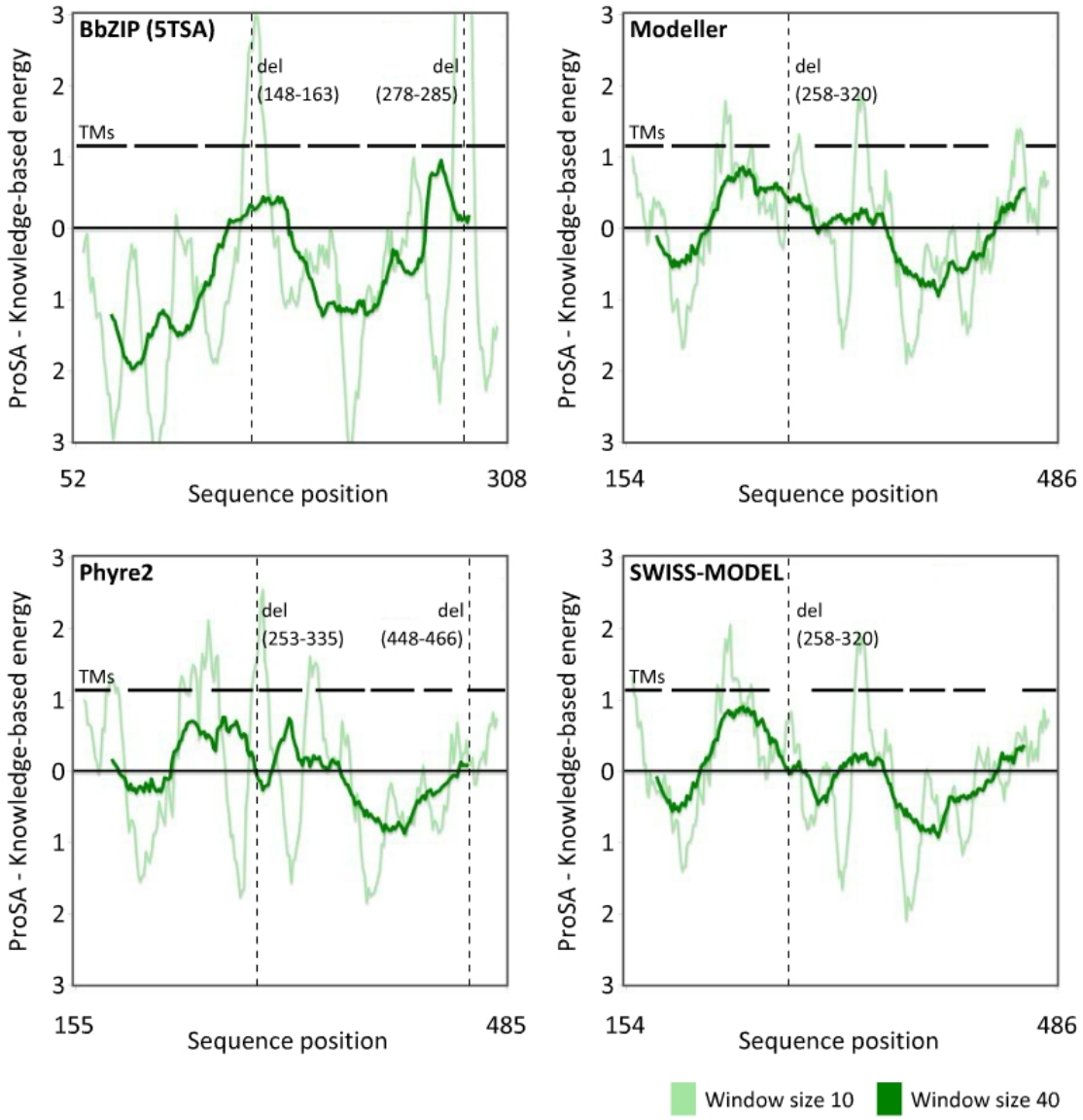


Figure 4.27: **ProSA-web energy plots of protein structures.** ProSA-web energy plots for BbZIP and hZIP14 protein models, smoothed by averaging over 10 and 40 residues, as indicated. Positions of TM helices are marked by horizontal black lines. The BbZIP ProSA-web energy plot indicates higher model quality (lower energy) than the homology models, particularly for the TM helical regions; the homology models have similar ProSA-web energy plots to each other, particularly for TM helices 6, 7 and 8.

4.2.11.2 ANOLEA Assessment Ranks the SWISS-MODEL Homology Model Most Highly

The ANOLEA software assesses non-local atomic interactions (residues more than 11 residues apart) based on a database of non-redundant proteins [245], and returns an energy plot smoothed over a 5-residue window, and an overall Z-score describing model quality [199].

The Z-scores returned by the ANOLEA server ranked the BbZIP crystal structure as being of much higher quality than the homology models (the lowest Z-score), as expected (table 4.4). Like the ProSA-web software, ANOLEA ranked the Phyre2 homology model the lowest, and the SWISS-MODEL structure the highest quality. The ANOLEA energy plots follow this trend for most of the TMs, with a large decrease in quality of the Phyre2 model over TMs 3 and 7 (figure 4.28). This difference between the ProSA-web and ANOLEA assessments may reflect which aspects of the protein structures are assessed by each tool; ANOLEA focuses on atomic interactions, while ProSA-web looks at the overall protein fold, and does not pick up ”close contacts or other violations of basic steric principles” [242].

Both the ProSA-web and ANOLEA tools use a knowledge-based approach, relying on databases of experimentally-determined structures to assess how realistic the submitted structure is. The vast majority of experimentally-determined, high-quality structures available are globular proteins, that function in aqueous environments, in which the constraints on protein-folding are very different from the hydrophobic environment of the lipid bilayer. The ProSA-web software, for example, is known to return higher Z-scores for membrane proteins. The QMEANBrane assessment program was trained exclusively on membrane proteins, and offers an in-depth analysis of a range of contributors

to model quality [200].

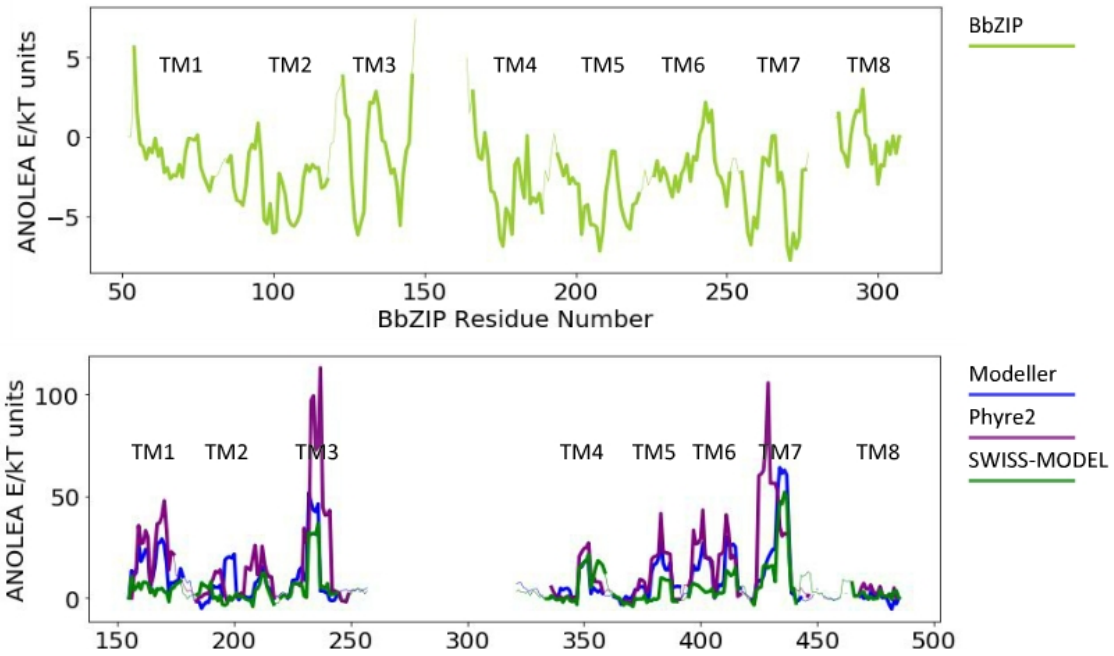


Figure 4.28: **ANOLEA energy plots of protein structures.** ANOLEA energy plots smoothed by averaging over 5 residues; bold lines indicate alpha-helical regions. Upper panel: ANOLEA energy plot for the BbZIP 5TSA crystal structure. Lower panel: ANOLEA energy plots for each homology model.

4.2.11.3 QMEANBrane Assessment Ranks the SWISS-MODEL Homology Model Most Highly

The QMEANBrane local quality assessments of the BbZIP crystal structure and the homology models are shown in figures 4.29 A and B, respectively. As with ProSA-web and ANOLEA, the BbZIP structure is assessed as having a higher quality (described as 'predicted similarity', i.e. to the real structure) than the homology models, the SWISS-MODEL structure as being the best of the homology models, and the Phyre2 structure as the worst.

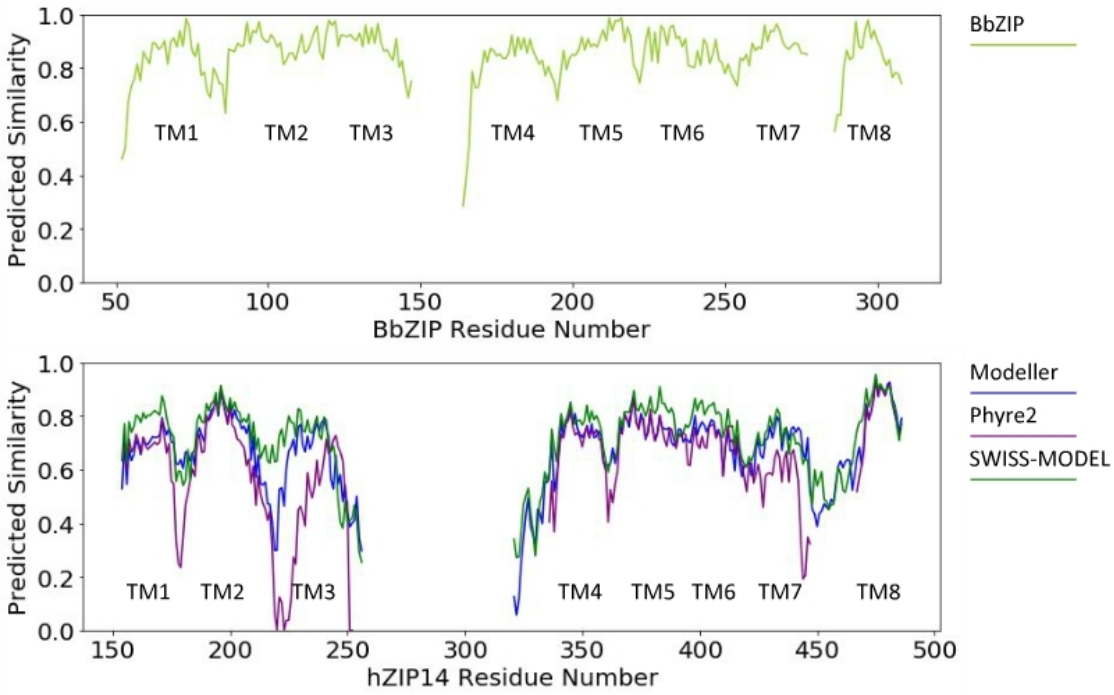


Figure 4.29: **QMEANBrane local quality prediction ranks the SWISS-MODEL structure most highly across the majority of residues.** Graphs of local quality predictions (predicted similarity to the correct protein structure) at each residue of the protein model. Upper panel: Local quality prediction for the BbZIP 5TSA crystal structure. Lower panel: Local quality prediction for each homology model.

Position in the Membrane

The QMEANBrane results include a prediction of which residues are within the plane of the lipid bilayer for each structure (figure 4.30). The majority of helices are tilted relative to the membrane in each structure, as described for BbZIP by Zhang et al. [241].

The predicted membranous regions for each structure are not identical; like the BbZIP crystal structure, the SWISS-MODEL structure is predicted by the QMEANBrane tool to extend both above and below the membrane, while the Modeller and Phyre2 models are predicted to protrude into the cytosol, but not through the outer leaflet of the membrane to the outside of the cell. According

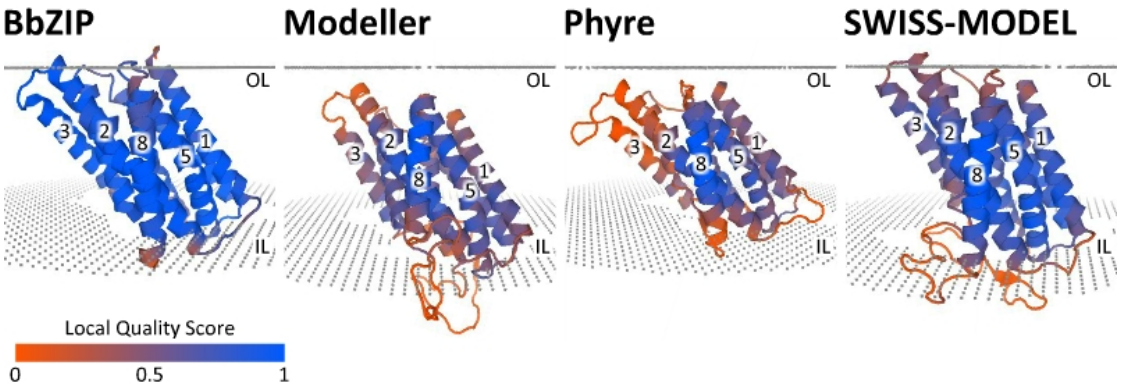


Figure 4.30: **QMEANBrane predicted orientation within the membrane.** Alpha-helices are numbered, model colouring indicates QMEANBrane quality prediction, as indicated by the key; OL=outer edge of outer layer of membrane, IL=outer edge of inner layer of membrane. A: BbZIP 5TSA crystal structure. B: hZIP14 Modeller homology model. C: Phyre2 homology model. D: SWISS-MODEL homology model.

to this prediction, the SWISS-MODEL homology model is more similar to the BbZIP crystal structure than the other two models.

Quality of the Phyre2 Model

The lower quality of the Phyre2 model, as assessed by QMEANBrane, is particularly pronounced for TMs 3 and 7 (figure 4.29). One obvious difference in the Phyre2 TM 7 is the decreased length of the alpha-helix, with Phyre2 TM 7 covering residues 423-437, and the SWISS-Model and Modeller TM 7s covering residues 422-441 and 422-443 respectively; the Phyre2 TM 7 is approximately one third shorter. Further analysis of the QMEANBrane components suggests that this contributes to the decreased quality score of the Phyre2 TM 7 (figure 4.31).

Figure 4.31 (upper panel) shows that the secondary structure agreement (the extent to which the QMEANBrane tool finds the secondary structure in that region of the model plausible) for the Phyre2 TM 7 decreases after residue

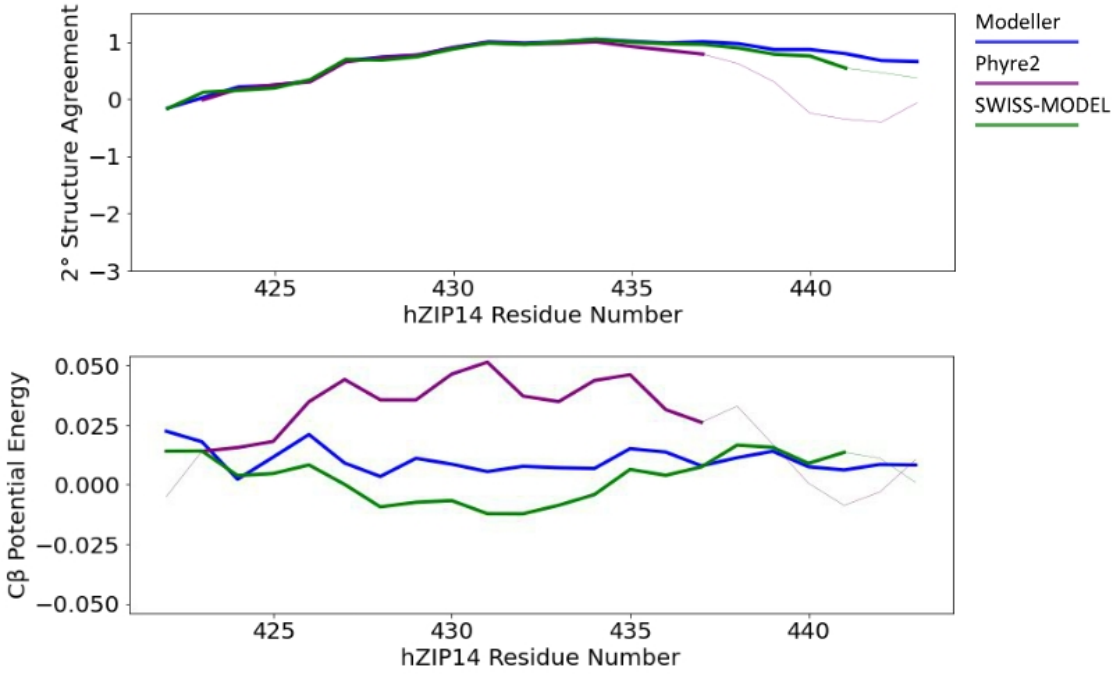


Figure 4.31: **QMEANBrane assessment components of the TM 7 region.** Graphs depicting the QMEANBrane assessment of secondary structure agreement (upper panel) and Cβ potential energy (lower panel) for the TM 7 region. Thick lines represent the alpha-helical region of TM 7 for each model.

437, where the alpha-helix ends, suggesting that the continuation of the alpha-helix may be favoured beyond this point. However, the Cβ potential energy component (which relates to the positions of the residue side-chains) indicates that side-chain interactions are not optimal across the length of the Phyre2 TM 7 helix, with peaks in potential energy (representing unfavourable interactions) occurring every 4 residues, at positions 427, 431 and 435 (figure 4.31, lower panel). Figure 4.32 C shows that in each homology model, these residues are in essentially identical positions, facing TM 3. Where the Phyre2 model differs, is the composition of TM 3, due to the alternative structural alignment produced by the Phyre2 web server.

In the BbZIP crystal structure, TMs 3 and 7 are in close proximity, with small residues such as glycine and alanine at the points where the two helices are

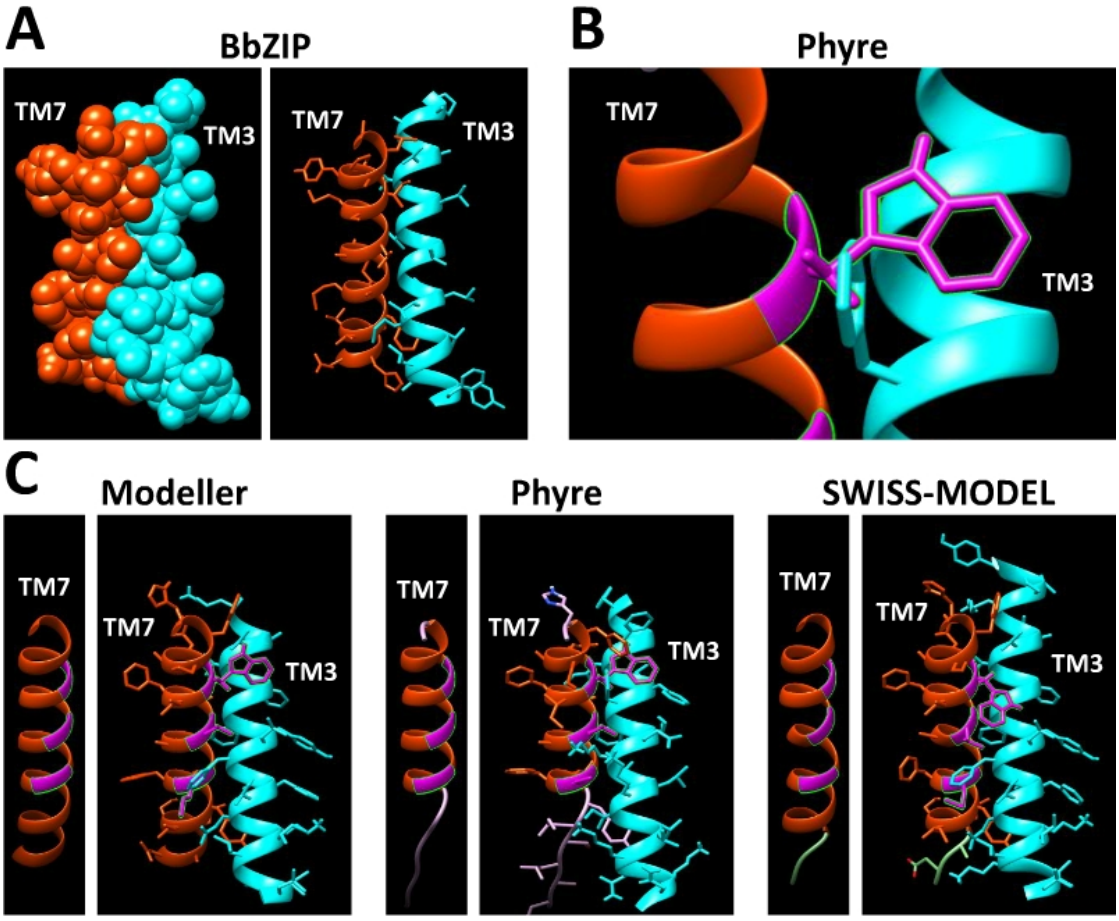


Figure 4.32: **Close-packing between TMs 7 and 3 causes a clash in the Phyre2 homology model.** The interactions between TMs 7 and 3 in the BbZIP 5TSA crystal structure and the three homology models; hZIP14 residues 427, 431 and 435 are highlighted in magenta. A: BbZIP TMs 7 and 3 with side-chains represented as spheres and sticks, respectively. B: Close-up of the clash between hZIP14 W27 (TM 7) and F235 (TM 3), according to the Phyre2 homology model. C: hZIP14 TM 7 and interactions between TMs 7 and 3, for each homology model.

closest, and larger, bulkier residues such as tryptophan and phenylalanine at the beginnings and ends of the helices (figure 4.32 A), a pattern repeated in the Modeller and SWISS-MODEL structures (figure 4.32 C). In the Phyre2 model, due to the alternative alignment for TM 3, large residues from TM 3 are facing TM 7 (figure 4.32 C), resulting in proximities that are not energetically favourable, and even a physical clash between residues W427 and F235 (figure

4.32 B). The poor quality assessment for the Cβ component of Phyre2 TM 7 appears to be largely a side-effect of problems with TM 3. This clash between TMs 3 and 7 is in agreement with the results of the ANOLEA assessment, which focuses on side-chain interactions, and also assessed TMs 3 and 7 in the Phyre2 model as being low quality (section 4.2.11.2).

All the Phyre2 TM 3 quality assessment components indicate problems, including the lack of agreement with QMEANBrane's secondary structure prediction, and unfavourable side-chain interactions for residues facing TM 7 (figure 4.33). This complements the quality assessment of TM 7 and the conclusion that the interactions between TM 3 and TM 7 in the Phyre2 model are unlikely to be correct, due to an incorrect alignment of TM3 for this model.

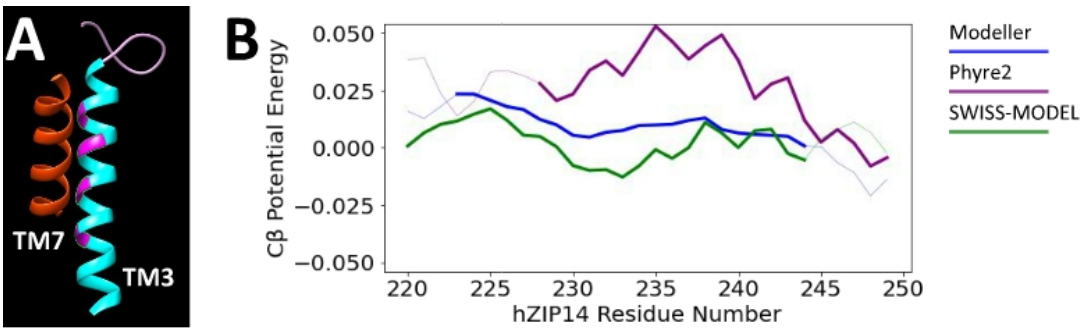


Figure 4.33: **QMEANBrane assesses TM 3 side-chain interactions as unfavourable.** A: Ribbon structure of Phyre2 TMs 3 and 7. hZIP14 residues 232, 235, 239 and 243 are highlighted in magenta. B: Graphs depicting the QMEANBrane assessment of Cβ potential energy for the TM 3 region. Thick lines represent the alpha-helicals region of TM 7 for each model.

The local quality assessment for TM8 was not expected to differ noticeably between the Phyre2 structure and the other homology models, as the slight difference in the alignment for TM 8 did not alter the positions of TM 8 residues in the structures, excepting the first two residues (figure 4.24).

The QMEANBrane assessment of the Phyre2 model was detailed, not only identifying the regions of the structure that were predicted to be incorrect, but

also the causes of the problems in the structure.

4.2.11.4 Quality of the SWISS-MODEL Model

The Modeller and SWISS-MODEL structures are based on the same sequence alignment, and the root mean square deviation (RMSD, the average distance between aligned amino acids) between the alpha-helical regions of the two models is only 0.424 Å. This RMSD value is relatively small, similar to those observed between pairs of experimentally-determined structures for the same protein, which are typically well under 1 Å (for structures with resolutions of less than 2 Å [246]).

The superior QMEANBrane quality assessment of the SWISS-MODEL structure is primarily related to more favourable side-chain interactions, and in small part to more favourable interactions of all atoms. Figure 4.34 shows quality assessment results for TM 1, where the difference between the quality of the SWISS-MODEL and Modeller structures is predicted by QMEANBrane to be greatest. The majority of the TM 1 assessment difference between the SWISS-MODEL and Modeller structures is due to lower potential energy (i.e. more stable positions in the structure) of the residue side-chains, and of all atoms. Generally, the SWISS-MODEL structure was rated more favourably than the Modeller structure for a range of the QMEANBrane assessment components. In agreement with ProSA-web and ANOLEA, the QMEANBrane analysis found the SWISS-MODEL hZIP14 model to be a more plausible structure than both the Phyre2 and Modeller structures, and will be used as the basis for the design of further experiments.

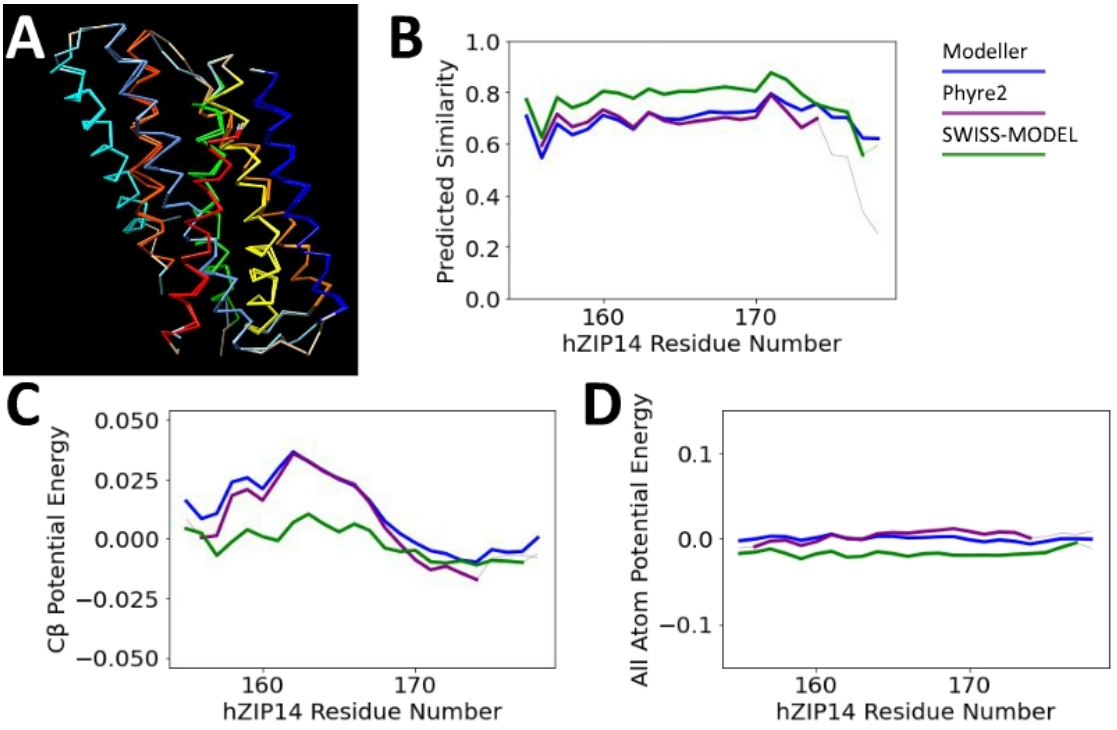


Figure 4.34: **The SWISS-MODEL structure is rated more favourably than the Modeller structure.** A: Wire structures of the SWISS-MODEL and Modeller structures, alpha-helices are coloured from blue (N-terminus) to red (C-terminus). B,C,D: Graphs depicting the QMEANBrane quality assessment, or components, for the TM1 region (residues 155-178) of the three models. Thick lines represent alpha-helical regions of TM 1 for each model. B: Graph of the local quality assessment for TM 1. C: Graph of the potential energy of the Cβ-interactions (i.e. side-chain interactions) for TM 1. D: Graph of the potential energy of long-range interactions for TM 1.

4.2.12 Positions of Mutations in hZIP14 Structure

Mutations F98V, S104I and R128W are located in the N-terminal extracellular domain (figure 4.35). S104 is located on the outer surface of each ECD monomer, facing away from the dimer interface, and R128 is located C-terminal to the final alpha-helix of the ECD, before the first helix of the transmembrane domain (figure 4.35 A). F98 is on the third alpha-helix, buried within the globular structure of the ECD, and interacts with residues on the second alpha-helix and the 3-4 inter-helical loop (figure 4.35 B).

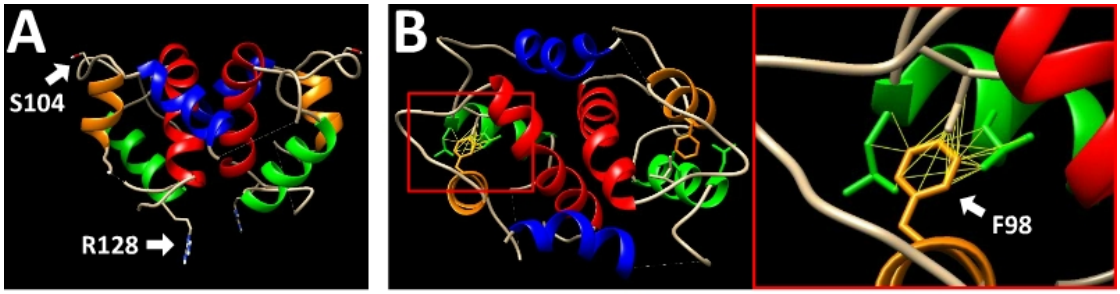


Figure 4.35: **Locations of N-terminal disease-causing mutations.** A: R-groups of S104 and R128 in the ECD. B: The R-group of F98 in the ECD, and interacting amino acids; close-up highlighted in red, yellow lines indicate interactions between atoms.

The P379L mutation is located in the center of TM 5, causing a kink in the alpha-helix, both P379L and G383R are on the side of the helix facing away from TM 4 (figure 4.36 A). The L441R and N469K mutations are on the outer surface of the TMD, on TMs 7 and 8, respectively (figure 4.36 B).

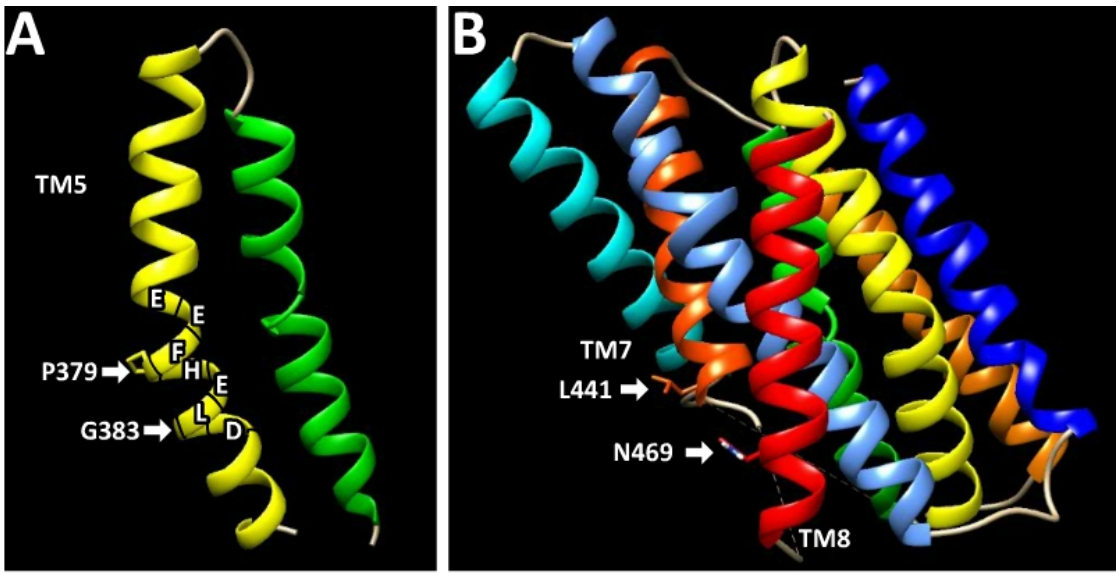


Figure 4.36: **Locations of C-terminal disease-causing mutations.** A: R-groups of P379L and G383 in the TMD; metal transport motif indicated. B: R-groups of L441 and N469 in the TMD.

4.3 Discussion

In this work, homology models of the N-terminal extracellular domain (ECD) and the C-terminal transmembrane domain (TMD) were developed, in order to understand the effects of the disease-causing mutations, and in turn, to use these studies of the disease-causing mutants to better understand hZIP14.

The template for the ECD model was the crystal structure of the *Pteropus alecto* hZIP4 ECD (P_ECD, PDB ID: 4X82), with a resolution of 2.76 Å, solved as a homodimer [219], while the TMD model was based on the crystal structure of the *Bordetella bronchiseptica* ZIP (BbZIP, PDB ID: 5TSA), with a resolution of 2.4 Å [241].

4.3.1 Accuracy of the Structural Alignments

The first step in homology modelling is developing an accurate structural alignment. For the ECD, the target and template sequences were both mammalian species in the same protein family, fairly closely related. The pairwise and profile alignments were almost identical, and so I have high confidence in the accuracy of that alignment. The structural alignment for the TMD was more challenging; there is a great evolutionary distance between the bacteria *Bordetella bronchiseptica* and *Homo sapiens*, and the sequence identity (the percentage of residues in the alignment that are identical in both proteins) for the initial pairwise alignment was only 18.8%. Sequence identities of $\sim$20-25% have been described as being in the 'twilight zone', an empirically determined level of sequence identity *above* which proteins are fairly likely to be homologous, and below which the similarity between proteins (sequence identity) is more likely to be due to random chance [235].

Since the sequence identity between BbZIP and hZIP14 was below this threshold, this raises the question of whether these two proteins are genuinely in the same family and share a function. The Pfam database defines BbZIP and hZIP14 as being members of the PF02535 family of zinc transporters [243]. The Pfam protein family database makes use of hidden Markov Model (HMM) profiles, based on high quality alignments, which capture the essential characteristics of a protein family conserved over evolution [247]. These HMM profiles are then used to search for more distantly related proteins, which are then used to refine the HMM profile, in an iterative process that maintains the quality of alignments as well as providing high coverage of protein sequences [247]. These types of iterative processes have been described as 'sequence-space hopping' [248], and are effective because homology is transitive, "if A is related to B and B is related to C, A must be related to (homologous to) C" [249]. In addition, HMM profiling has been shown to be far more effective than other search methods at identifying remote homologues without false positives, with the result that the assignment of both BbZIP and hZIP14 to the same protein family is highly reliable [250–253].

Low sequence identities between template and target proteins also have implications for homology modelling. The similarities between protein structures decrease as amino acids are substituted, for the protein backbone, and even more so for the side-chain positions [235, 254]. Although in the final structural alignment the sequence identity was much higher than 18.8% when focusing on the regions of the BbZIP sequence in the crystal structure (table 4.3), nonetheless, limitations imposed by low sequence identity over particular regions should be taken into account when analysing the significance of structural features. Modelling based on a distant homologue is accepted as suitable for understanding protein function and making sense of disease-

causing mutants, while closer homologues have been known to generate models suitable for drug design [228].

The final structural alignment for TMs 4-8 is essentially based on the MSA for the PF02535 Pfam family [243], albeit with some minor adjustments to deal with gaps and insertions. These helices cover the most important functional region of the TMD, as well as the disease-causing mutations in the TMD. For the remainder of the protein a profile-profile alignment was created between large MSAs for BbZIP and hZIP14. In the same manner, I created an alignment between hZIP14 and hZIP4, in order to compare my profile-derived alignment to the BbZIP-hZIP4 alignment published by Zhang et al. [241]. The profile-derived alignment of BbZIP-hZIP14, through my intermediary hZIP4-hZIP14 alignment, was in agreement with the alignment by Zhang et al. [241] for TM 2. For TM 1, I agreed with the Zhang et al. [241] alignment over my profile alignment due to the preservation of properties of its most conserved residues (figure 4.18). TM3 seems to have been difficult to align, as substantially different results were produced for each method used. For the other TMs, similar results were typically found by at least two alignment methods, with differences mainly due to small insertions and gaps at the ends of the TM regions. For TM3, different alignment methods produced conflicting results. To illustrate this point, residue 136 in TM3 of BbZIP was aligned with residue 242 in the pairwise alignment, residue 234 in the profile-profile alignment, residue 230 in the hZIP4-derived alignment, and residue 239 in the Phyre 2 alignment (figures 4.15, 4.20 and 4.23). In this instance, the profile-profile alignment was chosen to preserve the characteristics of the most highly conserved BbZIP residues.

In support of the final structural alignment of the TMD developed here, the Phyre2 web server, which used a different strategy (HHblits) to create its BbZIP-hZIP14 alignment [193], was essentially in agreement with my final structural

alignment for all TMs except 3, which caused significant structural clashes (figures 4.23 and 4.32). This error in the TM 3 region of the Phyre2 alignment does not prove that my structural alignment was correct, but TM3 of the Modeller and SWISS-MODEL structures did not stand out in the quality assessment stage as being of particularly low quality, relative to the rest of the structure. Whereas both ANOLEA and QMEANBrane detected problems for TMs 3 and 7 in the Phyre2 model, due to the steric clashes caused by misalignment of TM 3 (figures 4.28, 4.31 and 4.33).

4.3.2 Dimerisation of the ECD

The function of the ECD of hZIP14 is currently unknown, though homology between the ECDs of other ZIP proteins, such as ZIP4, indicate that ZIP ECDs typically form dimers [219]. The homology model of the hZIP14 ECD fully supports its dimerisation. As shown in figure 5.2, the surface of the ECD is covered in hydrophilic side-chains, as expected for a soluble protein, except for the region predicted to dimerise, which is highly hydrophobic. This hydrophobic region would be unusual for a soluble monomeric protein, as burying hydrophobic residues in protein interiors due to their decreased ability to interact with the polar solvent is a major contributor to protein folding and stability [255]. Protein-protein interfaces can include both hydrophobic and hydrophilic residues, but when hydrophobic regions are found on the surface of soluble proteins, they are typically binding sites for other compounds or proteins [256, 257]. This is more pronounced in homodimers, for which a sizeable proportion of dimerisation interfaces are comprised of a hydrophobic core surrounded by hydrophilic residues [258]. Upon rotamer optimization by SCWRL [195], C118 residues in each ECD monomer were predicted to form a disulphide bridge. This prediction is supported by the work of Zhang et al.

[219], who found that an S297C mutation in hZIP4 (equivalent to the C118 position in hZIP14) resulted in an intermolecular disulphide bridge between the two ECD monomers, and increased the detection of the dimer in non-reducing SDS-PAGE without significantly altering zinc transport, cell surface expression or glycosylation.

In Chapter 3, western blotting of hZIP14 and disease-causing variants revealed high molecular mass bands which were fainter or absent in N-terminal disease-causing mutants, a loss which was most pronounced for the highest molecular mass band (figures 3.6 and 3.7). The dimeric homology model of the hZIP14 ECD supports the proposition that these higher molecular mass bands are related to oligomerisation of hZIP14, and that mutations in the ECD disrupt this oligomerisation. It is unclear however, why there should be two high molecular mass bands. It should also be noted that the absence of these bands may reflect a decrease in oligomerisation *in vivo*, or decreased affinity of monomers leading to less persistence of oligomers in the denaturing, reducing conditions used in sample preparation of cell lysates, or a combination of these two factors.

The flow cytometry results in Chapter 3 for N-terminal disease-causing mutants, were consistent with the hypothesis that dimerisation of the ECD occurs, and is involved in effective cell surface localisation (discussed in section 3.3.4.1). Unlike the western blotting results, in which the higher molecular mass bands were faintest for hZIP14 F98V, followed by S104I and then R128W, the flow cytometry results for the N-terminal disease-causing mutants were very similar. It may be that hZIP14 cell surface expression is insensitive to small differences in the affinity of the ECD monomers, for example, this would be the case if any decrease in affinity beyond a certain threshold had the same impact on trafficking to the cell surface.

4.3.3 Structural Impacts of the ECD Disease-Causing Mutations

The positions of the disease-causing mutations in the ECD homology model go some way towards providing structural explanations for their deleterious effects. The homology model predicts that the F98V mutation interferes with a number of hydrophobic interactions that stabilize the structure of the ECD (figure 4.35). The S104I mutation disrupts a conserved glycosylation site [55]. Glycosylation is known to play an important role in protein folding, maturation and quality control [259]; it follows that the loss of this glycosylation site would cause increased retention in the endoplasmic reticulum, and decreased stability of the ECD could also impact dimerisation. In the ECD homology model, S104 is situated on the surface of the ECD facing away from the TMD, consistent with its being part of a glycosylation site.

The position of R128 in the model gives no obvious clues as to the potential effects of the R2128W mutation on hZIP14 function. According to the homology model, R128 is not technically part of the ECD, falling two residues C-terminal of the final ECD alpha-helix, in the 28-residue sequence between the ECD and TMD. One possibility is that R128 is involved in mediating an interaction between these two domains. This 28-residue sequence is composed almost entirely of polar and charged residues (figure 4.37), which may serve some functional role, though this region does not show high conservation even within ZIP14 proteins [205]. This section of hZIP14 connects two distinct protein domains, making it a domain linker. The sequences of domain linkers are typically poorly conserved, as they tend to be intrinsically disordered regions, which form ensembles of conformations, rather than well-defined secondary and tertiary structures [260]. This disorder is often of functional importance; disordered linkers can aid folding

```
             .+ ...-..-.--.-..-- + .  -
hZIP14  127  SRACTSENQENEENEQTEEGRPSAVEVW  154
```

Figure 4.37: **The amino acid sequence between the ECD and TMD.** Residues 126 to 154 of hZIP14; the symbols above indicate polar (.), positive (+) or negative (-) residues.

of connected domains by forming clear boundaries between them, and they vary in terms of length and flexibility, allowing or constraining movements of domains relative to one another [261, 262]. Intrinsically disordered regions typically contain more polar and charged residues, and few large hydrophobic residues, making them more soluble with little propensity to fold into stable structures [262]. This sequence, with its high concentration of negatively charged residues, matches the description of a 'polyelectrolyte' [262], though there may be a more specific classification for linkers with central negative residues flanked by positively-charged residues. It is possible that the replacement of R128 with a large, hydrophobic tryptophan residue has disrupted the function of this linker; its native function could be to enhance folding of these two separate domains (R128 is on the side of the linker closest to the ECD), or to facilitate dimerisation of the ECDs by regulating their movements relative to the TMD. Further exploration of these possibilities could be undertaken by mutagenesis of this linker region.

4.3.4 The Metal Transport Motif of the TMD

The TM 5 HEXPHEXGD metal transport motif particular to the LIV-1 subfamily of ZIP, is believed to function in concert with the conserved HNF residues on TM 4 [11]. The identity of the residues in these conserved positions differ as we widen out to the larger ZIP family, but these positions are nonetheless believed to be important in metal transport and solute specificity. Within the ZIP

family, the conserved residues that make up the core metal binding site can be described as TM4-1-2-x-x-3 and TM5-1-2-x-x-3 [263], which are 177-HNxxE and 207-QDxxE for BbZIP, and 347-HNxxD and 376-EExxH for hZIP14. As illustrated in figure 5.5 (page 227), these residues are aligned, showing that the homology model captures the most important functional region of the TMD.

4.3.5 Structural Impacts of the TMD Disease-Causing Mutations

The TMD mutations had a range of impacts on both iron uptake and cellular localisation. The most noticeable impact of the G383R mutation is that it prevents iron uptake (figure 3.10). In my homology model of the hZIP14 TMD, glycine residues on the interior of hZIP14 are closely packed against other residues; any side-chain in these spaces would cause steric clashes, requiring the helices to move in order to accommodate the increased volume of any other residue. This is well illustrated by the previously discussed interactions between TMs 3 and 7 (section 4.2.11.3), but occurs elsewhere in the model.

Figure 4.38 compares the proximities of four TM 4 and TM 6 glycines to their closest residues, with the proximity of G383 to its closest residues. Compared to other interior glycines, G383 is far away from its nearest TM helix. In the context of the whole structure, G383 is on the surface of a large cavity at the intracellular side of the TMD. The template BbZIP structure is believed to be in an open-inward conformation, as the main metal-binding site is exposed on the intracellular side of the protein [241]. For these reasons, the function of G383 may not be apparent in this structure; the importance of this glycine may be related to the outward-open conformation, or the structural changes that occur when hZIP14 switches between the inward-open and outward-open

conformations.

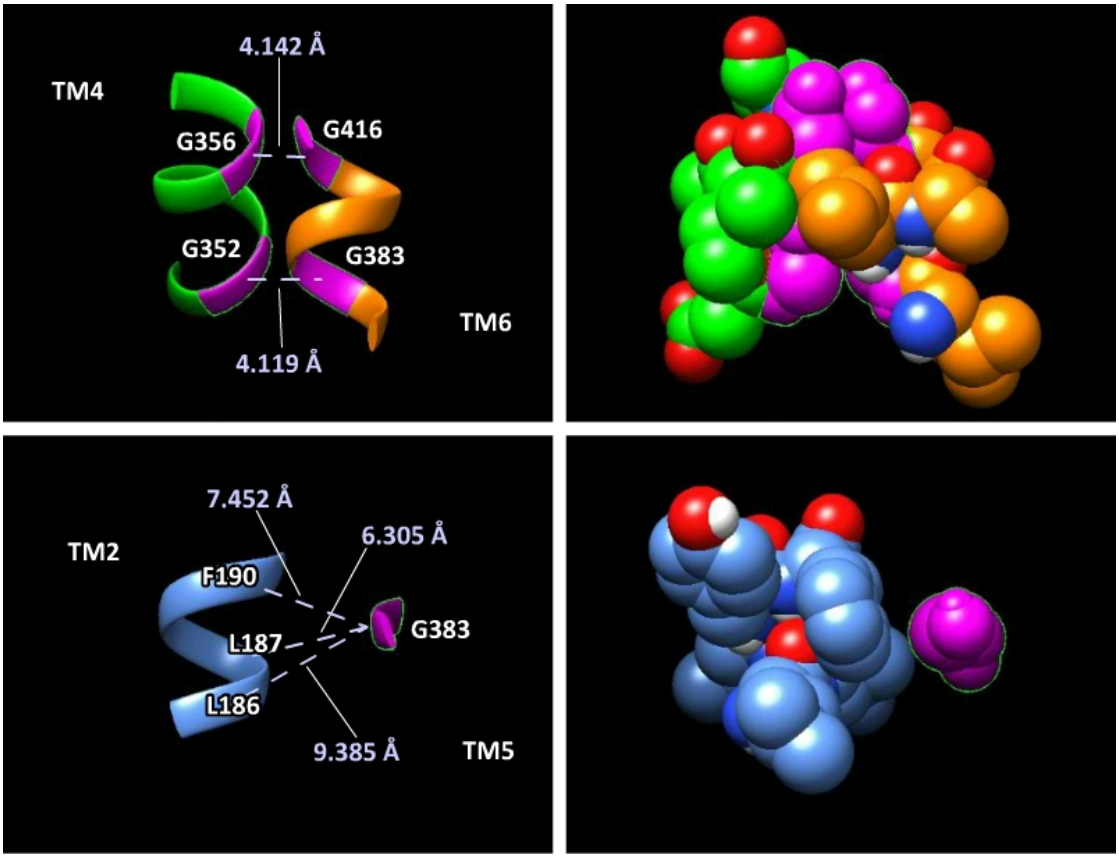


Figure 4.38: **Glycines in the TMD homology model.** Top figures: Glycines in TMs 4 and 6 in relation to nearby residues (ribbon representation on the left, spheres on the right). Bottom figures: G383 in relation to nearby residues in TM 2 (ribbon representation on the left, spheres on the right). Dotted lines show the distance between alpha-carbons of indicated residues.

Unlike G383, the structural importance of the highly conserved P379 is much clearer, it causes a kink in TM 5, in the center of the region forming the core metal transport binding site. However, as P379L had no clear effect on the iron uptake function of hZIP14, the need for this kink in TM 5 is not clear. P379L lies between residues 376-377 and 380 (in 376-**EE**XP**H**EXGD) (figure 4.36 A), so could be important for the correct relative positioning of these residues. P379 is so highly conserved in both the human and bacterial ZIPs, it is likely to be essential for normal function; transporters have a range of characteristics that

could be altered by a mutation, other than the rate at which they transport a specific metal, which may be important for the role of that transporter. Given the position of P379 in the metal binding site, possibilities could include a change in the pH range over which transport occurs, altered ability to transport other solutes, or a change in solute preference. Since P379L is known to disrupt manganese homeostasis [71], it is highly likely that this mutation prevents or decreases manganese transport directly by at least one of these mechanisms.

The homology model also offers no explanation of how the P379L mutation could mediate the short-term and long-term effects observed on hZIP14 cell surface levels (figures 3.37 and 3.12). However, there remains much to learn about the regulation of wild-type hZIP14 trafficking; it is possible that the P379L mutation alters an aspect of metal transport I have not measured, and that these effects are related to regulation of hZIP14 trafficking by metals.

Both L441 and N469 side-chains project out from the surface of the TMD (figure 4.36). Although they are found on different TM helices, 7 and 8 respectively, their locations are similar; they are on the same side of the TMD and are on the end of the TMD that is closest to the inside of the cell (figure 4.36). Unlike the other two TMD disease-causing mutations (P379L and G383R), these mutations showed effects on cellular localisation in transiently transfected cells that were apparent under immunofluorescence microscopy (figure 3.14). Differential staining with flow cytometry showed a sigmoidal, and sigmoidal-like relationship between interior and cell surface hZIP14 for the L441R and N469K mutations, respectively (figure 3.37). This gives rise to the possibility that this region of the protein is important in regulating trafficking. The expression-dependent aspect of this impact on trafficking is similar to that seen with the ECD disease-causing mutations, so it is possible that this region forms a binding site which is altered by the L441R and N469K mutations. According to the

homology model, L441 and N469 do not interact with residues involved in metal transport, which may explain why their mutants are still able to transport metals (3.11 and Hendrickx et al. [95]).

4.3.6 The Cytosolic TM 3-4 Inter-helical Loop

Based on the homology model, the TM 3-4 inter-helical loop (residues 242-334) is 93 residues long. If hZIP14 trafficking is regulated, the TM 3-4 cytosolic loop is a promising region to study, though the model's cytosolic TM 7-8 loop is 23 residues long. Although the TM 3-4 loop has been found to contain various regulatory features, it is not conserved between ZIP proteins, and little is known about its role in hZIP14 [264]. In a study by Zhao et al. [55], where they found that high intracellular iron protected endocytosed hZIP14 from proteasomal degradation, they reported that deletion of residues 238-326 did not prevent this regulation of hZIP14 degradation by iron. This deletion mutant was made based on a secondary structure prediction, but according to my model would remove 4 residues from the C-terminus of TM 3. Results from our lab found that this 238-326 deletion (with a C-terminal haemagglutinin tag, rather than a FLAG tag (DYKDDDDK [265])) caused very low levels of expression and ER retention, based on western blotting and immunofluorescence microscopy [A. Pistofidis and K. Bowers, unpublished data]. The findings published by Zhao et al. [55] show only western blotting results from the deletion mutant in different conditions, and are not compared to wild-type hZIP14, so decreased expression may not have been detected. Within mammalian ZIP14 proteins, there are relatively conserved regions in this loop which could be investigated [205]. This structural model of hZIP14 delineates boundaries for this loop, information which could be used to enable further study of its importance in hZIP14 regulation.

4.3.7 Other Models of hZIP14 and Related Proteins

Since this homology model was made, other computational models of hZIP14 have become available, as well as an additional, metal-free, crystal structure of *B. bronchiseptica* ZIP [233, 266–268]. Any differences between these models and the homology models will be discussed further in Chapter 5.

4.3.8 Summary of Implications of Homology Models

In this chapter, high resolution crystal structures were used as templates for homology modelling of the hZIP14 extracellular and transmembrane domains. An accurate structural alignment is crucial for the production of high quality models, and multiple strategies were used in order to create the most plausible alignments of hZIP14 with the ECD of *P. alecto* ZIP4 and the *B. bronchiseptica* ZIP. Although the quality of the homology model will be affected by the large evolutionary distance between BbZIP and hZIP14 (even assuming a perfect structural alignment), homology models are considered to be suitable for understanding and designing mutations [228].

These models were able to suggest likely mechanisms for the effects of some of the disease-causing mutants described in Chapter 3, though potential mechanisms for others, such as L441R and N469K, were less clear. Dimerisation of the ECD in particular seemed to be important for hZIP14 localisation, and the TMD metal binding site was in agreement with literature predictions. Modelling the domains also provided information on their boundaries, which could be used to further study unstructured regions, such as the negatively charged linker between the ECD and the TMD, or the large intra-helical loop between TMs 3 and 4.

Chapter 5

Functional Analysis of ZIP14

5.1 Introduction

The homology models created in chapter 4 are meant to aid in our understanding of the relationship between hZIP14 structure and function. The homology model of the hZIP14 extracellular domain (ECD) is based on a fairly closely related mammalian ZIP4 [219], and determining the structural alignment was relatively straightforward. In addition, the ECD homology model enabled predictions for the mechanisms leading to the effects of most of the N-terminal mutations, and provided insight into the ECD residues involved in dimerisation. In this chapter, mutations located in the ECD were designed to test the accuracy of the ECD homology model, and to test the effects of inhibiting dimerisation of the ECD.

The transmembrane domain homology model was based on a more distantly related bacterial ZIP [241], and developing the structural alignment was challenging. The resulting homology model raised many questions about

the mechanisms underpinning the effects of the C-terminal disease-causing mutations, and was in agreement with the literature regarding the structure of the core metal-binding region [241, 263], without offering any further insight into the transport mechanism used by hZIP14. In this chapter, mutations located in the TMD were designed to test the effect on solute specificity of hZIP14, and to test the accuracy of the TMD homology model.

5.1.1 Selection of Mutations in the ECD

According to the ECD homology model, the F98V mutation disrupts interactions at the core of the globular structure, with residues L59, L62 and I109, of which L62 is the most highly conserved (figure 4.35 B).

In order to test the accuracy of the ECD homology model, it was analysed to find other positions where mutations might have a similar effect to the F98V mutation. Interactions between R-groups of amino acids were identified using UCSF Chimera [197], and filtered for those involving the most conserved amino acids. This process identified three amino acids: L62, L97 and I109, of which L62 was the most highly conserved. Figure 5.1 shows the side-chain interactions of these three amino acids, and the extent to which these occur between conserved amino acids. L97 was not chosen for mutation, as the aim was to use the structure, rather than the sequence, to predict a mutation that would cause a similar effect to F98V. Since L97 is adjacent to F98, similar effects on protein function would not be surprising. L62 was chosen as it was at the centre of this network of interactions, interacting with both L97 and I109, as well as F98; there was a high likelihood of the L62A mutation having a similar effect to F98V, and this prediction was made on the basis of the ECD structure.

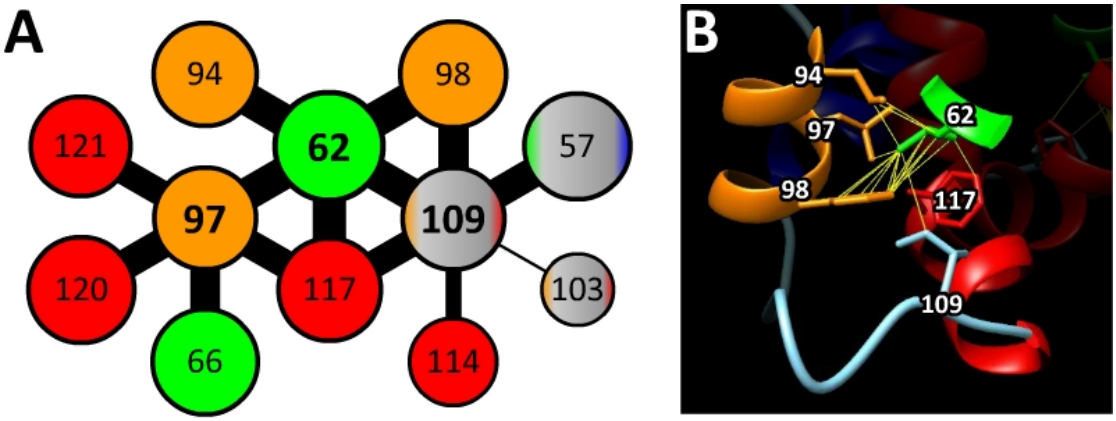


Figure 5.1: **Residue L62 forms multiple interactions with other conserved residues.** A: R-group interactions of residues 62, 97 and 109; Larger circles indicate more highly conserved amino acids, line thicknesses reflect the product of the normalised conservation scores of the linked amino acids, circles are coloured to match the alpha-helices of the ECD structure. B: R-group interactions of L62 in context of the ECD structure; yellow lines indicate interactions between atoms.

5.1.1.1 Mutations Predicted to Inhibit Dimerisation of the ECD

The surface of the dimerisation site of the N-terminal ECD is enriched in hydrophobic residues that form multiple interactions between the two monomers (figure 5.2 A). Of these interacting pairs the most highly conserved were L47 and L122, followed by L69 and L125. L47 and L122 are close to the center of the predicted binding site; mutating them could disrupt the structure of the ECD. L69 and L125 are at the border of the predicted binding site, so these were chosen to be mutated to aspartate. The aim was to disrupt dimerisation by replacing the hydrophobic interactions between L69 and L125, with the repulsion of two negative charges from aspartate. The cysteine predicted to form a disulphide bridge between the two monomers was also chosen to be mutated to aspartate, again, to introduce a repulsive force and inhibit dimerisation. As a final test of the importance of dimerisation of the ECD, a deletion mutant was designed,

removing the entire ECD (residues 40-130), while leaving the signal sequence intact to preserve the correct topology of hZIP14; this was done for isoforms A and B.

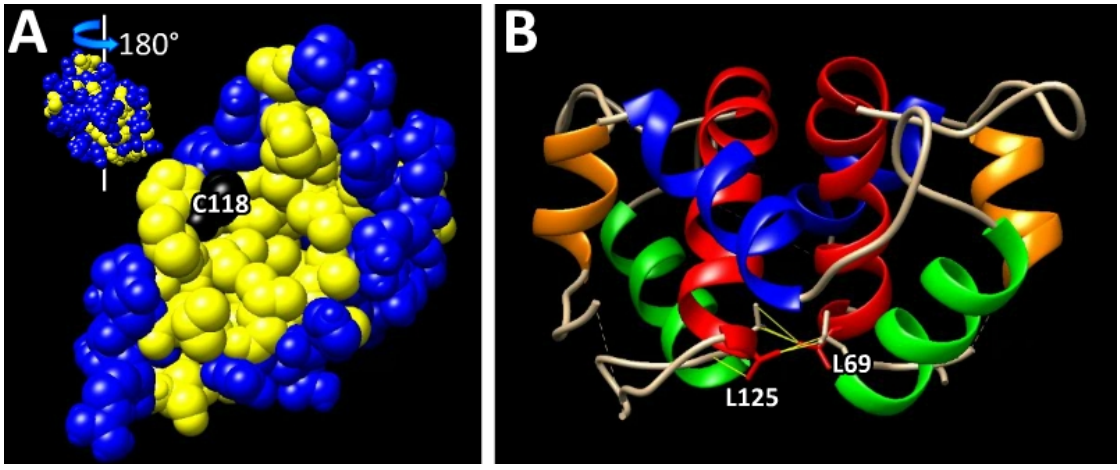


Figure 5.2: **The ECD dimerisation site is enriched in hydrophobic residues.** A: The hZIP14 ECD domain monomer; hydrophilic residues are coloured in blue, hydrophobic residues are coloured in yellow, C118 is in black. B: The hZIP14 ECD domain dimer; helices are coloured from the N-terminus to the C-terminus in blue to red, residues L69 and L125 are labelled, yellow lines indicate interactions between atoms.

5.1.2 Selection of Mutations in the TMD

In order to test the accuracy of the transmembrane domain homology model, the aim was to introduce single mutations predicted to have deleterious effects on hZIP14 function, and then pair these with a second mutation predicted to undo the effects of the first, with these predictions being based on the homology model.

UCSF Chimera was used to identify all interactions between R-groups of the TM helices [197], these were then segregated into interactions which were conserved (identical amino acids involved) between hZIP14 and BbZIP and those which were not. Of the interacting pairs which were identical in both

proteins, only three were in positions highly conserved in both proteins - the interactions between amino acids L206, A354 and A432 (figure 5.3).

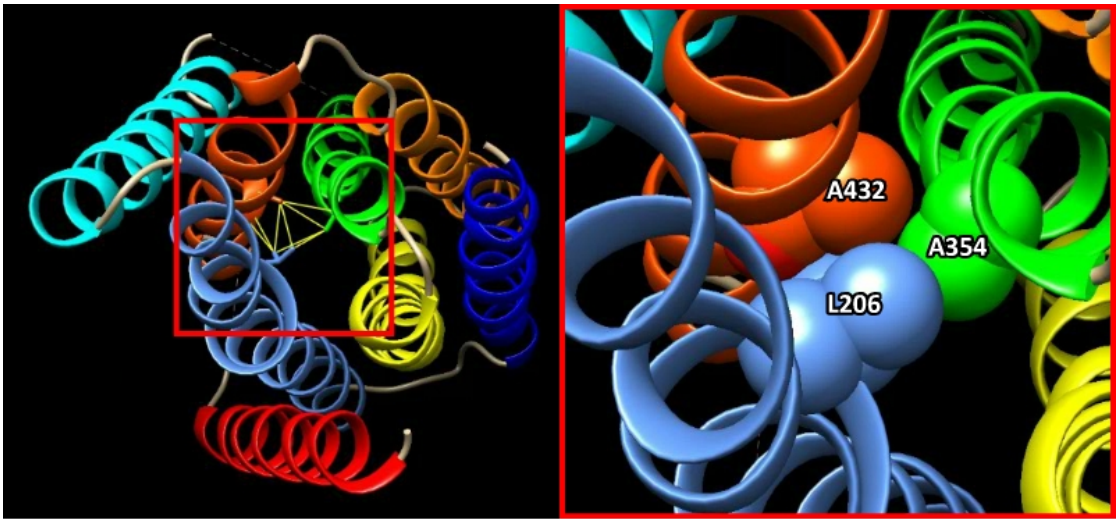


Figure 5.3: **Positions of L206, A354 and A432 in hZIP14.** L206, A354 and A432 in context of the hZIP14 structure; TM helices are coloured from the N-terminus to the C-terminus in blue to red, yellow lines indicate interactions between atoms, the red square shows a close-up of the residues in sphere form.

These amino acids, which correspond to L107, A184 and A265 in BbZIP, meet where TMs 4, 5 and 7 are closest (figure 5.3). In order to test this aspect of the structure, A354 was chosen to be mutated to a leucine - this mutation was predicted to be deleterious to hZIP14 function, as its increased size would be difficult to fit into this tightly packed space, altering the relative positions of the helices involved. The compensatory mutation was L206A, predicted to accommodate the increased size of the A354L mutation.

Of the interacting pairs that differed in identity from BbZIP but were highly conserved in the hZIP14 multiple sequence alignment, and excluding those directly involving the metal transport motif, a promising interaction occurred between E209 and Q365 (figure 5.4). These two amino acids were among the most highly conserved, and form a plausible interaction when viewed in the model. To test this interaction, Q365 was chosen to be mutated to glutamate,

introducing a negative charge predicted to repel the negative charge of E209. The compensatory mutation was E209Q, intended to restore the glutamate-glutamine interaction at this site.

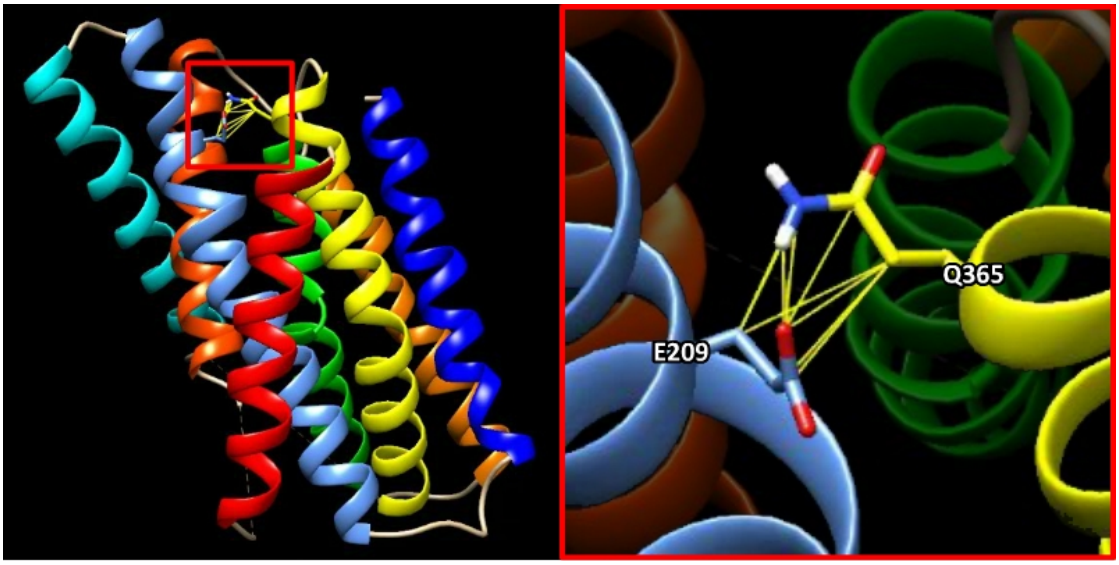


Figure 5.4: **Positions of E209 and Q365 in hZIP14.** E209 and Q365 in context of the hZIP14 structure; TM helices are coloured from the N-terminus to the C-terminus in blue to red, yellow lines indicate interactions between atoms, the red square shows a close-up of the residues (altered orientation).

Figure 5.5 A shows the binuclear metal binding site identified by Zhang et al. [241, 263], of which M1 is believed to be the most important for metal transport. Figure 5.5 B shows the positions of these metal binding sites relative to the hZIP14 homology model, and the key residues predicted to play a role in metal transport in TMs 4 and 5 [13, 263]. The metal transport motif of ZIPs 8 and 14 (EEXPHEXGD) differs from the rest of the LIV-1 family (HEXPHEXGD) [13]. This replacement of histidine with glutamate has been predicted to explain the ability of hZIP14 and hZIP8 to transport manganese and iron, in addition to zinc and cadmium [11, 13]. The position of E376 in the hZIP14 homology model is consistent with this prediction, due to its position in the M1 metal-binding site. It was therefore decided to mutate this residue to histidine to test the effect on

hZIP14 iron transport. The designed mutations are summarised in table 5.1.

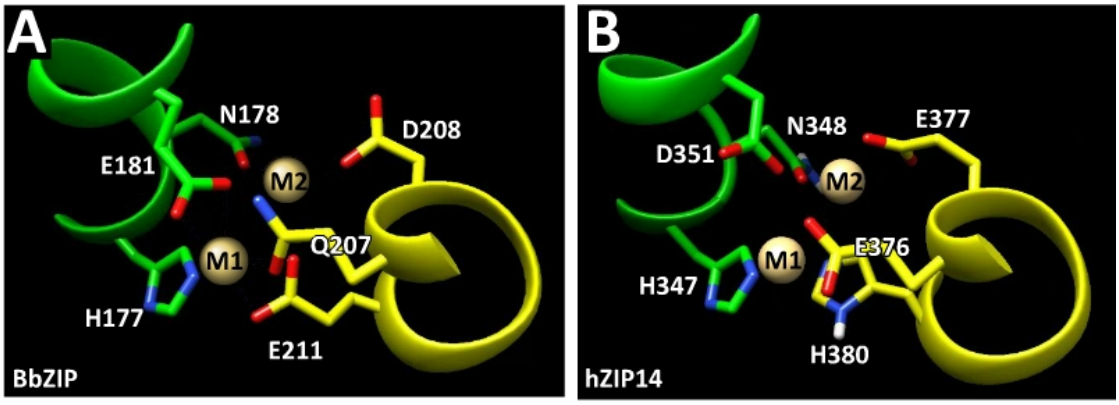


Figure 5.5: **The BbZIP binuclear metal binding site is maintained in the hZIP14 TMD homology model.** M1 and M2 binding sites and key residues are indicated; TM helices are coloured from the N-terminus to the C-terminus in blue to red. A: The binuclear metal binding site of BbZIP. B: The predicted binuclear metal binding site of hZIP14 in the homology model.

5.1.3 Experiments on Designed hZIP14 Mutants

The mutations designed to test the homology models, and functional aspects of hZIP14 predicted by the homology models, range from fairly obvious mutations at positions whose importance could have been predicted from multiple sequence alignments, to more speculative mutations based entirely on the structure, though conservation was considered, in order to increase the likelihood of identifying important amino acid interactions. The designed hZIP14 mutations were tested using methods developed in Chapter 3, in order to better understand hZIP14 function. Successful expression was confirmed by SDS-PAGE analysis and immunofluorescence microscopy. hZIP14 mutants were tested by differential staining and flow cytometry, and then selected mutations were used to create stable monoclonal cell lines for measurements of iron uptake.

Table 5.1: **Summary of mutations designed based on the hZIP14 homology model and the rationale for each.**

Test accuracy of ECD homology model	
Mimics the effect of the F98V mutation.	L62A
Inhibits dimerisation of the ECD	
Introduces a negative charge(s) to the dimerisation interface of each monomer.	L69D, L69D;L125D
Introduces a negative charge(s) to the dimerisation interface of each monomer, and prevents the formation of a disulphide bridge linking the monomers.	C118D
Removes the entire ECD (signal sequence left intact).	$\triangle$40-130 (A & B)
Test accuracy of the TMD homology model	
Introduces a larger amino acid into a tightly packed space.	A354L
Compensates for the A354L mutation by introducing a smaller amino acid.	A354L;L206A
Introduces a negative charge to repel E209.	Q365E
Compensates for the Q365E mutation by removing a negative charge, replacing a glutamine.	Q365E;E209Q
Test importance of E376 in solute selectivity	
Changes the metal transport motif to match other members of the LIV-1 family.	E376H

5.1.4 Co-evolution Analysis of the TMD Homology Model

In order to further investigate the accuracy of the TMD homology model, co-evolution analysis of hZIP14 was undertaken. Functionally important regions of proteins can often be identified by strong sequence conservation between species. However, a protein's tertiary structure shows greater conservation during evolution than a protein's sequence [269]; in the face of random

mutations over time, protein structures are maintained due to positive selection for compensatory mutations in affected positions [270]. In this way, amino acids co-evolve with other amino acids that are nearby in the three-dimensional structure; algorithms have been developed that assess large multiple sequence alignments for co-evolving pairs of amino acids, in order to predict contacts. There are several caveats to consider, the major issue being the difficulty in distinguishing between pairs of amino acids that have genuinely co-evolved, and pairs of amino acids that share a contact, resulting in the appearance of co-evolution [270]. Despite this consideration, patterns of co-evolving amino acids allow me to compare predicted contacts to residue proximities in the homology model, in order to assess its accuracy. The combination of the experimental results with the co-evolution analysis was used to come to a deeper understanding of the the structure-function relationship of hZIP14.

5.2 Results

In the previous chapter, models of the extracellular and transmembrane domains of hZIP14 were created, based on the crystal structures of homologous proteins. Based on these models, a number of mutations were designed to test three aspects of the model: dimerisation of the extracellular domain, interactions between helices of the transmembrane-domain, and the solute specificity of the protein. In this chapter, these mutants will be experimentally characterised, and the transmembrane model will be subjected to co-evolution analysis, in order to assess the accuracy of the models, and their fitness for explaining the mechanisms underpinning the effects of the disease-causing mutations.

5.2.1 Creation of Plasmids for Functional Analysis of Designed Mutations

As discussed in section 2.1.1, the point mutations listed in table 5.2 were introduced into the pIRESneo2 plasmid encoding wild-type human ZIP14, isoform A (Q15043-1 [208]), with a C-terminal hemagglutinin tag (hZIP14_A-HA).

Table 5.2: **Designed mutations and locations.** Mutations to be studied, based on the predicted structure of hZIP14.

Designed Mutations	Location
△40-130 (A & B)	Removal of N-terminal extracellular domain
L62A	N-terminal extracellular domain
L69D, L69D;L125D, C118D	N-terminal extracellular domain (dimerisation interface)
A354L	Transmembrane domain, helix 4
A354L;L206A	Transmembrane domain, interaction between helices 4 and 2
Q365E, E376H	Transmembrane domain, helix 5
Q365E;E209Q	Transmembrane domain, interaction between helices 5 and 2

The △40-130 mutations were made on pIRESneo2 plasmids encoding A and B isoforms of hZIP14-HA. The other mutations were introduced into the wild-type hZIP14_A-HA sequence; the A354L;L206A plasmid was made using site-directed mutagenesis of the A354L hZIP14_A-HA plasmid, likewise for the Q365E;E209Q plasmid and the Q365E hZIP14_A-HA plasmid. Plasmids were sequenced to ensure that the mutations were correct, and that no further mutations had been introduced.

The designed mutations were based on the homology models whose

development is described in section 4.2. For the N-terminal extracellular domain (ECD), the mutations aimed to remove the ECD (△40-130 - the first 39 amino acids were retained to prevent the loss of the signal sequence), to disrupt its folding (L62A), or to inhibit its dimerisation (C118D, L69D and L69D;L125D). For the transmembrane domain (TMD), the mutations were chosen to test the accuracy of the homology model by introducing deleterious mutations in conserved positions (A354L and Q365E), and then introducing compensatory mutations (A354L;L206A and Q365E;E209Q). In addition, the expanded solute profile of ZIP14 and ZIP8 relative to the rest of the LIV-1 family, is believed to be in large part due to E376. LIV-1 proteins contain a conserved HEXPHEXGD motif, and are not known to have physiological roles in the transport of iron or manganese [11]. The ability of ZIP14 and ZIP8 to transport manganese and iron in addition to zinc, is believed to be due to glutamate (E376 in human ZIP14) replacing the first histidine in this motif. The E376H mutation was designed to test this assertion.

5.2.2 Designed hZIP14-HA N-terminal Extracellular Domain Mutants Can Be Expressed in HeLa Cells

HeLa cells were transiently transfected with plasmids encoding the N-terminal hZIP14_A-HA designed mutants and the △40-130 (△ECD) hZIP14-HA mutants. Protein was extracted and analysed by SDS-PAGE electrophoresis and western blot (figure 5.6).

All mutants were expressed, with bands in the expected size ranges. hZIP14 is known to be glycosylated at three positions on the ECD, which introduces variability into the final size of the protein. Due to this glycosylation, the C118D, L62A, L69D and L69D;L125D mutants were all expected to vary in size, from

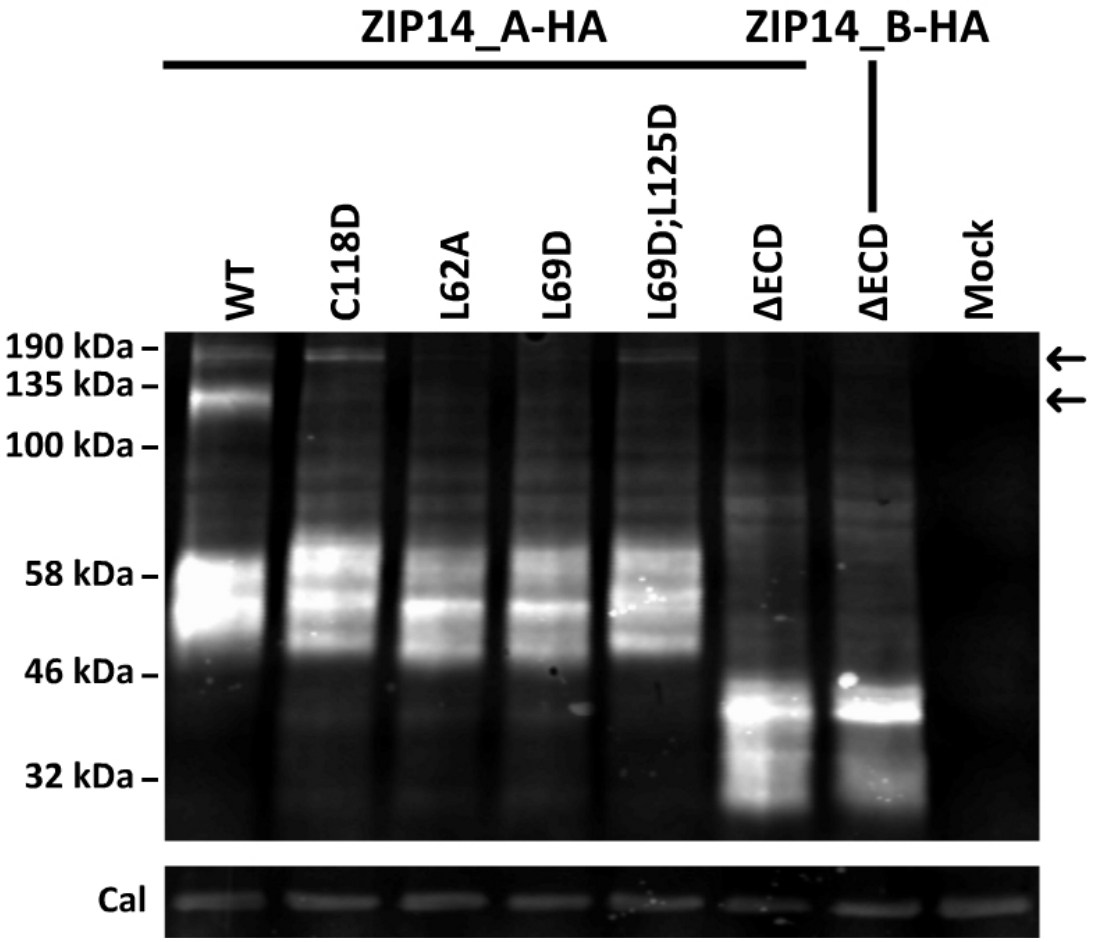


Figure 5.6: **N-terminal designed hZIP14_A-HA mutants were expressed in HeLa cells.** HeLa cells were transiently transfected with plasmids encoding hZIP14-HA variants, or mock-transfected. At 48 hours, transfected cells were lysed and the lysates (25 µg of protein) were analysed by SDS-PAGE (8% polyacrylamide) and western blotting, for hZIP14-HA (top panel) and calreticulin (bottom panel). Arrows indicate the positions of higher molecular mass bands seen.

a minimum of approximately ∼52 kDa; all had bands at approximately 58 kDa. The ΔECD mutants had bands slightly below 46 kDa, the predicted molecular masses are ∼40 kDa. Regarding the two higher molecular mass bands seen in the wild-type hZIP14_A-HA, these bands were not seen for the L62A, L69D or ΔECD mutants, in multiple western blots. For the C118D and L69D;L125D mutants, the ∼180 kDa molecular mass band was repeatedly observed, though

it was consistently fainter for the L69D;L125D than for the C118D mutant. The higher molecular mass bands may represent oligomers of hZIP14-HA, in which case, their absence would suggest that these mutations alter protein-protein interactions.

The L62A mutation was intended to disrupt the structure of the ECD monomer, while the L69D mutant and the L69D;L125D double mutant were both designed to disrupt hydrophobic interactions between the ECD monomers, and predicted to have similar effects. Surprisingly, the bands of the L69D mutant were more similar to those of the L62A mutant than the L69D;L125D double mutant; the size ranges of the bands for L62A and L69D were apparently identical, while the L69D;L125D bands were more like those of C118D.

The $\triangle$ECD mutants resulted in a range of bands; unlike the size ranges of bands for the other mutants which extend above the expected sizes (due to glycosylation increasing the molecular mass of the protein), the size range of the $\triangle$ECD mutants extends to below its expected size. As these mutants are lacking the extracellular domain, they are not expected to be glycosylated, and not expected to have increased molecular masses. The additional lower molecular mass bands may be $\triangle$ECD mutants that are only partially denatured, or in various stages of degradation.

The L62A mutation was predicted to mimic the effect of the F98V mutation, disrupting conserved interactions at the heart of the extracellular domain. For comparison, the L62A hZIP14_A-HA protein was also analysed by Western blot alongside the N-terminal disease-causing mutants (figure 5.7). The pattern of bands was most similar to the F98V mutant, in that the $\sim$52 kDa bands were very similar, and both proteins did not have either of the higher molecular mass bands seen for the wild-type protein, which were faintly visible for the S104I and R128W mutants.

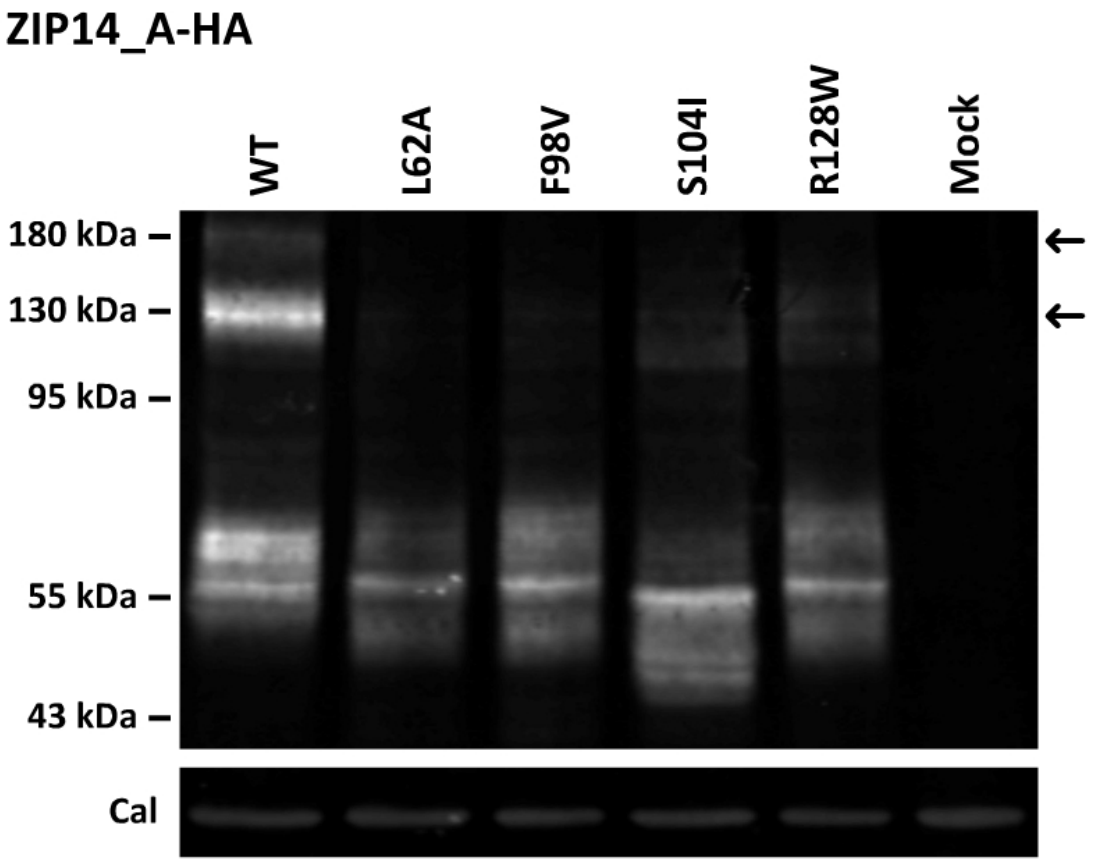


Figure 5.7: **Comparison of the L62A mutant with the N-terminal disease-causing mutations.** HeLa cells were transiently transfected with plasmids encoding hZIP14_A-HA variants, or mock-transfected. At 48 hours, transfected cells were lysed and the lysates (25 µg of protein) were analysed by SDS-PAGE (8% polyacrylamide) and western blotting, for hZIP14_A-HA (top panel) and calreticulin (bottom panel). Arrows indicate the positions of higher molecular mass bands seen. The pattern of bands displayed for the L62A mutant most closely matches that of the F98V mutant, relative to the other N-terminal disease-causing mutants.

5.2.2.1 Cellular Distribution of hZIP14-HA is Altered by the Designed Extracellular Domain Mutations

In section 3.2.5, disease-causing mutations involving the extracellular domain (ECD) caused hZIP14-HA protein to be retained in the endoplasmic reticulum. To see if this was the case for the designed mutations in the ECD, HeLa cells

were transiently transfected with plasmids encoding the N-terminal hZIP14_A-HA designed mutants and the △40-130 (△ECD) hZIP14-HA mutants. Cells were seeded onto coverslips and stained for hZIP14-HA variants. For each of the N-terminal mutants, cellular distribution varied, seemingly dependant on the level of expression (figures 5.8 and 5.9).

Figure 5.8 shows cells where the cellular distributions of the N-terminal mutants differed from the wild-type hZIP14_A-HA. In these images, the L62A, L69D, L69D;L125D and C118D mutants are not visible at the plasma membrane. They show an intracellular pattern resembling that of the N-terminal disease-causing mutants and calreticulin, with increased levels of the mutant hZIP14_A-HA proteins around the nucleus and in a reticular pattern (figures 3.13, 3.16 and 3.17). The △ECD mutants were slightly different, often seen with little or no protein at the membrane, bright intracellular puncta resembling wild-type hZIP14-HA, and as well as a bright line around the nucleus, similar to the N-terminal disease-causing mutants, an additional bright region of high intensity next to the nucleus, an aggregation of mutant hZIP14-HA. These localisation patterns were seen in cells that stained less brightly, and required higher exposure times in order to obtain clear images of the cells, indicating lower levels of expression of hZIP14-HA.

On the other hand, figure 5.9 shows images of cells expressing the same mutants, which more closely resemble wild-type hZIP14-HA. The △ECD mutants both look indistinguishable from wild-type, while the L62A, L69D, L69D;L125D and C118D mutants pictured are present at the cell membrane, have bright intracellular puncta, though they typically still had increased amounts surrounding the nucleus. These distributions were seen in cells that had stained more brightly, and required short exposure times, suggesting much higher levels of expression.

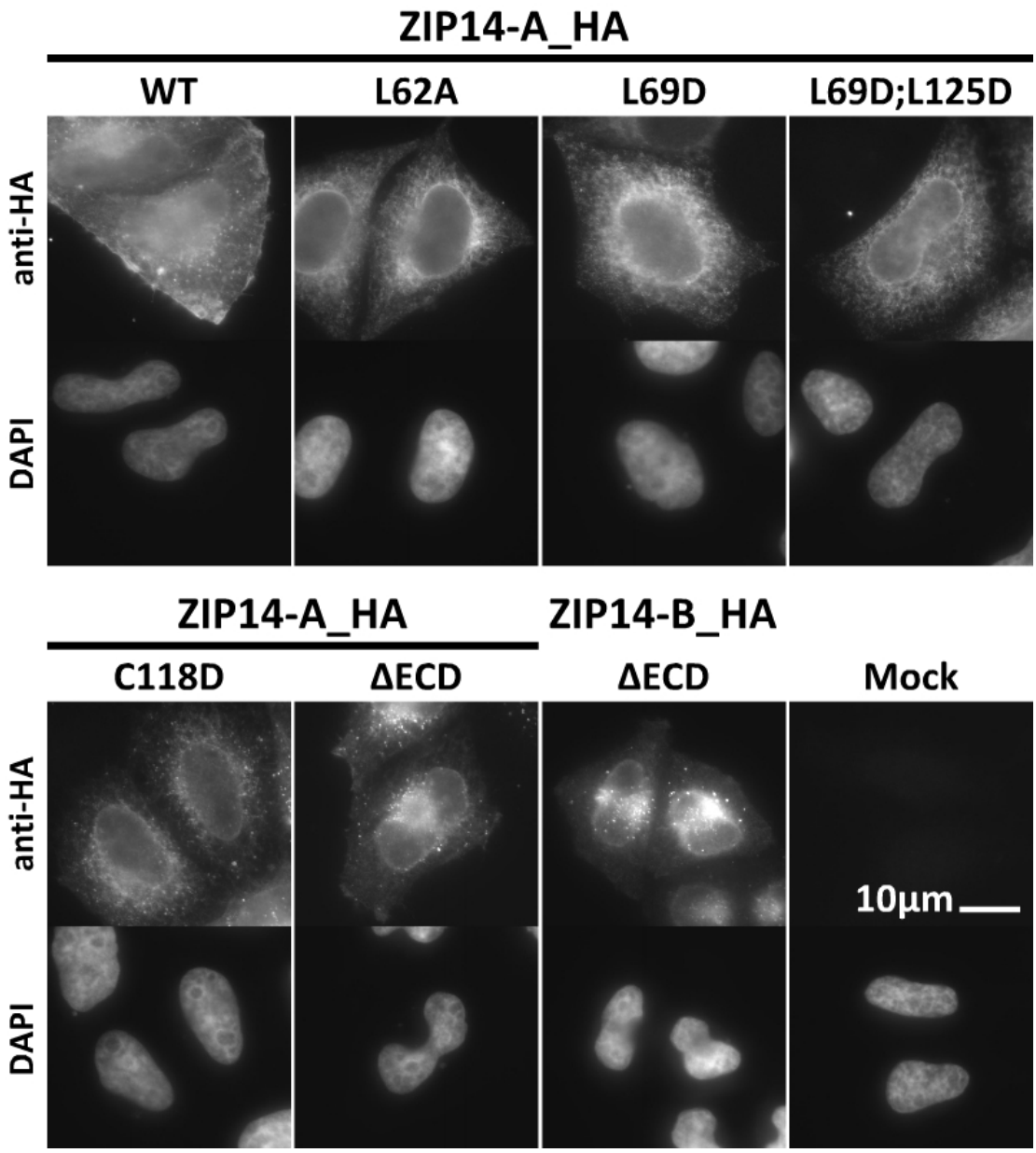


Figure 5.8: **Effects of N-terminal designed mutations on cellular distribution of hZIP14-HA.** HeLa cells were transiently transfected with plasmids encoding hZIP14_A-HA variants, or mock-transfected; ΔECD refers to the removal of the ECD, amino acids 40-130. Transfectants were seeded onto coverslips and stained with mouse anti-HA antibody, followed by anti-mouse Alexa Fluor 488. Coverslips were viewed by immunofluorescence microscopy.

The L62A mutation was predicted to mimic the effect of the F98V mutation, and has shown the same range of cellular distributions, appearing to be retained in the endoplasmic reticulum, but also localising to the cell surface in some

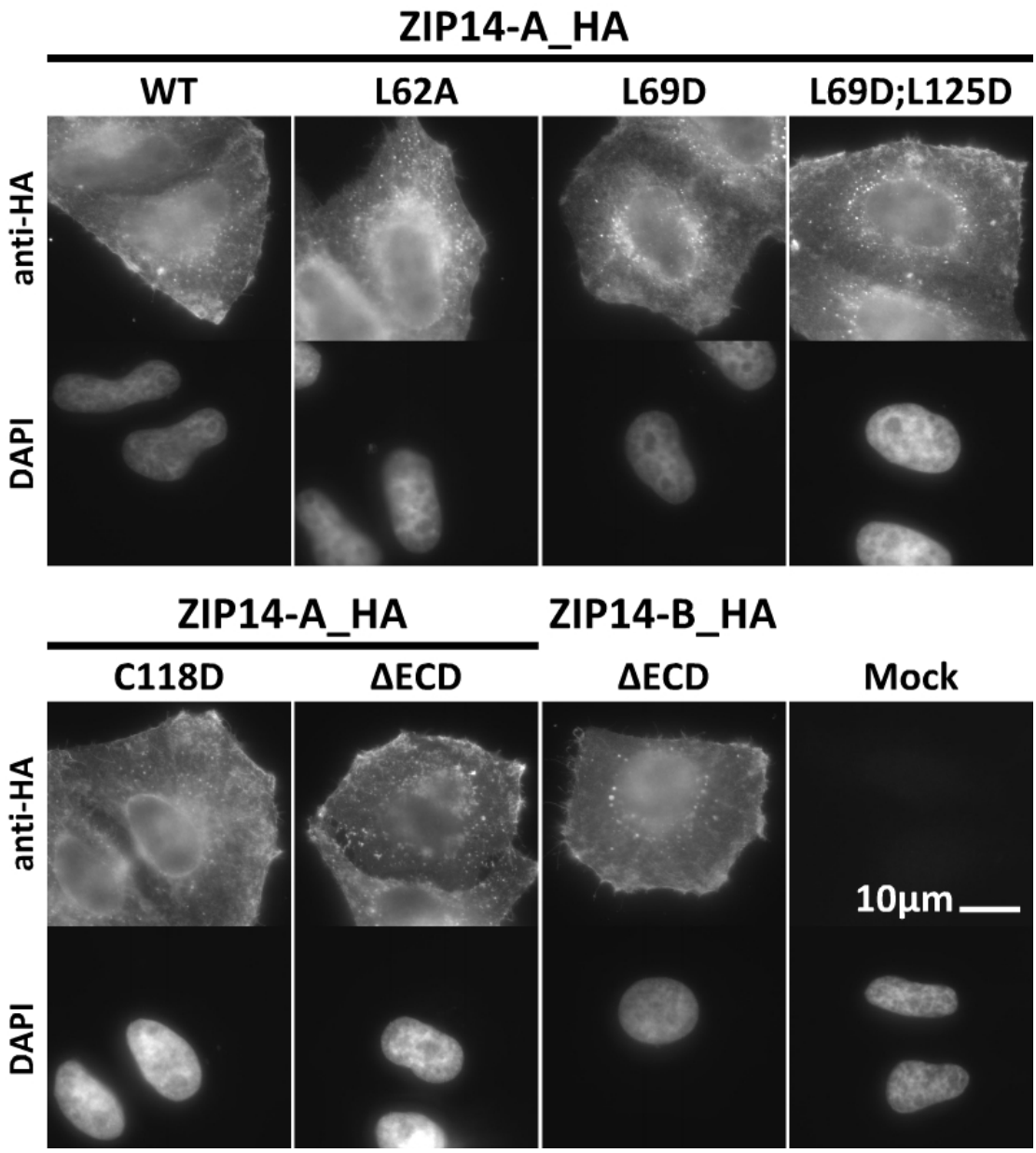


Figure 5.9: **Effects of N-terminal designed mutations on cellular distribution of hZIP14-HA are diminished at high levels of expression.** HeLa cells were transiently transfected with plasmids encoding hZIP14_A-HA variants, or mock-transfected; ΔECD refers to the removal of the ECD, amino acids 40-130. Transfectants were seeded onto coverslips and stained with mouse anti-HA antibody, followed by anti-mouse Alexa Fluor 488. Coverslips were viewed by immunofluorescence microscopy.

cells. It should be noted that immunofluorescence microscopy shows no clear differences between F98V and the other N-terminal disease-causing mutations,

S104I and R128W (figure 3.13).

The C118D, L69D and L69D;L125D mutations were all predicted to decrease dimerisation of the extracellular domains without affecting the structure of the monomer itself, and seem to have had similar effects to the disease-causing N-terminal domain mutations on the cellular localisation of hZIP14_A-HA; each of these mutants showed endoplasmic reticulum-like distributions at lower expression levels and at higher expression levels (as approximated by immunofluorescence microscopy), the distributions were more similar to wild-type. The appearance of the △ECD mutants, was surprisingly different from all the other N-terminal mutations, aggregating next to the nucleus. While a quantitative study of the differences in trafficking between these mutants is more suited to flow cytometry, immunofluorescence microscopy reveals that a mutation predicted to disrupt the structure of the ECD has a similar effect to mutations predicted to inhibit ECD dimerisation, and that removal of the extracellular domain has additional effects on hZIP14-HA trafficking, while still allowing wild-type-like distributions at high expression levels.

5.2.2.2 Effect of Expression on Surface Levels of Designed hZIP14-HA N-terminal ECD Mutants

Immunofluorescence microscopy of these mutants (figures 5.9 and 5.8) showed variability in cellular distributions, including differences in trafficking to the cell surface. Immunofluorescence microscopy resulted in similar findings for the N-terminal disease-causing mutants (figure 3.29), and flow cytometry was able to reveal a dependence of cell surface localisation on expression levels that differed between the wild-type and the disease-causing ECD mutants (figure 3.36). The sigmoidal distributions observed were hypothesized to be caused by decreased binding between the ECD monomers, resulting in decreased cell

surface localisation of the hZIP14_A-HA proteins.

The designed ECD mutations (L62A, L69D, L69D;L125D, and the △ECD mutations) are all predicted to impact dimerisation of hZIP14-HA, resulting in similar sigmoidal relationships between surface and interior hZIP14-HA protein. These mutants were all tested by flow cytometry, to characterise the relationship between expression levels and cell surface localisation of these mutants.

HeLa cells were transiently transfected with plasmids encoding hZIP14_A-HA with mutations designed to disrupt dimerisation of the extracellular domain, and differentially stained for surface and intracellular hZIP14-HA, as described in section 3.2.7.

The C118D, L69D and L69D;L125D mutations all displayed a sigmoidal relationship between internal and surface levels of hZIP14_A-HA, unlike the hyperbolic relationship observed for wild-type hZIP14_A-HA (figure 5.10). The relationship between internal and surface levels of hZIP14_A-HA for L69D was indistinguishable from that of L69D;L125D. The pattern for C118D showed small but consistent differences; at low levels of expression, the gradient was steeper and at a higher offset (closer to wild-type) than the L69D and L69D;L125D mutants, and the increase in gradient occurred at a lower level of interior hZIP14_A-HA protein. Taken together, this suggests that lower levels of expression are required for the C118D mutant to localise to the cell surface, compared to the L69D and L69D;L125D mutants. In addition, differences in the proportions of cells with high amounts of hZIP14_A-HA protein at the cell surface were also consistently seen between the mutants. HeLa cells transfected with C118D mutant had the largest proportion of cells with high amounts of hZIP14_A-HA C118D on the surface, followed by the L69D;L125D mutant. For cells transfected with the L69D mutant, only a small proportion of cells had high levels of L69D hZIP14_A-HA at the surface.

The L62A mutation, which was designed to disrupt the structure of the ECD similarly to F98V, showed the same type of sigmoidal relationship between interior and surface levels of protein as seen in the N-terminal disease-causing mutations (figures 5.11 and 3.36), F98V is pictured here for comparison. The effects of the dimerisation mutants are very similar to those of the other ECD mutations, disease-causing or designed, in that they decrease trafficking to the cell surface at low concentrations. One interpretation is that the primary purpose of the extracellular domain is to enhance trafficking to the surface of the cell, in a dimerisation-dependent manner.

The behaviour of the $\triangle$ECD mutants contradicts this interpretation, as they were both able to traffic to the cell surface (figure 5.11). At moderate expression levels, the $\triangle$ECD hZIP14_A-HA mutant trafficked to the cell surface more effectively than the dimerisation mutants (figure 5.10), and at high expression levels, more effectively than wild-type hZIP14_A-HA. At low expression levels, the interpretation is more complex. The lower left area of the density plot corresponds to the lowest expression levels; increasing expression at this point creates a diagonal pattern trending to the upper right, indicating that a proportion of the additional protein is being directed to the cell membrane. This trend stops abruptly, and further increases in expression are represented at the middle left of the density plot. A horizontal shift to the far left indicates an abrupt increase in the proportion of protein at the surface at this slightly higher level of expression. From this expression level onwards, $\triangle$ECD hZIP14_A-HA behaves very much like wild-type hZIP14_A-HA, trending towards slightly higher amounts of protein at the surface for a given expression level. $\triangle$ECD hZIP14_B-HA had similar results to the A isoform, but with even higher proportions at the cell surface, across all levels of expression.

The results for the $\triangle$ECD and dimerisation mutants indicate that the ECD is not

essential for trafficking to the cell surface, but when present, levels at the cell surface are decreased by mutations that prevent dimerisation of this domain.

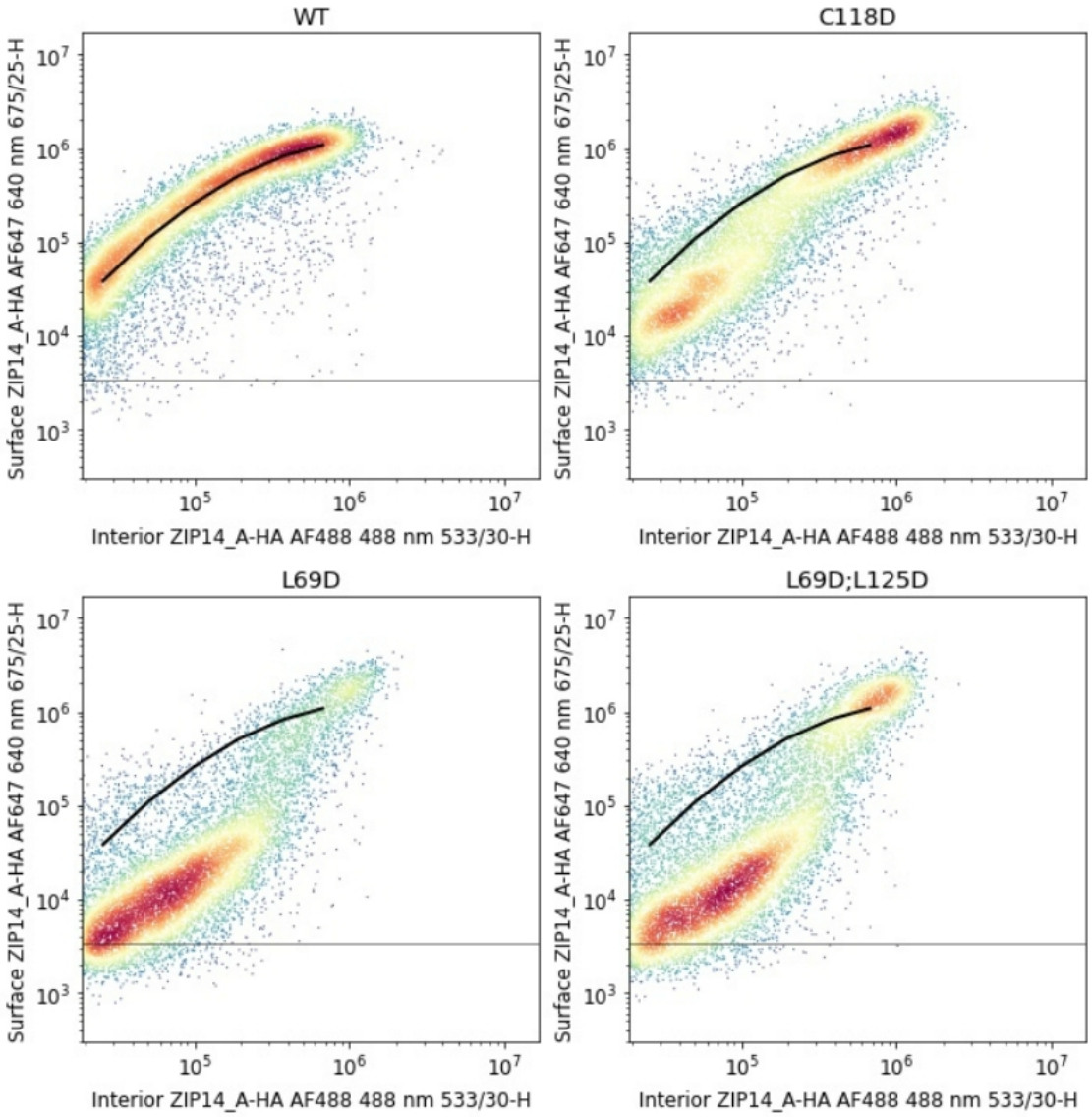


Figure 5.10: **Effect of dimerisation mutations on cell surface levels of hZIP14_A-HA relative to internal levels.** HeLa cells were transiently transfected with plasmids encoding hZIP14_A-HA variants predicted to disrupt the dimerisation interface of the extracellular domain (C118D, L69D and L69D;L125D), or mock-transfected. Transfectants in suspension were incubated with mouse anti-HA, then anti-mouse Alexa Fluor 647 antibody, in non-permeabilising conditions. Cells were then fixed and permeabilised, then incubated with rat anti-HA, then anti-rat Alexa Fluor 488 antibody. Cell readings were taken in the FL-1 and FL-4 channels of a BD Accuri^{TM} C6 flow cytometer. Density plots show results for transfected cells, and are representative of three independent experiments. The horizontal line demarcates background surface fluorescence observed in mock-transfected cells; the solid black line represents wild-type hZIP14_A-HA MFIs.

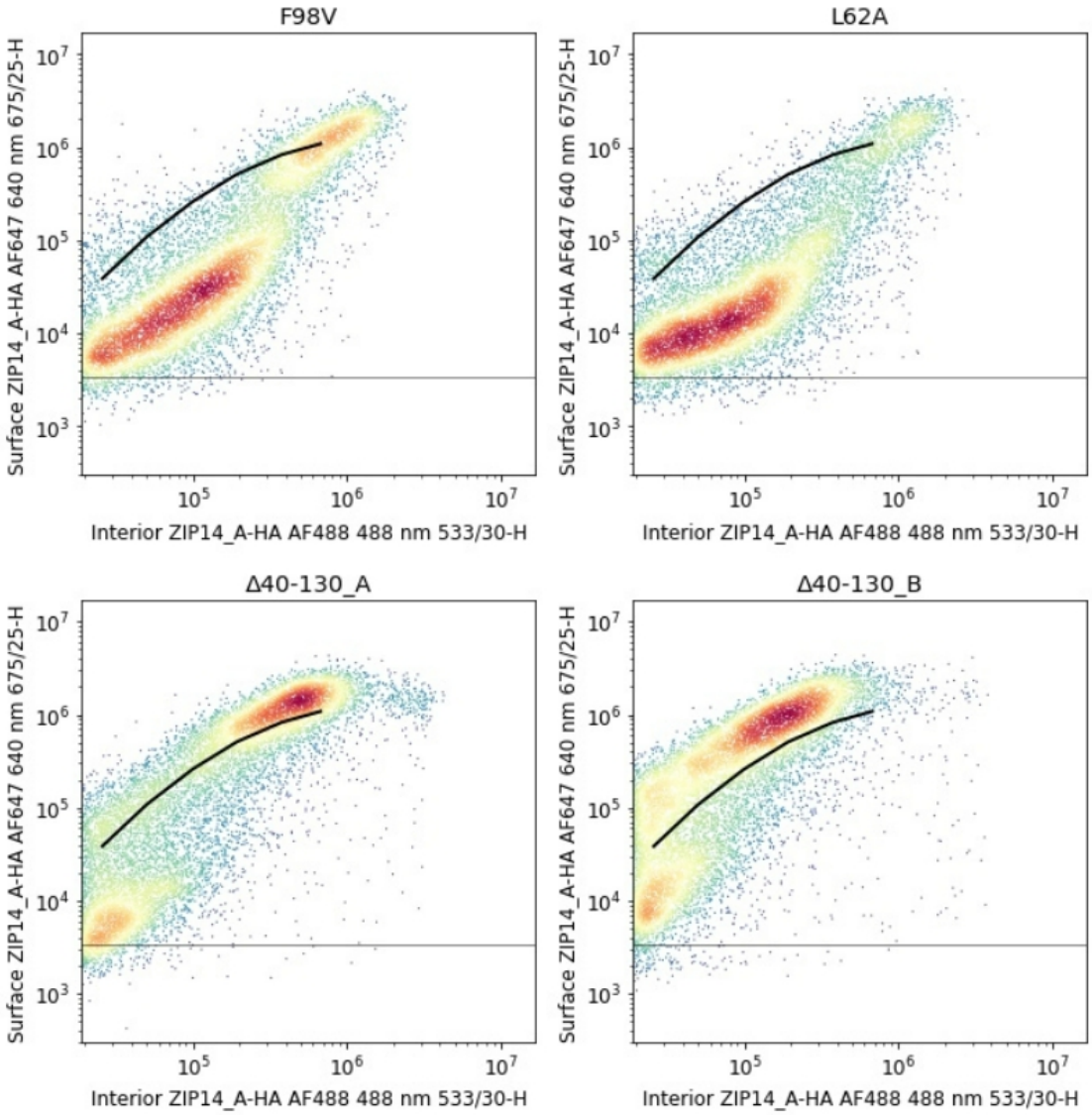


Figure 5.11: **Effect of designed ECD mutations on cell surface levels of hZIP14-HA relative to internal levels.** HeLa cells were transiently transfected with plasmids encoding hZIP14-HA variants predicted to remove or damage the structure of the extracellular domain, or mock-transfected. Transfectants in suspension were incubated with mouse anti-HA, then anti-mouse Alexa Fluor 647 antibody, in non-permeabilising conditions. Cells were then fixed and permeabilised, then incubated with rat anti-HA, then anti-rat Alexa Fluor 488 antibody. Cell readings were taken in the FL-1 and FL-4 channels of a BD Accuri^{TM} C6 flow cytometer. Density plots show results for transfected cells, and are representative of three independent experiments. The horizontal line demarcates background surface fluorescence observed in mock-transfected cells; the solid black line represents wild-type hZIP14_A-HA MFIs.

5.2.2.3 Creation of A Stable Monoclonal Cell Line Expressing L69D;L125D hZIP14_A-HA

To test the importance of dimerisation on iron uptake function, the L69D;L125D mutant was selected for further study - it appeared to have a greater impact on dimerisation than the C118D mutation, but was more able to traffic to the surface than the L69D mutant.

Stable cell lines were created, as described in section 2.1.2.1, and the line with the highest expression was retained. Figure 5.12 shows the total expression level of the L69D;L125D hZIP14_A-HA relative to the wild-type monoclonal cell line. Iron uptake results are discussed in section 5.2.4.

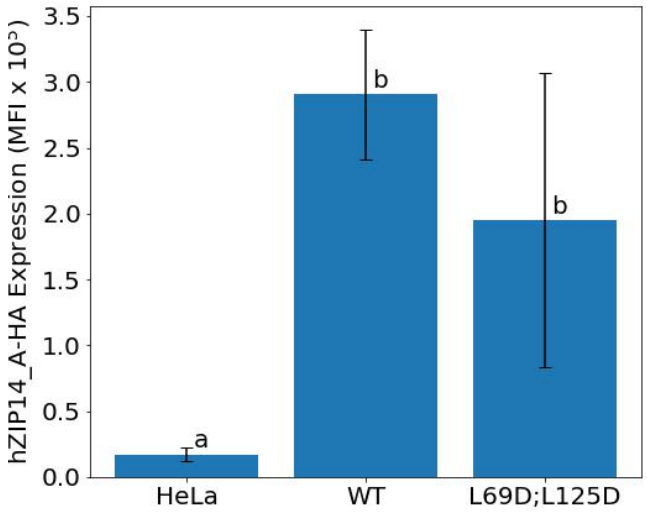


Figure 5.12: **Expression of wild-type and L69D;L125D hZIP14_A-HA by monoclonal cell lines.** HeLa cells were transfected with a pIRESneo2 plasmid encoding L69D;L125D hZIP14_A-HA, or mock transfected. Cells were treated with G418 to select for cells that had incorporated neomycin resistance, and monoclonal cell lines cultivated from the survivors. Samples of the cell lines were permeabilised, fixed and then stained for analysis by flow cytometry; the cell line with the highest expression of L69D;L125D hZIP14_A-HA was chosen. For each cell line, n=3. Error bars represent the standard deviation; different letters represent means that are significantly different from each other (one-way ANOVA with Tukey-Kramer post-hoc analysis, $P < 0.05$).

5.2.3 Designed hZIP14_A-HA Transmembrane Domain Mutants Can Be Expressed in HeLa Cells

Designed mutations in the transmembrane domain (TMD) were intended to test the validity of the TMD homology model, based on the *B. bronchiseptica* ZIP14 crystal structure (PDB ID:5TSA). Two pairs of conserved amino acids (E209 and Q365, and L206 and A354), predicted to interact based on the homology model, were chosen for testing. One amino acid of each pair was mutated to the other amino acid (Q365E and A354L) in order to disrupt the interaction between the two residues. Mutating Q365 to a glutamate should result in two negative charges (Q365E and E209) positioned in close proximity, and potentially repelling each other rather than interacting. Mutating A354 to the larger leucine residue, should clash with L206, disrupting the structure of hZIP14 in this region. The other amino acid was then mutated to its partner (Q365E;E209Q and A354L;L206A) - these double-mutants were predicted to revert to wild-type behaviour, as the double-mutations should replicate the conserved interactions. These mutations were expected to primarily affect metal uptake function rather than trafficking, though effects on trafficking were still a possibility, especially if overall protein stability was altered.

The E376H mutation was expected to decrease iron uptake function, and not to impact protein stability or trafficking. hZIP14 is a member of the zinc-transporting LIV-1 family, most of which contain a conserved HEXPHEXGD believed to be essential for zinc transport and solute specificity. ZIP14 and ZIP8 differ from other members of the LIV-1 family in their ability to transport iron, manganese and cadmium, in addition to zinc. The ability to transport these additional metals is believed to be a result of a difference in the metal-transporting motif, i.e., the glutamate residue in place of the initial histidine of

the motif [13]. The E376H mutation reverts the motif back to the form seen in the other members of the LIV-1 family, and was created to test the effect of this residue on solute specificity of hZIP14.

In order to check whether the designed TMD mutants could be successfully expressed in HeLa cells, HeLa cells were transiently transfected with plasmids encoding the designed TMD hZIP14_A-HA mutants, and the protein was extracted and analysed by SDS-PAGE electrophoresis and western blotting (figure 5.13).

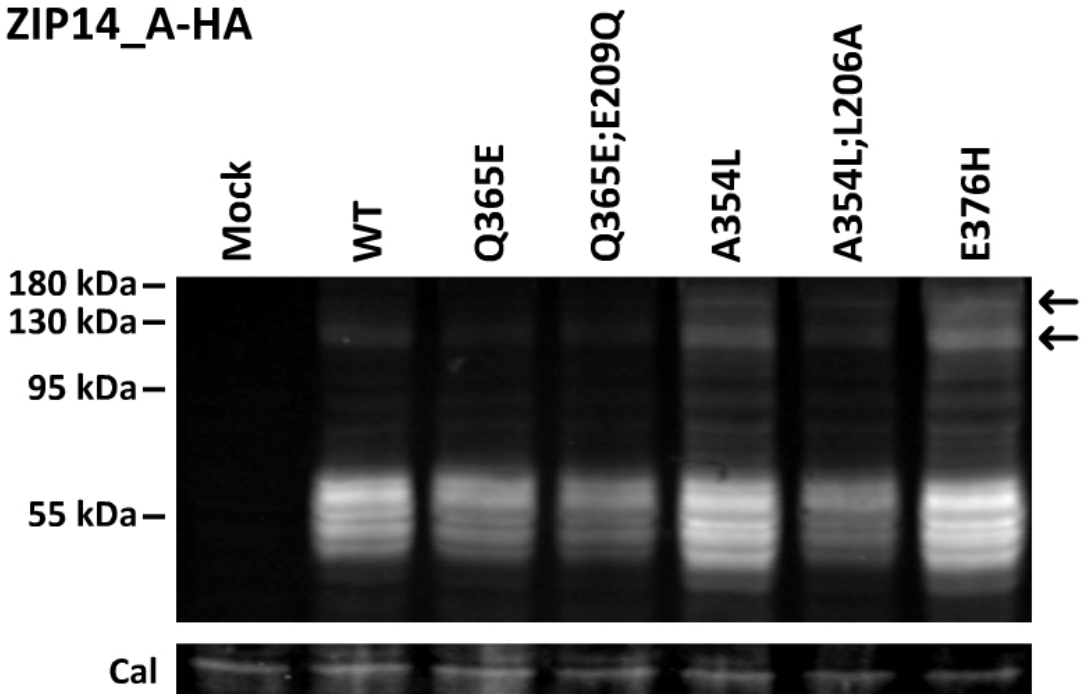


Figure 5.13: **Western blot analysis of transmembrane domain designed mutations.** HeLa cells were transiently transfected with plasmids encoding hZIP14_A-HA variants, or mock transfected. Transfected cells were lysed and the lysates (40 µg of protein) were analysed by SDS-PAGE (8% polyacrylamide) and western blotting, for hZIP14-HA (top panel) and calreticulin (Cal). Arrows indicate the positions of higher molecular mass bands seen.

All designed TMD mutants were expressed, with bands in the expected size ranges. Each mutant had a pattern of bands identical to wild-type hZIP14_A-HA, including the two higher molecular mass bands, with the brightness varying

according to the efficiency of the transfection. These results give no indication that the mutations had any impact on the structure or glycosylation of hZIP14_A-HA. This is in contrast to the disease-causing TMD mutations, which did alter the band patterns of the mutants, relative to the wild-type (figure 3.7).

5.2.3.1 The Designed TMD Mutations Did Not Alter Cellular Distribution of hZIP14_A-HA

In order to determine if the designed TMD mutations altered cellular localisation of hZIP14-HA, HeLa cells were transiently transfected with the designed TMD mutations and examined by immunofluorescence microscopy. The images showed that cellular distributions of the mutants could not be distinguished from that of wild-type hZIP14_A-HA; the mutants were seen on intracellular puncta, and at the cell membrane. As expected of the designed TMD hZIP14_A-HA mutants, there were no indications of retention in the endoplasmic reticulum, or failure to localise at the cell surface (figure 5.14).

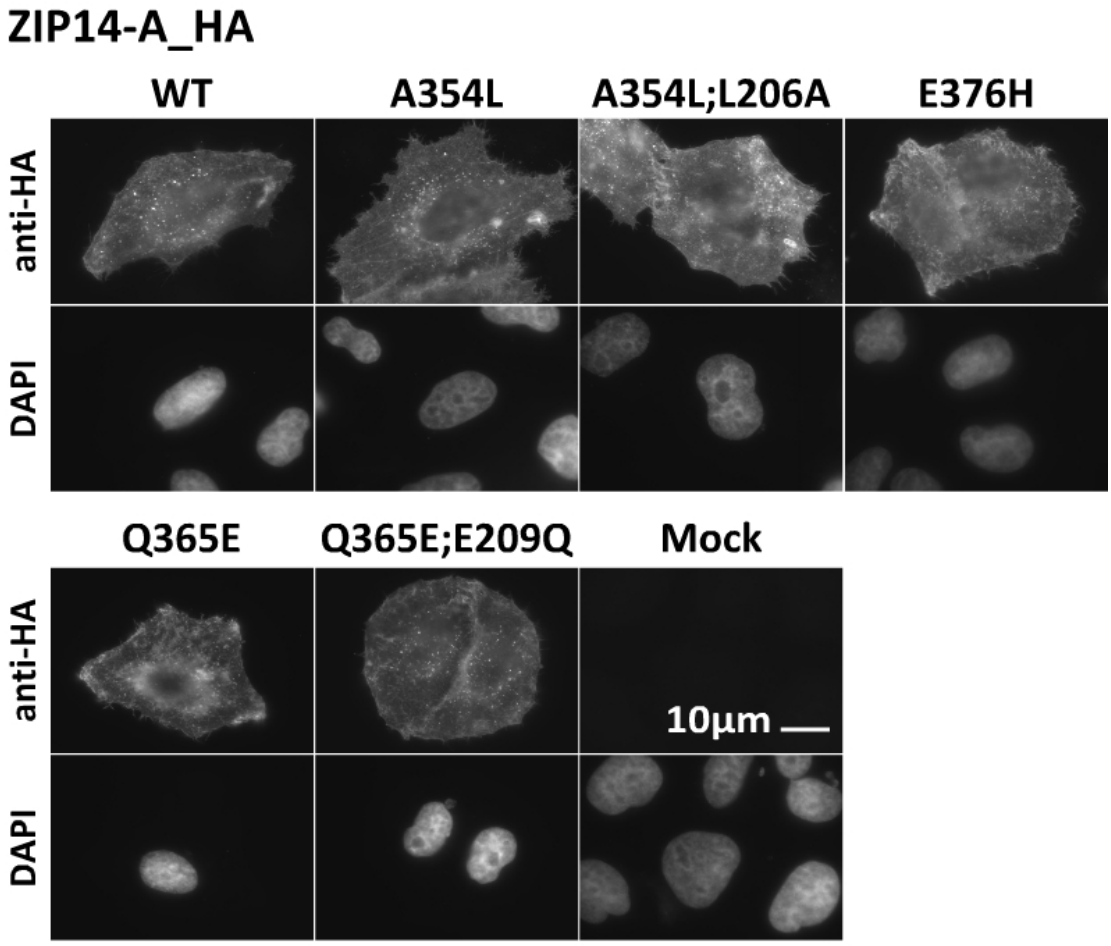


Figure 5.14: **Effects of TMD designed mutations on cellular distribution of hZIP14_A-HA.** HeLa cells were transiently transfected with plasmids encoding HA-tagged ZIP14 variants, or mock-transfected. Transfectants were seeded onto coverslips and stained with mouse anti-HA antibody, followed by anti-mouse Alexa Fluor 488. Coverslips were viewed by immunofluorescence microscopy.

5.2.3.2 The Designed TMD hZIP14_A-HA Mutations Had Subtle Effects on Cell Surface Expression

The designed TMD hZIP14_A-HA mutations were not expected to significantly affect the cellular localisation of the protein, and immunofluorescence microscopy did not detect any clear effects. However, differential staining of transiently transfected cells can quantitatively identify subtle effects on

expression of hZIP14-HA at the cell surface. To this end, HeLa cells were transiently transfected with plasmids encoding the designed TMD hZIP14_A-HA mutants, and differentially stained for intracellular and cell surface hZIP14_A-HA.

Figure 5.15 shows the relationships between internal and surface hZIP14_A-HA protein for the designed mutations in the transmembrane region. The E376H mutant displays less variation in cell surface levels than the wild-type; the cell surface levels are higher than wild-type at low levels of expression, and increase to a lesser extent as expression increases. A slight decrease in cell surface levels was observed for cells with the highest amounts of interior E376H hZIP14_A-HA, possibly caused by internalisation of the protein into the cell.

The relationship between interior and surface levels of the A354L hZIP14_A-HA mutant was almost identical to that of wild-type hZIP14_A-HA. The only consistent difference observed was the apparent internalisation of A354L protein at high levels of expression. The A354L;L206A double-mutant did differ from wild-type hZIP14_A-HA - rather than the hyperbolic curve observed for wild-type, there was an approximately straight line relationship, and slightly lower levels of A354L;L206A at the membrane than wild-type hZIP14_A-HA, for a given level of internal hZIP14_A-HA.

The Q365E and Q365E;E209Q mutants both displayed near straight-line relationships between the interior and cell surface levels of the proteins, similar to the A354L;L206A double-mutant, but with higher levels of protein at the surface, closely matching the wild-type levels.

The relationship for the A354L and E376H mutants is most similar to the G383R mutant - a hyperbolic curve similar to that of wild-type, and then a slight decrease in surface levels at the highest expression levels (figure 3.37).

All of the designed transmembrane domain mutations showed subtle effects on

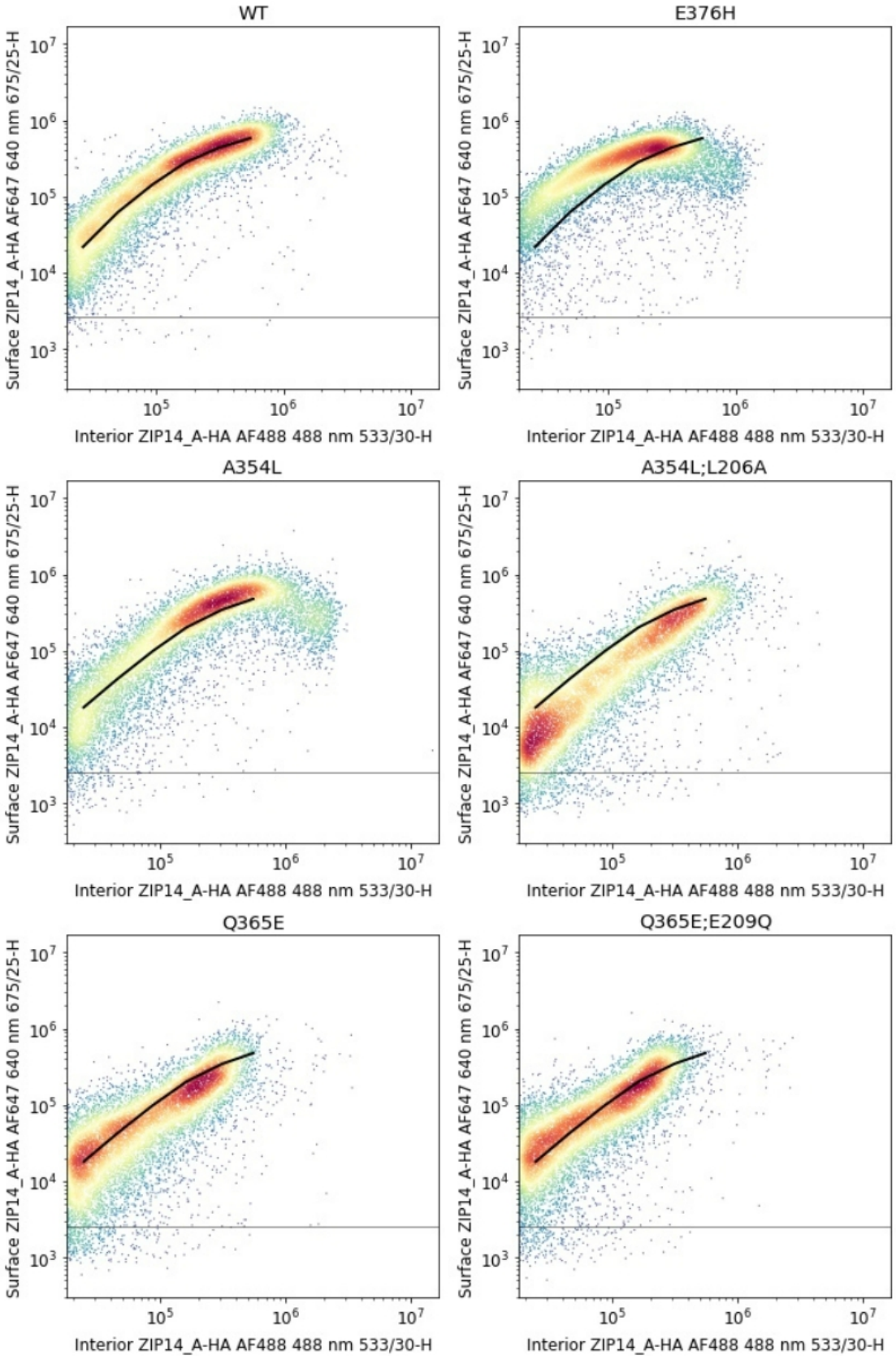

WT
E376H
A354L
A354L;L206A
Q365E
Q365E;E209Q
Surface ZIP14_A-HA AF647 640 nm 675/25-H
Interior ZIP14_A-HA AF488 488 nm 533/30-H
10^3
10^4
10^5
10^6
10^7

Figure 5.15: **Effects of designed TMD mutations on cell surface levels of hZIP14_A-HA relative to internal levels.** HeLa cells were transiently transfected with plasmid encoding hZIP14_A-HA variants, or mock-transfected. Transfectants in suspension were incubated with mouse anti-HA, then anti-mouse Alexa Fluor 647 antibody, in non-permeabilising conditions. Cells were then fixed and permeabilised, then incubated with rat anti-HA, then anti-rat Alexa Fluor 488 antibody. Cell readings were taken in the FL-1 and FL-4 channels of a BD Accuri^{TM} C6 flow cytometer. Density plots show results for transfected cells, and are representative of three independent experiments. The horizontal line demarcates background surface fluorescence observed in mock-transfected cells; the solid black line represents wild-type hZIP14_A-HA MFIs.

trafficking of hZIP14_A-HA to the surface, without preventing it, but none have the sigmoidal relationship so typical of the ECD mutations.

5.2.3.3 Creation of Stable Monoclonal Cell Lines Expressing Q365E, Q365E;E209Q and E376H hZIP14_A-HA

The designed TMD hZIP14_A-HA mutations were predicted to primarily affect metal uptake function; the E376H mutant was specifically created in order to study its impact on iron uptake. Of the paired mutants, the Q365E and Q365E;E209Q mutants were chosen for iron uptake experiments, as their trafficking behaviours appeared to be very similar to each other (figure 5.15).

Stable monoclonal cell lines for these mutants were created, as described in section 2.1.2.1, and the cell lines with the highest expression levels were retained for further experiments; expression levels are shown in figure 5.16.

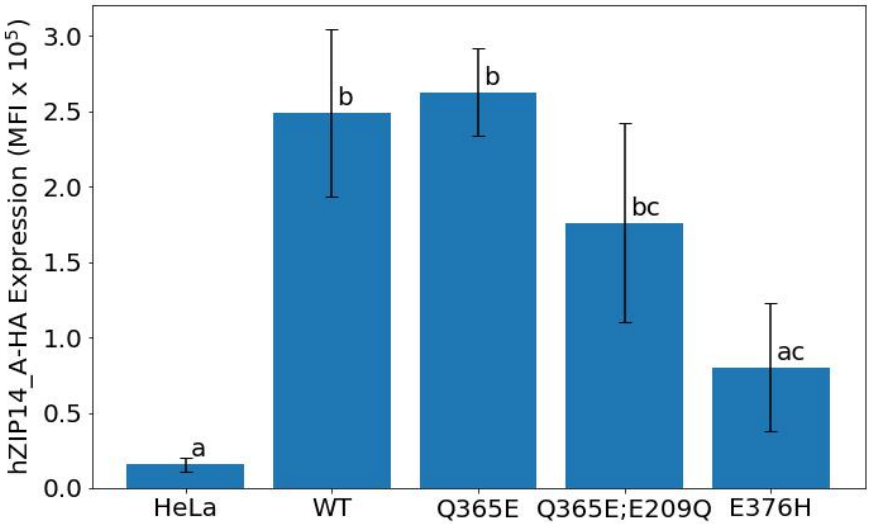


Figure 5.16: **Expression of wild-type, Q365E, Q365E;E209Q hZIP14_A-HA by monoclonal cell lines.** HeLa cells were transfected with pIRESneo2 plasmids encoding Q365E and Q365E;E209Q hZIP14_A-HA, or mock transfected. Cells were treated with G418 to select for cells that had incorporated neomycin resistance, and monoclonal cell lines cultivated from the survivors. Samples of the cell lines were permeabilised, fixed and then stained for analysis by flow cytometry; the cell lines with the highest expression levels were chosen. For E376H, n=3; for other cell lines, n=4. Error bars represent the standard deviation; different letters represent means that are significantly different from each other (one-way ANOVA with Tukey-Kramer post-hoc analysis, $P < 0.05$).

5.2.4 Iron Uptake by the Designed hZIP14_A-HA Mutants

In this section the impact of a subset of the designed mutations on iron uptake was studied. Mutants in both the N-terminal ECD (L69D;L125D) and the C-terminal TMD (Q365E, Q365E;E209Q and E376H) were used to create monoclonal HeLa cell lines stably expressing these mutants (for expression levels, see figures 5.12 and 5.16).

HeLa cells (the negative control) and monoclonal cell lines stably expressing wild-type, L69D;L125D, Q365E, Q365E;E209Q and E376H hZIP14_A-HA were incubated in 10µM $FeSO_4$ labelled with ^{55}Fe for 30 mins. Uptake of ^{55}Fe was

measured by liquid scintillation counting, and results were normalised to protein concentration, measured by BCA assay.

Average uptake of iron by cells expressing the L69D;L125D and E376H mutants was very similar to that of HeLa cells, the negative control (figure 5.17); expression of these mutants made little or no difference to the iron uptake capacity of the HeLa cells expressing them. Iron uptake by cells expressing the Q365E mutant was consistently higher than cells expressing wild-type hZIP14_A-HA, while uptake by cells expressing the Q365E;E209Q mutant (meant to undo the effect of the Q365E mutation) was on average similar to cells expressing wild-type hZIP14_A-HA (figure 5.17).

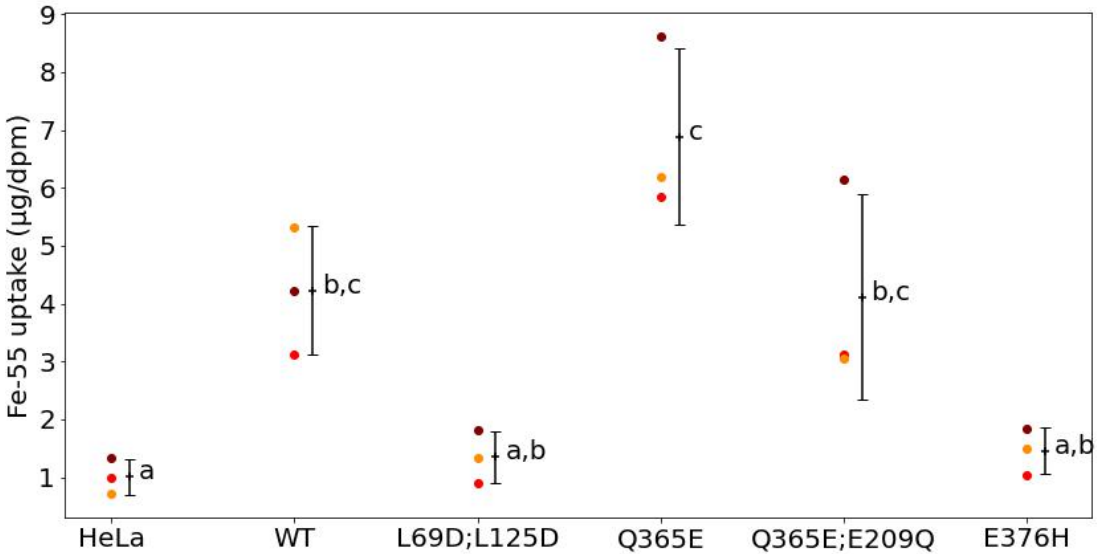


Figure 5.17: **Iron uptake by monoclonal cell lines expressing designed mutants of hZIP14_A-HA.** HeLa and HeLa cell lines expressing hZIP14_A-HA, wild-type and designed mutants, were incubated in 10 µM $FeSO_4$ labelled with ^{55}Fe. ^{55}Fe uptake was measured by liquid scintillation counting, and protein concentration by BCA assay. Data points represent averages of triplicates, coloured points represent three separate experiments. Error bars represent the standard deviation; different letters represent means that are significantly different from each other (one-way ANOVA with Tukey-Kramer post-hoc analysis, $P < 0.05$).

Surface Levels of Designed Mutants In the Monoclonal Cell Lines Clarify Effects on Iron Uptake

Flow cytometry was used to assess the levels of the designed hZIP14_A-HA mutants at the cell surface (monoclonal stable cell lines), in order to determine if this could explain differences in iron uptake between cell lines (figure 5.18). In the L69D;L125D stable cell line, the L69D;L125D mutant was not detectable at the cell surface, explaining the lack of iron uptake by this mutant.

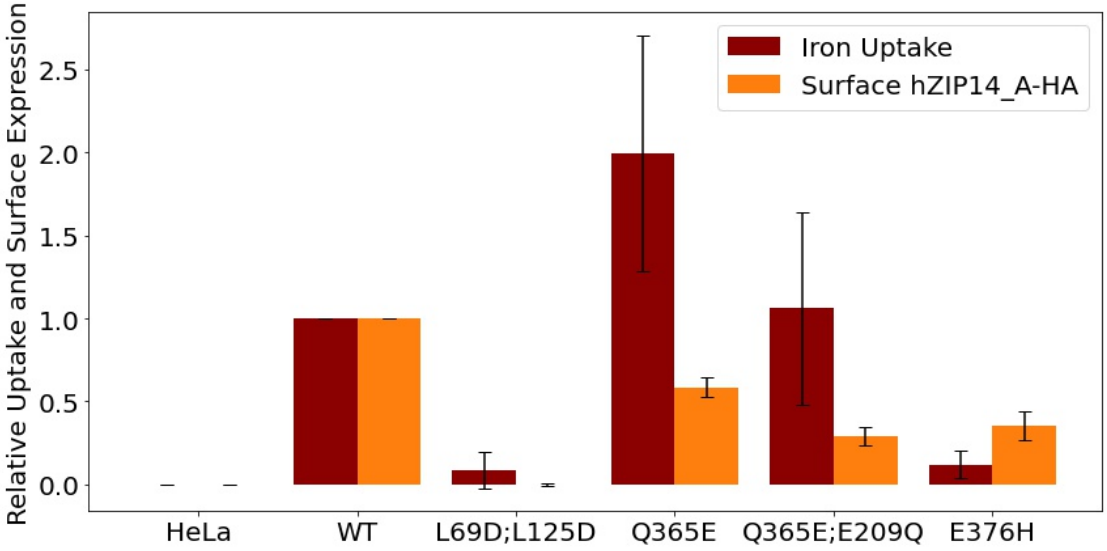


Figure 5.18: **Relative iron uptake compared to relative surface levels of designed hZIP14_A-HA mutants.** Relative iron uptake and surface levels of hZIP14_A-HA variants normalized relative to HeLa at 0 and WT at 1, for each experiment.

The E376H mutant, which similarly failed to increase iron uptake by HeLa cells, was expressed at the cell surface. The levels of E376H at the cell surface were low relative to levels of wild-type hZIP14_A-HA in the wild-type cell line, but iron uptake was even lower, indicating that E376 is an important residue for the ability of hZIP14 to transport iron.

Surface levels of the Q365E hZIP14_A-HA mutant were relatively lower than the levels of wild-type hZIP14_A-HA, and yet iron uptake was much

higher. Therefore the Q365E mutation enhanced iron uptake. Similarly, the Q365E;E209Q double mutation, meant to revert to wild-type behaviour, also enhanced iron uptake function. While the Q365E;E209Q double mutant took up, on average, a similar amount of iron to the wild-type cell line, cell surface levels were approximately a third of the wild-type cell line. In fact, both Q365E and its corresponding double mutant took up $\sim$ 3.5 times more iron than the wild-type hZIP14_A-HA, when accounting for cell surface levels. Therefore, the E209Q mutation did not undo the effect of the Q365E mutation.

Cell surface levels of the designed hZIP14_A-HA mutants showed that the failure of the L69D;L125D mutant to increase iron uptake can be explained by its absence at the cell surface. Taking cell surface levels into account also reveals that the E376H mutation decreases the ability of hZIP14_A-HA to transport iron, and that iron uptake by the Q365E and Q365E;E209Q mutants is, surprisingly, increased by approximately the same amount.

Cell Surface Localisation of Designed hZIP14_A-HA Mutants in Monoclonal Stably-Transfected Cell Lines

For figures 5.10, 5.11 and 5.15, HeLa cells were transiently transfected with hZIP14_A-HA mutants, and differentially stained for interior and surface hZIP14_A-HA. Those results showed that certain mutants were, on average, localised at the cell membrane, in greater (E376H), similar (Q365E and Q365E;E209Q) or lesser (L69D;L125D) proportions than wild-type hZIP14_A-HA.

The results for the transiently transfected cells cannot be directly compared to those shown in figure 5.19, where stably-transfected monoclonal cell lines were stained for surface and total (rather than interior) levels of hZIP14_A-HA mutants. However, there are important differences that should be mentioned.

Although the transiently transfected L69D;L125D mutant localised to the cell surface at lower levels than wild-type hZIP14_A-HA, it was detectable on the surface across the range of expression levels; in the stably-expressing cell lines, none was detected at the cell surface at all (figures 5.10 and 5.19).

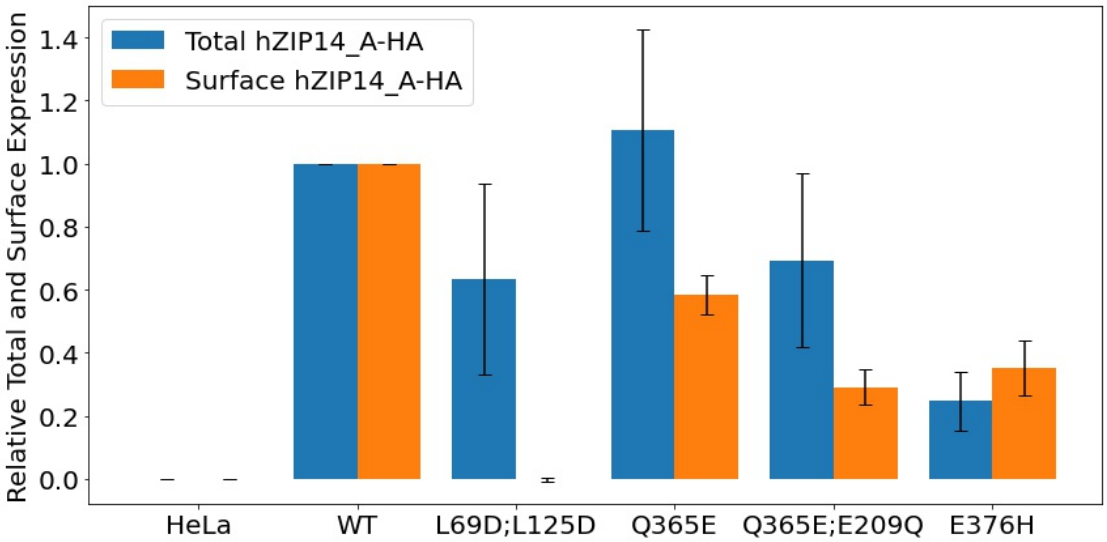


Figure 5.19: **Comparison of relative total and surface levels of designed hZIP14_A-HA mutants.** Surface and total levels of hZIP14_A-HA variants were normalized relative to HeLa at 0 and WT at 1, for each experiment.

Transiently transfected Q365E and Q365E;E209Q mutants localised to the cell surface in similar proportions to wild-type hZIP14_A-HA, and yet in the stable cell lines, cell surface levels were relatively low (figures 5.15 and 5.19).

The behaviour of E376H was more consistent, with surface levels that were slightly higher than wild-type, relative to interior or total expression, in the transiently and stably transfected cells, respectively (figures 5.15 and 5.19).

These results suggest that long-term regulation of hZIP14_A-HA trafficking may be occurring, and that the hZIP14_A-HA mutations alter this regulation.

5.2.5 Co-evolution Analysis of the hZIP14_A-HA Transmembrane Domain Homology Model

Based on the hZIP14 homology model, the A354L and Q365E mutations were predicted to be deleterious, and the double mutations (A354L;L206A and Q365E;E209Q) were predicted to cause a reversion to wild-type behaviour. The results of experiments on these mutants did not support the transmembrane homology model, in that the double mutants did not behave more like wild-type hZIP14_A-HA than the single mutants. A354L had little effect on relative proportions of protein localising at the cell surface, while the A354L;L206A mutant did differ from wild-type in this regard (figure 5.15). The Q365E and Q365E;E209Q mutations were very similar to each other, both in the proportions localising to the cell membrane at different levels of expression (figure 5.15), and in their rate of iron uptake relative to the expression of each mutant at the cell surface (figure 5.18).

An alternative method for assessing the accuracy of a protein structure is to use multiple sequence alignments to identify co-varying amino acid positions; covariance in amino acids is believed to reflect the preservation of key interactions between the side-chains of amino acids throughout evolution. Pairs of co-evolving amino acids, though perhaps distant according to the protein sequence, are likely to have close proximity in the fully folded protein. In order to carry out co-evolution analysis, the hZIP14 sequence was submitted to the PSIPRED DeepMetaPSICOV [271] and the OpenSEQ GREMLIN [272] servers. Each of these servers compiles its own multiple sequence alignment; the latter server was only able to carry out an analysis on the transmembrane region of hZIP14, excluding the majority of the TM3-4 intracellular loop (amino acid residues 258-320), due to greater requirements for variation in the multiple

sequence alignment.

Figure 5.20 is a distance matrix for the hZIP14 homology model; regions not based on the original *Bordetella bronchiseptica* crystal structure are not included, or amino acids less than 4 residues away in the protein sequence. The diagonal stretches of dots represent regions of alpha helices that are positioned next to each other. Figure 5.21 shows the results of co-evolution analysis of hZIP14, from both servers. The patterns are very similar, signifying an agreement on the overall structure of hZIP14.

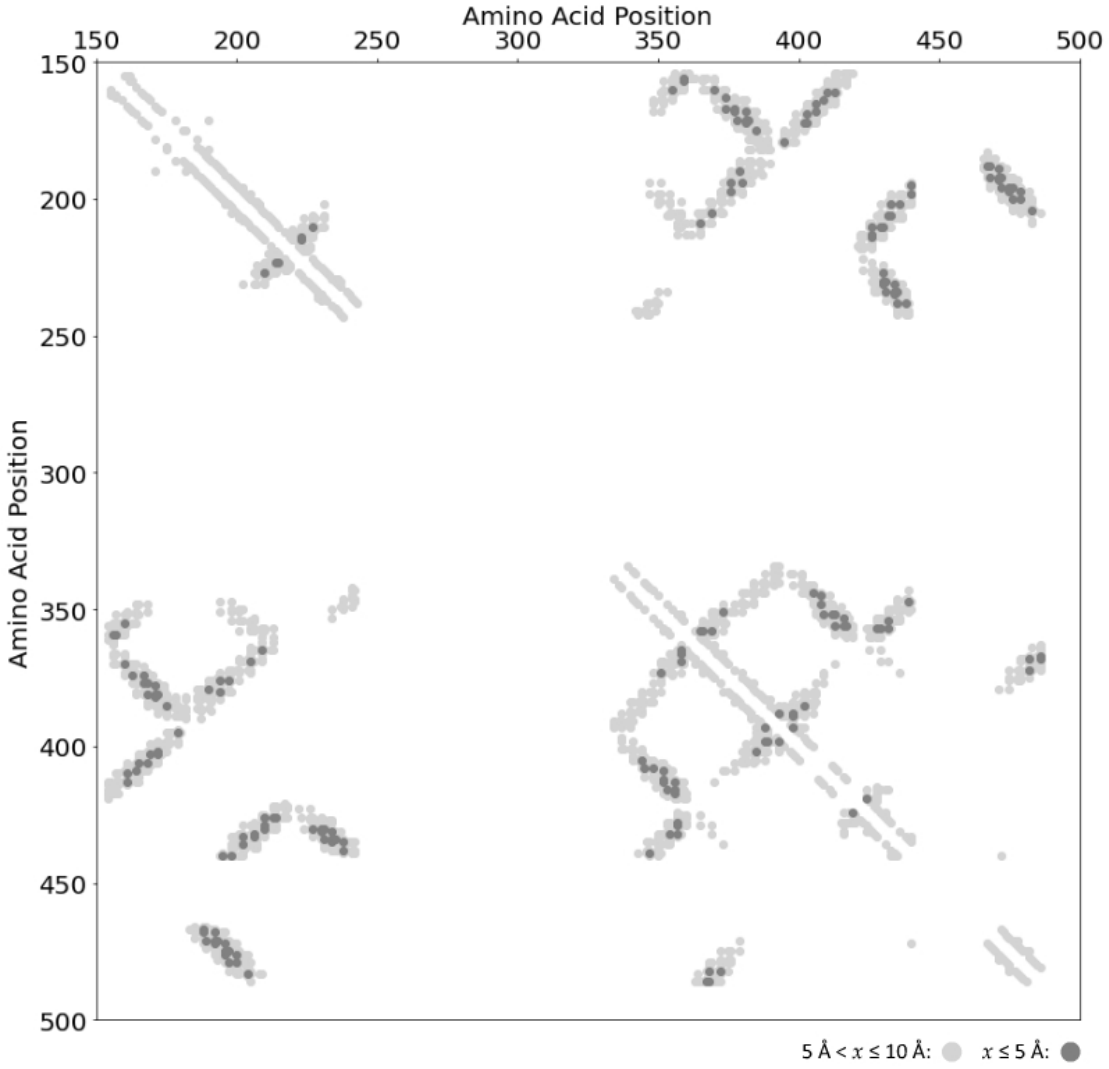


Figure 5.20: **Distance matrix of the hZIP14 homology model.** Points represent amino acid positions with β-carbons (α-carbons for glycine) less than 10 Å apart, in the hZIP14 homology model. Distances for amino acids less than 4 residues apart, and the TM 3-4 and TM 7-8 loops are not included. The distance matrix reflects the three-dimensional arrangement of the amino acids.

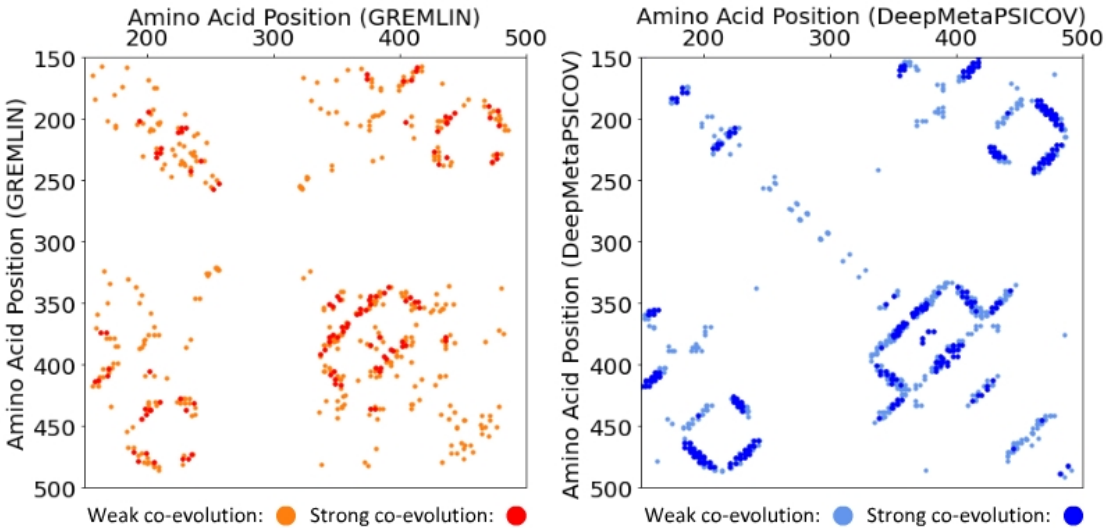


Figure 5.21: **Co-evolution matrices for hZIP14 agree on protein structure.** Co-evolution matrices produced by the GREMLIN and DeepMetaPSICOV web-servers for hZIP14. Co-evolution signals are a predictor of proximity between amino acids in the three-dimensional structure, and the strength of the signal predicts the importance of the amino acid interactions to protein structure and/or function.

5.2.5.1 Agreement Between the Co-evolution Data and the Homology Model

Overlays of the co-evolution data onto the distance matrix of the hZIP14 homology model show the extent to which the co-evolution data agree with the homology model (figure 5.22). The majority of helical interactions in the model are clearly represented in the data from both servers, though the exact amino acid pairs differ. The strongest co-evolution signals cover the interactions between TMs 1, 4, 5 and 6 (all combinations).

As an illustrative example, figure 5.23 shows how the co-evolution predictions map onto the homology model, regarding the interactions of TM 1. The combined analyses of the DeepMetaPSICOV and GREMLIN web-servers predict interactions between the N-terminus of TM 1 and the C-terminus of TM 4, and interactions between the full-length of TM 1 and both TMs 5 and 6. In addition, the predictions support the relative orientations of these helices, as the amino acid pairs predicted to co-evolve typically line the sides of the alpha helices that face each other.

The entire homology model (transmembrane domain) is supported, to some extent, by the co-evolution data, with three exceptions that will be discussed further in section 5.2.5.2. For more detailed information on the interactions of each transmembrane domain helix, see appendix D.4.

The co-evolution analyses provided by the DeepMetaPSICOV and GREMLIN web-servers strongly support the majority of interactions in the homology model, both in terms of the overall organisation of the alpha helices, and their relative orientations.

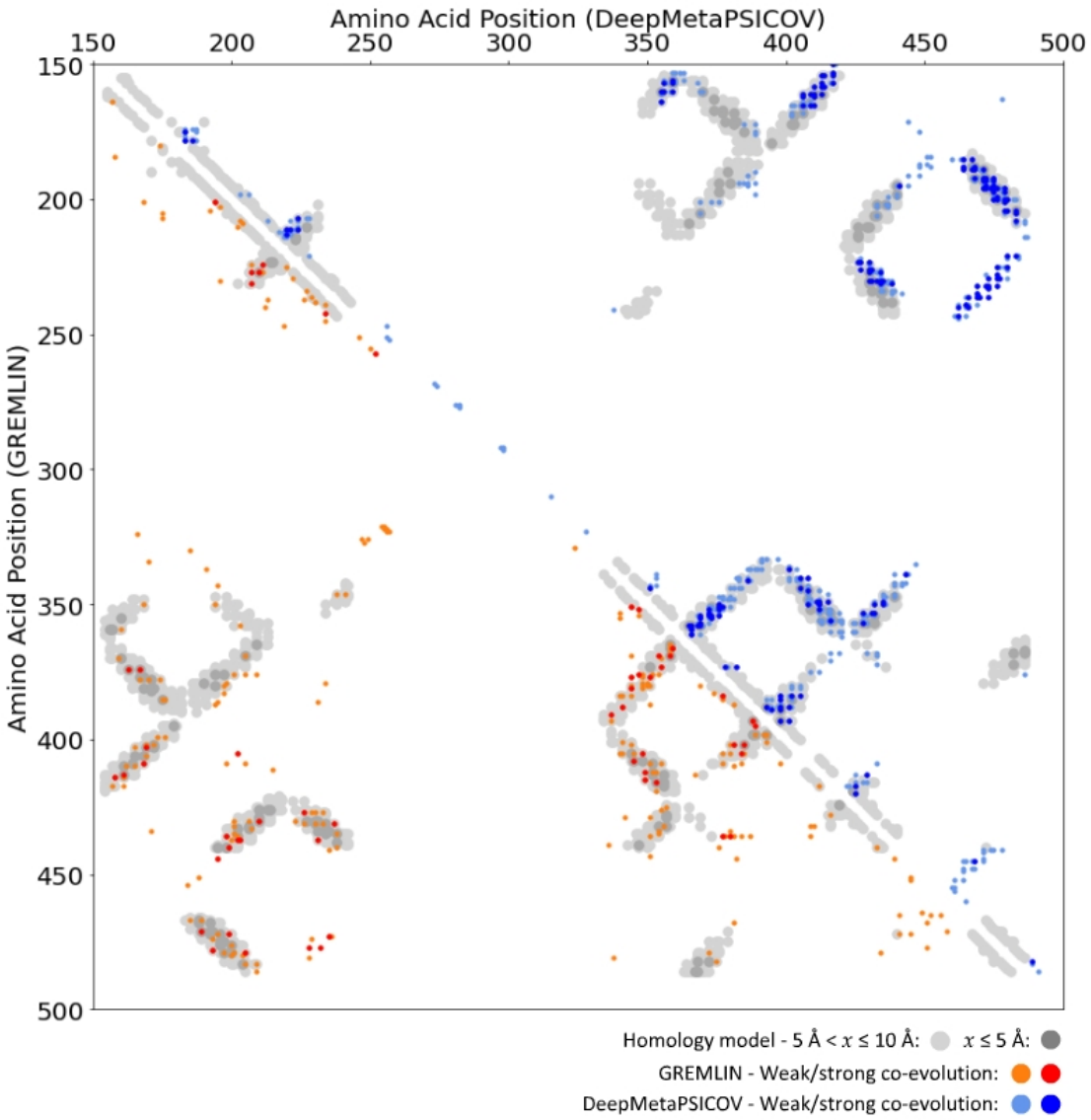


Figure 5.22: **Overlaid distance and co-evolution matrices for hZIP14.** Co-evolution matrices from the GREMLIN (lower-left) and DeepMetaPSICOV (upper-right) analyses of hZIP14 were overlaid onto the hZIP14 homology model distance matrix. The overlap shows areas of agreement between the co-evolution data and the homology model.

5.2.5.2 Disagreement Between the Co-evolution Data and the Homology Model

Despite the general agreement between the co-evolution data and the hZIP14 homology model, there are three predictions from the co-evolution data that are

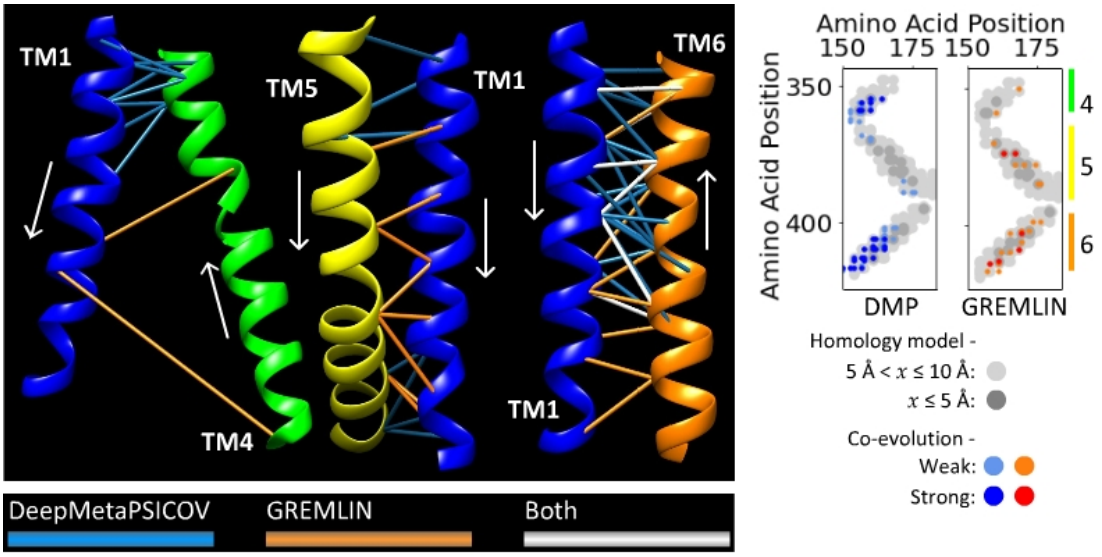


Figure 5.23: **Co-evolution data support interactions of TM 1.** Left panel: Segments of the hZIP14 homology model; lines join the α-carbons of amino acid pairs predicted to co-evolve according to the DeepMetaPSICOV analysis, the GREMLIN analysis, or both (blue lines, orange lines, or white lines, respectively - see key in lower left panel). White arrows show the direction from the N-terminus to the C-terminus of each TM helix. Right panel: Overlay of co-evolution data (from the DeepMetaPSICOV (DMP) or GREMLIN servers) and the hZIP14 homology model distance matrix for interactions between TM 1 and TMs 4, 5 and 6.

not found in the homology model. Both co-evolution analyses of hZIP14 predict a series of interactions between amino acids after the C-terminal end of TM 7 (according to the homology model) and other TM helices. In addition, some predictions from the co-evolution data suggest that certain helices are vertically shifted relative to others. Finally, there is strong support for interactions between TMs 3 and 8, in both the DeepMetaPSICOV and GREMLIN analyses, which are not found in the homology model.

Co-evolution Data Suggests a C-terminal Extension of the TM 7 α-Helix

The co-evolution data support the homology model, regarding interactions between TM 7 and TMs 2, 3, 4, 5, 6 and 8 (see residues 422 to 441 in figure

5.22, or appendix D.4 for more detail). However, the predicted diagonal pattern of interactions extends beyond the point at which TM 7 ends in the homology model, for interactions with TMs 2, 4 and 8 (figure 5.24).

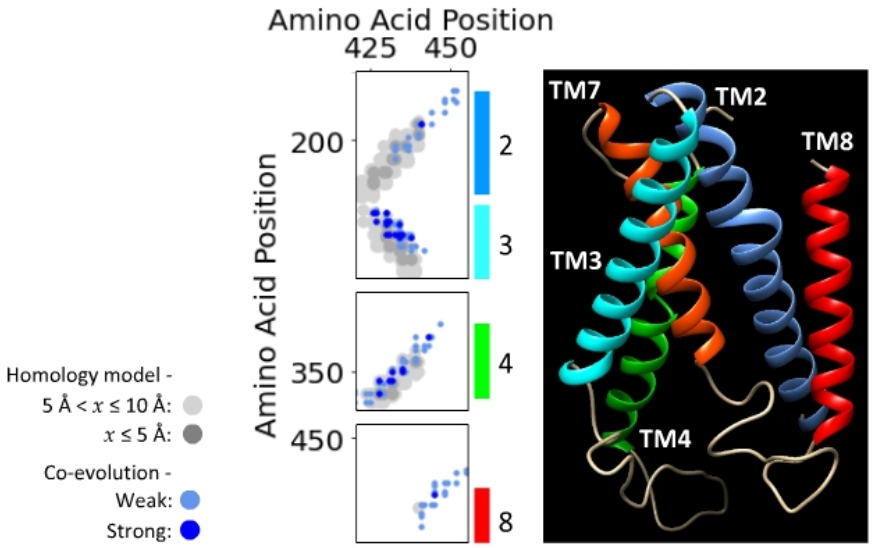


Figure 5.24: **Co-evolution data predict an extension of the TM 7 helix at the C-terminal end.** Left panel: Overlay of the DeepMetaPSICOV co-evolution data and the hZIP14 homology model distance matrix for interactions between TM 7 and TMs 2, 3, 4 and 8. Right panel: TMs 2, 3, 4, 7 and 8, and intracellular TM 3-4 and 7-8 loops, of the hZIP14 homology model.

According to the homology model, the N-terminus of TM 7 does not make contact with TM 8; TMs 7 and 8 are not close, and both TMs 2 and 3 come in between them (figure 5.24, right panel) - this does not contradict the co-evolution data. The C-terminal end of TM 7 protrudes between TMs 2 and 3, getting progressively closer to TM 8, but ending before the two helices can make contact. The co-evolution data suggest that TM 7 continues beyond this point, extending its contacts along the length of the N-terminal end of TM 2, slightly extending contact towards the C-terminal end of TM 4, and making contact with the N-terminal portion of TM 8 (figure 5.24, left panel). An extension of contact with TM 3 is not predicted by the co-evolution data. Figure 5.24 (right panel) shows that if TM 7 continued at the C-terminal end, it would likely project

towards TMs 2 and 8, and away from TM 3.

The co-evolution data show strong support for a longer TM 7 helix, and predict interactions with other helices that are consistent with the overall structure of the homology model.

Co-evolution Data Suggest Conformational Changes in hZIP14

As a carrier, hZIP14 is expected to undergo conformational changes during transport of its solutes. Co-evolution analyses are known to reflect conformational changes by identifying amino acid interactions that are essential in alternative protein conformations [273, 274].

A comparison of the interactions predicted between various TM helices in hZIP14 shows that the strengths of the associations vary (both by numbers of predicted pairs and by intensities of the co-evolution signals), with the strongest associations divided between two groups of helices, with one group composed of TMs 1, 4, 5 and 6, and the other composed of TMs 2, 3, 7 and 8. Figure 5.25 A shows the predicted interactions for TMs 2, 3, 7 and 8 (excluding the TM 3-8 predictions, discussed below); the co-evolution data supports the homology model, predicting multiple pairs of co-evolving amino acids that are consistent with the arrangements and orientations of these helices. Likewise, figure 5.25 B shows that the co-evolution data strongly support the interactions of TMs 1, 4, 5 and 6 in the homology model.

Interfacing these two sections are TMs 2 and 7 on one side, and TMs 4 and 5 on the other (figure 5.25 C). The co-evolution data predict multiple pairs of co-evolving amino acids between TMs 4 and 7, TMs 5 and 7, and TMs 2 and 5, shown in figures 5.26 A, B and C, respectively. Some of these interactions are in agreement with the homology model - the β-carbons are less than 10 Å

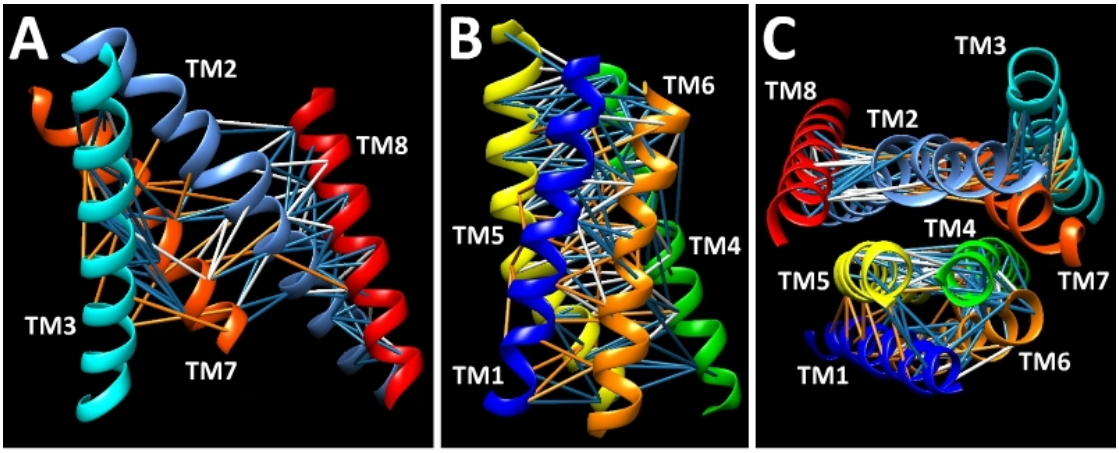


Figure 5.25: **Co-evolution data support interactions that divide hZIP14 into two sections.** Segments of the hZIP14 homology model; lines join the α-carbons of amino acid pairs predicted to co-evolve according to the DeepMetaPSICOV analysis, the GREMLIN analysis, or both (blue lines, orange lines, or white lines, respectively). A: TMs 3, 7, 2 and 8 of the hZIP14 homology model. B: TMs 5, 1, 6 and 4 of the hZIP14 homology model. C: All TM helices of the hZIP14 homology model, extracellular view.

apart, especially for the TM 4-7 interactions. However, many of the co-evolution predictions suggest that amino acids on TMs 4 and 5, have co-evolved with residues on TMs 2 and 7 that are slightly towards the extracellular end of the protein, though the helices are oriented correctly.

Taken together, the data suggest that if the section containing TMs 1, 4, 5 and 6 were rotated, with the interaction at the centre of TM 4 and the C-terminal end of TM 7 (according to the homology model, see above) acting as an axis, such that TM 5 was brought closer to the extracellular end of the protein, this conformation would be supported by the co-evolution data (figure 5.26). Such a conformation would be consistent with the interactions between TM 4 and a C-terminal extended TM 7, that are predicted by the co-evolution data (figure 5.24).

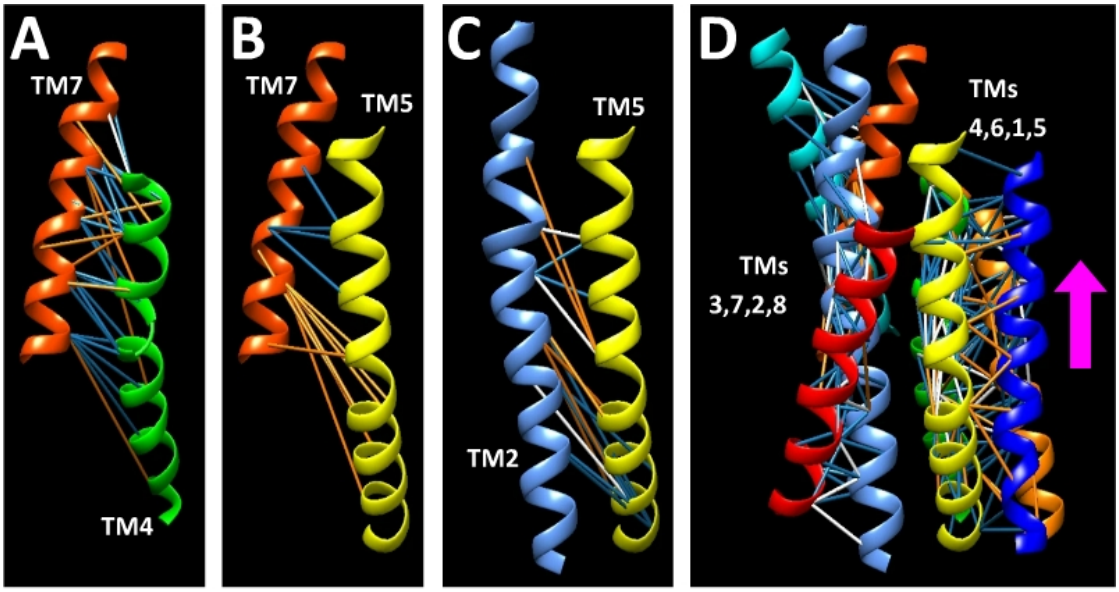


Figure 5.26: **Co-evolution data suggest movement of the central helices.** Segments of the hZIP14 homology model (identical orientations); lines join the α-carbons of amino acid pairs predicted to co-evolve according to the DeepMetaPSICOV analysis, the GREMLIN analysis, or both (blue lines, orange lines, or white lines, respectively). A: Co-evolution predicted interactions between TMs 7 and 4. B: Co-evolution predicted interactions between TMs 7 and 5. C: Co-evolution predicted interactions between TMs 2 and 5. D: Co-evolution predicted interactions within sections of hZIP14; pink arrow indicates movement of the right section relative to the left.

Co-evolution Data Predict Dimerisation of hZIP14

Both the GREMLIN and DeepMetaPSICOV servers predict a series of interactions across the lengths of TMs 3 and 8, suggesting that these two helices are in close proximity in the three-dimensional structure; this arrangement contradicts the hZIP14 homology model (figure 5.27).

The amino acids involved in the predicted interactions predominantly project outwards from the hZIP14 structure (figure 5.28), which is compatible with a dimeric arrangement in which each TM 3 interacts with the TM 8 of the other monomer. A hypothetical arrangement is shown in figure 5.29 that satisfies the interactions predicted by the co-evolution data between TMs 3 and 8.

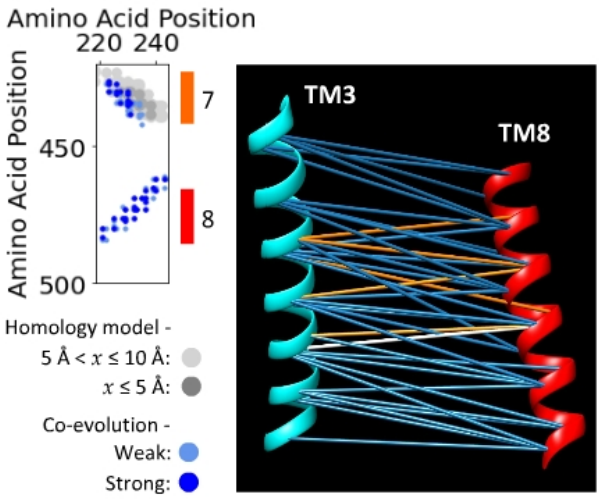


Figure 5.27: **Co-evolution data predict numerous interactions between TMs 3 and 8.** Left panel: Overlay of the DeepMetaPSICOV co-evolution data and the hZIP14 homology model distance matrix for interactions between TM 3 and TMs 7 and 8. Right panel: TMs 3 and 8 of the hZIP14 homology model; lines join the α-carbons of amino acid pairs predicted to co-evolve according to the DeepMetaPSICOV analysis, the GREMLIN analysis, or both (blue lines, orange lines, or white lines, respectively). The co-evolution data appear to contradict the homology model by suggesting that a number of TM 3 amino acids co-evolve with TM 8 amino acids.

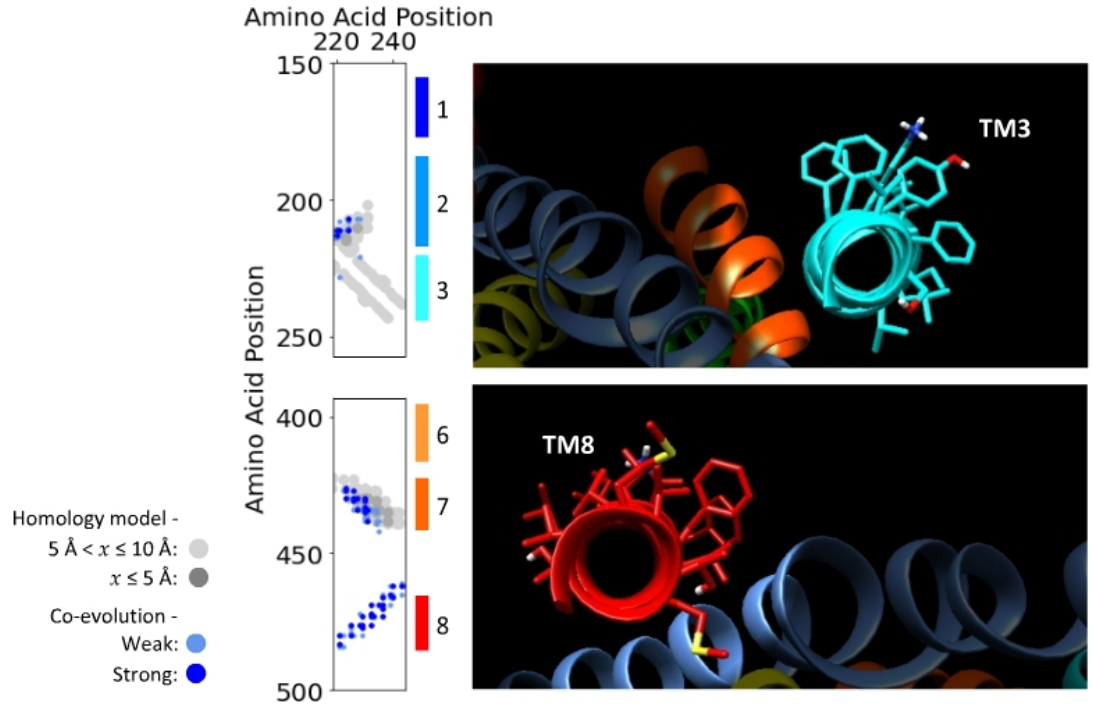


Figure 5.28: **TM 3 and 8 amino acids predicted to interact project outwards.** Left panel: Overlay of the DeepMetaPSICOV co-evolution data and the hZIP14 homology model distance matrix for interactions between TM 3 and TMs 2, 7 and 8. Upper right panel: Amino acid R-groups of the TM 3 residues predicted to interact with TM 8. Lower right panel: Amino acid R-groups of the TM 8 residues predicted to interact with TM 3. The co-evolution data predict which faces of TMs 3 and 8 interact with each other.

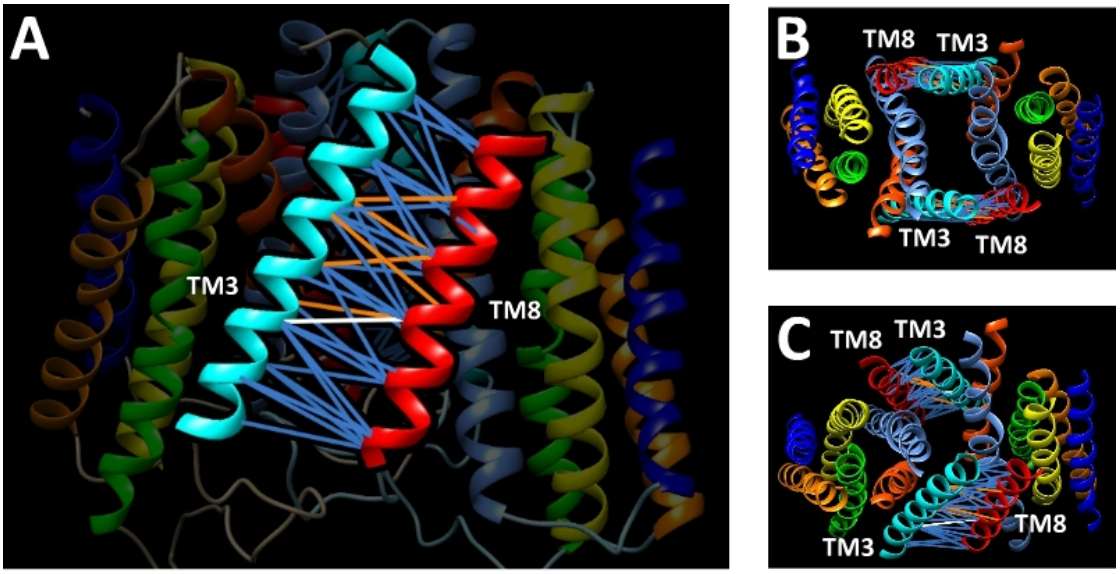


Figure 5.29: **Possible dimeric arrangement of the hZIP14 TM regions supported by the co-evolution analyses.** hZIP14 homology models arranged as a dimer; lines join the α-carbons of amino acid pairs predicted to co-evolve according to the DeepMetaPSICOV analysis, the GREMLIN analysis, or both (blue lines, orange lines, or white lines, respectively). A: Dimeric arrangement of hZIP14 homology models, with TMs 3 and 8 highlighted, and predicted pairs of interacting amino acids. B and C: Dimeric arrangement of hZIP14 homology models viewed from different angles, with predicted interactions between TMs 3 and 8. The proposed dimeric arrangement is consistent with the TM 3-8 interactions predicted by the co-evolution analyses.

5.2.5.3 Co-evolution Analysis Supports the Structure of the hZIP14 Homology Model

The co-evolution analyses produced by the GREMLIN and DeepMetaPSICOV servers differ in regards to exact amino acid interactions, but agree in terms of overall structure, predicting the same interactions between transmembrane helices, with the same orientations; the DeepMetaPSICOV predictions tend to be more detailed, and show a clearer pattern of interactions. The co-evolution analyses support the majority of the homology model structure, additionally suggesting an alternative protein conformation, and the possibility of dimerisation mediated by TMs 3 and 8. The co-evolution data indicate that the transmembrane region of the homology model may be accurate enough to predict the mechanisms underpinning the effects of the disease-causing mutations in this domain.

5.3 Discussion

In this chapter, mutations based on hZIP14 homology models were introduced into plasmids encoding wild-type hZIP14-HA, and the effects of these mutations on cellular localisation and iron uptake were tested. Mutations were designed to affect dimerisation of the extracellular domain, to alter solute selectivity, or to test the accuracy of the homology models.

Mutations in the extracellular domain altered cellular localisation in an expression-dependent manner, supporting the hypothesis that dimerisation of the extracellular domain affects the proportions of hZIP14 at the cell surface.

The TMD mutations intended to alter hZIP14 function had mild effects on cell surface localisation, which were not reversed by the paired mutations

predicted to compensate for their effects. The E376H mutation decreased iron uptake function, as predicted, while in the paired mutations tested (Q365E and Q365E;E209Q), both the single and double mutants surprisingly increased iron uptake by similar amounts.

Unlike the designed TMD mutations, co-evolution data supported the majority of the homology model, also indicating an alternative conformation, and dimerisation of the transmembrane domain.

5.3.1 Designed ECD mutations Had a Range of Impacts on Cellular Localisation

Based on the homology model of the extracellular domain, mutations were designed with the intention of decreasing the stability of the ECD (L62A), inhibiting dimerisation of the ECD (L69D and L69D;L125D) and preventing the formation of the intermolecular disulphide bridge predicted to link the ECD monomers. Deletion mutants (△40-130 (A & B)) lacking the N-terminal extracellular domain were also created, to test the effect of its absence on hZIP14 cellular localisation (see tables 5.1 and 5.2 for the rationale behind the designed mutations, and a summary of their locations in the homology models).

The designed ECD point mutations altered cellular localisation in a similar manner as has been previously seen for the disease-causing N-terminal mutations - the mutants were retained in the endoplasmic reticulum, particularly at low expression levels, and there was a sigmoidal relationship between internal and surface hZIP14 (figures 3.13, 3.36, 5.8, 5.10 and 5.11).

The L69D, L69D;L125D and C118D mutations were all intended to inhibit dimerisation of the ECD without affecting its structure, with L69D expected to have a milder impact than L69D;L125D, as only one hydrophobic residue had

been replaced with a negatively charged aspartate (section 5.1.1.1). Both the flow cytometry data and western blotting results suggest that these predictions were not entirely correct. The pattern of bands for L69D is the same as that of L62A (the mutant meant to destabilise ECD structure) (figure 5.6), and according to the flow cytometry results, the proportion of cells expressing high levels of L69D hZIP14_A-HA was decreased relative to L69D;L125D, again making it more similar to the L62A mutant, where only a small proportion of transfected cells were highly expressing (figures 5.10 and 5.11). According to the ECD homology model, L69 makes a number of intramolecular contacts; this was not considered when selecting the mutations for the predicted dimerisation interface, and could explain the the destabilising effect of this mutant. It is unclear, however, why additionally mutating L125D should decrease the impact of the L69D mutation. One explanation is that the homology model is not accurate enough to make predictions about the relative impacts of particular mutations.

The effects of the C118D mutation were milder than those of the L62A, L69D and L69D;L125D mutations - the sigmoidal curve relating internal and surface levels of the mutant was relatively shallow, and in western blotting, one of the two higher molecular mass bands was clearly visible (figures 5.10 and 5.11). Although C118 is part of the conserved CPALLY motif found in most LIV-1 proteins [11, 219], the cysteine is not invariant - as an example, the closely related ZIP4 has a serine residue in this position. The results for the C118D mutant suggest that the intermolecular disulphide link affects hZIP14 trafficking and enhances cell surface localisation, but is not *necessary* for cell surface localisation.

This conclusion is demonstrated even more effectively by the $\triangle$40-130 mutants. Counter to expectations, the ECD deletion mutants differed from both the

designed and disease-causing mutants, and were more similar to wild-type hZIP14_A. Under immunofluorescence microscopy, the △40-130 mutants did not have the reticular pattern of expression observed with the other ECD mutants. At low expression levels they surrounded, and aggregated next to the nucleus (figure 5.8). When examined by flow cytometry after differential staining, the △40-130 mutants did not have a sigmoidal relationship between internal and surface levels of their proteins, rather, as expression levels increased, there was an abrupt change in the partitioning of mutant hZIP14 between the inside and the surface of the cells (figure 5.11). On comparing the △40-130 mutants to the mutations that were meant to merely decrease ECD dimerisation, we can see that in terms of cell surface localisation, a faulty ECD is worse than no ECD at all.

On the basis of these findings, I hypothesize that improperly folded ECDs prevent exit from the ER. All of the ECD mutations that destabilise the ECD cause the mutants to accumulate in the endoplasmic reticulum (ER). This was demonstrated in Chapter 3 by strong co-localisation of the N-terminal disease-causing mutants with the ER protein, calreticulin, and L62A showed the same pattern of expression (figures 3.16, 3.17, 3.19, 3.20 and 3.22). Multiple mutations in the extracellular domain of the LIV-1 protein ZIP4 have also been recently shown to alter folding and lead to the protein being retained in the ER [275]. Furthermore, this hypothesis is consistent with the well-established role of the ER in quality control, which is essential for preventing improperly folded proteins from accumulating and harming the cell [276].

We also hypothesize that dimerisation of the ECD, when present, is required for exit from the ER, based on the expression-dependent effects on cell-surface localisation seen for the various ECD mutations that inhibit its dimerisation without, I believe, significantly altering its structure - hZIP14 cannot reach

the cell surface, without first leaving the ER. A need for oligomerisation to occur before exit from the ER has been noted for other transporters, with oligomerisation being proposed to act as a marker for correct folding [277]. Among the markers of mis-folding detected by ER proteins are exposed hydrophobic patches and exposed unpaired cysteine residues, both of which are found in the dimerisation site of the ECD (figure 5.2). Dimerisation may also help proteins to fold correctly - for those mutants in which the ECD was destabilised (N-terminal disease-causing mutations and L62A), cell surface localisation reached that of wild-type hZIP14 at high levels of expression, suggesting that at higher concentrations, dimerisation of the mutants allowed them to pass ER quality control mechanisms (figures 3.36 and 5.11).

Although it appears that hZIP14 exit from the ER is prevented when the ECD misfolds or fails to dimerise, regulation of exit from the ER is unlikely to be the purpose of the ECD. For wild-type hZIP14, the relationship between expression and cell surface levels is hyperbolic (figure 3.36). This type of relationship does not need a concentration-dependent checkpoint, and in any case implies that dimerisation is occurring effectively across the range of expression levels observed. The decreased cell surface levels of the $\triangle$40-130 mutant at low levels of expression (figure 5.11) strongly suggest that the actual purpose of the ECD is to enhance interactions between the transmembrane domains, which appear to have an additional effect on cell surface localisation (section 5.3.2). The physiologically relevant levels of hZIP14 expression may include those at which cell surface expression of the $\triangle$40-130 mutant was decreased, in which case the association of the ECD monomers may be essential for normal trafficking of hZIP14.

The finding of increased cell surface levels for ZIP14 ECD deletion mutants (at lower levels of expression) is in agreement with results for the closely related

protein ZIP4, where deletion of the ECD increased cell surface levels [219]. In a study by Kambe and Andrews [278], it was found that in the polarised cell lines studied, zinc deficiency led to cleavage of the ZIP4 ECD, increased levels of ZIP4 at the cell surface and increased uptake of zinc. These results suggest that cleavage of the ZIP4 ECD is used to enable this important transporter of dietary zinc to increase zinc uptake in response to zinc insufficiency. On the other hand, in wild-type ZIP4, in which there is no disulphide bridge linking the ECD monomers, a mutation introducing this feature was reported to have no impact on cell surface levels or zinc uptake function [219]. It is possible that structural features shared by ZIP proteins may serve different regulatory purposes, allowing them to respond appropriately to specific situations.

5.3.2 Dimerisation of the Transmembrane Domain Is Required For Cell Surface Localisation

The co-evolution analyses predict interactions between TMs 3 and 8, an arrangement which is compatible with my homology model, if two TMD domains dimerise via these TM helices. Several members of the ZIP family have been suggested to form homodimers, and ZIPs 6 and 10 form a heterodimer [219–222]. This includes ZIPs 7 and 13, whose N-termini are not homologous with the ZIP4 ECD and are predicted to be highly disordered, therefore these ZIPs may have dimerisation sites in their transmembrane domains [219, 266, 279–281]. In an attempt to use co-evolution data to model the transmembrane domain of the LIV-1 protein ZIP4, Antala et al. [282] found that no single structure could satisfy all of the predicted contacts, unless the transmembrane domain formed a dimer - they were able to predict the arrangement of helices in the dimer, though there were several options for how those helices could segregate

between monomers. Dimerisation of the ZIP4 TMD was later demonstrated by chemical cross-linking [283]. In a recently solved crystal structure of *Bordetella bronchseptica* ZIP (BbZIP - the template for the hZIP14 TMD homology model), BbZIP was captured in the form of a dimer, with the TM 3 of each monomer interacting with the TM 8 of the other [267]. In addition, my experimental data are consistent with TMs 3, 7, 2 and 8 forming a dimerisation interface.

Mutations predicted to inhibit dimerisation of the ECD led to sigmoidal relationships between internal and surface hZIP14, and the L441R and N469K mutations, which lie within the predicted dimerisation site of the TMD, also resulted in sigmoidal relationships (figures 3.36, 3.37, 5.10 and 5.11). Aside from the similar effects of these mutations, the sigmoidal curve, in and of itself indicates co-operativity, that there are at least two ways in which increasing the concentration of these hZIP14 mutants increases cell surface localisation, and that they work together. One possible mechanism of co-operativity is that dimerisation of the ECD increases dimerisation of the TMD, and vice versa. In contrast, the flow cytometry results for the $\triangle$40-130 mutants showed an expression-dependant effect on cell surface localisation, but it was not a sigmoidal curve - the loss of the ECD meant the loss of co-operativity between the two dimerisation domains (figure 5.11). Instead, there was an abrupt increase in the proportion of hZIP14 at the cell surface above a threshold concentration, likely reflecting the affinity of the TMD dimerisation site.

Of the two domains, dimerisation of the TMD is likely to be the deciding factor on cell surface localisation, as the L441R and N469K mutations decreased cell surface localisation to a greater extent than the ECD mutations, with the majority of transfected cells having no hZIP14 detectable at the cell surface (figure 3.37). Even at high expression levels, cell surface expression of these mutants did not reach that of wild-type. In addition, the flow cytometry results for the $\triangle$40-130

hZIP14_A-HA mutant resembled wild type, at moderate and high concentrations, indicating that dimerisation of the TMD is sufficient for cell surface localisation (figure 5.11). The large TM 3-4 cytosolic loop, for which no tertiary structure has been predicted, may be important in this regard. This intracellular loop appears to be lacking any well-defined secondary structure, though there are regions that show moderate degrees of conservation (residues 268-287 and 319-333) [205, 266]. The histidine-rich sequence (residues 251-257), and the putative endocytosis motif (residues 330-333) are also worth considering [264]. Dimerisation of the TMD may bring key elements of the TM 3-4 loops together; the L441R and N469K mutations could prevent this by inhibiting dimerisation, or by preventing optimal alignment of the TMD domains relative to each other. Whether TMD dimerisation increases recycling or trafficking to the membrane, or decreases endocytosis, remains to be determined.

5.3.3 hZIP14 May Form Higher-Order Oligomers

The C118D hZIP14_A-HA mutant displayed two characteristics that differentiated it from the disease-causing N-terminal mutants, aside from its comparatively mild impact on cell surface localisation. Firstly, the C118D hZIP14_A-HA transfectants were predominantly highly-expressing, while the other N-terminal mutant transfectants expressed low levels of their respective mutants (figures 3.36 and 5.10). Secondly, western blotting of C118D hZIP14_A-HA showed an absence of the lower of the two high molecular mass bands, while the upper band remained (figure 5.6). This was not seen in any of the disease-causing N-terminal mutants, where the lower of the two bands was typically faint or absent, and in wild-type and all the disease-causing mutants, the upper band was always fainter than the lower band (figures 3.6 and 3.7).

Higher molecular mass bands for ZIP14 have been observed by other researchers, who suggested they may be oligomeric forms of ZIP14 [13, 35]. On the basis of the experimental evidence, I propose that 1) the highest molecular mass band represents an hZIP14 D2 tetramer (a dimer of dimers), held together by dimerisation of both the ECD and TMD domains with separate proteins, as depicted in figure 5.30 B, and 2) the lower of these two molecular mass bands represents incompletely dissociated tetramers, resulting in hZIP14 dimers, linked by the hydrophobic interactions and disulphide bridges connecting the ECDs of each monomer.

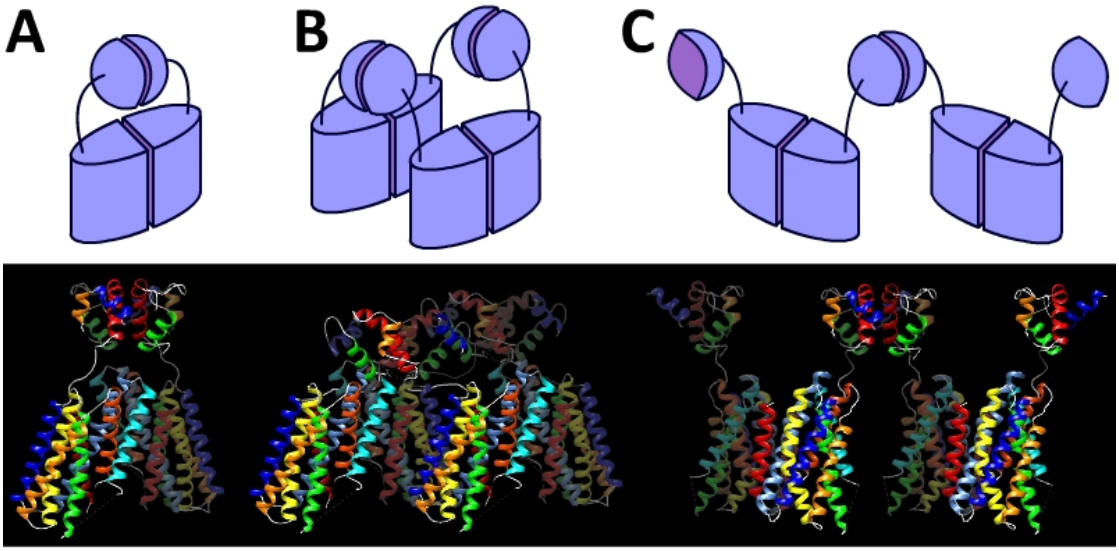


Figure 5.30: **Possible quaternary structures for hZIP14.** Upper diagrams represent possible oligomeric arrangements for hZIP14 domains, lower diagrams are the corresponding predicted ribbon structures. A: A closed C1 dimer. B: A closed D2 tetramer. C: An open tetramer - could form an extended multimer.

The rationale for this proposal is illustrated in figure 5.31, using F98V as an example of an N-terminal point mutation. In wild-type hZIP14_A-HA transfectants, many cells expressed high levels of hZIP14_A-HA, and both dimerisation interfaces would have been correctly folded; high concentrations and correctly folded proteins would lead to the majority of hZIP14_A-HA forming

tetramers. During sample preparation for western blotting, the use of SDS and beta-mercaptoethanol would denature or partially dissociate the majority of tetramers, but due to their high concentrations and the decreased effectiveness of SDS at denaturing hydrophobic membrane proteins [214], some would escape denaturation and remain as tetramers. Partial dissociation would preferentially disrupt dimerisation of the TMD domains, as these rely only on non-covalent interactions, resulting in dimers that remain linked by the disulphide bridge connecting the ECD domains. In this schema, the resultant protein sample contains mostly fully dissociated and denatured hZIP14_A-HA, a moderate amount of hZIP14_A-HA dimers, and a smaller amount of hZIP14_A-HA tetramers (figure 5.31).

In cells transfected with N-terminal disease-causing mutants, the instability of the extracellular domain would cause the majority of transfectants to retain their mutants in the ER, where they are targeted for degradation, with the result being that most of the cells have low levels of expression. In the small proportion of transfectants with expression levels above a certain threshold, dimerisation of the ECD would stabilise its folding and the monomers would be linked by a disulphide bridge. These proteins could form tetramers - but the majority of the hZIP14_A-HA N-terminal mutants in the protein sample would be in the ER, as monomers, or possibly dimers interacting through the TMD domains. During sample preparation, these few tetramers would be almost completely dissociated to undetectable levels, with the small amount of dimers (linked by the disulphide bridge between ECD monomers) reflecting decreased proportions of tetramer in the original cell sample (figure 5.31).

In cells expressing the C118D mutant, dimerisation of the ECD is relatively transient without a disulphide bridge to permanently link the domains, but nonetheless occurs, and C118D hZIP14_A-HA exit from the ER is only

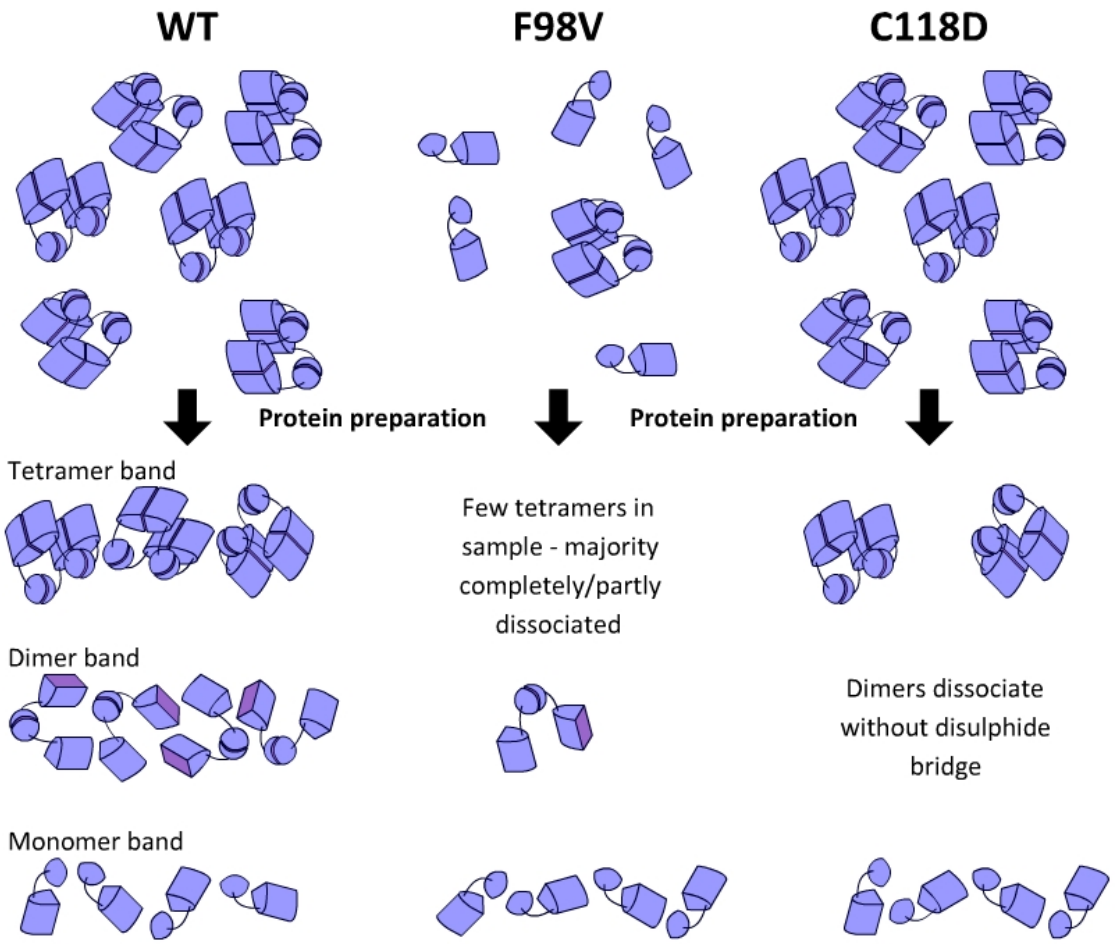


Figure 5.31: **Oligomers and dissociation in western blotting.** Diagram representing possible oligomers present in WT, F98V and C118D protein samples, and their dissociation during sample preparation; C118D monomers are connected only by non-covalent interactions.

moderately impaired. The C118D mutants would be able to form tetramers, similar to those of wild-type, but without two disulphide bridges connecting the four ECDs. During sample preparation, as with wild-type, the majority of tetramers would dissociate, leaving some unaffected. Partial dissociation of tetramers, without the covalent link connecting the ECD domains, would rapidly become total dissociation, resulting in the loss of the dimer band, while the tetramer band would remain due to the increased avidity of the monomers resulting from the multiple binding sites (figure 5.31). While this hypothesis draws together all the experimental results collected on the various

hZIP14 mutants studied in this thesis, it could be best tested by studying the oligomeric status of hZIP14 directly, rather than by western blotting or using the relationships between internal and cell-surface protein as a proxy for dimerisation.

5.3.3.1 The Potential of hZIP14 to Form Oligomers

Oligomerisation is relatively common for proteins; estimates suggest that more than half of proteins form oligomers, with homodimers being the most common, followed by homotetramers [284, 285].

The existence of neither hZIP14 dimers or oligomers has been unambiguously confirmed, though its homology to other ZIP proteins whose dimerisation has been experimentally confirmed strongly supports the dimerisation of both the TMD and ECD domains [219, 267]. No members of the ZIP family are reported to be tetramers, how likely is hZIP14 to form a tetramer? The well-known requirement for the formation of a tetramer is two distinct binding sites, and with this minimal requirement proteins are capable of forming many types of oligomers, constrained only by the positions of those binding sites, and the relative orientations that can be assumed by the monomers [285, 286]. Although the inability of membrane proteins to rotate within the plane of the bilayer limits the oligomeric arrangements they can assume, the possession of more than one binding site still opens up a range of possibilities for homomerisation. Rhodopsin, the G-protein coupled receptor that senses light in rod photoreceptor cells, has two dimerisation sites that allow it to form rows of dimers [287, 288]. Some members of the SLC6 family, which transport neurotransmitters, have multiple dimerisation sites, and rather than forming well-defined oligomers with fixed numbers of subunits, exist in the membrane as a cluster of various oligomers, ranging from dimers to pentamers, in addition to

monomers [289].

The two dimerisation sites of hZIP14 are connected by a domain linker, therefore the main constraints on its oligomerisation (apart from its inability to rotate within the plane of the membrane bilayer) are the length and flexibility of the domain linker. According to George and Heringa [261], the average length of domain linkers is 10.0 $\pm$ 5.8 residues. On sorting linkers into small, medium and large, large linkers had an average length of 21.0 $\pm$ 7.6 residues; the hZIP14 linker is 28 residues long. Domain linkers may contain multiple small, polar or charged amino acids, as well as proline and/or glycine [262]. The proportions and distributions of these residues combine to determine the characteristics of the linker, which range from expanded, flexible linkers that allow a wide range of motion between the connected domains, or rigid, inflexible linkers that enforce limits on the distances and movements accessible to the two domains [290]. Not only is the hZIP14 domain linker relatively long, it has a high density of polar and negatively charged residues, both factors which increase the flexibility and extension of linkers, increasing the range of motion between connected domains [262, 291]. For comparison, the sequences of linkers for other human members of the ZIP family are detailed in figure 5.32, the combination of length with the high density of polar and negatively charged residues suggest that it may be the most flexible of the human ZIP domain linkers. The varying characteristics of the linkers for ZIP proteins may function to control the potential quaternary structures of their respective proteins. In a similar vein, designing multimeric proteins through the genetic fusion of homodimeric protein domains via appropriate linkers is an active area of research, though the linkers are generally designed to be relatively rigid in order to limit potential interactions between the domains, and control the resulting multimeric structure [290, 292, 293]. Given the length and flexibility

of the hZIP14 domain linker, hZIP14 could take part in a number of oligomeric arrangements. Figure 5.30 shows some of the possibilities; homodimers or homotetramers are the most obvious options, but technically, an hZIP14 oligomer could have the open symmetry represented by figure 5.30 C, which has the potential to seed polymeric chains of hZIP14. This type of oligomer is formed by acetyl CoA carboxylase (ACC), a multi-domain protein with two separate dimerisation sites, each on a different domain [294]. The carboxyl transferase domains form a permanent dimer, while the low affinity dimerisation sites of the biotin carboxylase domains may remain unbound, bind to each other to form closed dimers, or associate with other ACC dimers to form extended polymers (the catalytically active form) - depending on the local concentrations of various metabolites.

```
                .  .....+   .-.
hZIP4    306    SGACTSQSRPPVQDQ                  320
                .+ .
hZIP5    194    SRVCIGAPAPAPPG                   207
                 +..  +..-++ -   +...
hZIP6    300    ARSCLIHTSEKKAEIPPKTYS            320
                + .--+ +++.+
hZIP8    111    HPCEDRPKHKTR                     122
                .+ . -+ -+   -- .+-+.   --- .
hZIP10   373    SRLCIEHFDKLLVEDINKDKNLVPEDEANIG  403
                ....+  +-.. +
hZIP12   346    SCSCHLPKDQQAKLPP                 361
                .+ ...-..-.--.-..-- + .  -
hZIP14   127    SRACTSENQENEENEQTEEGRPSAVEVW     154
```

Figure 5.32: **Domain linkers connecting N- and C-terminal domains of human ZIP proteins.** Sequences of the domain linkers connecting the N- and C- terminal domains, for those human ZIPs with well structured N-terminal domains; the symbols above each sequence indicate polar (.), positive (+) or negative (-) residues.

Tetramerisation of hZIP14 is compatible with the flow cytometry data - the dependence of cell surface localisation on expression levels in mutants that

inhibit ECD dimerisation, could be explained by increased tetramerisation between hZIP14 dimers (connected by either the TMD or ECD domains) as expression levels increase. In this case, the purpose of ECD dimerisation would be to enhance interaction of the TMD domains through the increased avidity of dimers interacting through two available binding sites. Higher order hZIP14 oligomers are also a possibility, but are unlikely to have been observed in western blotting, as any large oligomers that were not dissociated during sample preparation, are unlikely to have transferred to the nitrocellulose membrane due to retardation in the polyacrylamide gel.

5.3.4 The Transport Cycle of hZIP14

The co-evolution analysis of hZIP14 was in strong agreement with the TMD homology model for the relative positions of TMs 2, 3, 7 and 8 (the dimerisation helices), and TMs 1, 4, 5 and 6 (the metal-binding helices) (figure 5.25). Some of the predicted contacts between these two bundles were in agreement with the homology model, but many predicted that the metal-binding helices interacted with positions on the dimerisation helices closer to the extracellular side of the protein, suggesting an alternative conformation of hZIP14 (figure 5.26). Since the BbZIP template was in the open-inward metal-binding conformation, these alternative contacts are likely to be satisfied by the open-outward conformation. The transition between conformations would require a slight rotation between TMs 4 and 7, with movement of the metal-binding helices towards the extracellular side of the protein. This would expose the metal-binding site to the outside of the cell, where it could bind metal ions prior to uptake. This type of transport mechanism has been referred to as a 'two-domain elevator mechanism', and has recently been predicted for BbZIP by Wiuf et al. [267], on the basis of their BbZIP open-inward metal-free crystal structure,

which showed that, compared to the BbZIP open-inward metal-binding crystal structure, the metal-binding helices had moved relative to the dimerisation helices, and that each group of helices were fairly static within their respective bundles.

The two-domain elevator mechanism is used by a number of dimeric secondary active transporters, where the 'oligomerisation' or 'gate' domain typically acts as a static scaffold against which the solute-binding 'core' domain moves, so that access to the solute-binding residues alternates between the two sides of the membrane [295, 296]. So far, transporters found to use this recently identified mechanism have all been dimers, and dimerisation has been shown to be necessary for transport in some cases [297, 298]. Both the N469K mutants and the L441R mutants, believed to inhibit dimerisation of the TMD, are still able to transport metals - transport of zinc (albeit from an intracellular location) by L441R hZIP14 has been documented [95], and N469K hZIP14_A-HA transported iron at a higher rate than wild-type, relative to cell surface levels (figure 3.11). Transport by these mutants of hZIP14 may have been limited to those proteins whose TMD domains were dimerised; however, if hZIP14 is able to transport metals while not dimerised, further study could expand understanding of the requirements for transport by the elevator mechanism.

5.3.5 Choice of Designed TMD Paired Mutations

Designing deleterious mutations can be done with only a multiple sequence alignment, by choosing highly conserved residues, and changing them to amino acids with very different physico-chemical properties. In order to test the TMD structure, I decided on an ambitious challenge, to design pairs of mutations that would have no effect on protein function when combined, where one mutation

of the pair *would* have a negative effect on protein function. The literature was searched for software that would predict deleterious mutations on the basis of a structure alone, which led to the Amino Acid Interaction Server (INTAA) [299]. This server analyses submitted PDB files, and returns a matrix describing the strength of amino acid interactions, focusing on residues, side-chains or backbones as selected. This software could potentially have been used to aid in the selection of sites for mutation, on the basis of the homology model alone. However, most pairwise interaction energies predicted by this software were very similar.

In the absence of software specifically designed for identifying sites for mutagenesis and predicting the effects, the choices of paired TMD mutations were made on the basis of the homology model, in combination with conservation scores from the multiple sequence alignment. Since the co-evolution analysis was largely in agreement with the homology model, it seems that the paired TMD mutations were not an effective method for testing the validity of the homology model. An alternative strategy for selecting mutations could have been to use programs such as QMEANBrane to assess the effects of various point mutations at different positions in the structure [200].

5.3.6 Effects of Designed Mutations on Iron Uptake

The E376H mutation was predicted to decrease iron uptake, as this mutation would bring hZIP14 in line with other LIV-1 proteins which have a histidine in the equivalent position, restoring the LIV-1 HEXPHEXGD motif, and removing the difference believed to be responsible for the ability of ZIP8 and ZIP14 to transport iron and manganese in addition to zinc (discussed in section 5.1.2). Iron uptake by HeLa cells expressing E376H hZIP14_A-HA was not significantly

different from the negative control (figure 5.17). Although the expression of E376H was low (both in total and at the cell surface), iron uptake was still low relative to wild-type (figure 5.18). This decrease in iron uptake was in line with expectations, but in order to determine whether the E376H mutation altered solute specificity, rather than simply decreasing or preventing transport, both zinc and manganese uptake by this mutant must be tested. A cell line expressing higher levels of E376H would also be preferable.

The Q365E and Q365E;E209Q mutants, relative to cell surface levels, transported iron at a higher rate than wild-type hZIP14_A-HA (figure 5.18). An increase in function is not typically the expected effect of a mutation (though this was also the case for N469K hZIP14_A-HA), but may be explained by looking at the positions of these mutations in relation to the previously discussed elevator mechanism through which hZIP14 may transport metals. The highly conserved E209 and Q365 residues are predicted by the TMD homology model to form an interaction between TMs 2 and 5, part of the metal-binding helix bundle and the dimerisation helix bundle, respectively. According to the Amino Acid Interaction Server, a program that analyses protein structures to predict residues' contributions to protein stability, these two amino acids had the most stable (low energy) interaction in the TMD homology model [299]. Since the co-evolution data predict that TMs 2 and 5 move past each other during the transport cycle; mutating these two residues may have removed a stabilising interaction between TMs 2 and 5, making this motion easier, and allowing a faster rate of transport. Co-evolution data predict a contact between E209 and E376, another TM 5 residue, part of the TM 5 metal-binding motif, and much further away than Q365 (figure 5.33). This suggests that E209 has a conserved role in the conformational changes that occur in the hZIP14 transport cycle, in which case mutating it to glutamine was unlikely to reverse the effect of the

Q365E mutation.

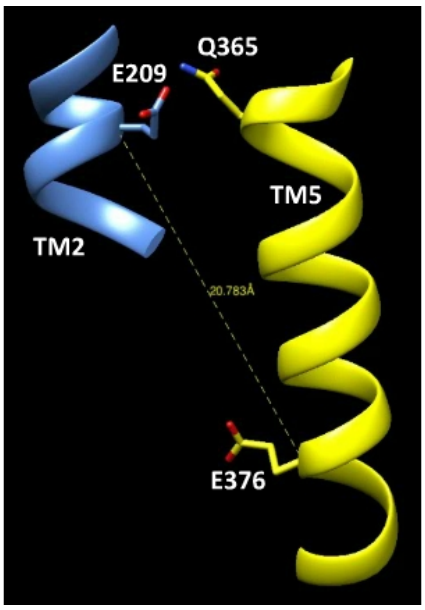


Figure 5.33: **Co-evolution predicted contact of E209 relative to Q365.** TMs 2 and 5, with residues 209, 365 and 376 in stick-form; distance marker represents the distance of 20.783 Å between the alpha-carbons of E209 and E376.

5.3.7 Altered Metal Uptake May Affect Regulation of hZIP14 Trafficking

As discussed in Chapter 3, section 3.3.4.1, hZIP14 trafficking may be regulated by metals, in both the long and short term. The western blot results for the designed TMD mutations were identical to those of wild-type hZIP14_A-HA, giving no indication of significant structural differences between these mutants (figure 5.13). The flow cytometry results for these mutants showed a range of effects on the partitioning of hZIP14 between the inside and the surface of the cell, at different levels of expression (figure 5.15). The E376H mutant had similar cell surface expression across the range of expression levels, the A354L mutant was similar to wild-type, but at high expression levels A354L hZIP14_A-HA appeared to be decreased at the cell surface (similar to the effect of the

P379L mutant, see figure 3.37), and the A354L;L206A density plot showed an approximately proportional relationship. The relationship between internal and surface hZIP14 for the Q365E and Q365E;E209Q mutants was an inverse sigmoidal curve, and these mutants both transported iron at a rate $\sim$ 3.5 times that of wild-type hZIP14_A-HA (normalised by cell surface levels). It is possible that the various relationships between internal and surface hZIP14 are related to differences in metal transport, which could be tested by incubating the transient transfectants in different concentrations of metals prior to differential staining.

5.3.8 Regulation of hZIP14_A and hZIP14_B May Differ

The $\triangle$40-130 mutations were introduced into plasmids encoding both the A and B isoforms of hZIP14-HA, though as the only positive control was wild-type hZIP14_A-HA, the most meaningful comparison to be made for $\triangle$40-130 hZIP14_B-HA, was to $\triangle$40-130 hZIP14_A-HA. Differences between these two proteins were detected in western blotting, and by flow cytometry (figures 5.6 and 5.11).

In particular, the relative surface levels of $\triangle$40-130 hZIP14_B were higher than those of $\triangle$40-130 hZIP14_A. hZIP14 A and B are nearly identical; they are the same length, and differ by only 20 residues over a stretch of 44 residues throughout TM1, the TM1-2 cytosolic loop and the N-terminal half of TM 2 (figures 5.34 and 5.35). In the homology model, the positions of these altered residues may interact with TMs 5, 6 and 8, and residue 199 (tyrosine or leucine for isoforms A and B, respectively) projects out into the predicted dimerisation interface. Many of the alternate residues share physicochemical characteristics, but these small differences have the potential to alter the metal binding site through effects on TM 5 (potentially affecting solute affinity,

specificity or preference), the rate of transport through effects on the interactions between TMs 5 and 2, and dimerisation through a change to the affinity of the dimerisation site. Since metal uptake by hZIP14 has the potential to alter trafficking, any of these effects could explain the increased cell surface localisation of △40-130 hZIP14_B relative to △40-130 hZIP14_A.

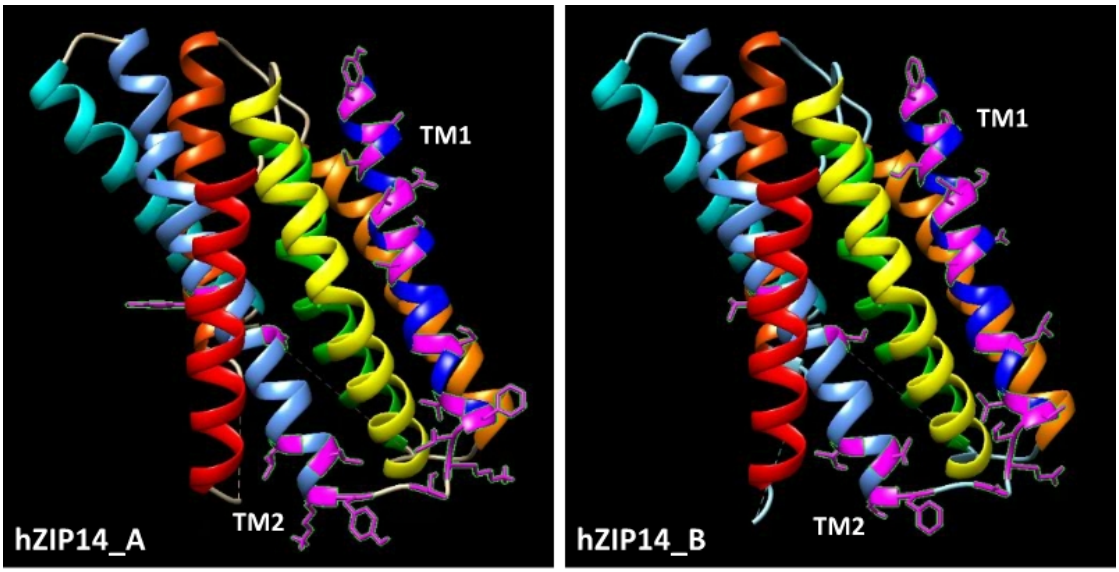


Figure 5.34: **Differences between hZIP14_A and B.** Ribbon depiction of hZIP14 homology models for hZIP14 A and B. Residues that differ between the isoforms are coloured in magenta.

```
                 * * *  * * ****  * *   * *  * * ***** ****
hZIP14_A   156  YGLLCVTVISLCSLLGASVVPFMKKTFYKRLLLYFIALAIGTLY  199
hZIP14_B   156  FGFLSVSLINLASLLGVLVLPCTEKAFFSRVLTYFIALSIGTLL  199
```

Figure 5.35: **The sequence that differs between the A and B isoforms of hZIP14.** Asterisks indicate positions with identical amino acids.

5.3.9 AlphaFold2 Prediction for the Structure of hZIP14

The recently released database of human protein structures predicted by the AlphaFold2 neural network includes a predicted structure for the entire human

ZIP14 sequence; AlphaFold2 dramatically outperformed its competitors in the 14th Critical Assessment of protein Structure Prediction (CASP14, 2020) [233].

5.3.9.1 The AlphaFold2 hZIP14 ECD Structure

The AlphaFold2 prediction for hZIP14 is in the monomeric form, with a disordered linker connecting the N-terminal and C-terminal domains (the linker is portrayed as an alpha-helix with a low per-residue confidence score, which is argued to be an indicator of intrinsic disorder [281, 300]).

A comparison of the AlphaFold2 hZIP14 ECD to my ECD homology model shows that the two structures are nearly identical in the alpha-helical regions (average RMSD of 1.091 Å) (figure 5.36 A). The third alpha helix (residues 94-100) shows the biggest difference, however, this does not substantially alter the interactions of F98, or change the explanation for the effect of that mutation. As shown in figure 5.36 B, F98 is predicted to be at the core of the ECD monomer in both models, interacting with conserved hydrophobic residues. The positions of S104 and R128 are also unchanged in the AlphaFold2 hZIP14 model. The AlphaFold2 model also predicts structures for the inter-helical loops, which were largely disregarded in my ECD model due to the decreased sequence similarity between the *Pteropus alecto* ZIP4 and human ZIP14 ECDs in these regions.

5.3.9.2 The AlphaFold2 hZIP14 TMD Structure Differs From the TMD Homology Model

The AlphaFold2 and homology TMD models were very similar; with the exception of the N469K mutation, none of the differences affected predictions made on the basis of the homology model, about the positions of the disease-causing or designed TMD mutations.

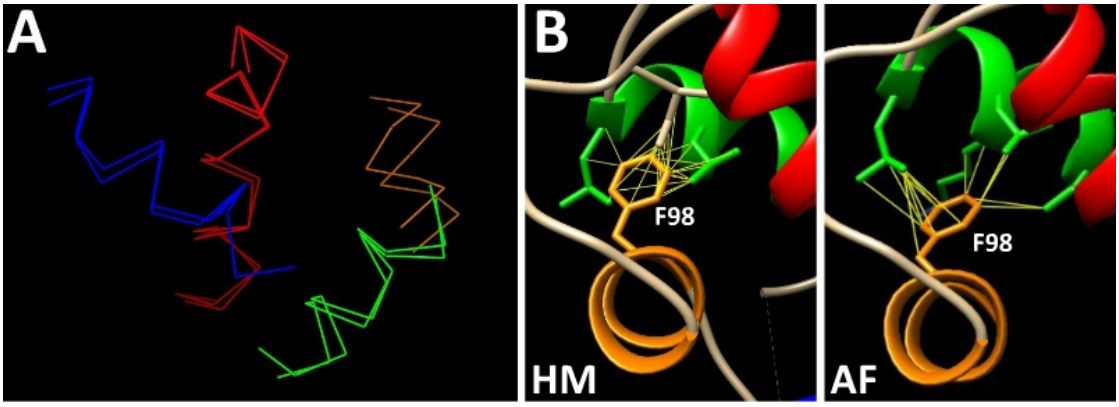


Figure 5.36: **AlphaFold2 prediction for the hZIP14 ECD is similar to the homology model.** A: Wire-frame depiction of the alpha-helical regions of the hZIP14 ECD homology and AlphaFold2 models. B: Interactions of F98 with nearby residues, in the homology model (HM) and the AlphaFold2 model (AF).

As shown in table 5.3, the differences in the positions of the helices involved in metal-binding (TMs 1, 4, 5 and 6) are very small, and are more pronounced for the helices involved in dimerisation (TMs 2, 3, 7 and 8), especially TM 3. While the general organisation of the helices is the same for both models, the AlphaFold2 TM 3 is in a very different position, interacting with TM 8 rather than TM 7 (figures 5.37 A and B). A comparison of the distance matrix for the AlphaFold2 model with the hZIP14 co-evolution analyses by the PSIPRED DeepMetaPSICOV [271] and OpenSEQ GREMLIN [272] servers, shows excellent agreement (figure 5.38). There are three main differences between the distance matrices of the AlphaFold2 and homology models (figures 5.22 and 5.38). Firstly, the AlphaFold2 TM 7 is longer at the C-terminus, as predicted in section 5.2.5.2. Secondly, the AlphaFold2 model does not have the predicted contacts between TMs 3 and 7. Thirdly, the AlphaFold2 model does have the predicted contacts between TMs 3 and 8.

AlphaFold-Multimer, which was trained using similar methods to AlphaFold2, can be used to predict the structures of protein complexes [196]. The predicted

Table 5.3: **Differences in position between the TM helices of the homology model and AlphaFold2 TMD structures./hflNoteYes.**

hZIP14 TM	Average RMSD Between AlphaFold2 and Homology Model TMs (Å)
1	1.967
2	5.945
3	27.327
4	2.154
5	1.357
6	1.338
7	4.105
8	7.174

homodimer is shown in figure 5.37 C. The arrangement of helices in the monomers of the AlphaFold-Multimer hZIP14 homodimer is in agreement with the AlphaFold2 monomer, rather than the TMD homology model (figure 5.37). However, in the homodimer predicted on the basis of the hZIP14 homology model, and the AlphaFold-Multimer homodimer, the positions of TM helices are essentially the same, except each TM 3 belongs to a different monomer (figures 5.29 and 5.37 C).

The distance matrix for the AlphaFold-Multimer homodimer shows that all the predicted contacts are made, with the addition of contacts between identical residues in TMs 2 and 7 from each monomer - these could not be predicted by co-evolution analysis (figure 5.39). Despite the difference in the position of TM 3, the AlphaFold-Multimer homodimer still makes the contacts previously predicted for a homodimer of the TMD homology model (section 5.2.5.2), and supports the explanation proposed for the effect of the L441R mutation. In the AlphaFold-Multimer homodimer, L441 is at the center of the dimerisation

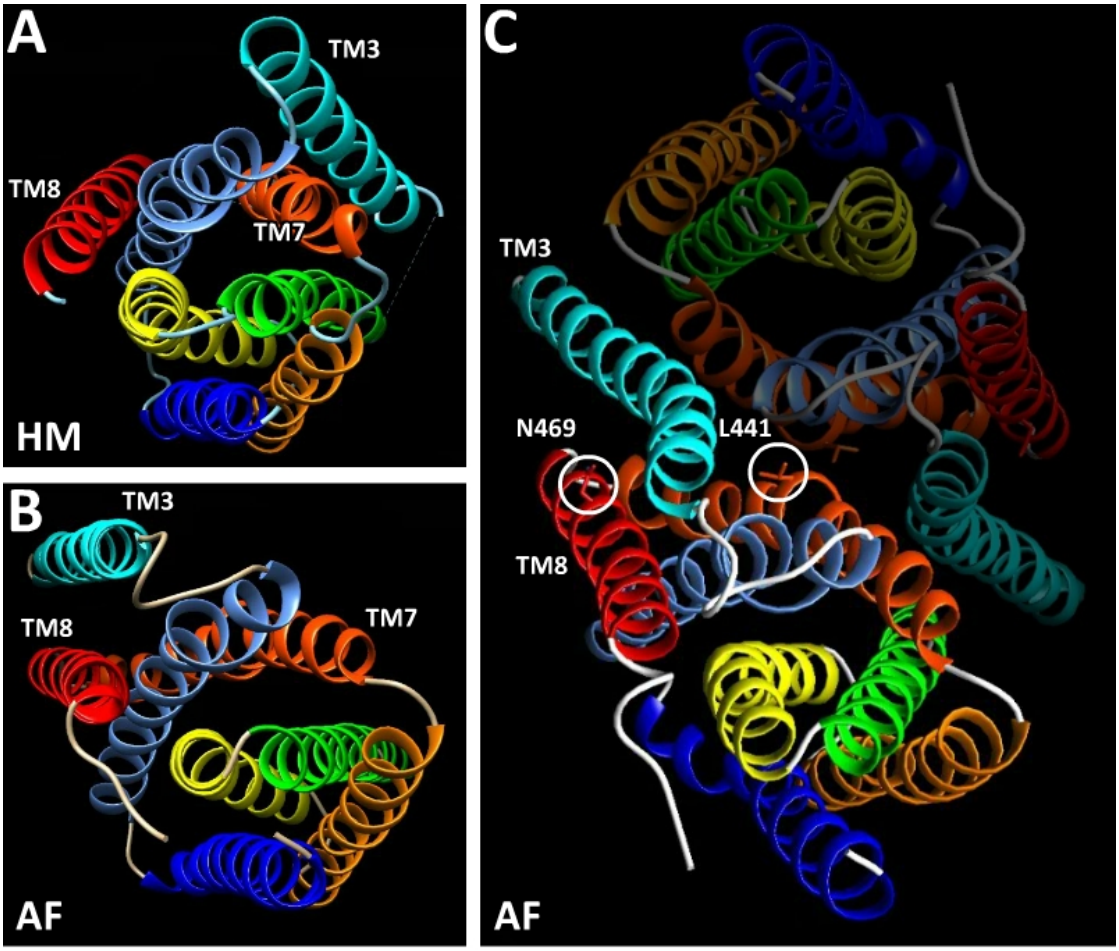


Figure 5.37: **AlphaFold2 predicts a different arrangement of TM helices than the homology model.** A: Homology model of hZIP14 TMD. B: AlphaFold2 model of hZIP14. C: AlphaFold-Multimer model of the hZIP14 TMD as a dimer; second monomer greyed out. Ribbon structures viewed from the extracellular perspective, coloured from the N-terminus to the C-terminus (blue to red).

interface (figure 5.37 C); mutating this leucine to an arginine, a large and positively charged residue (as in the L441R mutation) is highly likely to inhibit dimerisation of the TMD. For the purposes of understanding the effects of the disease-causing mutations, the key difference between the TMD homology model and the AlphaFold2 (and AlphaFold-Multimer) model is the position of N469. In the TMD homology model, N469 is on the surface of the TMD, facing the predicted dimerisation site, which was believed to explain the sigmoidal

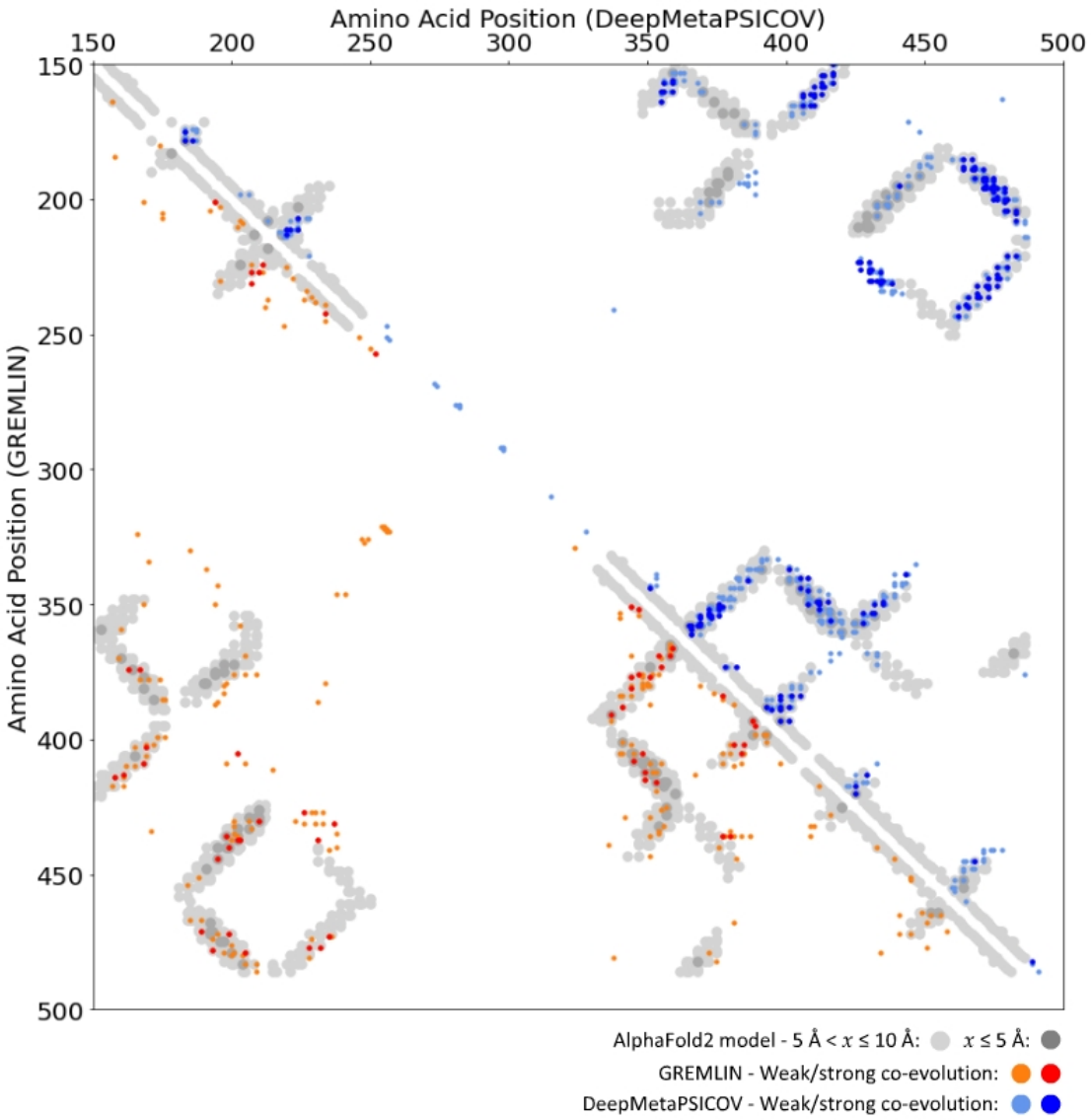


Figure 5.38: **Overlaid distance and co-evolution matrices for the AlphaFold2 hZIP14 TMD domain.** Co-evolution matrices from the GREMLIN (lower-left) and DeepMetaPSICOV (upper-right) analyses of hZIP14 were overlaid onto the AlphaFold2 hZIP14 TMD model distance matrix. The overlap shows areas of agreement between the co-evolution data and the AlphaFold2 model.

relationship between internal and surface hZIP14 caused by this mutation. In the AlphaFold models, N469 is no longer part of the dimerisation site (figure 5.37 C), which would seem to suggest that the AlphaFold models are incompatible with the theory that N469 is important for dimerisation.

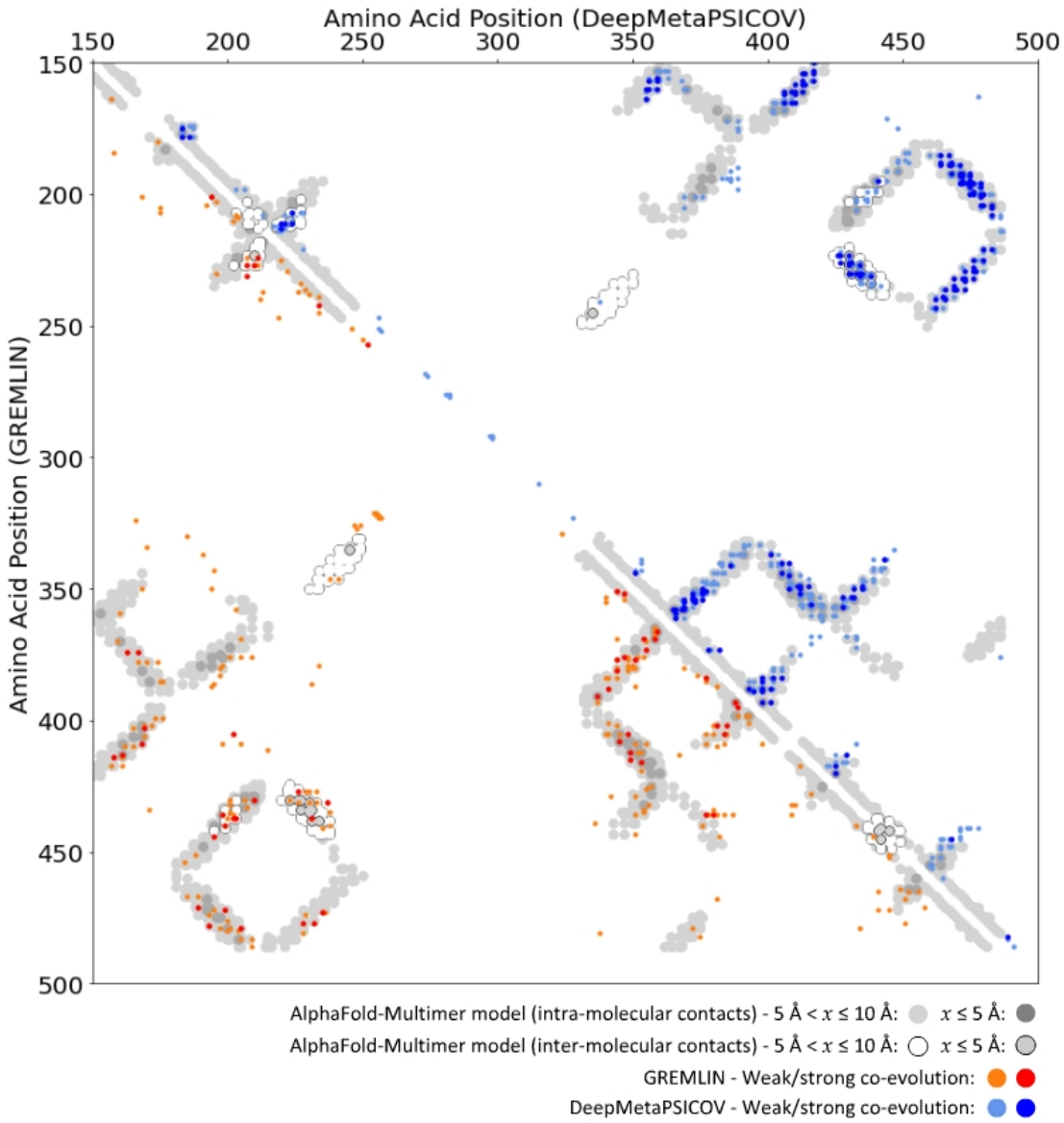


Figure 5.39: **Overlaid distance and co-evolution matrices for the AlphaFold-Multimer hZIP14 TMD domain homodimer.** Co-evolution matrices from the GREMLIN (lower-left) and DeepMetaPSICOV (upper-right) analyses of hZIP14 were overlaid onto the AlphaFold-Multimer hZIP14 TMD homodimer model distance matrix. The overlap shows areas of agreement between the co-evolution data and the AlphaFold-Multimer model.

AlphaFold2 has predicted the structures for a number of proteins in the ZIP family, and the more closely related LIV-1 family. The arrangement of helices in the template structure (BbZIP) has been described as 3+2+3, with TMs 1, 2 and 3, and TMs 6, 7 and 8 making two symmetrically arranged groups of 3

helices, and TMs 4 and 5 making the symmetrical bundle of 2 helices at the center of the protein [241]. AlphaFold2 predicts this 3+2+3 arrangement for BbZIP, IRT1 (one of the founding members of the ZIP family), and every human ZIP protein, except for hZIP8 and hZIP14, with hZIP8 predicted to have the same arrangement of TMs as predicted for hZIP14. ZIPs 8 and 14 are more closely related to each other than to the other members of the ZIP and LIV-1 families, and share a functional difference, in that they transport manganese and iron in addition to zinc. As discussed in section 5.3.6, the substitution of the initial histidine in the HEXPHEXGD motif for a glutamate has been predicted to explain the difference in solute profile [11], and the mutation of E376 to histidine almost abolished hZIP14 iron uptake (figure 5.18), though transport of iron and manganese may still require additional differences.

Since TMs 2, 3, 7 and 8 apparently share approximately the same position in the dimeric arrangements of mammalian and bacterial ZIPs, it is possible that a mutation affecting TM 3-8 interactions could inhibit dimerisation even if the mutated residue is not technically part of the dimerisation site. This has been seen in the UapA transporter, where mutations in TM 7 impaired oligomerisation, though not technically part of the dimerisation site [277]. Apart from the milder effect on cell surface localisation of the N469K mutation, relative to the L441R mutation, the N469K mutation had a tendency to cause a reticular pattern of localisation under immunofluorescence microscopy, though L441R cell surface levels were much lower (figures 3.37 and 3.15). This could indicate that as well as inhibiting dimerisation, N469K also disrupted folding and increased ER retention of hZIP14, which is more consistent with the AlphaFold2 model. An interesting consideration for the AlphaFold2 model is how it might relate to the elevator-transport mechanism proposed for hZIP14. With TM 3 located adjacent to TM 8 instead of TM 7, the metal-binding site

is considerably more exposed in the open-inward conformation when hZIP14 is a monomer, and as a dimer, TM 3 of each monomer makes contacts with TM 4, possibly affecting the transition between the open-inward and open-outward conformations. These differences between the monomeric and dimeric interactions may be involved in the observation that N469K, once at the cell surface, transported iron at a faster rate than wild-type hZIP14.

The core of the AlphaFold2 model is essentially the same as the homology model, and as such, most of the predictions and mutations based on the homology model still hold. However, the AlphaFold2 model covers regions that a homology model cannot, and the impressive performance of AlphaFold2 in CASP14 means that its novel prediction for the position of TM 3 should not be lightly dismissed, especially as homology modelling is unable to predict such a drastic rearrangement of secondary structures. Designing experiments to distinguish between these two arrangements of helices may be challenging, though retention of N469K in the ER suggests that TM 3:8 contacts may be differentiated from dimerisation contacts by their effects on protein folding.

5.3.10 Summary of Effects of the Designed hZIP14 Mutations

In this chapter, mutations designed on the basis of the homology models for the ECD and TMD were tested for their effects on cellular localisation and iron uptake. The TMD homology model was also compared to the results of co-evolution analysis, to assess its accuracy.

The results for the designed ECD mutations suggested that correct folding and dimerisation of this domain were important requirements for exit from the ER. It is also likely that dimerisation of the ECD and TMD domains

showed co-operativity, potentially due to dimerisation of each domain enhancing dimerisation of the other. Removal of the ECD domain indicated that the TMD was sufficient for cell surface localisation, potentially dependent on its dimerisation, and also revealed a potential difference in trafficking between hZIP14 isoforms A and B. The results for C118D mutant, which prevented the formation of a disulphide bridge linking ECD monomers, suggested that hZIP14 may also form higher-order oligomers. This would be compatible with my data and the structural features of hZIP14, but would be a novel finding for a ZIP protein.

The E376H mutation decreased iron uptake as predicted, though the designed, paired mutations for the TMD did not have the expected effects, in that the double mutants did not revert to wild-type characteristics. However, the TMD homology model structure was generally supported by the co-evolution analyses of hZIP14.

By comparing the homology model with the contacts predicted by the co-evolution data, I was able to predict that the hZIP14 TMD formed a dimer, and the conformational change that would occur during the transport cycle of hZIP14. These predictions were both consistent with recent literature findings on homologous proteins, and helped to explain the effects of L441R and N469K on cell surface localisation relative to expression, and suggested explanations for the unexpected increase in iron uptake by the Q365E and Q365E;E209Q mutants.

The recent release of the AlphaFold Protein Structure Database provided another hZIP14 model against which my results could be assessed. The AlphaFold2 model was generally in agreement with the homology models designed in Chapter 4, although TM 3 of the TMD was in a decidedly different location. This resulted in a negligible difference with the predicted arrangement

of TMs in the TMD homodimer, but there were potential implications for the mechanism explaining the effects of the N469K mutation, regarding its increased uptake of iron, and its increased retention in the ER.

The results described in this chapter greatly increase knowledge of how hZIP14 structure relates to its function, as well as suggesting interesting avenues for further study. Many of the findings are consistent with recent discoveries for similar proteins, though hZIP14 appears to have some interesting differences. The potential for hZIP14 to form tetramers and the possible role of the domain linker in hZIP14 quaternary structure, the co-operativity between the extracellular and transmembrane domains, and the potential mechanisms behind the mutations that increased iron uptake, are not topics that have been covered in the literature, though they may provide useful contributions towards fully understanding hZIP14 and related proteins.

Chapter 6

Discussion

In this research, documented disease-causing mutations in human ZIP14 (hZIP14) have been shown to have a range of effects on cellular localisation and iron uptake function (Chapter 3). In order to understand how these mutations might cause these varying effects, homology modelling was undertaken to predict protein structures for the extracellular and transmembrane domains of hZIP14 (Chapter 4). Further mutations were chosen to test the accuracy of the models, predicted dimerisation of the extracellular domain (ECD), and the involvement of E376 in solute specificity (Chapter 5). The results of these experiments, combined with evaluation of the homology models by co-evolution analysis, were largely in support of the homology models, and provided a basis for further predictions relating the structure of ZIP14 to its function.

6.1 Differences Between Isoforms of hZIP14

Reverse-transcriptase PCR of cDNA from HeLa, HEK293 and HepG2 (originating from uterine, kidney and liver tissues, respectively) showed

expression of both hZIP14_A and B in HepG2 cells, and only hZIP14_A in HEK293 cells (figure 3.4). Differences in the tissue distribution of ZIP14_A and B are documented in the literature, as well as differences in the kinetic properties of the transporters; the significance of these differences is not known [14, 19, 36].

The homology models of hZIP14 predict that differences between isoforms A and B are restricted to TMs 1 and 2 of the transmembrane domain (figure 5.34); the locations of the alternate residues are in close proximity to the predicted metal translocation pore, and the predicted TMD dimerisation site. The differential staining of intracellular and surface hZIP14 followed by flow cytometry, found differences in the proportions of the hZIP14_A and B ECD-deletion mutants at the cell surface, suggesting that the trafficking of these two isoforms may differ (figure 5.11).

Ideally, wild-type hZIP14_B should also have been tested to see if differences in cell surface levels of the two isoforms were present in the presence of the extracellular domain. If differences in cell surface localisation were present in the wild-type isoforms, mutations in their alternate residues could have been used to probe the causes of these differences. Such information could further our understanding of ZIP14 structural features, and the physiological roles of the two isoforms.

6.2 The Transport Mechanism of ZIP14

Radioactively-labelled iron uptake relative to cell surface levels of hZIP14 protein was measured in monoclonal stable cell lines (sections 3.2.4.2 and 5.2.4, and figures 3.11 and 5.18). Although cell surface levels of the N-terminal mutants were decreased relative to wild-type, the disease-causing N-terminal

mutant tested (R128W) also showed no ability to transport iron. The effects of the TMD mutations varied, and were not easily interpreted. While decreased iron uptake by hZIP14 G383R and E376H was expected due to the conservation of these residues in the predicted metal translocation pore, P379 is also a highly conserved residue within this motif, and yet iron uptake relative to cell surface levels of hZIP14 P379L did not appear to be affected; yet other TMD mutations appeared to increase iron uptake function. The co-evolution analysis of the hZIP14 TMD homology model allowed for some speculation on possible mechanisms underpinning these effects (sections 5.3.4 and 5.3.6).

Increased iron uptake was measured for the hZIP14 N469K mutant, which is predicted to inhibit dimerisation of the TMD. The L441R mutant, also predicted to inhibit dimerisation, is reported to be able to transport zinc within the cell [95]. This is of interest as two-domain elevator transporters are believed to function as dimers [295, 297, 298]. Directly establishing experimental evidence of ZIP14 TMD dimerisation and then exploring its effects on metal uptake, both in terms of rate and solute specificity, may provide useful insights into ZIP14 specifically, and perhaps two-domain elevator transporters more generally.

Given more time, there are further optimisations that could have increased the reliability of the iron uptake results. Establishing monoclonal cell lines with similar levels of hZIP14 at the cell surface would have been beneficial. This could have been achieved by using flow cytometry to sort transfected cells on the basis of surface staining for hZIP14. Additionally, although iron uptake was normalised relative to levels of protein, it is possible that differences in the confluence of sample wells could have contributed variability to the results. Optimisation of the protocol did not account for the possibility of increased iron transport, so for cells expressing mutants with high iron uptake capability, iron uptake may have been limited, particularly if confluence was high. Further

tests should have been performed to ensure that iron uptake of all mutants reflected their iron transport capacity, and not other limiting factors. In addition, the effect of the hZIP14 mutations on uptake of zinc and manganese should also have been measured; the radioactive isotopes of these metals pose additional safety concerns and require additional resources to detect. However, other methods for quantifying metal uptake, such as the use of fluorescent metal-binding indicators, or inductively coupled plasma mass spectrometry, could be investigated.

6.3 Potential Dimerisation of the Extracellular Domain

The N-terminal disease-causing mutations caused retention of hZIP14 in the ER, decreasing cell surface levels (sections 3.2.5.1, 3.2.5.1 and 3.2.6). Further study showed these effects to be dependent on expression levels, such that hZIP14 affected by these mutations could be expressed at the cell surface at wild-type levels, at very high levels of expression (figure 3.36). As the research progressed, several pieces of evidence strongly suggested that the hZIP14 extracellular domain formed a homodimer, and that one mechanism through which N-terminal disease-causing mutations decreased cell surface levels of hZIP14, was inhibition of this dimerisation (sections 4.3.2, 4.2.2 and 5.3.1). The results on the effects of the designed ECD mutations revealed that the ECD was not strictly necessary for cell surface expression. Mutations predicted to impair folding or dimerisation of the ECD, decreased cell surface levels of hZIP14 at low expression levels, but mutants lacking the ECD domain were still able to localise at the cell surface at relatively low levels of expression (figure 5.11).

Taken together, this research suggests that one role of the ECD may be to enhance cell surface localisation at low levels of expression. In the future, it would be useful to characterise the range of hZIP14 expression levels that occur in different cell types, and in response to substances previously indicated to regulate hZIP14 expression. This would allow me to identify physiologically meaningful expression levels, and could potentially be done by altering genomic DNA of relevant cell lines to add, for example, an HA tag to the C-terminus of hZIP14, in order to monitor expression of endogenous hZIP14.

A number of ZIP proteins have extracellular domains with homology to hZIP14; several pathological mutations occur in the ECD of the closely-related protein ZIP4, suggesting it plays an important functional role [219]. Experiments using flow cytometry to study transiently transfected cells in order to test for concentration-dependent effects of the ECD domains on ZIP function, may be a useful avenue for further research, particularly for ZIPs that are regulated through changes in expression levels.

6.3.1 Potential Dimerisation of the Transmembrane Domain

The sigmoidal curves relating internal and surface hZIP14, found for the N-terminal mutations, indicated co-operativity was involved in cell surface localisation of hZIP14 (figures 3.36, 5.10 and 5.11). Two TMD mutations, L441R and N469K, also had sigmoidal relationships between internal and surface hZIP14, and were later shown to be at the interface of a predicted TMD dimerisation site based on co-evolution analysis of the TMD homology model (figure 3.37 and section 5.3.2). Dimerisation of the TMD has been shown experimentally for a number of other ZIPs, and is also consistent with the two-domain elevator mechanism of transport indicated by the co-evolution analysis

[219, 221, 222, 280].

The L441R and N469K mutations had a strong impact on cell surface levels of hZIP14, while L441R did not display a reticular pattern of localisation (figures 3.37 and 3.14). If these mutations exert their effects on localisation through inhibition of TMD dimerisation, the implication is that dimerisation of the TMD (unlike dimerisation of the ECD) may impact localisation at the cell surface independently of effects on exit from the ER. Mutations designed to inhibit TMD dimerisation, in the presence and absence of the extracellular domain, could be used to test this hypothesis; the choice of locations for these mutations should consider the contradictory structural model of hZIP14 created by AlphaFold2 (section 5.3.9).

Results from SDS-PAGE analysis and western blotting, particularly of the hZIP141_A C118D mutant, suggested potential tetramerisation of hZIP14 (figure 5.6 and section 5.3.3). This has not been proposed for any other members of the ZIP family, though dimerisation has been confirmed for several [219, 221, 222, 280]. The predicted existence of two separate dimerisation sites, connected by a domain linker likely to be disordered, indicates that tetramerisation, and indeed, the formation of higher-order oligomers, is possible. Western blotting of hZIP14 has previously revealed potential oligomers [13, 35]; evidence of oligomers in other ZIPs may have been ignored. In a study of the closely-related ZIP4, and its dimerisation in relation to the histidine-rich extracellular TM 2-3 loop, western blots of chemically cross-linked ZIP4 consistently showed two higher molecular mass bands [283]. In another study of ZIP4, fluorescence correlation spectroscopy was used to study the oligomeric status of cell surface ZIP4, concluding it was dimeric [223]. However, as this experiment was performed in high zinc conditions and the majority of ZIP4 had been endocytosed; it is likely that cell surface ZIP4 tetramers would not

form in such conditions. The possibility that certain ZIPs may form tetramers is worth investigating, and may reveal additional regulatory mechanisms affecting members of the ZIP family.

6.3.2 Further Possibilities For the Study of hZIP14 Cell Surface Localisation

The differential staining protocol used to study the relationship between expression levels of the hZIP14 variants and their presence at the cell surface, produced a comparison between internal and surface protein levels which was challenging to interpret (figure 3.35). As the relative fluorescence intensity resulting from the pairs of antibodies used was unknown, it was not possible to determine total expression of hZIP14, or the actual proportion of hZIP14 at the surface. One possible solution would have been to separately stain samples of cells, permeabilised and non-permeabilised, with the pairs of antibodies, so that relative fluorescence intensities for the same amount of hZIP14 could be determined for each pair of antibodies. Another possibility would be to add an additional protein marker to hZIP14, such as the FLAG octapeptide or green fluorescent protein (GFP), so that total hZIP14 could be determined. A large marker, such as GFP, would need to be assessed for independent impacts on localisation of hZIP14.

The N-terminal mutations, ECD deletions and the L441R and N469K mutations all showed expression-dependent effects on cell surface levels of hZIP14 (figures 3.36, 5.10 and 5.11), and co-operativity between the two domains was proposed, and argued to be a result of dimerisation of each domain enhancing dimerisation of the other (section 5.3.2). This could be tested by combining mutations, for example, the L441R and N469K mutations with deletion of the

ECD. If the sigmoidal relationship between internal and surface levels of these mutants remained, that would effectively disprove the hypothesis, though it would be ideal to establish direct experimental evidence of dimerisation of each domain first.

6.4 Summary

The documentation of disease-causing mutations in human ZIP14 offered potential insights into structurally important regions of human ZIP14. In this research, the effects of these mutations on cellular localisation and iron transport function were used as a starting point to further our understanding of ZIP14. Computational modelling, site-directed mutagenesis and co-evolution analysis were employed to build on the initial findings, resulting in a deeper understanding of human ZIP14, as well as additional research questions, that may contribute to the study and understanding of the ZIP family as a whole.

References

[1] M. J. Salgueiro, M. Zubillaga, A. Lysionek, M. I. Sarabia, R. Caro, T. De Paoli, A. Hager, R. Weill, and J. Boccio, "Zinc as an essential micronutrient: A review," *Nutrition Research*, vol. 20, no. 5, pp. 737–755, 2000.

[2] E. S. Gurzau, C. Neagu, and A. E. Gurzau, "Essential metals—case study on iron," *Ecotoxicology and Environmental Safety*, vol. 56, no. 1, pp. 190–200, 2003.

[3] P. Chen, J. Bornhorst, and M. Aschner, "Manganese metabolism in humans," *Frontiers in Bioscience - Landmark*, vol. 23, no. 9, 2018.

[4] M. L. Ackland and A. Michalczyk, "Zinc deficiency and its inherited disorders -a review," *Genes Nutr*, vol. 1, no. 1, pp. 41–49, 2006.

[5] S. Dev and J. L. Babitt, "Overview of iron metabolism in health and disease," *Hemodialysis International*, vol. 21, pp. S6–S20, 2017.

[6] J. W. Finley and C. D. Davis, "Manganese deficiency and toxicity: Are high or low dietary amounts of manganese cause for concern?" *BioFactors*, vol. 10, no. 1, 1999.

[7] L. M. Plum, L. Rink, and H. Haase, "The essential toxin: impact of zinc on

human health," *Int J Environ Res Public Health*, vol. 7, no. 4, pp. 1342–1365, 2010.

[8] C. W. Siah, J. Ombiga, L. A. Adams, D. Trinder, and J. K. Olynyk, "Normal Iron Metabolism and the Pathophysiology of Iron Overload Disorders," *Clinical Biochemist Reviews*, vol. 27, no. 1, pp. 5–16, 2006. [Online]. Available: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1390789/https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1390789/pdf/cbr27_1p005.pdf

[9] R. A. Hauser, T. A. Zesiewicz, A. S. Rosemurgy, C. Martinez, and C. W. Olanow, "Manganese intoxication and chronic liver failure," *Ann Neurol*, vol. 36, no. 6, pp. 871–875, 1994.

[10] L. A. Gaither and D. J. Eide, "Eukaryotic zinc transporters and their regulation," 2001.

[11] K. M. Taylor and R. I. Nicholson, "The LZT proteins; the LIV-1 subfamily of zinc transporters," *Biochimica et Biophysica Acta (BBA) - Biomembranes*, vol. 1611, no. 1-2, pp. 16–30, 2003.

[12] K. M. Taylor, H. E. Morgan, K. Smart, N. M. Zahari, S. Pumford, I. O. Ellis, J. F. R. Robertson, and R. I. Nicholson, "The Emerging Role of the LIV-1 Subfamily of Zinc Transporters in Breast Cancer," *Molecular Medicine*, vol. 13, no. 7-8, pp. 396–406, 2007. [Online]. Available: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1936980/

[13] K. M. Taylor, H. E. Morgan, A. Johnson, and R. I. Nicholson, "Structure-function analysis of a novel member of the LIV-1 subfamily of zinc transporters, ZIP14," *FEBS Lett*, vol. 579, no. 2, pp. 427–432, 2005. [Online]. Available: https://www.ncbi.nlm.nih.gov/pubmed/15642354

[14] K. Girijashanker, L. He, M. Soleimani, J. M. Reed, H. Li, Z. Liu, B. Wang, T. P. Dalton, and D. W. Nebert, "Slc39a14 gene encodes ZIP14, a metal/bicarbonate symporter: similarities to the ZIP8 transporter," *Mol Pharmacol*, vol. 73, no. 5, pp. 1413–1423, 2008.

[15] N. Zhao, J. Gao, C. A. Enns, and M. D. Knutson, "ZRT/IRT-like protein 14 (ZIP14) promotes the cellular assimilation of iron from transferrin," *J Biol Chem*, vol. 285, no. 42, pp. 32 141–32 150, 2010. [Online]. Available: https://www.ncbi.nlm.nih.gov/pubmed/20682781

[16] H. Nam, C. Y. Wang, L. Zhang, W. Zhang, S. Hojyo, T. Fukada, and M. D. Knutson, "ZIP14 and DMT1 in the liver, pancreas, and heart are differentially regulated by iron deficiency and overload: implications for tissue iron uptake in iron-related disorders," *Haematologica*, vol. 98, no. 7, pp. 1049–1057, 2013.

[17] K. Thorsen, F. Mansilla, T. Schepeler, B. Oster, M. H. Rasmussen, L. Dyrskjot, R. Karni, M. Akerman, A. R. Krainer, S. Laurberg, C. L. Andersen, and T. F. Orntoft, "Alternative splicing of SLC39A14 in colorectal cancer is regulated by the Wnt pathway," *Mol Cell Proteomics*, vol. 10, no. 1, p. M110 002998, 2011. [Online]. Available: https://www.ncbi.nlm.nih.gov/pubmed/20938052

[18] L. A. Lichten and R. J. Cousins, "Mammalian zinc transporters: nutritional and physiologic regulation," *Annu Rev Nutr*, vol. 29, pp. 153–176, 2009.

[19] K. Tuschl, E. Meyer, L. E. Valdivia, N. Zhao, C. Dadswell, A. Abdul-Sada, C. Y. Hung, M. A. Simpson, W. K. Chong, T. S. Jacques, R. L. Woltjer, S. Eaton, A. Gregory, L. Sanford, E. Kara, H. Houlden, S. M. Cuno, H. Prokisch, L. Valletta, V. Tiranti, R. Younis, E. R. Maher, J. Spencer, A. Straatman-Iwanowska, P. Gissen, L. A. Selim, G. Pintos-Morell,

W. Coroleu-Lletget, S. S. Mohammad, S. Yoganathan, R. C. Dale, M. Thomas, J. Rihel, O. A. Bodamer, C. A. Enns, S. J. Hayflick, P. T. Clayton, P. B. Mills, M. A. Kurian, and S. W. Wilson, "Mutations in SLC39A14 disrupt manganese homeostasis and cause childhood-onset parkinsonism-dystonia," *Nat Commun*, vol. 7, p. 11601, 2016. [Online]. Available: https://www.ncbi.nlm.nih.gov/pubmed/27231142https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4894980/pdf/ncomms11601.pdf

[20] J. P. Liuzzi, L. A. Lichten, S. Rivera, R. K. Blanchard, T. Beker-Aydemir, M. D. Knutson, T. Ganz, and R. J. Cousins, "Interleukin-6 regulates the zinc transporter Zip14 in liver and contributes to the hypozincemia of the acute-phase response," *Proceedings of the National Academy of Sciences*, vol. 102, no. 19, pp. 6843–6848, 2005.

[21] T. B. Aydemir, S. M. Chang, G. J. Guthrie, A. B. Maki, M. S. Ryu, A. Karabiyik, and R. J. Cousins, "Zinc transporter ZIP14 functions in hepatic zinc, iron and glucose homeostasis during the innate immune response (endotoxemia)," *PLoS One*, vol. 7, no. 10, p. e48679, 2012.

[22] K. Thambiayya, K. Wasserloos, V. E. Kagan, D. Stoyanovsky, and B. R. Pitt, "A critical role for increased labile zinc in reducing sensitivity of cultured sheep pulmonary artery endothelial cells to LPS-induced apoptosis," *Am J Physiol Lung Cell Mol Physiol*, vol. 302, no. 12, pp. 1287–95, 2012. [Online]. Available: https://www.ncbi.nlm.nih.gov/pubmed/22523284

[23] A. Sayadi, A. T. Nguyen, F. A. Bard, and E. A. Bard-Chapeau, "Zip14 expression induced by lipopolysaccharides in macrophages attenuates inflammatory response," *Inflamm Res*, vol. 62, no. 2, pp. 133–143, 2013.

[24] H. Fujishiro, M. Yoshida, Y. Nakano, and S. Himeno, "Interleukin-6

enhances manganese accumulation in SH-SY5Y cells: implications of the up-regulation of ZIP14 and the down-regulation of ZnT10," *Metallomics*, vol. 6, no. 4, pp. 944–949, 2014.

[25] M. Galvez-Peralta, Z. Wang, S. Bao, D. L. Knoell, and D. W. Nebert, "Tissue-Specific Induction of Mouse ZIP8 and ZIP14 Divalent Cation/Bicarbonate Symporters by, and Cytokine Response to, Inflammatory Signals," *Int J Toxicol*, vol. 33, no. 3, pp. 246–258, 2014.

[26] K. Tominaga, T. Kagata, Y. Johmura, T. Hishida, M. Nishizuka, and M. Imagawa, "SLC39A14, a LZT protein, is induced in adipogenesis and transports zinc," *FEBS J*, vol. 272, no. 7, pp. 1590–1599, 2005. [Online]. Available: https://www.ncbi.nlm.nih.gov/pubmed/15794747

[27] Y. Xin, H. Gao, J. Wang, Y. Qiang, M. U. Imam, Y. Li, J. Wang, R. Zhang, H. Zhang, Y. Yu, H. Wang, H. Luo, C. Shi, Y. Xu, S. Hojyo, T. Fukada, J. Min, and F. Wang, "Manganese transporter Slc39a14 deficiency revealed its key role in maintaining manganese homeostasis in mice," *Cell Discov*, vol. 3, p. 17025, 2017. [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5519003/pdf/celldisc201725.pdf

[28] L. F. Jorge-Nebert, M. Galvez-Peralta, J. Landero Figueroa, M. Somarathna, S. Hojyo, T. Fukada, and D. W. Nebert, "Comparing gene expression during cadmium uptake and distribution: untreated versus oral Cd-treated wild-type and ZIP14 knockout mice," *Toxicol Sci*, vol. 143, no. 1, pp. 26–35, 2015.

[29] S. Hojyo, T. Fukada, S. Shimoda, W. Ohashi, B. H. Bin, H. Koseki, and T. Hirano, "The zinc transporter SLC39A14/ZIP14 controls G-protein coupled receptor-mediated signaling required for systemic growth," *PLoS One*, vol. 6, no. 3, p. e18059, 2011.

[30] T. B. Aydemir, C. Troche, J. Kim, M. H. Kim, O. Y. Teran, C. Leeuwenburgh, and R. J. Cousins, “Aging amplifies multiple phenotypic defects in mice with zinc transporter Zip14 (Slc39a14) deletion,” *Exp Gerontol*, vol. 85, pp. 88–94, 2016.

[31] T. B. Aydemir, C. Troche, M. H. Kim, and R. J. Cousins, “Hepatic ZIP14-mediated Zinc Transport Contributes to Endosomal Insulin Receptor Trafficking and Glucose Metabolism,” *J Biol Chem*, vol. 291, no. 46, pp. 23 939–23 951, 2016.

[32] G. J. Guthrie, T. B. Aydemir, C. Troche, A. B. Martin, S. M. Chang, and R. J. Cousins, “Influence of ZIP14 (slc39A14) on intestinal zinc processing and barrier function,” *Am J Physiol Gastrointest Liver Physiol*, vol. 308, no. 3, pp. 171–8, 2015.

[33] C. Troche, T. B. Aydemir, and R. J. Cousins, “Zinc transporter Slc39a14 regulates inflammatory signaling associated with hypertrophic adiposity,” *American journal of physiology. Endocrinology and metabolism*, vol. 310, no. 4, pp. E258–E268, 2 2016. [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/26646099https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4971811/

[34] J. P. Liuzzi, F. Aydemir, H. Nam, M. D. Knutson, and R. J. Cousins, “Zip14 (Slc39a14) mediates non-transferrin-bound iron uptake into cells,” *Proceedings of the National Academy of Sciences*, vol. 103, no. 37, pp. 13 612–13 617, 2006.

[35] J. J. Pinilla-Tenas, B. K. Sparkman, A. Shawki, A. C. Illing, C. J. Mitchell, N. Zhao, J. P. Liuzzi, R. J. Cousins, M. D. Knutson, and B. Mackenzie, “Zip14 is a complex broad-scope metal-ion transporter whose functional

properties support roles in the cellular uptake of zinc and nontransferrin-bound iron," *Am J Physiol Cell Physiol*, vol. 301, no. 4, pp. 862–71, 2011.

[36] D. W. Nebert, M. Galvez-Peralta, E. B. Hay, H. Li, E. Johansson, C. Yin, B. Wang, L. He, and M. Soleimani, "ZIP14 and ZIP8 zinc/bicarbonate symporters in Xenopus oocytes: characterization of metal uptake and inhibition," *Metallomics*, vol. 4, no. 11, pp. 1218–1225, 2012.

[37] Z. I. Cabantchik and R. Greger, "Chemical probes for anion transporters of mammalian cell membranes," *American Journal of Physiology-Cell Physiology*, vol. 262, no. 4, pp. C803–C827, 1992.

[38] S. Bröer, "Xenopus laevis Oocytes," *Methods in molecular biology (Clifton, N.J.)*, vol. 637, 2010.

[39] D. G. Barceloux, "Manganese," pp. 293–307, 1999.

[40] D. Rahelić, M. Kujundžić, Romić, K. Brkić, and M. Petrovečki, "Serum concentration of zinc, copper, manganese and magnesium in patients with liver cirrhosis," *Collegium Antropologicum*, vol. 30, no. 3, 2006.

[41] E. M. G. Jacobs, J. C. M. Hendriks, B. L. J. H. van Tits, P. J. Evans, W. Breuer, D. Y. Liu, E. H. J. M. Jansen, K. Jauhiainen, B. Sturm, J. B. Porter, B. Scheiber-Mojdehkar, L. von Bonsdorff, Z. I. Cabantchik, R. C. Hider, and D. W. Swinkels, "Results of an international round robin for the quantification of serum non-transferrin-bound iron: Need for defining standardization and a clinically relevant isoform," *Analytical Biochemistry*, vol. 341, no. 2, pp. 241–250, 2005.

[42] M. Saaranen, M. Kantola, S. Saarikoski, and T. Vanha-Perttula, "Human seminal plasma cadmium: comparison with fertility and smoking habits," *Andrologia*, vol. 21, no. 2, 1989.

[43] T. L. Wright, P. Brissot, W. L. Ma, and R. A. Weisiger, “Characterization of non-transferrin-bound iron clearance by rat liver,” *J Biol Chem*, vol. 261, no. 23, pp. 10 909–10 914, 1986.

[44] M. Valko, K. Jomova, C. J. Rhodes, K. Kuca, and K. Musilek, “Redox- and non-redox-metal-induced formation of free radicals and their role in human disease,” *Arch Toxicol*, vol. 90, no. 1, pp. 1–37, 2016.

[45] S. Jenkitkasemwong, C. Y. Wang, R. Coffey, W. Zhang, A. Chan, T. Biel, J. S. Kim, S. Hojyo, T. Fukada, and M. D. Knutson, “SLC39A14 Is Required for the Development of Hepatocellular Iron Overload in Murine Models of Hereditary Hemochromatosis,” *Cell Metab*, vol. 22, no. 1, pp. 138–150, 2015.

[46] R. Coffey and M. D. Knutson, “The plasma membrane metal-ion transporter ZIP14 contributes to nontransferrin-bound iron uptake by human β-cells,” *American Journal of Physiology - Cell Physiology*, vol. 312, no. 2, pp. C169–C175, 2017.

[47] A. K. Tripathi, S. Haldar, J. Qian, A. Beserra, S. Suda, A. Singh, U. Hopfer, S. G. Chen, M. D. Garrick, J. R. Turner, M. D. Knutson, and N. Singh, “Prion protein functions as a ferrireductase partner for ZIP14 and DMT1,” *Free Radic Biol Med*, vol. 84, pp. 322–330, 2015. [Online]. Available: https://www.ncbi.nlm.nih.gov/pubmed/25862412

[48] R. S. Ohgami, D. R. Campagna, E. L. Greer, B. Antiochos, A. McDonald, J. Chen, J. J. Sharp, Y. Fujiwara, J. E. Barker, and M. D. Fleming, “Identification of a ferrireductase required for efficient transferrin-dependent iron uptake in erythroid cells,” *Nature Genetics*, vol. 37, no. 11, 2005.

[49] R. S. Ohgami, D. R. Campagna, A. McDonald, and M. D. Fleming, “The Steap proteins are metalloreductases,” *Blood*, vol. 108, no. 4, 2006.

[50] R. Daher and Z. Karim, “Iron metabolism: State of the art,” *Transfus Clin Biol*, vol. 24, no. 3, pp. 115–119, 2017.

[51] J. Huotari and A. Helenius, “Endosome maturation,” *The EMBO Journal*, vol. 30, no. 17, pp. 3481–3500, 2011.

[52] H. Gunshin, B. Mackenzie, U. V. Berger, Y. Gunshin, M. F. Romero, W. F. Boron, S. Nussberger, J. L. Gollan, and M. A. Hediger, “Cloning and characterization of a mammalian proton-coupled metal-ion transporter,” *Nature*, vol. 388, p. 482, 1997.

[53] M. T. Worthington, L. Browne, E. H. Battle, and R. Q. Luo, “Functional properties of transfected human DMT1 iron transporter,” *American Journal of Physiology-Gastrointestinal and Liver Physiology*, vol. 279, no. 6, pp. G1265–G1273, 2000. [Online]. Available: https://doi.org/10.1152/ajpgi.2000.279.6.G1265

[54] C. Y. Wang and M. D. Knutson, “Hepatocyte divalent metal-ion transporter-1 is dispensable for hepatic iron accumulation and non-transferrin-bound iron uptake in mice,” *Hepatology*, vol. 58, no. 2, pp. 788–798, 2013.

[55] N. Zhao, A. S. Zhang, C. Worthen, M. D. Knutson, and C. A. Enns, “An iron-regulated and glycosylation-dependent proteasomal degradation pathway for the plasma membrane metal transporter ZIP14,” *Proc Natl Acad Sci U S A*, vol. 111, no. 25, pp. 9175–9180, 2014. [Online]. Available: https://www.ncbi.nlm.nih.gov/pubmed/24927598

[56] C. Y. Wang, S. Jenkitkasemwong, S. Duarte, B. K. Sparkman, A. Shawki, B. Mackenzie, and M. D. Knutson, "ZIP8 is an iron and zinc transporter whose cell-surface expression is up-regulated by cellular iron loading," *J Biol Chem*, vol. 287, no. 41, pp. 34 032–34 043, 2012. [Online]. Available: https://www.ncbi.nlm.nih.gov/pubmed/22898811

[57] J. Gao, N. Zhao, M. D. Knutson, and C. A. Enns, "The hereditary hemochromatosis protein, HFE, inhibits iron uptake via down-regulation of Zip14 in HepG2 cells," *J Biol Chem*, vol. 283, no. 31, pp. 21 462–21 468, 2008.

[58] H. Coppin, V. Darnaud, L. Kautz, D. Meynard, M. Aubry, J. Mosser, M. Martinez, and M.-P. Roth, "Gene expression profiling of Hfe-/-liver and duodenum in mouse strains with differing susceptibilities to iron loading: identification of transcriptional regulatory targets of Hfe and potential hemochromatosis modifiers," *Genome Biology*, vol. 8, no. 10, p. R221, 2007.

[59] M. Lunova, P. Schwarz, R. Nuraldeen, K. Levada, D. Kuscuoglu, M. Stutzle, M. Vujic Spasic, J. Haybaeck, P. Ruchala, M. Jirsa, J. C. Deschemin, S. Vaulont, C. Trautwein, and P. Strnad, "Hepcidin knockout mice spontaneously develop chronic pancreatitis owing to cytoplasmic iron overload in acinar cells," *J Pathol*, vol. 241, no. 1, pp. 104–114, 2017.

[60] M. Aschner, K. M. Erikson, E. Herrero Hernandez, and R. Tjalkens, "Manganese and its role in Parkinson's disease: from transport to neuropathology," *Neuromolecular Med*, vol. 11, no. 4, pp. 252–266, 2009.

[61] C. Andreini, I. Bertini, G. Cavallaro, G. L. Holliday, and J. M. Thornton, "Metal ions in biological catalysis: from enzyme databases to general principles," *J Biol Inorg Chem*, vol. 13, no. 8, pp. 1205–1218, 2008.

[62] J. W. Finley, P. E. Johnson, and L. A. K. Johnson, “Sex affects manganese absorption and retention by humans from a diet adequate in manganese,” *American Journal of Clinical Nutrition*, vol. 60, no. 6, 1994.

[63] J. W. Finley, “Manganese absorption and retention by young women is associated with serum ferritin concentration,” *Am J Clin Nutr*, vol. 70, no. 1, pp. 37–43, 1999.

[64] J. M. Davis, “Inhalation health risks of manganese: an EPA perspective,” *Neurotoxicology*, vol. 20, no. 2-3, pp. 511–518, 1999.

[65] C. Au, A. Benedetto, and M. Aschner, “Manganese transport in eukaryotes: the role of DMT1,” *Neurotoxicology*, vol. 29, no. 4, pp. 569–576, 2008.

[66] T. E. Kehl-Fie and E. P. Skaar, “Nutritional immunity beyond iron: a role for manganese and zinc,” *Curr Opin Chem Biol*, vol. 14, no. 2, pp. 218–224, 2010.

[67] M. Bianchi, M. J. Niemiec, U. Siler, C. F. Urban, and J. Reichenbach, “Restoration of anti-Aspergillus defense by neutrophil extracellular traps in human chronic granulomatous disease after gene therapy is calprotectin-dependent,” *J Allergy Clin Immunol*, vol. 127, no. 5, pp. 1243–52, 2011.

[68] S. M. Damo, T. E. Kehl-Fie, N. Sugitani, M. E. Holt, S. Rathi, W. J. Murphy, Y. Zhang, C. Betz, L. Hench, G. Fritz, E. P. Skaar, and W. J. Chazin, “Molecular basis for manganese sequestration by calprotectin and roles in the innate immune response to invading bacterial pathogens,” *Proc Natl Acad Sci U S A*, vol. 110, no. 10, pp. 3841–3846, 2013.

[69] M. Juneja, U. Shamim, A. Joshi, A. Mathur, B. Uppili, S. Sairam, S. Ambawat, R. Dixit, and M. Faruq, “A novel mutation in SLC39A14

causing hypermanganesemia associated with infantile onset dystonia," *J Gene Med*, vol. 20, no. 4, p. e3012, 2018.

[70] L. H. Rodan, M. Hauptman, A. M. D'Gama, A. E. Qualls, S. Cao, K. Tuschl, F. Al-Jasmi, J. Hertecant, S. J. Hayflick, M. Wessling-Resnick, E. T. Yang, G. T. Berry, A. Gropman, A. D. Woolf, and P. B. Agrawal, "Novel founder intronic variant in SLC39A14 in two families causing Manganism and potential treatment strategies," *Mol Genet Metab*, vol. 124, no. 2, pp. 161–167, 2018.

[71] A. Zeglam, A. Abugrara, and M. Kabuka, "Autosomal-recessive iron deficiency anemia, dystonia and hypermanganesemia caused by new variant mutation of the manganese transporter gene SLC39A14," *Acta Neurol Belg*, 2018. [Online]. Available: https://link.springer.com/content/pdf/10.1007%2Fs13760-018-1024-7.pdf

[72] L. Marti-Sanchez, J. D. Ortigoza-Escobar, A. Darling, M. Villaronga, H. Baide, M. Molero-Luis, M. Batllori, M. I. Vanegas, J. Muchart, L. Aquino, R. Artuch, A. Macaya, M. A. Kurian, and P. Duenas, "Hypermanganesemia due to mutations in SLC39A14: further insights into Mn deposition in the central nervous system," *Orphanet J Rare Dis*, vol. 13, no. 1, p. 28, 2018.

[73] D. Garg, S. Yoganathan, U. Shamim, K. Mankad, P. Gulati, V. Bonifati, A. Botre, U. Kalane, A. G. Saini, N. Sankhyan, K. Srivastava, V. K. Gowda, M. Juneja, M. Kamate, H. Padmanabha, D. Panigrahi, S. Pachapure, V. Udani, A. Kumar, S. Pandey, M. Thomas, S. Danda, S. A. Iqbalahmed, A. Subramanian, H. Pemde, V. Singh, M. Faruq, and S. Sharma, "Clinical Profile and Treatment Outcomes of Hypermanganesemia with Dystonia 1 and 2 among 27 Indian Children," *Movement Disorders*

Clinical Practice, vol. n/a, no. n/a, 7 2022. [Online]. Available: https://doi.org/10.1002/mdc3.13516

[74] C. K. Fung and N. Zhao, "The Combined Inactivation of Intestinal and Hepatic ZIP14 Exacerbates Manganese Overload in Mice," *International Journal of Molecular Sciences*, 2022. [Online]. Available: http://hdl.handle.net/10150/665501

[75] I. F. Scheiber, Y. Wu, S. E. Morgan, and N. Zhao, "The intestinal metal transporter ZIP14 maintains systemic manganese homeostasis," *Journal of Biological Chemistry*, vol. 294, no. 23, pp. 9147–9160, 2019.

[76] T. B. Aydemir, T. L. Thorn, C. H. Ruggiero, M. Pompilus, M. Febo, and R. J. Cousins, "Intestine-specific deletion of metal transporter Zip14 (Slc39a14) causes brain manganese overload and locomotor defects of manganism," *American Journal of Physiology - Gastrointestinal and Liver Physiology*, vol. 318, no. 4, 2020.

[77] K. J. Thompson and M. Wessling-Resnick, "ZIP14 is degraded in response to manganese exposure," *Biometals*, 2019. [Online]. Available: https://doi.org/10.1007/s10534-019-00216-1https://link.springer.com/content/pdf/10.1007%2Fs10534-019-00216-1.pdf

[78] L. Liu, S. Oza, D. Hogan, J. Perin, I. Rudan, J. E. Lawn, S. Cousens, C. Mathers, and R. E. Black, "Global, regional, and national causes of child mortality in 2000-13, with projections to inform post-2015 priorities: an updated systematic analysis," *Lancet*, vol. 385, no. 9966, pp. 430–440, 2015.

[79] R. S. MacDonald, "The role of zinc in growth and cell proliferation," *J Nutr*, vol. 130, no. 5S Suppl, pp. 1500s–8s, 2000.

[80] B. L. Vallee and K. H. Falchuk, “The biochemical basis of zinc physiology,” *Physiological Reviews*, vol. 73, no. 1, pp. 79–118, 1993. [Online]. Available: http://www.physiology.org/doi/abs/10.1152/physrev.1993.73.1.79

[81] K. Wang, B. Zhou, Y.-M. Kuo, J. Zemansky, and J. Gitschier, “A Novel Member of a Zinc Transporter Family Is Defective in Acrodermatitis Enteropathica,” *The American Journal of Human Genetics*, vol. 71, no. 1, pp. 66–73, 2002. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0002929707600354https://www.ncbi.nlm.nih.gov/pmc/articles/PMC419995/pdf/AJHGv71p66.pdf

[82] D. Leupold, J. R. Poley, and W. N. Meigel, “Zinc therapy in acrodermatitis enteropathica,” *Helv Paediatr Acta*, vol. 31, no. 2, pp. 109–115, 1976.

[83] J. C. King, D. M. Shames, and L. R. Woodhouse, “Zinc Homeostasis in Humans,” *The Journal of Nutrition*, vol. 130, no. 5, pp. 1360S–1366S, 2000.

[84] J. P. Liuzzi, J. A. Bobo, L. A. Lichten, D. A. Samuelson, and R. J. Cousins, “Responsive transporter genes within the murine intestinal-pancreatic axis form a basis of zinc homeostasis,” *Proceedings of the National Academy of Sciences of the United States of America*, vol. 101, no. 40, pp. 14 355–14 360, 2004.

[85] F. Wang, B.-E. Kim, M. J. Petris, and D. J. Eide, “The Mammalian Zip5 Protein Is a Zinc Transporter That Localizes to the Basolateral Surface of Polarized Cells,” *Journal of Biological Chemistry*, vol. 279, no. 49, pp. 51 433–51 441, 2004. [Online]. Available: http://www.jbc.org/content/279/49/51433.abstract

[86] N. Thyresson, “Acrodermatitis enteropathica. Report of a case healed with zinc therapy,” *Acta Derm Venereol*, vol. 54, no. 5, pp. 383–385, 1974.

[87] R. L. Bailey, K. P. West Jr, and R. E. Black, “The Epidemiology of Global Micronutrient Deficiencies,” *Annals of Nutrition and Metabolism*, vol. 66(suppl 2), no. Suppl. 2, pp. 22–33, 2015.

[88] E. Mayo-Wilson, J. A. Junior, A. Imdad, S. Dean, X. H. Chan, E. S. Chan, A. Jaswal, and Z. A. Bhutta, “Zinc supplementation for preventing mortality, morbidity, and growth failure in children aged 6 months to 12 years of age,” *Cochrane Database Syst Rev*, no. 5, p. Cd009384, 2014.

[89] M. Y. Yakoob, E. Theodoratou, A. Jabeen, A. Imdad, T. P. Eisele, J. Ferguson, A. Jhass, I. Rudan, H. Campbell, R. E. Black, and Z. A. Bhutta, “Preventive zinc supplementation in developing countries: impact on mortality and morbidity due to diarrhea, pneumonia and malaria,” *BMC Public Health*, vol. 11 Suppl 3, p. S23, 2011.

[90] H. Merza, N. Sood, and R. Sood, “Idiopathic hyperzincemia with associated copper deficiency anemia: a diagnostic dilemma,” *Clinical Case Reports*, vol. 3, no. 10, pp. 819–822, 2015.

[91] E. D. Gallery, J. Blomfield, and S. R. Dixon, “Acute zinc toxicity in haemodialysis,” *Br Med J*, vol. 4, no. 5836, pp. 331–333, 1972.

[92] T. Ogiso, N. Ogawa, and T. Miura, “Inhibitory Effect of High Dietary Zinc on Copper Absorption in Rats. II. Binding of Copper and Zinc to Cytosol Proteins in the Intestinal Mucosa,” *CHEMICAL & PHARMACEUTICAL BULLETIN*, vol. 27, no. 2, pp. 515–521, 1979.

[93] S. Fessatou, M. K. Fagerhol, J. Roth, A. Stamoulakatou, V. Kitra, M. Hadarean, G. Paleologos, H. Chandrinou, B. Sampson, and

I. Papassotiriou, “Severe anemia and neutropenia associated with hyperzincemia and hypercalprotectinemia,” *J Pediatr Hematol Oncol*, vol. 27, no. 9, pp. 477–480, 2005.

[94] B. Sampson, I. Z. Kovar, A. Rauscher, S. Fairweather-Tait, J. Beattie, H. J Mcardle, R. Ahmed, and C. Green, “A Case of Hyperzincemia with Functional Zinc Depletion: A New Disorder?” *Pediatric Research*, vol. 42, p. 219, 1997.

[95] G. Hendrickx, V. M. Borra, E. Steenackers, T. A. Yorgan, C. Hermans, E. Boudin, J. J. Waterval, I. D. C. Jansen, T. B. Aydemir, N. Kamerling, G. J. Behets, C. Plumeyer, P. C. D’Haese, B. Busse, V. Everts, M. Lammens, G. Mortier, R. J. Cousins, T. Schinke, R. J. Stokroos, J. J. Manni, and W. Van Hul, “Conditional mouse models support the role of SLC39A14 (ZIP14) in Hyperostosis Cranialis Interna and in bone homeostasis,” *PLoS Genet*, vol. 14, no. 4, p. e1007321, 2018.

[96] P. Lichtlen and W. Schaffner, “The ”metal transcription factor” MTF-1: biological facts and medical implications,” *Swiss Med Wkly*, vol. 131, no. 45-46, pp. 647–652, 2001.

[97] J. C. Philcox, P. Coyle, A. Michalska, K. H. Choo, and A. M. Rofe, “Endotoxin-induced inflammation does not cause hepatic zinc accumulation in mice lacking metallothionein gene expression,” *Biochem J*, vol. 308 (Pt 2), pp. 543–546, 1995.

[98] A. K. Baltaci, K. Yuce, and R. Mogulkoc, “Zinc Metabolism and Metallothioneins,” 2018.

[99] S. Jenkitkasemwong, C. Y. Wang, B. Mackenzie, and M. D. Knutson,

"Physiologic implications of metal-ion transport by ZIP14 and ZIP8," *Biometals*, vol. 25, no. 4, pp. 643–655, 2012.

[100] M. J. Liu, S. Bao, M. Galvez-Peralta, C. J. Pyle, A. C. Rudawsky, R. E. Pavlovicz, D. W. Killilea, C. Li, D. W. Nebert, M. D. Wewers, and D. L. Knoell, "ZIP8 regulates host defense through zinc-mediated inhibition of NF-kappaB," *Cell Rep*, vol. 3, no. 2, pp. 386–400, 2013.

[101] T. P. Dalton, M. L. Miller, X. Wu, A. Menon, E. Cianciolo, R. A. McKinnon, P. W. Smith, L. J. Robinson, and D. W. Nebert, "Refining the mouse chromosomal location of Cdm, the major gene associated with susceptibility to cadmium-induced testicular necrosis," *Pharmacogenetics*, vol. 10, no. 2, pp. 141–151, 2000.

[102] T. P. Dalton, L. He, B. Wang, M. L. Miller, L. Jin, K. F. Stringer, X. Chang, C. S. Baxter, and D. W. Nebert, "Identification of mouse SLC39A8 as the transporter responsible for cadmium-induced toxicity in the testis," *Proc Natl Acad Sci U S A*, vol. 102, no. 9, pp. 3401–3406, 2005.

[103] T. B. Aydemir, J. P. Liuzzi, S. McClellan, and R. J. Cousins, "Zinc transporter ZIP8 (SLC39A8) and zinc influence IFN-gamma expression in activated human T cells," *J Leukoc Biol*, vol. 86, no. 2, pp. 337–348, 2009.

[104] B. Y. Besecker, M. C. Exline, J. Hollyfield, G. Phillips, R. A. Disilvestro, M. D. Wewers, and D. L. Knoell, "A comparison of zinc metabolism, inflammation, and disease severity in critically ill infected and noninfected adults early after intensive care unit admission," *Am J Clin Nutr*, vol. 93, no. 6, pp. 1356–1364, 2011.

[105] N. A. Begum, M. Kobayashi, Y. Moriwaki, M. Matsumoto, K. Toyoshima,

and T. Seya, “Mycobacterium bovis BCG cell wall and lipopolysaccharide induce a novel gene, BIGM103, encoding a 7-TM protein: identification of a new protein family having Zn-transporter and Zn-metalloprotease signatures,” *Genomics*, vol. 80, no. 6, pp. 630–645, 2002.

[106] R. J. Cousins, J. P. Liuzzi, and L. A. Lichten, “Mammalian zinc transport, trafficking, and signals,” *J Biol Chem*, vol. 281, no. 34, pp. 24 085–24 089, 2006.

[107] X. Mao, B. E. Kim, F. Wang, D. J. Eide, and M. J. Petris, “A histidine-rich cluster mediates the ubiquitination and degradation of the human zinc transporter, hZIP4, and protects against zinc cytotoxicity,” *J Biol Chem*, vol. 282, no. 10, pp. 6992–7000, 2007.

[108] R. J. Cousins, R. K. Blanchard, M. P. Popp, L. Liu, J. Cao, J. B. Moore, and C. L. Green, “A global view of the selectivity of zinc deprivation and excess on genes expressed in human THP-1 mononuclear cells,” *Proceedings of the National Academy of Sciences of the United States of America*, vol. 100, no. 12, 2003.

[109] D. L. Manning, R. A. McClelland, J. M. Knowlden, S. Bryant, J. M. W. Gee, C. D. Green, J. F. Robertson, R. W. Blamey, R. L. Sutherland, C. J. Ormandy, and R. I. Nicholson, “Differential expression of oestrogen regulated genes in breast cancer,” *Acta Oncologica*, vol. 34, no. 5, pp. 641–646, 1995.

[110] K. M. Mayle, A. M. Le, and D. T. Kamei, “The intracellular trafficking pathway of transferrin,” 2012.

[111] E. Macia, M. Ehrlich, R. Massol, E. Boucrot, C. Brunner, and

T. Kirchhausen, “Dynasore, a Cell-Permeable Inhibitor of Dynamin,” *Developmental Cell*, vol. 10, no. 6, 2006.

[112] E. Cocucci, R. Gaudin, and T. Kirchhausen, “Dynamin recruitment and membrane scission at the neck of a clathrin-coated pit,” *Molecular Biology of the Cell*, vol. 25, no. 22, 2014.

[113] J. J. Thottacherry, M. Sathe, C. Prabhakara, and S. Mayor, “Spoiled for choice: Diverse endocytic pathways function at the cell surface,” 2019.

[114] G. Preta, J. G. Cronin, and I. M. Sheldon, “Dynasore - not just a dynamin inhibitor,” *Cell Communication and Signaling*, vol. 13, no. 1, p. 24, 2015.

[115] L. P. Clemente, M. Rabenau, S. Tang, J. Stanka, E. Cors, J. Stroh, C. Culmsee, and S. von Karstedt, “Dynasore Blocks Ferroptosis through Combined Modulation of Iron Uptake and Inhibition of Mitochondrial Respiration,” *Cells*, vol. 9, no. 10, 2020.

[116] J. Luzio, M. J. Parkinson, S. Gray, and N. Bright, “The delivery of endocytosed cargo to lysosomes,” *Biochemical Society Transactions*, vol. 37, no. 5, p. 1019, 2009.

[117] E. J. Bowman, A. Siebers, and K. Altendorf, “Bafilomycins; A class of inhibitors of membrane ATPases from microorganisms, animal cells, and plant cells,” *Proceedings of the National Academy of Sciences of the United States of America*, vol. 85, no. 21, 1988.

[118] T. Yoshimori, A. Yamamoto, Y. Moriyama, M. Futai, and Y. Tashiro, “Bafilomycin A1, a specific inhibitor of vacuolar-type H+-ATPase, inhibits acidification and protein degradation in lysosomes of cultured cells,” *Journal of Biological Chemistry*, vol. 266, no. 26, 1991.

[119] T. Asano, M. Komatsu, Y. Yamaguchi-Iwai, F. Ishikawa, N. Mizushima, and K. Iwai, “Distinct Mechanisms of Ferritin Delivery to Lysosomes in Iron-Depleted and Iron-Replete Cells,” *Molecular and Cellular Biology*, vol. 31, no. 10, 2011.

[120] D. H. Lee, “Proteasome inhibitors: Valuable new tools for cell biologists,” 1998.

[121] L. Meng, R. Mohan, B. H. Kwok, M. Elofsson, N. Sin, and C. M. Crews, “Epoxomicin, a potent and selective proteasome inhibitor, exhibits in vivo antiinflammatory activity,” *Proceedings of the National Academy of Sciences of the United States of America*, vol. 96, no. 18, 1999.

[122] J. Hanna, A. Guerra-Moreno, J. Ang, and Y. Micoogullari, “Protein Degradation and the Pathologic Basis of Disease,” 2019.

[123] R. R. Paudel, D. Lu, S. Roy Chowdhury, E. Y. Monroy, and J. Wang, “Targeted Protein Degradation via Lysosomes,” *Biochemistry*, 9 2022. [Online]. Available: https://doi.org/10.1021/acs.biochem.2c00310

[124] M. Schneider, K. Ackermann, M. Stuart, C. Wex, U. Protzer, H. M. Schätzl, and S. Gilch, “Severe Acute Respiratory Syndrome Coronavirus Replication Is Severely Impaired by MG132 due to Proteasome-Independent Inhibition of M-Calpain,” *Journal of Virology*, vol. 86, no. 18, 2012.

[125] A. F. Kisselev, W. A. Van Der Linden, and H. S. Overkleeft, “Proteasome inhibitors: An expanding army attacking a unique target,” 2012.

[126] F. Steinberg, M. Gallon, M. Winfield, E. C. Thomas, A. J. Bell, K. J. Heesom, J. M. Tavare, and P. J. Cullen, “A global analysis of SNX27-retromer assembly and cargo specificity reveals a function in glucose

and metal ion transport," *Nat Cell Biol*, vol. 15, no. 5, pp. 461–471, 2013. [Online]. Available: https://www.ncbi.nlm.nih.gov/pubmed/23563491

[127] K. E. McNally, R. Faulkner, F. Steinberg, M. Gallon, R. Ghai, D. Pim, P. Langton, N. Pearson, C. M. Danson, H. Nagele, L. L. Morris, A. Singla, B. L. Overlee, K. J. Heesom, R. Sessions, L. Banks, B. M. Collins, I. Berger, D. D. Billadeau, E. Burstein, and P. J. Cullen, "Retriever is a multiprotein complex for retromer-independent endosomal cargo recycling," *Nat Cell Biol*, vol. 19, no. 10, pp. 1214–1225, 2017.

[128] M. Chandra, A. K. Kendall, and L. P. Jackson, "Toward Understanding the Molecular Role of SNX27/Retromer in Human Health and Disease," 2021.

[129] P. Temkin, B. Lauffer, S. Jager, P. Cimermancic, N. J. Krogan, and M. von Zastrow, "SNX27 mediates retromer tubule entry and endosome-to-plasma membrane trafficking of signaling receptors," *Nature cell biology*, vol. 13, no. 6, pp. 715–721, 2011. [Online]. Available: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3113693/

[130] F. Steinberg, K. J. Heesom, M. D. Bass, and P. J. Cullen, "SNX17 protects integrins from degradation by sorting between lysosomal and recycling pathways," *J Cell Biol*, vol. 197, no. 2, pp. 219–230, 2012.

[131] K. E. Chen, M. D. Healy, and B. M. Collins, "Towards a molecular understanding of endosomal trafficking by Retromer and Retriever," 2019.

[132] R. Ghai, A. Bugarcic, H. Liu, S. J. Norwood, S. Skeldal, E. J. Coulson, S. S. Li, R. D. Teasdale, and B. M. Collins, "Structural basis for endosomal trafficking of diverse transmembrane cargos by PX-FERM proteins," *Proc Natl Acad Sci U S A*, vol. 110, no. 8, pp. 643–52, 2013.

[133] T. Clairfeuille, C. Mas, A. S. Chan, Z. Yang, M. Tello-Lafoz, M. Chandra, J. Widagdo, M. C. Kerr, B. Paul, I. Merida, R. D. Teasdale, N. J. Pavlos, V. Anggono, and B. M. Collins, "A molecular code for endosomal recycling of phosphorylated cargos by the SNX27-retromer complex," *Nat Struct Mol Biol*, vol. 23, no. 10, pp. 921–932, 2016.

[134] H. Haase and W. Maret, "Fluctuations of cellular, available zinc modulate insulin signaling via inhibition of protein tyrosine phosphatases," *J Trace Elem Med Biol*, vol. 19, no. 1, pp. 37–42, 2005.

[135] ——, "Intracellular zinc fluctuations modulate protein tyrosine phosphatase activity in insulin/insulin-like growth factor-1 signaling," *Exp Cell Res*, vol. 291, no. 2, pp. 289–298, 2003.

[136] T. B. Aydemir, H. S. Sitren, and R. J. Cousins, "The zinc transporter Zip14 influences c-Met phosphorylation and hepatocyte proliferation during liver regeneration in mice," *Gastroenterology*, vol. 142, no. 7, pp. 1536–46, 2012.

[137] M.-H. Kim, T. B. Aydemir, J. Kim, and R. J. Cousins, "Hepatic ZIP14-mediated zinc transport is required for adaptation to endoplasmic reticulum stress," *Proceedings of the National Academy of Sciences*, 2017.

[138] H. Haase, J. L. Ober-Blobaum, G. Engelhardt, S. Hebel, A. Heit, H. Heine, and L. Rink, "Zinc signals are essential for lipopolysaccharide-induced signal transduction in monocytes," *J Immunol*, vol. 181, no. 9, pp. 6491–6502, 2008.

[139] S. Zhang and Z.-Y. Zhang, "PTP1B as a drug target: recent developments in PTP1B inhibitor discovery," *Drug Discovery Today*,

vol. 12, no. 9, pp. 373–381, 2007. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1359644607001365

[140] A. Barthel, E. A. Ostrakhovitch, P. L. Walter, A. Kampkotter, and L. O. Klotz, "Stimulation of phosphoinositide 3-kinase/Akt signaling by copper and zinc ions: mechanisms and consequences," *Arch Biochem Biophys*, vol. 463, no. 2, pp. 175–182, 2007.

[141] X. Tang and N. F. Shay, "Zinc has an insulin-like effect on glucose transport mediated by phosphoinositol-3-kinase and Akt in 3T3-L1 fibroblasts and adipocytes," *J Nutr*, vol. 131, no. 5, pp. 1414–1420, 2001.

[142] T. Maxel, K. Smidt, A. Larsen, M. Bennetzen, K. Cullberg, K. Fjeldborg, S. Lund, S. B. Pedersen, and J. Rungby, "Gene expression of the zinc transporter ZIP14 (SLC39a14) is affected by weight loss and metabolic status and associates with PPARgamma in human adipose tissue and 3T3-L1 pre-adipocytes," *BMC Obesity*, vol. 2, p. 46, 2015.

[143] T. Maxel, P. F. Svendsen, K. Smidt, J. K. Lauridsen, B. Brock, S. B. Pedersen, J. Rungby, and A. Larsen, "Expression Patterns and Correlations with Metabolic Markers of Zinc Transporters ZIP14 and ZNT1 in Obesity and Polycystic Ovary Syndrome," *Front Endocrinol (Lausanne)*, vol. 8, p. 38, 2017.

[144] S. F. Schmidt, M. Rohm, S. Herzig, and M. Berriel Diaz, "Cancer Cachexia: More Than Skeletal Muscle Wasting," 2018.

[145] G. Wang, A. K. Biswas, W. Ma, M. Kandpal, C. Coker, P. M. Grandgenett, M. A. Hollingsworth, R. Jain, K. Tanji, S. López-Pintado, A. Borczuk, D. Hebert, S. Jenkitkasemwong, S. Hojyo, R. V. Davuluri, M. D. Knutson, T. Fukada, and S. Acharyya, "Metastatic cancers promote cachexia

through ZIP14 upregulation in skeletal muscle," *Nature Medicine*, vol. 24, no. 6, 2018.

[146] A. R. Shakri, T. J. Zhong, W. Ma, C. Coker, S. Kim, S. Calluori, H. Scholze, M. Szabolcs, T. Caffrey, P. M. Grandgenett, M. A. Hollingsworth, K. Tanji, M. D. Kluger, G. Miller, A. K. Biswas, and S. Acharyya, "Upregulation of ZIP14 and altered zinc homeostasis in muscles in pancreatic cancer cachexia," *Cancers*, vol. 12, no. 1, 2020.

[147] N. Ebner, S. D. Anker, and S. von Haehling, "Recent developments in the field of cachexia, sarcopenia, and muscle wasting: highlights from the 11th Cachexia Conference," in *Journal of Cachexia, Sarcopenia and Muscle*, vol. 10, no. 1, 2019.

[148] R. Medzhitov, "Recognition of microorganisms and activation of the immune response," *Nature*, vol. 449, p. 819, 2007.

[149] F. Y. Liew, D. Xu, E. K. Brint, and L. A. O'Neill, "Negative regulation of toll-like receptor-mediated immune responses," *Nat Rev Immunol*, vol. 5, no. 6, pp. 446–458, 2005.

[150] C. Alexander and E. T. Rietschel, "Bacterial lipopolysaccharides and innate immunity," *J Endotoxin Res*, vol. 7, no. 3, pp. 167–202, 2001.

[151] L. A. Lichten, J. P. Liuzzi, and R. J. Cousins, "Interleukin-1beta contributes via nitric oxide to the upregulation and functional activity of the zinc transporter Zip14 (Slc39a14) in murine hepatocytes," *Am J Physiol Gastrointest Liver Physiol*, vol. 296, no. 4, pp. 860–7, 2009.

[152] H. Haase and L. Rink, "Multiple impacts of zinc on immune function," *Metallomics*, vol. 6, no. 7, pp. 1175–1180, 2014.

[153] R. Hallgren, N. Feltelius, and U. Lindh, “Redistribution of minerals and trace elements in chronic inflammation–a study on isolated blood cells from patients with ankylosing spondylitis,” *J Rheumatol*, vol. 14, no. 3, pp. 548–553, 1987.

[154] K. Subramanian Vignesh, J. A. Landero Figueroa, A. Porollo, J. A. Caruso, and G. S. Deepe Jr., “Granulocyte macrophage-colony stimulating factor induced Zn sequestration enhances macrophage superoxide and limits intracellular pathogen survival,” *Immunity*, vol. 39, no. 4, pp. 697–710, 2013.

[155] H. Botella, G. Stadthagen, G. Lugo-Villarino, C. de Chastellier, and O. Neyrolles, “Metallobiology of host-pathogen interactions: an intoxicating new insight,” *Trends Microbiol*, vol. 20, no. 3, pp. 106–112, 2012.

[156] P. Lee, H. Peng, T. Gelbart, and E. Beutler, “The IL-6- and lipopolysaccharide-induced transcription of hepcidin in HFE-, transferrin receptor 2-, and beta 2-microglobulin-deficient hepatocytes,” *Proc Natl Acad Sci U S A*, vol. 101, no. 25, pp. 9263–9265, 2004.

[157] S. Rivera, E. Nemeth, V. Gabayan, M. A. Lopez, D. Farshidi, and T. Ganz, “Synthetic hepcidin causes rapid dose-dependent hypoferremia and is concentrated in ferroportin-containing organs,” *Blood*, vol. 106, no. 6, pp. 2196–2199, 2005.

[158] K. M. Papp-Wallace and M. E. Maguire, “Manganese Transport and the Role of Manganese in Virulence,” *Annual Review of Microbiology*, vol. 60, no. 1, pp. 187–209, 2006.

[159] B. D. Corbin, E. H. Seeley, A. Raab, J. Feldmann, M. R. Miller, V. J. Torres,

K. L. Anderson, B. M. Dattilo, P. M. Dunman, R. Gerads, R. M. Caprioli, W. Nacken, W. J. Chazin, and E. P. Skaar, “Metal chelation and inhibition of bacterial growth in tissue abscesses,” *Science*, vol. 319, no. 5865, pp. 962–965, 2008.

[160] C. Andreini, I. Bertini, and A. Rosato, “Metalloproteomes: a bioinformatic approach,” *Acc Chem Res*, vol. 42, no. 10, pp. 1471–1479, 2009.

[161] K. S. Min, M. Takano, K. Amako, H. Ueda, and K. Tanaka, “Involvement of the essential metal transporter Zip14 in hepatic Cd accumulation during inflammation,” *Toxicol Lett*, vol. 218, no. 1, pp. 91–96, 2013.

[162] N. Nesseler, Y. Launey, C. Aninat, F. Morel, Y. Malledant, and P. Seguin, “Clinical review: The liver in sepsis,” *Crit Care*, vol. 16, no. 5, p. 235, 2012.

[163] D. L. Knoell, M. W. Julian, S. Bao, B. Besecker, J. E. Macre, G. D. Leikauf, R. A. DiSilvestro, and E. D. Crouser, “Zinc deficiency increases organ damage and mortality in a murine model of polymicrobial sepsis,” *Crit Care Med*, vol. 37, no. 4, pp. 1380–1388, 2009.

[164] K. T. Crowell, B. E. Phillips, S. L. Kelleher, D. I. Soybel, and C. H. Lang, “Immune and metabolic responses in early and late sepsis during mild dietary zinc restriction,” *J Surg Res*, vol. 210, pp. 47–58, 2017.

[165] B. Besecker, S. Bao, B. Bohacova, A. Papp, W. Sadee, and D. L. Knoell, “The human zinc transporter SLC39A8 (Zip8) is critical in zinc-mediated cytoprotection in lung epithelia,” *American Journal of Physiology-Lung Cellular and Molecular Physiology*, vol. 294, no. 6, pp. L1127–L1136, 2008.

[166] M. G. Toscano, D. Ganea, and A. M. Gamero, “Cecal ligation puncture procedure,” *J Vis Exp*, no. 51, 2011.

[167] I. Wessels and R. J. Cousins, “Zinc dyshomeostasis during polymicrobial sepsis in mice involves zinc transporter Zip14 and can be overcome by zinc supplementation,” *Am J Physiol Gastrointest Liver Physiol*, vol. 309, no. 9, pp. 768–78, 2015. [Online]. Available: https://www.ncbi.nlm.nih.gov/pubmed/26272258

[168] D. Rittirsch, M. A. Flierl, and P. A. Ward, “Harmful molecular mechanisms in sepsis,” *Nat Rev Immunol*, vol. 8, no. 10, pp. 776–787, 2008.

[169] S. Bao, M.-J. Liu, B. Lee, B. Besecker, J.-P. Lai, D. C. Guttridge, and D. L. Knoell, “Zinc modulates the innate immune response in vivo to polymicrobial sepsis through regulation of NF-κB,” *American Journal of Physiology - Lung Cellular and Molecular Physiology*, vol. 298, no. 6, pp. L744–L754, 2010.

[170] M. J. Liu, S. Bao, J. R. Napolitano, D. L. Burris, L. Yu, S. Tridandapani, and D. L. Knoell, “Zinc regulates the acute phase response and serum amyloid A production in response to sepsis through JAK-STAT3 signaling,” *PLoS One*, vol. 9, no. 4, p. e94934, 2014.

[171] M. S. Hayden and S. Ghosh, “Shared Principles in NF-κB Signaling,” *Cell*, vol. 132, no. 3, pp. 344–362.

[172] J. P. Liuzzi, C. P. Wong, E. Ho, and A. Tracey, “Regulation of hepatic suppressor of cytokine signaling 3 by zinc,” *J Nutr Biochem*, vol. 24, no. 6, pp. 1028–1033, 2013.

[173] I. Wessels, H. J. Fischer, and L. Rink, “Update on the multi-layered levels of zinc-mediated immune regulation,” *Seminars in Cell and Developmental Biology*, vol. 115, 2021.

[174] M. Yu, W. W. Lee, D. Tomar, S. Pryshchep, M. Czesnikiewicz-Guzik, D. L. Lamar, G. Li, K. Singh, L. Tian, C. M. Weyand, and J. J. Goronzy, "Regulation of T cell receptor signaling by activation-induced zinc influx," *J Exp Med*, vol. 208, no. 4, pp. 775–785, 2011.

[175] E. K. LeGrand and J. Alcock, "Turning Up The Heat: Immune Brinksmanship In The Acute-phase Response," *The Quarterly Review of Biology*, vol. 87, no. 1, pp. 3–18, 2012.

[176] H. R. Stennicke and G. S. Salvesen, "Biochemical characteristics of caspases-3, -6, -7, and -8," *J Biol Chem*, vol. 272, no. 41, pp. 25 719–25 723, 1997.

[177] S. Katz, M. A. Jimenez, W. E. Lehmkuhler, and J. L. Grosfeld, "Liver bacterial clearance following hepatic artery ligation and portacaval shunt," *Journal of Surgical Research*, vol. 51, no. 3, pp. 267–270, 1991.

[178] T. C. Vary and S. R. Kimball, "Regulation of hepatic protein synthesis in chronic inflammation and sepsis," *American Journal of Physiology - Cell Physiology*, vol. 262, no. 2 31-2, 1992.

[179] V. Aburto-Luna, S. Trevino, G. Santos-Lopez, D. Moroni-Gonzalez, O. Calva-Cruz, P. Aguilar-Alonso, B. A. Leon-Chavez, and E. Brambila, "Hepatic mobilization of zinc after an experimental surgery, and its relationship with inflammatory cytokines release, and expression of metallothionein and Zip14 transporter," *Inflamm Res*, vol. 66, no. 2, pp. 167–175, 2017.

[180] C. Lang, C. Murgia, M. Leong, L. W. Tan, G. Perozzi, D. Knight, R. Ruffin, and P. Zalewski, "Anti-inflammatory effects of zinc and alterations in

zinc transporter mRNA in mouse models of allergic inflammation," *Am J Physiol Lung Cell Mol Physiol*, vol. 292, no. 2, pp. 577–84, 2007.

[181] X. Tian, Y. Zheng, Y. Li, Z. Shen, L. Tao, X. Dou, J. Qian, and H. Shen, "Psychological stress induced zinc accumulation and up-regulation of ZIP14 and metallothionein in rat liver," *BioMed Central Gastroenterology*, vol. 14, 2014.

[182] J. Sambrook and D. W. Russell, "Molecular cloning: a laboratory manual (3-volume set)," 2001.

[183] G. V. Heimer and C. E. Taylor, "Improved mountant for immunofluorescence preparations," *Journal of Clinical Pathology*, vol. 27, no. 3, 1974.

[184] J. Schindelin, I. Arganda-Carreras, E. Frise, V. Kaynig, M. Longair, T. Pietzsch, S. Preibisch, C. Rueden, S. Saalfeld, B. Schmid, J. Y. Tinevez, D. J. White, V. Hartenstein, K. Eliceiri, P. Tomancak, and A. Cardona, "Fiji: An open-source platform for biological-image analysis," 2012.

[185] C. L. Kielkopf, W. Bauer, and I. L. Urbatsch, "Sodium dodecyl sulfate-polyacrylamide gel electrophoresis of proteins," *Cold Spring Harbor Protocols*, vol. 2021, no. 12, 2021.

[186] U. K. Laemmli, "Cleavage of Structural Proteins during the Assembly of the Head of Bacteriophage T4," *Nature*, vol. 227, no. 5259, pp. 680–685, 1970.

[187] "BD Accuri(TM) C6 Plus System User's Guide," San Jose, CA. [Online]. Available: https://www.bdbiosciences.com/content/dam/bdb/marketing-documents/BD-Accuri-C6-Plus-Users-Guide.pdf

[188] S. M. Castillo-Hair, J. T. Sexton, B. P. Landry, E. J. Olson, O. A. Igoshin, and J. J. Tabor, "FlowCal: A User-Friendly, Open Source Software

Tool for Automatically Converting Flow Cytometry Data from Arbitrary to Calibrated Units," *ACS Synthetic Biology*, vol. 5, no. 7, pp. 774–780, 2016.

[189] F. Madeira, Y. M. Park, J. Lee, N. Buso, T. Gur, N. Madhusoodanan, P. Basutkar, A. R. N. Tivey, S. C. Potter, R. D. Finn, and R. Lopez, "The EMBL-EBI search and sequence analysis tools APIs in 2019," *Nucleic Acids Research*, vol. 47, no. W1, pp. W636–W641, 2019.

[190] R. C. Edgar, "MUSCLE: a multiple sequence alignment method with reduced time and space complexity," *BMC bioinformatics*, vol. 5, p. 113, 2004.

[191] J. Pei and N. V. Grishin, "AL2CO: calculation of positional conservation in a protein sequence alignment," *Bioinformatics*, vol. 17, no. 8, pp. 700–712, 2001.

[192] N. Eswar, B. Webb, M. A. Marti-Renom, M. S. Madhusudhan, D. Eramian, M.-Y. Shen, U. Pieper, and A. Sali, "Comparative protein structure modeling using Modeller," *Current protocols in bioinformatics*, vol. Chapter 5, pp. Unit–5.6, 2006.

[193] L. A. Kelley, S. Mezulis, C. M. Yates, M. N. Wass, and M. J. E. Sternberg, "The Phyre2 web portal for protein modeling, prediction and analysis," *Nature Protocols*, vol. 10, no. 6, pp. 845–858, 2015.

[194] A. Waterhouse, M. Bertoni, S. Bienert, G. Studer, G. Tauriello, R. Gumienny, F. T. Heer, T. A. P. de Beer, C. Rempfer, L. Bordoli, R. Lepore, and T. Schwede, "SWISS-MODEL: homology modelling of protein structures and complexes," *Nucleic Acids Res*, vol. 46, no. W1, pp. W296–w303, 2018.

[195] G. G. Krivov, M. V. Shapovalov, and R. L. Dunbrack Jr., “Improved prediction of protein side-chain conformations with SCWRL4,” *Proteins*, vol. 77, no. 4, pp. 778–795, 2009.

[196] R. Evans, M. O’Neill, A. Pritzel, N. Antropova, A. Senior, T. Green, A. Žídek, R. Bates, S. Blackwell, J. Yim, O. Ronneberger, S. Bodenstein, M. Zielinski, A. Bridgland, A. Potapenko, A. Cowie, K. Tunyasuvunakool, R. Jain, E. Clancy, P. Kohli, J. Jumper, and D. Hassabis, “Protein complex prediction with AlphaFold-Multimer,” *bioRxiv*, p. 2021.10.04.463034, 1 2021. [Online]. Available: http://biorxiv.org/content/early/2021/10/04/2021.10.04.463034.abstract

[197] E. F. Pettersen, T. D. Goddard, C. C. Huang, G. S. Couch, D. M. Greenblatt, E. C. Meng, and T. E. Ferrin, “UCSF Chimera - A visualization system for exploratory research and analysis,” *Journal of Computational Chemistry*, vol. 25, no. 13, 2004.

[198] M. Wiederstein and M. J. Sippl, “ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins,” *Nucleic Acids Research*, vol. 35, no. suppl_2, pp. W407–W410, 2007.

[199] F. Melo, D. Devos, E. Depiereux, and E. Feytmans, “ANOLEA: a www server to assess protein structures,” *Proc Int Conf Intell Syst Mol Biol*, vol. 5, pp. 187–190, 1997.

[200] G. Studer, M. Biasini, and T. Schwede, “Assessing the local structural quality of transmembrane protein models using statistical potentials (QMEANBrane),” *Bioinformatics*, vol. 30, no. 17, pp. 505–11, 2014. [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4147910/pdf/btu457.pdf

[201] P. J. Cock, T. Antao, J. T. Chang, B. A. Chapman, C. J. Cox, A. Dalke, I. Friedberg, T. Hamelryck, F. Kauff, B. Wilczynski, and M. J. De Hoon, "Biopython: Freely available Python tools for computational molecular biology and bioinformatics," *Bioinformatics*, vol. 25, no. 11, 2009.

[202] T. Hamelryck and B. Manderick, "PDB file parser and structure class implemented in Python," *Bioinformatics*, vol. 19, no. 17, 2003.

[203] J. U. Bowie, J. F. Reidhaar-Olson, W. A. Lim, and R. T. Sauer, "Deciphering the message in protein sequences: Tolerance to amino acid substitutions," *Science*, vol. 247, no. 4948, 1990.

[204] A. Kitts and S. Sherry, "5. The Single Nucleotide Polymorphism Database (dbSNP) of Nucleotide Sequence Variation. in The NCBI Handbook," in *NCBI Handbook*, 2002, no. Md.

[205] K. T. Chang, J. Guo, A. di Ronza, and M. Sardiello, "Aminode: Identification of Evolutionary Constraints in the Human Proteome," *Scientific Reports*, vol. 8, no. 1, p. 1357, 2018.

[206] K. L. Brouwer, D. Keppler, K. A. Hoffmaster, D. A. Bow, Y. Cheng, Y. Lai, J. E. Palm, B. Stieger, and R. Evers, "In vitro methods to support transporter evaluation in drug discovery and development," 2013.

[207] V. Dvorak, T. Wiedmer, A. Ingles-Prieto, P. Altermatt, H. Batoulis, F. Bärenz, E. Bender, D. Digles, F. Dürrenberger, L. H. Heitman, A. P. IJzerman, D. B. Kell, S. Kickinger, D. Körzö, P. Leippe, T. Licher, V. Manolova, R. Rizzetto, F. Sassone, L. Scarabottolo, A. Schlessinger, V. Schneider, H. J. Sijben, A. L. Steck, H. Sundström, S. Tremolada, M. Wilhelm, M. Wright Muelas, D. Zindel, C. M. Steppan, and G. Superti-

Furga, “An Overview of Cell-Based Assay Platforms for the Solute Carrier Family of Transporters,” 2021.

[208] T. U. Consortium, “UniProt: the universal protein knowledgebase in 2021,” *Nucleic Acids Research*, vol. 49, no. D1, pp. D480–D489, 1 2021. [Online]. Available: https://doi.org/10.1093/nar/gkaa1100

[209] M. P. Jackson and E. W. Hewitt, “Cellular proteostasis: Degradation of misfolded proteins by lysosomes,” *Essays in Biochemistry*, vol. 60, no. 2, 2016.

[210] C. M. Haynes, E. A. Titus, and A. A. Cooper, “Degradation of misfolded proteins prevents ER-derived oxidative stress and cell death,” *Molecular Cell*, vol. 15, no. 5, 2004.

[211] M. W. McLane, G. Hatzidimitriou, J. Yuan, U. McCann, and G. Ricaurte, “Heating induces aggregation and decreases detection of serotonin transporter protein on western blots,” *Synapse*, vol. 61, no. 10, pp. 875–876, 10 2007. [Online]. Available: https://doi.org/10.1002/syn.20438

[212] J. Kaur and A. K. Bachhawat, “A modified Western blot protocol for enhanced sensitivity in the detection of a membrane protein,” *Analytical Biochemistry*, vol. 384, no. 2, pp. 348–349, 2009. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0003269708006829

[213] Y. Shi, R. A. Mowery, J. Ashley, M. Hentz, A. J. Ramirez, B. Bilgicer, H. Slunt-Brown, D. R. Borchelt, and B. F. Shaw, “Abnormal SDS-PAGE migration of cytosolic proteins can identify domains and mechanisms that control surfactant binding,” *Protein Science*, vol. 21, no. 8, pp. 1197–1209, 8 2012. [Online]. Available: https://doi.org/10.1002/pro.2107

[214] A. Rath, M. Glibowicka, V. G. Nadeau, G. Chen, and C. M. Deber, “Detergent binding explains anomalous SDS-PAGE migration of membrane proteins,” *Proceedings of the National Academy of Sciences*, vol. 106, no. 6, p. 1760, 2009.

[215] L. Vincenz-Donnelly and M. S. Hipp, “The endoplasmic reticulum: A hub of protein quality control in health and disease,” 2017.

[216] F. Cunningham, J. E. Allen, J. Allen, J. Alvarez-Jarreta, M. R. Amode, I. M. Armean, O. Austine-Orimoloye, A. G. Azov, I. Barnes, R. Bennett, A. Berry, J. Bhai, A. Bignell, K. Billis, S. Boddu, L. Brooks, M. Charkhchi, C. Cummins, L. Da Rin Fioretto, C. Davidson, K. Dodiya, S. Donaldson, B. El Houdaigui, T. El Naboulsi, R. Fatima, C. G. Giron, T. Genez, J. G. Martinez, C. Guijarro-Clarke, A. Gymer, M. Hardy, Z. Hollis, T. Hourlier, T. Hunt, T. Juettemann, V. Kaikala, M. Kay, I. Lavidas, T. Le, D. Lemos, J. C. Marugán, S. Mohanan, A. Mushtaq, M. Naven, D. N. Ogeh, A. Parker, A. Parton, M. Perry, I. Pilizota, I. Prosovetskaia, M. P. Sakthivel, A. I. A. Salam, B. M. Schmitt, H. Schuilenburg, D. Sheppard, J. G. Perez-Silva, W. Stark, E. Steed, K. Sutinen, R. Sukumaran, D. Sumathipala, M. M. Suner, M. Szpak, A. Thormann, F. F. Tricomi, D. Urbina-Gómez, A. Veidenberg, T. A. Walsh, B. Walts, N. Willhoft, A. Winterbottom, E. Wass, M. Chakiachvili, B. Flint, A. Frankish, S. Giorgetti, L. Haggerty, S. E. Hunt, G. R. Iisley, J. E. Loveland, F. J. Martin, B. Moore, J. M. Mudge, M. Muffato, E. Perry, M. Ruffier, J. Tate, D. Thybert, S. J. Trevanion, S. Dyer, P. W. Harrison, K. L. Howe, A. D. Yates, D. R. Zerbino, and P. Flicek, “Ensembl 2022,” *Nucleic Acids Research*, vol. 50, no. D1, 2022.

[217] ECACC, “Fundamental Techniques in Cell Culture Laboratory Handbook 4th Edition,” in *Merck KGaA*, 2018.

[218] R. E. E, E. D. J, and G. M. Lou, “Altered selectivity in an Arabidopsis } metal transporter,” *Proceedings of the National Academy of Sciences*, vol. 97, no. 22, pp. 12 356–12 360, 10 2000. [Online]. Available: https://doi.org/10.1073/pnas.210214197

[219] T. Zhang, D. Sui, and J. Hu, “Structural insights of ZIP4 extracellular domain critical for optimal zinc transport,” *Nat Commun*, vol. 7, p. 11979, 2016. [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4915132/pdf/ncomms11979.pdf

[220] M. C. Franz, A. Simonin, S. Graeter, M. A. Hediger, and G. Kovacs, “Development of the first fluorescence screening assay for the slc39a2 zinc transporter,” *Journal of Biomolecular Screening*, vol. 19, no. 6, 2014.

[221] S. U. Gerold, K. M. Taylor, I. A. Muraina, D. Brethour, T. Nimmanon, S. Ziliotto, P. Kille, and C. Hogstrand, “Zinc transporter ZIP10 forms a heteromer with ZIP6 which regulates embryonic development and cell migration,” *Biochemical Journal*, vol. 473, no. 16, 2016.

[222] C. L. Pocanschi, S. Ehsani, M. Mehrabian, H. Wille, W. Reginold, W. S. Trimble, H. Wang, A. Yee, C. H. Arrowsmith, Z. Bozóky, L. E. Kay, J. D. Forman-Kay, J. M. Rini, and G. Schmitt-Ulms, “The ZIP5 Ectodomain Co-Localizes with PrP and May Acquire a PrP-Like Fold That Assembles into a Dimer,” *PLoS ONE*, vol. 8, no. 9, 2013.

[223] M. E. Ahern, E. M. Bafaro, A. Cowan, and R. E. Dempski, “Quantifying the Oligomeric State of hZIP4 on the Surface of Cells,” *Biochemistry*, vol. 58, no. 13, 2019.

[224] P. B. Chetri, H. Khan, and T. Tripathi, “Methods to determine the

oligomeric structure of proteins," in *Advances in Protein Molecular and Structural Biology Methods*, 2022.

[225] K. A. Dill, S. B. Ozkan, M. S. Shell, and T. R. Weikl, "The protein folding problem," 2008.

[226] J. U. Bowie, "Solving the membrane protein folding problem," 2005.

[227] E. P. Carpenter, K. Beis, A. D. Cameron, and S. Iwata, "Overcoming the challenges of membrane protein crystallography," 2008.

[228] Y. Zhang, "Protein structure prediction: when is it useful?" 2009.

[229] X. Qu, R. Swanson, R. Day, and J. Tsai, "A Guide to Template Based Structure Prediction," *Current Protein & Peptide Science*, vol. 10, no. 3, 2009.

[230] D. Fischer, D. Rice, J. U. Bowie, and D. Eisenberg, "Assigning amino acid sequences to 3-dimensional protein folds," *The FASEB Journal*, vol. 10, no. 1, 1996.

[231] D. J. Osguthorpe, "Ab initio protein folding," 2000.

[232] Z. Du, H. Su, W. Wang, L. Ye, H. Wei, Z. Peng, I. Anishchenko, D. Baker, and J. Yang, "The trRosetta server for fast and accurate protein structure prediction," 2021.

[233] J. Jumper, R. Evans, A. Pritzel, T. Green, M. Figurnov, O. Ronneberger, K. Tunyasuvunakool, R. Bates, A. Žídek, A. Potapenko, A. Bridgland, C. Meyer, S. A. Kohl, A. J. Ballard, A. Cowie, B. Romera-Paredes, S. Nikolov, R. Jain, J. Adler, T. Back, S. Petersen, D. Reiman, E. Clancy, M. Zielinski, M. Steinegger, M. Pacholska, T. Berghammer, S. Bodenstein, D. Silver, O. Vinyals, A. W. Senior, K. Kavukcuoglu, P. Kohli, and

D. Hassabis, “Highly accurate protein structure prediction with AlphaFold,” *Nature*, vol. 596, no. 7873, 2021.

[234] A. Kryshtafovych, T. Schwede, M. Topf, K. Fidelis, and J. Moult, “Critical assessment of methods of protein structure prediction (CASP)—Round XIV,” *Proteins: Structure, Function and Bioinformatics*, vol. 89, no. 12, 2021.

[235] S. Y. Chung and S. Subbiah, “A structural explanation for the twilight zone of protein sequence homology,” *Structure*, vol. 4, no. 10, 1996.

[236] L. Mullan, “Pairwise sequence alignment - It’s all about us!” *Briefings in Bioinformatics*, vol. 7, no. 1, 2006.

[237] C. B. Do and K. Katoh, “Protein multiple sequence alignment,” 2008.

[238] J. Pei, “Multiple protein sequence alignment,” 2008.

[239] R. L. Dunbrack, “Sequence comparison and protein structure prediction,” 2006.

[240] G. Studer, C. Rempfer, A. M. Waterhouse, R. Gumienny, J. Haas, and T. Schwede, “QMEANDisCo—distance constraints applied on model quality estimation,” *Bioinformatics*, vol. 36, no. 6, 2020.

[241] T. Zhang, J. Liu, M. Fellner, C. Zhang, D. Sui, and J. Hu, “Crystal structures of a ZIP zinc transporter reveal a binuclear metal center in the transport pathway,” *Sci Adv*, vol. 3, no. 8, p. e1700344, 2017. [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5573306/pdf/1700344.pdf

[242] M. J. Sippl, “Recognition of errors in three-dimensional structures of proteins,” *Proteins: Structure, Function, and*

Bioinformatics, vol. 17, no. 4, pp. 355–362, 1993. [Online]. Available: https://doi.org/10.1002/prot.340170404https://onlinelibrary.wiley.com/doi/abs/10.1002/prot.340170404

[243] S. El-Gebali, J. Mistry, A. Bateman, S. R. Eddy, A. Luciani, S. C. Potter, M. Qureshi, L. J. Richardson, G. A. Salazar, A. Smart, E. L. Sonnhammer, L. Hirsh, L. Paladin, D. Piovesan, S. C. Tosatto, and R. D. Finn, "The Pfam protein families database in 2019," *Nucleic Acids Research*, vol. 47, no. D1, pp. D427–D432, 2018.

[244] S. F. Altschul, W. Gish, W. Miller, E. W. Myers, and D. J. Lipman, "Basic local alignment search tool," *J Mol Biol*, vol. 215, no. 3, pp. 403–410, 1990.

[245] F. Melo and E. Feytmans, "Novel knowledge-based mean force potential at atomic level 11Edited by F. E. Cohen," *J Mol Biol*, vol. 267, no. 1, pp. 207–222, 1997.

[246] T. P. Flores, C. A. Orengo, D. S. Moss, and J. M. Thornton, "Comparison of conformational characteristics in structurally similar protein pairs," *Protein science : a publication of the Protein Society*, vol. 2, no. 11, pp. 1811–1826, 1993.

[247] E. L. Sonnhammer, S. R. Eddy, and R. Durbin, "Pfam: A comprehensive database of protein domain families based on seed alignments," *Proteins: Structure, Function and Genetics*, vol. 28, no. 3, 1997.

[248] B. Rost, "Twilight zone of protein sequence alignments," *Protein Engineering*, vol. 12, no. 2, 1999.

[249] W. R. Pearson, "[15] Effective protein sequence comparison," *Methods in Enzymology*, vol. 266, 1996.

[250] J. Park, K. Karplus, C. Barrett, R. Hughey, D. Haussler, T. Hubbard, and C. Chothia, "Sequence comparisons using multiple sequences detect three times as many remote homologues as pairwise methods," *Journal of Molecular Biology*, vol. 284, no. 4, 1998.

[251] M. Madera and J. Gough, "A comparison of profile hidden Markov model procedures for remote homology detection," 2002.

[252] P. Skewes-Cox, T. J. Sharpton, K. S. Pollard, and J. L. DeRisi, "Profile hidden Markov models for the detection of viruses within metagenomic sequence data," *PLoS ONE*, vol. 9, no. 8, 2014.

[253] D. J. Studholme, N. D. Rawlings, A. J. Barrett, and A. Bateman, "A comparison of Pfam and MEROPS: Two database, one comprehensive, and one specialised," *BMC Bioinformatics*, vol. 4, 2003.

[254] A. M. Lesk, C. H. Chothia, D. M. Blow, A. R. Fersht, and G. Winter, "The response of protein structures to amino-acid sequence changes," *Philosophical Transactions of the Royal Society of London. Series A, Mathematical and Physical Sciences*, vol. 317, no. 1540, 1986.

[255] L. Lins and R. Brasseur, "The hydrophobic effect in protein folding," *The FASEB Journal*, vol. 9, no. 7, 1995.

[256] L. Young, R. L. Jernigan, and D. G. Covell, "A role for surface hydrophobicity in protein-protein recognition," *Protein Science*, vol. 3, no. 5, 1994.

[257] F. B. Sheinerman, R. Norel, and B. Honig, "Electrostatic aspects of protein-protein interactions," 2000.

[258] T. A. Larsen, A. J. Olson, and D. S. Goodsell, "Morphology of protein-protein interfaces," *Structure*, vol. 6, no. 4, 1998.

[259] N. G. Jayaprakash and A. Surolia, “Role of glycosylation in nucleating protein folding and stability,” 2017.

[260] A. K. Dunker, C. J. Brown, J. D. Lawson, L. M. Iakoucheva, and Z. Obradović, “Intrinsic disorder and protein function,” *Biochemistry*, vol. 41, no. 21, 2002.

[261] R. A. George and J. Heringa, “An analysis of protein domain linkers: Their classification and role in protein folding,” *Protein Engineering*, vol. 15, no. 11, 2002.

[262] R. Van Der Lee, M. Buljan, B. Lang, R. J. Weatheritt, G. W. Daughdrill, A. K. Dunker, M. Fuxreiter, J. Gough, J. Gsponer, D. T. Jones, P. M. Kim, R. W. Kriwacki, C. J. Oldfield, R. V. Pappu, P. Tompa, V. N. Uversky, P. E. Wright, and M. M. Babu, “Classification of intrinsically disordered regions and proteins,” 2014.

[263] T. Zhang, D. Sui, C. Zhang, L. Cole, and J. Hu, “Asymmetric functions of a binuclear metal center within the transport pathway of a human zinc transporter ZIP4,” *FASEB Journal*, vol. 34, no. 1, 2020. [Online]. Available: https://faseb.onlinelibrary.wiley.com/doi/abs/10.1096/fj.201902043R

[264] K. Bowers and S. K. S. Srai, “The trafficking of metal ion transporters of the Zrt- and Irt-like protein family,” *Traffic*, vol. 19, no. 11, pp. 813–822, 2018.

[265] A. Einhauer and A. Jungbauer, “The FLAG™ peptide, a versatile fusion tag for the purification of recombinant proteins,” 2001.

[266] M. Varadi, S. Anyango, M. Deshpande, S. Nair, C. Natassia, G. Yordanova, D. Yuan, O. Stroe, G. Wood, A. Laydon, A. Zídek, T. Green, K. Tunyasuvunakool, S. Petersen, J. Jumper, E. Clancy, R. Green,

A. Vora, M. Lutfi, M. Figurnov, A. Cowie, N. Hobbs, P. Kohli, G. Kleywegt, E. Birney, D. Hassabis, and S. Velankar, "AlphaFold Protein Structure Database: Massively expanding the structural coverage of protein-sequence space with high-accuracy models," *Nucleic Acids Research*, vol. 50, no. D1, 2022.

[267] A. Wiuf, J. H. Steffen, E. R. Becares, C. Grønberg, D. R. Mahato, S. G. F. Rasmussen, M. Andersson, T. Croll, K. Gotfryd, and P. Gourdon, "The two-domain elevator-type mechanism of zinc-transporting ZIP proteins," *Science Advances*, vol. 8, no. 28, p. eabn4331, 9 2022. [Online]. Available: https://doi.org/10.1126/sciadv.abn4331

[268] S. Bienert, A. Waterhouse, T. A. De Beer, G. Tauriello, G. Studer, L. Bordoli, and T. Schwede, "The SWISS-MODEL Repository-new features and functionality," *Nucleic Acids Research*, vol. 45, no. D1, 2017.

[269] A. Ingles-Prieto, B. Ibarra-Molero, A. Delgado-Delgado, R. Perez-Jimenez, J. M. Fernandez, E. A. Gaucher, J. M. Sanchez-Ruiz, and J. A. Gavira, "Conservation of protein structure over four billion years," *Structure*, vol. 21, no. 9, 2013.

[270] S. Tetchner, T. Kosciolek, and D. T. Jones, "Opportunities and limitations in applying coevolution-derived contacts to protein," 2014.

[271] D. W. A. Buchan and D. T. Jones, "The PSIPRED Protein Analysis Workbench: 20 years on," *Nucleic Acids Research*, vol. 47, no. W1, pp. W402–W407, 2019.

[272] H. Kamisetty, S. Ovchinnikov, and D. Baker, "Assessing the utility of coevolution-based residue–residue contact predictions in a sequence-

and structure-rich era," *Proceedings of the National Academy of Sciences*, vol. 110, no. 39, p. 15674, 2013.

[273] F. Morcos, B. Jana, T. Hwa, and J. N. Onuchic, "Coevolutionary signals across protein lineages help capture multiple protein conformations," *Proceedings of the National Academy of Sciences*, vol. 110, no. 51, p. 20533, 2013.

[274] P. Sfriso, M. Duran-Frigola, R. Mosca, A. Emperador, P. Aloy, and M. Orozco, "Residues Coevolution Guides the Systematic Identification of Alternative Functional Conformations in Proteins," *Structure*, vol. 24, no. 1, pp. 116–126, 2016. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0969212615004657

[275] E. Kuliyev, C. Zhang, D. Sui, and J. Hu, "Zinc transporter mutations linked to acrodermatitis enteropathica disrupt function and cause mistrafficking," *Journal of Biological Chemistry*, vol. 296, 2021.

[276] Z. Sun and J. L. Brodsky, "Protein quality control in the secretory pathway," 2019.

[277] Y. Alguel, A. D. Cameron, G. Diallinas, and B. Byrne, "Transporter oligomerization: Form and function," *Biochemical Society Transactions*, vol. 44, no. 6, 2016.

[278] T. Kambe and G. K. Andrews, "Novel Proteolytic Processing of the Ectodomain of the Zinc Transporter ZIP4 (SLC39A4) during Zinc Deficiency Is Inhibited by Acrodermatitis Enteropathica Mutations," *Molecular and Cellular Biology*, vol. 29, no. 1, pp. 129–139, 2009.

[279] K. M. Taylor, S. Hiscox, R. I. Nicholson, C. Hogstrand, and P. Kille, "Cell biology: Protein kinase CK2 triggers cytosolic zinc signaling pathways by

phosphorylation of zinc channel ZIP7," *Science Signaling*, vol. 5, no. 210, 2012.

[280] B. H. Bin, T. Fukada, T. Hosaka, S. Yamasaki, W. Ohashi, S. Hojyo, T. Miyai, K. Nishida, S. Yokoyama, and T. Hirano, "Biochemical characterization of human ZIP13 protein: A homo-dimerized zinc transporter involved in the spondylocheiro dysplastic Ehlers-Danlos syndrome," *Journal of Biological Chemistry*, vol. 286, no. 46, 2011.

[281] K. M. Ruff and R. V. Pappu, "AlphaFold and Implications for Intrinsically Disordered Proteins," 2021.

[282] S. Antala, S. Ovchinnikov, H. Kamisetty, D. Baker, and R. E. Dempski, "Computation and Functional Studies Provide a Model for the Structure of the Zinc Transporter hZIP4," *J Biol Chem*, vol. 290, no. 29, pp. 17 796–17 805, 2015.

[283] H. Chun, T. Korolnek, C.-J. Lee, H. J. Coyne, D. R. Winge, B.-E. Kim, and M. J. Petris, "An extracellular histidine-containing motif in the zinc transporter ZIP4 plays a role in zinc sensing and zinc-induced endocytosis in mammalian cells," *Journal of Biological Chemistry*, vol. 294, no. 8, pp. 2815–2826, 2019.

[284] J. M. Matthews and M. Sunde, "Dimers, oligomers, everywhere," 2012.

[285] E. D. Levy, E. B. Erba, C. V. Robinson, and S. A. Teichmann, "Assembly reflects evolution of protein complexes," *Nature*, vol. 453, no. 7199, 2008.

[286] G. Villar, A. W. Wilber, A. J. Williamson, P. Thiara, J. P. Doye, A. A. Louis, M. N. Jochum, A. C. Lewis, and E. D. Levy, "Self-assembly and evolution of homomeric protein complexes," *Physical Review Letters*, vol. 102, no. 11, 2009.

[287] P. S. Park, “Supramolecular organization of rhodopsin in rod photoreceptor cell membranes,” 2021.

[288] X. Periole, A. M. Knepp, T. P. Sakmar, S. J. Marrink, and T. Huber, “Structural determinants of the supramolecular organization of G protein-coupled receptors in bilayers,” *Journal of the American Chemical Society*, vol. 134, no. 26, 2012.

[289] K. Jayaraman, A. K. Das, D. Luethi, D. Szöllősi, G. J. Schütz, M. E. Reith, H. H. Sitte, and T. Stockner, “SLC6 transporter oligomerization,” 2021.

[290] X. Chen, J. L. Zaro, and W. C. Shen, “Fusion protein linkers: Property, design and functionality,” 2013.

[291] C. S. Sørensena and M. Kjaergaarda, “Effective concentrations enforced by intrinsically disordered linkers are governed by polymer physics,” *Proceedings of the National Academy of Sciences of the United States of America*, vol. 116, no. 46, 2019.

[292] Y. Hsia, R. Mout, W. Sheffler, N. I. Edman, I. Vulovic, Y. J. Park, R. L. Redler, M. J. Bick, A. K. Bera, A. Courbet, A. Kang, T. J. Brunette, U. Nattermann, E. Tsai, A. Saleem, C. M. Chow, D. Ekiert, G. Bhabha, D. Veesler, and D. Baker, “Design of multi-scale protein complexes by hierarchical building block fusion,” *Nature Communications*, vol. 12, no. 1, 2021.

[293] D. Permana, H. E. Putra, and D. Djaenudin, “Designed protein multimerization and polymerization for functionalization of proteins,” 2022.

[294] J. Wei and L. Tong, “How Does Polymerization Regulate Human Acetyl-CoA Carboxylase 1?” 2018.

[295] B. Byrne, “It takes two to transport via an elevator,” *Cell Research*, vol. 27, no. 8, 2017.

[296] G. Diallinas, “Transporter Specificity: A Tale of Loosened Elevator-Sliding,” 2021.

[297] A. Kourkoulou, P. Grevias, G. Lambrinidis, E. Pyle, M. Dionysopoulou, A. Politis, E. Mikros, B. Byrne, and G. Diallinas, “Specific residues in a purine transporter are critical for dimerization, ER exit, and function,” *Genetics*, vol. 213, no. 4, 2019.

[298] Y. Alguel, S. Amillis, J. Leung, G. Lambrinidis, S. Capaldi, N. J. Scull, G. Craven, S. Iwata, A. Armstrong, E. Mikros, G. Diallinas, A. D. Cameron, and B. Byrne, “Structure of eukaryotic purine/H+ symporter UapA suggests a role for homodimerization in transport activity,” *Nature Communications*, vol. 7, 2016.

[299] J. Galgonek, J. Vymětal, D. Jakubec, and J. Vondrášek, “Amino Acid Interaction (INTAA) web server,” *Nucleic Acids Research*, vol. 45, no. W1, 2017.

[300] M. Akdel, D. E. V. Pires, E. P. Pardo, J. Jänes, A. O. Zalevsky, B. Mészáros, P. Bryant, L. L. Good, R. A. Laskowski, G. Pozzati, A. Shenoy, W. Zhu, P. Kundrotas, V. R. Serra, C. H. M. Rodrigues, A. S. Dunham, D. Burke, N. Borkakoti, S. Velankar, A. Frost, K. Lindorff-Larsen, A. Valencia, S. Ovchinnikov, J. Durairaj, D. B. Ascher, J. M. Thornton, N. E. Davey, A. Stein, A. Elofsson, T. I. Croll, and P. Beltrao, “A structural biology community assessment of AlphaFold 2 applications,” *bioRxiv*, 2021.

Appendix A

Alignments

A.1 Pairwise Alignment of the N-terminal Extracellular Domain

```
P_ECD      1  MAILAWLEPRPLLAVLVLVLTMRMAQPAHLLTLLSSGQGALDRVALGGLL  50
hZIP14     0  --------------------------------------------------  0

P_ECD     51  NTLAARVHCTSGPCGKCLSVDDLLALGRPEEPGHLARLSAAAALYLSDPE  100
hZIP14     0  --------------------------------------------------  0

P_ECD    101  GTCEDIRAGRWASRADHLLALLEGPKALAPGLSRLLQRIQAQTTGQPSAG  150
hZIP14     0  --------------------------------------------------  0

P_ECD    151  EACVDPPQLLREAGVAGAPGSPGPVLATLLEHVGRGSC--------FHTL  192
hZIP14     1  -------------------------MKLLLLHPAFQSCLLLTLLGLWRTT  25
```

```
P_ECD    193  P-------------TPQYFVDFVFQQSHGNTPNISVAELAALMQRLGVG-  228
hZIP14    26  PEAHASSLGAPAISAASFLQDLIHRYGEGD--SLTLQQLKALLNHLDVGV  73

P_ECD    229  --GVTETHSDHHHQEKRVNRQGPTPLTAPNSSSDTWDTVCLSARDVMAVY  276
hZIP14    74  GRGNVTQHVQGHR----------------NLST------CFSSGDLFTAH  101

P_ECD    277  GLSEQTGVTPEAWAQLSPALLQQQLSGAC----------------------  305
hZIP14   102  NFSEQSRIGSSELQEFCPTILQQLDSRACTSENQENEENEQTEEGRPSAV  151

P_ECD    306  --                                                  305
hZIP14   152  EV                                                  153
```

A.2 Profile Alignment of the N-terminal Extracellular Domain

```
P_ECD      1  MAILAWLEPRPLLAVLVLVLTMRMAQPAHLLTLLSSGQGALDRVALGGLL  50
hZIP14     1  -------------------------MKLLLLHPAFQSCLLL--TLLGLWR  23

P_ECD     51  NTLAARVHCTSGPCGKCLSVDDLLALGRPEEPGHLARLSAAAALYLSDPE  100
hZIP14    24  T-TPEAHASSLGA------------------------------------  35

P_ECD    101  GTCEDIRAGRWASRADHLLALLEGPKALAPGLSRLLQRIQAQTTGQPSAG  150
hZIP14    35  ------------------------------------------------  35

P_ECD    151  EACVDPPQLLREAGVAGAPGSPGPVLATLLEHVGRGSCFHTLPTPQYFVD  200
hZIP14    36  ---------------------------------------PAISAASFLQ  45
```

```
P_ECD    201  FVFQQSHGNTPNISVAELAALMQRLGVGGVTETHSDHHHQEKRVNRQGPT  250
hZIP14    46  DLIHRYG-EGDSLTLQQLKALLNHLDVGVG---RGNVTQHVQG-------   84

P_ECD    251  PLTAPNSSSDTWDTVCLSARDVMAVYGLSEQTGVTPEAWAQLSPALLQQQ  300
hZIP14    85  --HRNLS-------TCFSSGDLFTAHNFSEQSRIGSSELQEFCPTILQQL  125

P_ECD    301  LSGAC                                               305
hZIP14   126  DSRAC                                               130
```

A.3 Pairwise Alignments of the Transmemembrane Domain

BbZIP vs hZIP14

```
BbZIP      1  ----MNQP---SSLAADLRGAWHAQAQSHPLITLGLAASAA---------   34
hZIP14     1  MKLLLLHPAFQSCLLLTLLGLWRTTPEAH-ASSLGAPAISAASFLQDLIH  49

BbZIP     35  ------GVVLLLVAGIVNAL-TGENRVHVGYAVLGG-------AAGFAAT   70
hZIP14    50  RYGEGDSLTLQQLKALLNHLDVGVGRGNVTQHVQGHRNLSTCFSSGDLFT  99

BbZIP     71  A--------LGA----------LMALGLRAISARTQD-------------   89
hZIP14   100  AHNFSEQSRIGSSELQEFCPTILQQLDSRACTSENQENEENEQTEEGRPS  149

BbZIP     90  --AMLGFAAGMMLAASAFSL-------ILPGLD--------------AAG  116
hZIP14   150  AVEVWGYG---LLCVTVISLCSLLGASVVPFMKKTFYKRLLLYFIALAIG  196
```

```
BbZIP   117  TI------------VGPGP------AAAAVVALGLG--------LGVLLM  140
hZIP14  197  TLYSNALFQLIPEAFGFNPLEDYYVSKSAVVFGGFYLFFFTEKILKILLK  246

BbZIP   141  LGLDYFTPHEH-----------ERTG-----------HQGPE--------  160
hZIP14  247  QKNEHHHGHSHYASESLPSKKDQEEGVMEKLQNGDLDHMIPQHCSSELDG  296

BbZIP   161  -AARVNR---------------------------------VWLFVLTIIL  176
hZIP14  297  KAPMVDEKVIVGSLSVQDLQASQSACYWLKGVRYSDIGTLAWMITLSDGL  346

BbZIP   177  HNLPEGMAIGVSFATGDLRIGLPLTSAIAIQDVPEGLAVALALRAVGLPI  226
hZIP14  347  HNFIDGLAIGASFTVSVFQ-GISTSVAILCEEFPHELGDFVILLNAGMSI  395

BbZIP   227  GRAVLVAVASGLMEPLGALVGVGISSGFALAYPISMGLAAGAMIFVVS--  274
hZIP14  396  QQALFFNFLSACCCYLGLAFGILAGSHFSANWIFAL---AGGMFLYISLA  442

BbZIP   275  ------HEVIPETHRNGH----ETTATVGLMAGFALMMFLDTALG-----  309
hZIP14  443  DMFPEMNEVCQEDERKGSILIPFIIQNLGLLTGFTIMVVLTMYSGQIQIG  492
```

BbZIP vs hZIP4

```
BbZIP     0  ------------------------------------------------  0
hZIP4     1  MASLVSLELGLLLAVLVVTATASPPAGLLSLLTSGQGALDQEALGGLLNT  50

BbZIP     0  ------------------------------------------------  0
hZIP4    51  LADRVHCANGPCGKCLSVEDALGLGEPEGSGLPPGPVLEARYVARLSAAA  100

BbZIP     0  ------------------------------------------------  0
hZIP4   101  VLYLSNPEGTCEDARAGLWASHADHLLALLESPKALTPGLSWLLQRMQAR  150
```

```
BbZIP     0  --------------------------------------------------  0
hZIP4   151  AAGQTPKMACVDIPQLLEEAVGAGAPGSAGGVLAALLDHVRSGSCFHALP  200

BbZIP     0  --------------------------------------------------  0
hZIP4   201  SPQYFVDFVFQQHSSEVPMTLAELSALMQRLGVGREAHSDHSHRHRGASS  250

BbZIP     0  ------------------------------------------MNQPSS  6
hZIP4   251  RDPVPLISSSNSSSVWDTVCLSARDVMAAYGLSEQAGVTPEAWAQLSPAL  300

BbZIP     7  LAADLRGAWHAQAQSHP----------------LITLGLAASAA-GVVLL  39
hZIP4   301  LQQQLSGA--CTSQSRPPVQDQLSQSERYLYGSLATLLICLCAVFGLLLL  348

BbZIP    40  LVAG---------------IVNALTGENRVHVGYAVLGGAAGFAATALGA  74
hZIP4   349  TCTGCRGVTHYILQTFLSLAVGAVTGDAVLHLTPKVLG------------  386

BbZIP    75  LMALGLRAISARTQDAMLGFAAGM---MLAASAFSLILPGLDAAGTIVGP  121
hZIP4   387  LHTHSEEGLSPQPTWRLLAMLAGLYAFFLFENLFNLLLP-RDPEDLEDGP  435

BbZIP   122  -GPAAAAVVALGLGLGVLLMLGLDYF----TPHEHERTG--------HQG  158
hZIP4   436  CGHSSHS--HGGHSHGVSLQLAPSELRQPKPPHEGSRADLVAEESPELLN  483

BbZIP   159  PEAARVN---RVWLFVLTI--ILHNLPEGMAIGVSFATGDLRIGLPLTSA  203
hZIP4   484  PEPRRLSPELRLLPYMITLGDAVHNFADGLAVGAAFAS-SWKTGLATSLA  532

BbZIP   204  IAIQDVPEGLAVALALRAVGLPIGRAVLVAVASGLMEPLGALVGVGISSG  253
hZIP4   533  VFCHELPHELGDFAALLHAGLSVRQALLLNLASALTAFAGLYV-------  575
```

```
BbZIP   254  FALAYPIS-------MGLAAGAMIFVVSHEVIPE--------------TH  282
hZIP4   576  -ALAVGVSEESEAWILAVATGLFLYVALCDMLPAMLKVRDPRPWLLFLLH  624

BbZIP   283  RNGHETTATVGLMAGFALMMFLDTALG----                     309
hZIP4   625  --------NVGLLGGWTVLLLLSLYEDDITF                     647
```

hZIP14 vs hZIP4

```
hZIP14    0  --------------------------------------------------  0
hZIP4     1  MASLVSLELGLLLAVLVVTATASPPAGLLSLLTSGQGALDQEALGGLLNT  50

hZIP14    0  --------------------------------------------------  0
hZIP4    51  LADRVHCANGPCGKCLSVEDALGLGEPEGSGLPPGPVLEARYVARLSAAA  100

hZIP14    0  --------------------------------------------------  0
hZIP4   101  VLYLSNPEGTCEDARAGLWASHADHLLALLESPKALTPGLSWLLQRMQAR  150

hZIP14    1  -------------------------------MKLLLLHPAFQSCLLLTL  18
hZIP4   151  AAGQTPKMACVDIPQLLEEAVGAGAPGSAGGVLAALLDHVRSGSCF----  196

hZIP14   19  LGLWRTTPEAHASSLGAPAISAASFLQDLIHRYGEGDSLTLQQLKALLNH  68
hZIP4   197  ----------HA--LPSPQY----FVDFVFQQHSSEVPMTLAELSALMQR  230

hZIP14   69  LDVGVGRGNVTQHVQGHRNLST-------------------CFSSGDLFT  99
hZIP4   231  L--GVGREAHSDHSHRHRGASSRDPVPLISSSNSSSVWDTVCLSARDVMA  278

hZIP14  100  AHNFSEQSRIGSSELQEFCPTILQQLDSRACTSENQENEENEQTEEGRPS  149
hZIP4   279  AYGLSEQAGVTPEAWAQLSPALLQQQLSGACTSQSRPPVQDQLSQSER--  326
```

```
hZIP14  150  AVEVWGYGLLCVTVISLCSLLGASVVP---FMKKTFYKRLLLYFIALAIG  196
hZIP4   327  ----YLYGSLATLLICLCAVFGLLLLTCTGCRGVTHY--ILQTFLSLAVG  370

hZIP14  197  TLYSNALFQLIPEAFGFNPLEDYYVS-----KSAVVFGGFYLFFFTEKIL  241
hZIP4   371  AVTGDAVLHLTPKVLGLHTHSEEGLSPQPTWRLLAMLAGLYAFFLFENLF  420

hZIP14  242  KILLKQKNE-----------HHH-GHSHYASESLPSKKDQEEGVMEKLQN  279
hZIP4   421  NLLLPRDPEDLEDGPCGHSSHSHGGHSH--------------GVSLQLAP  456

hZIP14  280  GDLDHMIPQHCSSELDGKAPMVDEKVIVGSLSVQDLQASQSACYWLKGVR  329
hZIP4   457  SELRQPKPPHEGSRAD----------LVAEESPELLNPEPR--------R  488

hZIP14  330  YS-DIGTLAWMITLSDGLHNFIDGLAIGASFTVSVFQGISTSVAILCEEF  378
hZIP4   489  LSPELRLLPYMITLGDAVHNFADGLAVGAAFASSWKTGLATSLAVFCHEL  538

hZIP14  379  PHELGDFVILLNAGMSIQQALFFNFLSACCCYLGLAFGILAG-SHFSANW  427
hZIP4   539  PHELGDFAALLHAGLSVRQALLLNLASALTAFAGLYVALAVGVSEESEAW  588

hZIP14  428  IFALAGGMFLYISLADMFPEMNEVCQEDERKGSILIPFIIQNLGLLTGFT  477
hZIP4   589  ILAVATGLFLYVALCDMLPAMLKV--RDPRPWLL---FLLHNVGLLGGWT  633

hZIP14  478  IMVVLTMYSGQIQIG                                    492
hZIP4   634  VLLLLSLYEDDITF-                                    647
```

A.4 Pfam-Derived Alignment of the Transmemembrane Domain

```
BbZIP    145  YFTPHE-------------------------------------------  150
hZIP14   151  -----VEVWGYGLLCVTVISLCSLLGASVVPFMKKTFYKRLLLYFIALAI  195

BbZIP    151  -------------------------------------------------  151
hZIP14   196  GTLYSNALFQLIPEAFGFNPLEDYYVSKSAVVFGGFYLFFFTEKILKILL  245

BbZIP    151  -------------------------------------------------  151
hZIP14   246  KQKNEHHHGHSHYASESLPSKKDQEEGVMEKLQNGDLDHMIPQHCSSELD  295

BbZIP    151  -------HER----------------------TGHQGPEAARVNRVWLFVL  172
hZIP14   296  GKAPMVD---EKVIVGSLSVQDLQASQSACYWLKGVRYSDIGTLAWMITL  342

BbZIP    173  TIILHNLPEGMAIGVSFATGDLRIGLPLTSAIAIQDVPEGLAVALALRAV  222
hZIP14   343  SDGLHNFIDGLAIGA-SFTVSVFQGISTSVAILCEEFPHELGDFVILLNA  391

BbZIP    223  GLPIGRAVLVAVASGLMEPLGALVGVGISSGFALAYPISMGLAAGAMIFV  272
hZIP14   392  GMSIQQALFFNFLSACCCYLGLAFGIL--AGSHFSANWIFALAGGMFLYI  439

BbZIP    273  VSHEVIP---ETHRNG--H------ETTATVGLMAGFAVMMFLD        305
hZIP14   440  SLADMFPEMNEVCQEDERKGSILIPFIIQNLGLLTGFTIMVVLT        483
```

A.5 Profile-Derived Alignment of the Transmemembrane Domain

```
BbZIP      1  MNQ----------------------------------P-S-S--------  6
hZIP14     1  MKLLLLHPAFQSCLLLTLLGLWRTTPEAHASSLGAPAISAASFLQDLIHR  50

BbZIP      7  -------------------------------L----AADLRGAWHAQAQS-  21
hZIP14    51  YGEGDSLTLQQLKALLNHLDVGVGRGNVTQ-HVQG--------------H  85

BbZIP     22  HPL---ITLGLAASAAGV--VLLLV-AGI--V--NA-LT-G-E-NRVHVG  57
hZIP14    86  RNLSTCFSSGDLFTAHNFSEQSRIGSSELQEFCPTILQQLDSRACTSENQ  135

BbZIP     58  YA--------------VLG---GAAGFAATALGALMALGL-----RAISA  85
hZIP14   136  ENEENEQTEEGRPSAVEV-WGYGLLCVTVISLCSLLGASVVPFMKKTFYK  184

BbZIP     86  RTQDAMLGFAAGMMLAASAFSLILP--GLDAAGTI-VGPGPAA-AAVVAL  131
hZIP14   185  RLLLYFIALAIGTLYSNALFQLIPEAFGFNPLEDYYV-----SKSAVVFG  229

BbZIP    132  GLGLGVLLMLGLDYFTP--------HEHERTG--H---------------  156
hZIP14   230  GFYLFFFTEKILKILLKQKNEHHHGHSHYASESLPSKKDQEEGVMEKLQN  279

BbZIP    157  ----------------------------------------Q-G-P-E-  160
hZIP14   280  GDLDHMIPQHCSSELDGKAPMVDEKVIVGSLSVQDLQASQSACYWL-KGV  328

BbZIP    161  -AA-RVNRVWLFVLTIILHNLPEGMAIGVSFATGDLRIGLPLTSAIAIQD  208
hZIP14   329  RYSDIGTLAWMITLSDGLHNFIDGLAIGASFTVS-VFQGISTSVAILCEE  377
```

```
BbZIP   209  VPEGLAVALALRAVGLPIGRAVLVAVASGLMEPLGALVGVGISSGFALAY  258
hZIP14  378  FPHELGDFVILLNAGMSIQQALFFNFLSACCCYLGLAFGILAG-SHFS-A  425

BbZIP   259  PISMGLAAGAMIFVVSHEVIPETH-RNGHETTAT-------------VGL  294
hZIP14  426  NWIFALAGGMFLYISLADMFPEMNEVCQEDE---RKGSILIPFIIQNLGL  472

BbZIP   295  MAGFALMMFLDTALG-----                                309
hZIP14  473  LTGFTIMVVLTMYSGQIQIG                                492
```

A.6 hZIP4-Derived Alignment of the Transmemembrane Domain

The first step was to create a profile-profile alignment of hZIP14 and hZIP4:

```
hZIP14    1  MKLLL-------------------------------------------  5
hZIP4     1  MAS--LVSLELGLLLAVLVVTATASPPAGLLSLLTSGQGALDQEALGGLL  48

hZIP14    6  -----------LHPA--------------FQS------------------  12
hZIP4    49  NTLADRVHCAN----GPCGKCLSVEDALGLGEPEGSGLPPGPVLEARYVA  94

hZIP14   13  ------CLLLTL------------LGLW-R--TTPEAHASSLGA------  35
hZIP4    95  RLSAAAVLYLSNPEGTCEDARAGL---WASHAD-----------HLLALL  130

hZIP14   35  -----------------------------------------------  35
hZIP4   131  ESPKALTPGLSWLLQRMQARAAGQTPKMACVDIPQLLEEAVGAGAPGSAG  180

hZIP14   36  -----------------PAISAASFLQDLIHRYGEGDSLTLQQLKALLNH  68
```

```
hZIP4    181  GVLAALLDHVRSGSCFHALPSPQYFVDFVFQQHSSEVPMTLAELSALMQR  230

hZIP14    69  LDVGVG----------RGNVTQ--HVQG--------------HRNLSTCF  92
hZIP4    231  LGVGREAHSDHSHRHR------GAS---SRDPVPLISSSNSSSVWDTVCL  271

hZIP14    93  SSGDLFTAHNFSEQSRIGSSELQEFCPTILQQLDSRACTSENQENEENEQ  142
hZIP4    272  SARDVMAAYGLSEQAGVTPEAWAQLSPALLQQQLSGACTSQSR----P--  315

hZIP14   143  TEEGRPSAVEVWGYGLLCVTVISLCSLLGASVVPFMK-KTFYKRLLLYFI  191
hZIP4    316  PVQDQLSQSERYLYGSLATLLICLCAVFGLLLLTCTGCRGVTHYILQTFL  365

hZIP14   192  ALAIGTLYSNALFQLIPEAFGFNPLEDYYV-------SKSAVVFGGFYLF  234
hZIP4    366  SLAVGAVTGDAVLHLTPKVLGLHTHSEE--GLSPQPTWRLLAMLAGLYAF  413

hZIP14   235  FFTEKILKILLKQKN-EH------HHG-HS--HYASESLP--SKKDQE--  270
hZIP4    414  FLFENLFNLLLPRD-PEDLEDGPCGHSSHSHGGHSHGVSLQLAPSELRQP  462

hZIP14   271  EGVMEKLQNGDLDHMIPQHCSSELDGKAPMVDEKVIVG-SLSVQDLQ---  316
hZIP4    463  K---------------------------------PPHEG--SRADLVAEE  477

hZIP14   317  -----ASQSACYWLKGVR-YSDIGTLAWMITLSDGLHNFIDGLAIGASFT  360
hZIP4    478  SPELL--NPEPR-----RLSPELRLLPYMITLGDAVHNFADGLAVGAAFA  520

hZIP14   361  VSVFQGISTSVAILCEEFPHELGDFVILLNAGMSIQQALFFNFLSACCCY  410
hZIP4    521  SSWKTGLATSLAVFCHELPHELGDFAALLHAGLSVRQALLLNLASALTAF  570

hZIP14   411  LGLAFGILAG-SHFSANWIFALAGGMFLYISLADMFPEMNEV-CQEDERK  458
```

```
hZIP4   571  AGLYVALAVGVSEESEAWILAVATGLFLYVALCDMLPAMLKVRDPR-P--  617

hZIP14  459  GSILIPFIIQ-NLGLLTGFTIMVVLTMYSGQIQIG                492
hZIP4   618  ---WLLFLLHN-VGLLGGWTVLLLLSLYEDDITF-                647
```

The second step was to combine this alignment with the BbZIP-hZIP4 alignment by [241] to generate the hZIP4-derived alignment:

```
TM 1
BbZIP    54  VHVGYAVLGGAAGFAATALGALMALGL  80
hZIP14  152  --WGYGLLCVTVISLCSLLGASVVPFM  178

TM 2
BbZIP    85  ARTQDAMLGFAAGMMLAASAFSLILPGLDAAGTI  118
hZIP14  184  KRLLLYFIALAIGTLYSNALFQLIPEAFGFNPLE  217

TM 3
BbZIP   123  PAAAAVVALGLGLGVLLMLGLDYF  146
hZIP14  222  -----SKSAVVFGGFYLFFFTEKI  240

TM 4
BbZIP   166  RVWLFVLTIILHNLPEGMAIGVSF  189
hZIP14  336  LAWMITLSDGLHNFIDGLAIGASF  359

TM 5
BbZIP   194  LRIGLPLTSAIAIQDVPEGLAVALALRA  221
hZIP14  363  VFQGISTSVAILCEEFPHELGDFVILLN  390

TM 6
```

```
BbZIP   226  IGRAVLVAVASGLMEPLGALVGVGIS  251
hZIP14  395  IQQALFFNFLSACCCYLGLAFGILAG  420

TM 7
BbZIP   255  ALAYPISMGLAAGAMIFVVSHE  276
hZIP14  422  HFSANWIFALAGGMFLYISLAD  443

TM 8
BbZIP   287  ETTATVGLMAGFALMMFLDTA  307
hZIP14  465  FIIQNLGLLTGFTIMVVLTMY  485
```

A.7 Final Structural Alignment of the Transmemembrane Domain

This alignment excludes the N-terminus and the TM3-4 intracellular loop.

```
BbZIP    52  NRVHVGYAVLGGAAGFAATALGALMALGLR-AISARTQDAMLGFAAGMML  100
hZIP14  154  ----WGYGLLCVTVISLCSLLGASVVPFMKKTFYKRLLLYFIALAIGTLY  199

BbZIP   101  AASAFSLILPGLDAAGTIVG--PGPAAAAVVALGLGLGVLLMLGLDYFT-  147
hZIP14  200  SNALFQLIPEAFGFNPLEDYYV---SKSAVVFGGFYLFFFTEKILKILLK  246

BbZIP   164  ------------------------VNRVWLFVLTIILHNLPEGMAIGVSF  189
hZIP14  247  QKNEHHHGHSHACYWLKGVRYSDIGTLAWMITLSDGLHNFIDGLAIGASF  359

BbZIP   190  ATGDLRIGLPLTSAIAIQDVPEGLAVALALRAVGLPIGRAVLVAVASGLM  239
```

```
hZIP14  360  -TVSVFQGISTSVAILCEEFPHELGDFVILLNAGMSIQQALFFNFLSACC  408

BbZIP   240  EPLGALVGVGISSGFALAYPISMGLAAGAMIFVVSHEV-----------  277
hZIP14  409  CYLGLAFGIL--AGSHFSANWIFALAGGMFLYISLADMFPEMNEVCQEDE  456

BbZIP   286  -H------ETTATVGLMAGFALMMFLDTAL                      308
hZIP14  457  RKGSILIPFIIQNLGLLTGFTIMVVLTMYS                      486
```

Appendix B

Alignment Development

B.1 Percentage Identity and Similarity Calculation for the Transmembrane Domain

Alignment files were in FASTA format, regions to check were in the ‘checks’ array and region selected was determined by the ‘ind’ variable:

```
aas = ['A','C','D','E','F','G','H','I','K','L','M','N','P','Q','R',
'S','T','V','W','Y']
sim_grps = {'A':'S','C':'','D':'EN','E':'DKQ','F':'WY','G':'',
'H':'NY','I':'LMV','K':'EQR','L':'IMV','M':'ILV','N':'DHS','P':'',
'Q':'EKR','R':'KQ','S':'ANT','T':'S','V':'ILM','W':'FY','Y':'FHW'}

checks = ['',(54,80),(85,118),(123,146),(166,189),(194,221),
(226,251),(255,276),(287,307)]
ind = 2
```

```
check = checks[ind]


names = []
seqs = []


fh = open('../file.fa')
ct = fh.readlines()


for i in range(len(ct)):
    ct[i].rstrip()
    ct[i].rstrip()
    if(ct[i][0] == '>'):
        names.append(ct[i])
        seqs.append('')
    else:
        seqs[len(seqs)-1] += ct[i]


sims = 0
idents = 0
tot = 0
cnt = 0


print('TM'+str(ind))


for j in range(min(len(seqs[0]),len(seqs[1]))):
    ltr0 = seqs[0][j]
    ltr1 = seqs[1][j]
```

```
    if(ltr0 != '-'):
        cnt += 1
    if(cnt >= check[0] and cnt <= check[1]):
        print(ltr0+","+ltr1)
        if(ltr0 != '-'):
            tot += 1
        if(ltr0 != '-' and ltr1 != '-'):
            if(ltr0 == ltr1):
                idents += 1
                sims += 1
            elif(ltr1 in sim_grps[ltr0]):
                sims += 1
print ('tot aas = '+str(tot)+"\n"+'% identity = '+str((idents/tot)
*100)+"\n"+'% similarity = '+str((sims/tot)*100))
```

B.2 Conservation Scores for the Transmembrane Domain

B.2.1 BbZIP

M 2.216 N -1.411 Q -1.328 P -1.444 S -1.434 S -1.395 L -1.000 A -1.311 A -1.373 D -1.178 L -1.311 R -1.295 G -1.332 A -1.078 W -0.740 H -1.241 A -1.334 Q -1.104 A -1.053 Q -1.308 S -1.364 H -0.851 P -0.978 L -1.061 I -1.171 T -0.745 L -0.847 G -0.641 L -0.103 A -0.741 A -0.587 S -0.822 A -0.469 A -0.560 G -0.620 V -0.177 V -0.619 L -0.857 L -0.563 L -0.847 V -0.835 A -1.111 G -0.953 I -0.841 V -0.796 N -1.233 A -1.199 L -0.848 T -1.457 G -1.247 E -1.356 N -1.307 R -1.204 V -1.303 H -1.226 V -0.602 G -1.366 Y -1.414 A 0.811 V -0.461

L -1.122 G 0.657 G 1.393 A -0.648 A -0.494 G 0.442 F 0.200 A -0.594 A -0.218 T 2.216 A 0.531 L -0.329 G 2.216 A 0.924 L -0.416 M -0.653 A 0.259 L -0.291 G -0.670 L -0.707 R -0.371 A -0.879 I -0.156 S -0.856 A -1.131 R -0.289 T -0.795 Q -0.420 D 1.144 A -0.828 M 0.561 L 0.485 G 1.541 F -0.396 A 0.521 A 0.594 G 2.216 M 0.947 M 2.216 L 0.636 A 0.894 A 2.216 S -0.160 A -0.359 F 1.541 S 1.564 L 2.216 I 0.838 L 0.188 P 2.216 G 0.826 L 0.422 D -0.057 A -0.330 A 0.496 G -0.956 T -1.142 I -0.814 V -0.975 G -1.021 P -1.001 G -1.091 P -1.133 A -1.296 A 0.191 A -0.099 A -0.682 V -0.525 V -0.056 A -0.175 L -0.722 G 0.923 L 0.061 G -0.645 L -0.118 G 2.216 V 0.108 L -0.769 L -0.318 M 0.506 L -0.555 G -0.626 L -0.080 D 0.287 Y -0.830 F -0.761 T -0.402 P 1.291 H 0.534 E -0.667 H 0.470 E -0.966 R -1.333 T -1.101 G 0.030 H -1.051 Q -0.979 G 0.553 P -0.918 E -1.062 A -1.404 A -1.185 R -0.903 V -0.984 N -0.772 R 0.597 V 0.169 W -0.128 L 1.300 F 0.971 V 0.776 L 0.027 T 1.260 I 1.123 I -0.080 L 0.762 H 1.717 N 1.717 L 0.250 P 2.216 E 1.717 G 1.717 M 0.742 A 1.467 I 0.628 G 2.216 V 0.830 S 0.562 F 0.127 A 0.248 T -0.777 G -0.459 D -0.340 L -0.954 R -1.247 I -0.724 G 1.018 L -1.100 P 0.046 L 0.189 T 0.522 S -0.386 A 1.311 I 2.216 A 0.392 I 1.012 Q 2.216 D 1.356 V -0.007 P 2.216 E 1.319 G 1.712 L 0.744 A 0.711 V 0.952 A 0.503 L -0.089 A 0.136 L 0.860 R -0.706 A -0.219 V -0.559 G 0.292 L -0.409 P -0.760 I -0.943 G -1.299 R -1.030 A 1.030 V -0.695 L -0.655 V -0.133 A 0.228 V -0.466 A -0.381 S 0.749 G 2.216 L -0.233 M 0.018 E 1.312 P 0.788 L 0.146 G 0.468 A 0.439 L -0.298 V -0.139 G 0.492 V -0.104 G -0.505 I -0.050 S -0.049 S -0.267 G -1.271 F -0.482 A -0.911 L -1.267 A -0.508 Y 0.265 P 1.311 I -0.609 S 0.873 M 0.770 G 0.308 L 0.130 A 1.222 A 1.395 G 1.706 A 0.865 M 1.706 I 0.667 F 0.703 V 0.953 V 1.328 S 0.016 H 0.369 E 0.575 V 0.921 I 0.865 P 2.216 E 0.838 T 0.222 H 0.597 R 1.196 N -0.366 G 0.966 H 0.256 E -0.013 T -1.071 T -1.150 A 1.304 T 1.142 V -0.920 G 1.184 L 0.605 M -0.139 A -0.635 G 2.216 F 1.339 A -0.523 L 0.377 M 0.804 M 0.713 F

-0.442 L 0.317 D 1.510 T -0.187 A -0.667 L 1.058 G 0.166

B.2.2 hZIP14

M -1.279 K -1.061 L -0.807 L -1.066 L -1.288 L -1.431 H -1.728 P -1.616 A -1.736 F -1.763 Q -1.761 S -1.617 C -1.748 L -1.442 L -1.044 L -1.006 T -1.075 L -0.960 L -1.023 G -1.467 L -1.626 W -1.246 R -1.860 T -1.474 T -1.754 P -1.291 E -1.583 A -1.439 H -1.898 A -1.250 S -1.627 S -1.424 L -1.755 G -1.887 A -1.622 P -1.750 A -1.421 I -1.279 S -0.444 A 0.347 A -0.495 S -1.593 F -0.126 L 0.515 Q -0.668 D -0.474 L 0.532 I -0.462 H -1.774 R 0.513 Y 0.430 G 0.477 E -0.793 G -0.769 D -1.455 S -0.152 L 0.350 T 0.348 L -0.125 Q -1.415 Q 0.530 L 1.270 K -0.489 A -1.424 L 1.270 L 0.512 N -1.292 H -1.587 L -0.008 D -1.287 V -0.084 G -0.463 V -1.444 G -1.263 R -1.277 G -1.258 N -0.803 V -1.466 T -1.060 Q -1.613 H -1.580 V -2.011 Q -2.132 G -1.461 H -1.597 R -1.401 N -1.060 L -0.483 S -0.774 T -1.614 C -0.495 F -0.162 S -0.103 S -0.146 G -1.248 D -1.225 L -0.102 F 0.348 T 0.044 A -0.779 H 0.401 N -0.493 F -0.498 S -0.380 E -1.210 Q -1.223 S -0.348 R -0.960 I -0.105 G -0.471 S -1.749 S -1.898 E -0.512 L 0.345 Q -0.394 E -1.367 F -0.041 C 1.270 P 1.270 T 0.509 I -0.164 L -0.001 Q 1.270 Q 1.270 L 1.270 D -0.446 S -0.791 R -0.792 A -0.049 C 0.520 T -1.488 S -1.620 E -1.402 N -1.042 Q -1.424 E -1.014 N -1.258 E -1.266 E -0.781 N -1.441 E -0.500 Q -1.436 T -0.804 E -1.431 E 1.270 G 1.270 R -0.126 P 1.270 S -0.087 A -1.619 V -1.058 E 1.270 V 1.270 W 1.270 G 1.270 Y -0.096 G 1.270 L -0.183 L 0.501 C 0.368 V 1.270 T 0.450 V -0.130 I 0.501 S 0.450 L 0.501 C -0.064 S 1.270 L 1.270 L -1.275 G 1.270 A 0.393 S -1.011 V 1.270 V 0.450 P 1.270 F 0.393 M 0.450 K -0.141 K 0.501 T -0.433 F 0.532 Y -0.103 K -0.445 R 1.270 L 0.393 L 1.270 L -0.064 Y 0.454 F 0.532 I 1.270 A 1.270 L 1.270 A 0.393 I 1.270 G 1.270 T 1.270 L 1.270 Y -0.446 S 1.270 N 0.501 A 1.270 L 0.501 F 1.270 Q 1.270 L 1.270 I 0.501 P 1.270 E 1.270 A 0.538 F 1.270 G 0.023 F 1.270 N

-0.172 P 0.509 L -1.046 E -0.781 D -0.426 Y -0.814 Y 1.270 V 1.270 S -0.014 K 0.548 S 1.270 A 0.455 V 0.539 V 0.548 F 1.270 G 0.539 G 1.270 F 1.270 Y 1.270 L 1.270 F 1.270 F 1.270 F 1.270 T 0.538 E 1.270 K 1.270 I 0.345 L 1.270 K 0.539 I -0.461 L -0.469 L 1.270 K 1.270 Q -0.083 K 0.466 N -1.190 E -1.052 H -0.542 H -1.263 H -0.412 G -0.040 H 0.494 S -0.143 H 0.510 Y -1.081 A -1.426 S -1.261 E -1.061 S -1.267 L -0.848 P -1.276 S -0.124 K -1.286 K -0.852 D 0.437 Q -1.286 E 0.493 E -0.127 G 0.488 V -0.473 M -0.515 E 0.485 K 1.270 L 1.270 Q -0.064 N 0.468 G 0.468 D -0.129 L -0.523 D -0.546 H -0.843 M -0.852 I -0.507 P -1.044 Q -0.826 H -1.050 C -1.754 S -1.086 S -1.270 E 0.345 L -1.227 D -0.523 G -1.269 K -0.542 A -1.089 P -0.780 M -1.460 V -1.467 D -0.811 E -0.819 K 0.347 V -1.234 I -1.452 V -1.257 G -1.036 S -1.039 L -1.456 S -1.061 V -0.811 Q 0.436 D -0.171 L -1.082 Q -0.531 A -0.558 S -1.066 Q -1.071 S -0.137 A -0.526 C 1.270 Y -0.056 W 1.270 L 0.389 K 1.270 G -0.058 V -0.844 R -1.086 Y 1.270 S 0.472 D 0.436 I 1.270 G 1.270 T 1.270 L 1.270 A 1.270 W 1.270 M 1.270 I 1.270 T 1.270 L 1.270 S 1.270 D 1.270 G 1.270 L 1.270 H 1.270 N 1.270 F 1.270 I 1.270 D 1.270 G 1.270 L 1.270 A 1.270 I 1.270 G 1.270 A 1.270 S 1.270 F 1.270 T 1.270 V -0.143 S 1.270 V 1.270 F 1.270 Q 1.270 G 1.270 I 0.376 S 1.270 T 1.270 S 1.270 V 0.419 A 1.270 I 1.270 L 0.473 C 1.270 E 1.270 E 1.270 F 1.270 P 1.270 H 1.270 E 1.270 L 1.270 G 0.447 D -0.065 F 0.385 V -0.065 I -0.065 L -0.065 L 1.270 N -0.065 A -0.065 G -0.065 M -0.065 S -0.430 I -0.491 Q -0.418 Q 0.507 A -0.464 L -0.060 F -0.060 F 0.507 N -0.060 F 0.476 L -0.091 S 0.476 A 1.270 C 0.476 C 0.476 C 0.476 Y 0.476 L -0.110 G -0.060 L -0.129 A -0.114 F 0.476 G 1.270 I 0.476 L -0.084 A 0.476 G 0.476 S 0.345 H -0.796 F 0.476 S -0.059 A -0.486 N 0.435 W 0.476 I 0.476 F 1.270 A 0.476 L 0.476 A 0.476 G 0.476 G 0.476 M 0.476 F 1.270 L 0.476 Y 0.476 I 0.476 S -0.089 L -0.059 A -0.031 D 0.476 M 0.476 F 1.270 P 0.503 E 1.270 M 1.270 N 0.503 E 0.500 V 1.270 C -0.118 Q -0.515 E 1.270 D -0.126 E -0.498

R -1.274 K -1.063 G 0.451 S -0.055 I -1.302 L -0.497 I -1.028 P -1.054 F 0.502 I -0.830 I 0.503 Q 1.270 N 1.270 L -0.112 G 1.270 L 1.270 L 1.270 T 1.270 G 1.270 F 0.502 T -0.825 I 1.270 M 1.270 V 0.435 V 0.356 L 1.270 T 1.270 M -0.798 Y 0.385 S 0.504 G 0.504 Q 0.503 I 0.498 Q 0.498 I -0.125 G 0.498

Appendix C

Quality Assessment and Analysis of Homology Models

C.1 Quality Assessment Results for the N-Terminal Extracellular Domain

C.1.1 Chain A

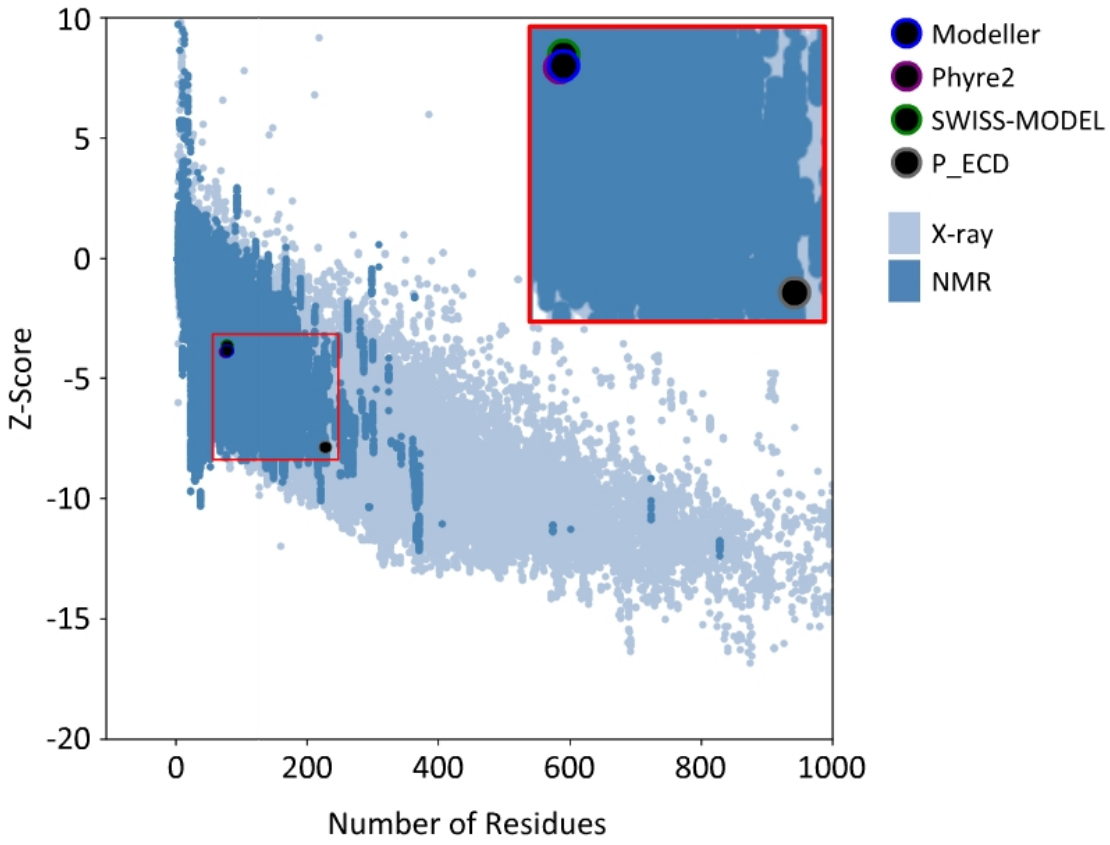


Figure C.1: **ProSA-web Z-scores of overall model quality for the N-terminal extracellular domain of *P. alecto* ZIP14 (P_ECD) and the hZIP14 homology models, relative to those of experimentally determined structures.** The Z-scores of P_ECD and the hZIP14 homology models fall within the range of those observed for experimentally determined structures, indicating the models are of good quality.

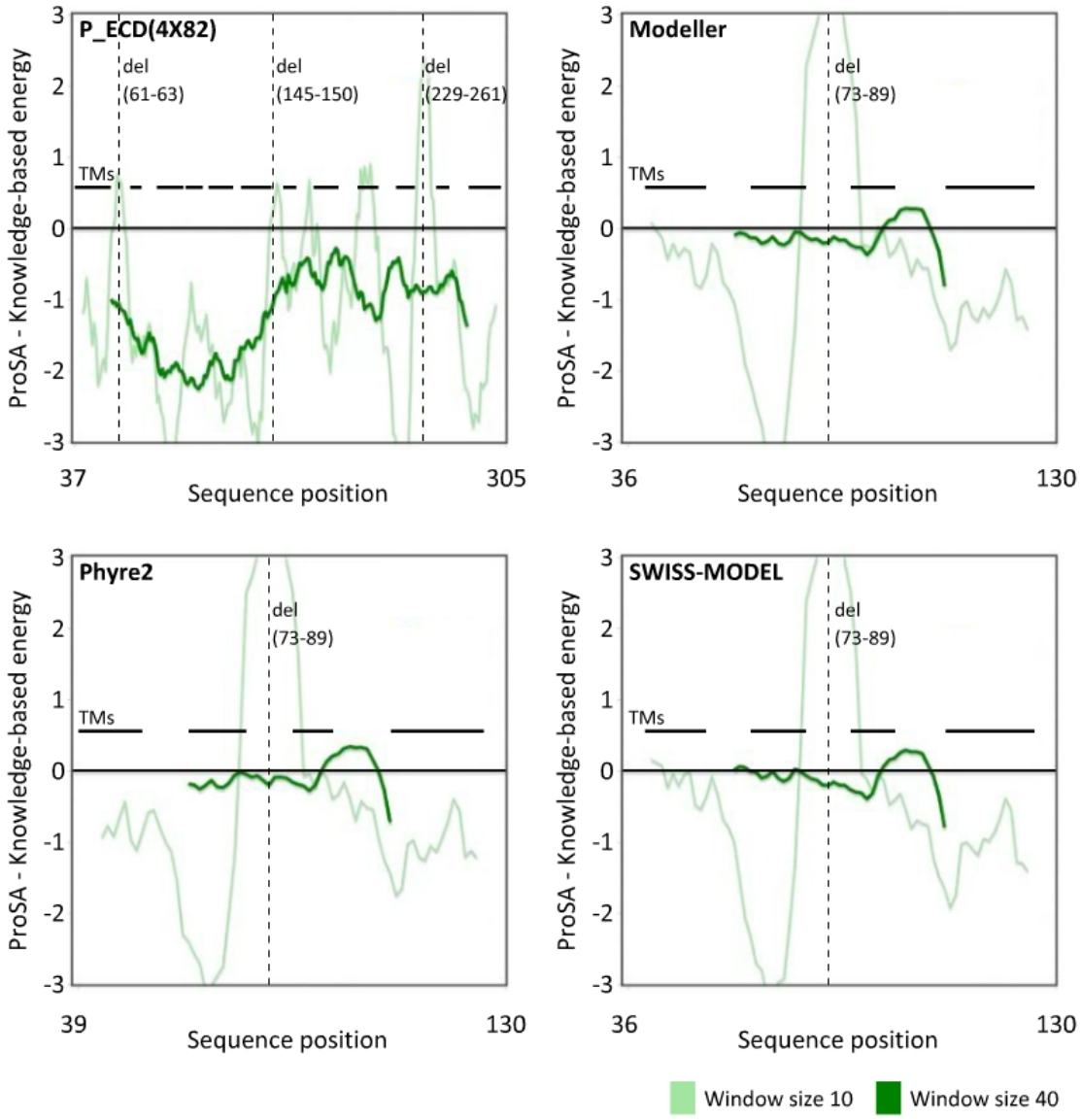


Figure C.2: **ProSA-web energy plots of ECD protein structures.** ProSA-web energy plots for P_ECD and the hZIP14 homology models, smoothed by averaging over 10 and 40 residues, as indicated. Positions of alpha-helices are marked by horizontal black lines.

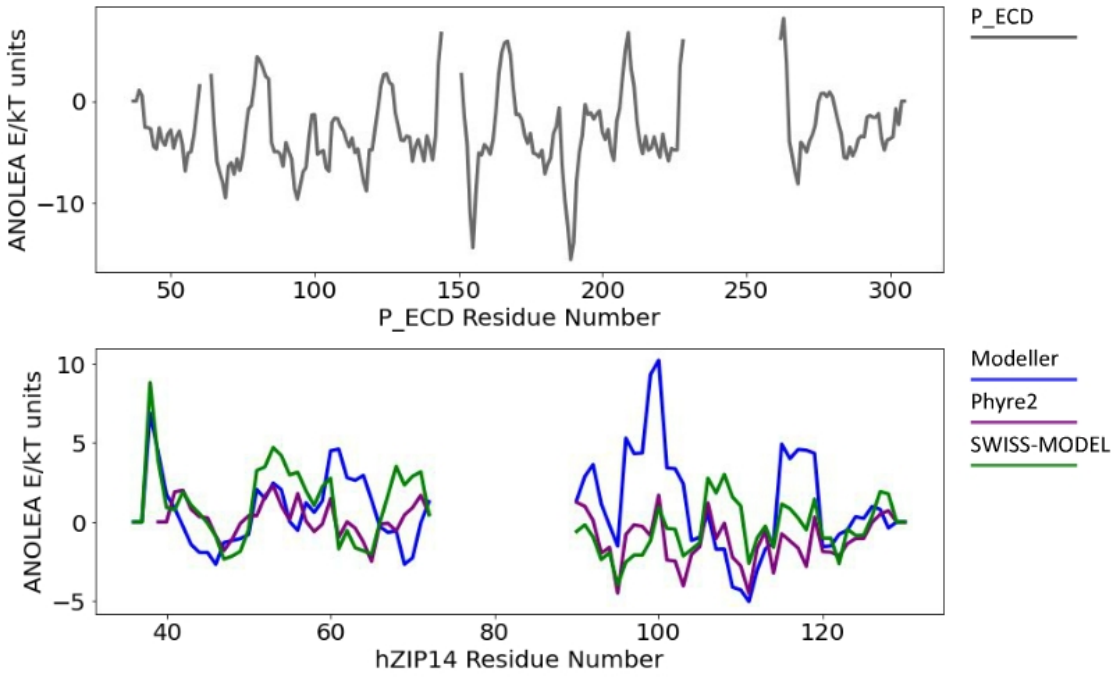


Figure C.3: **ANOLEA energy plots of ECD protein structures.** ANOLEA energy plots smoothed by averaging over 5 residues. Upper panel: ANOLEA energy plot for the P_ECD crystal structure. Lower panel: ANOLEA energy plots for each homology model.

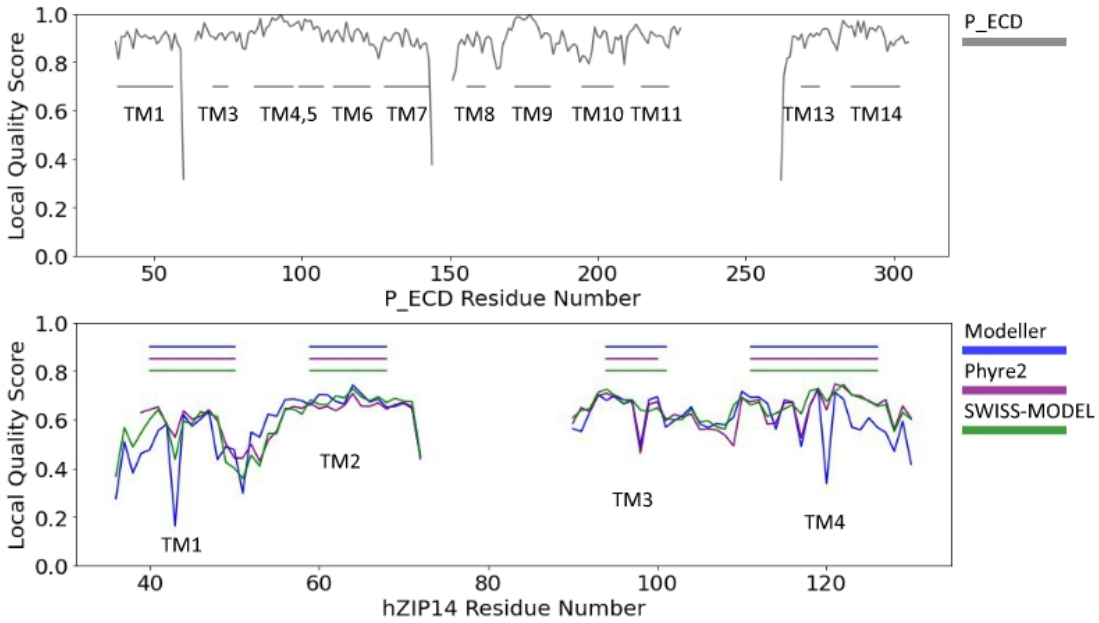


Figure C.4: **QMEANBrane local quality prediction for ECD protein structures.** Graphs of local quality predictions (predicted similarity to the correct protein structure) at each residue of the protein model. Upper panel: Local quality prediction for the P_ECD crystal structure. Lower panel: Local quality prediction for each homology model.

C.1.2 Chain B

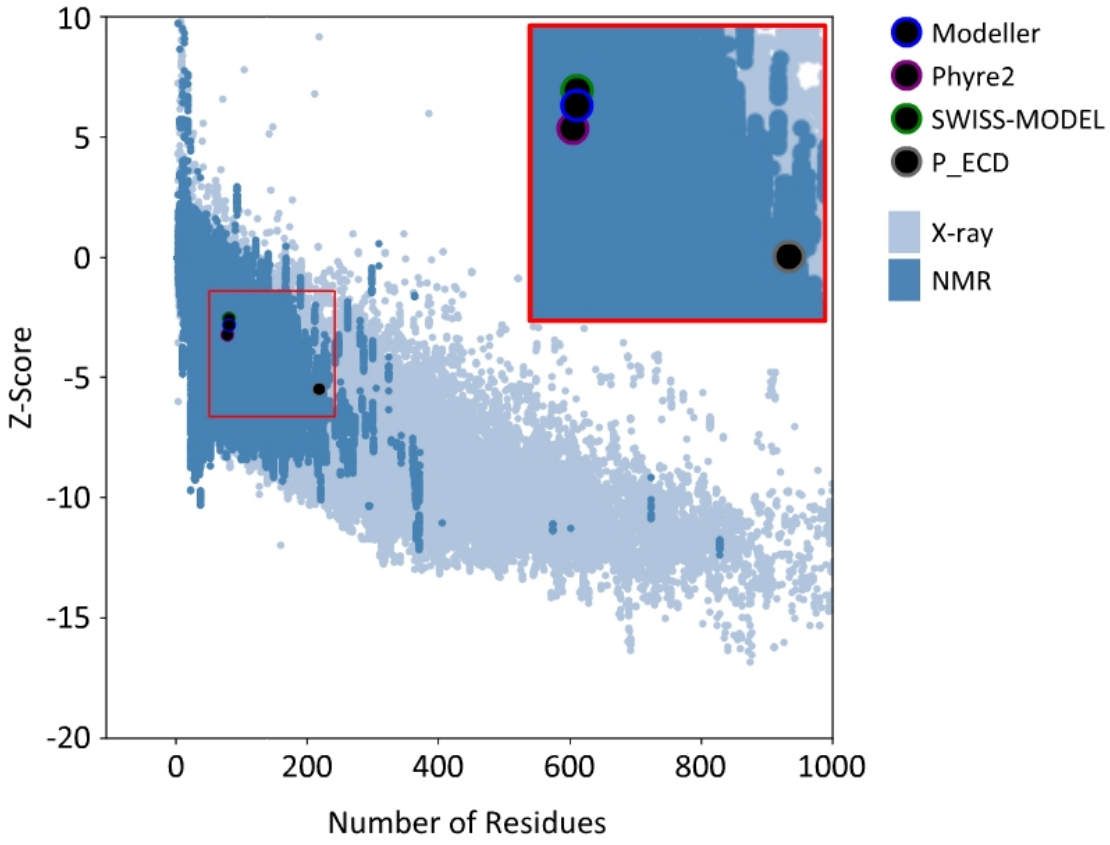


Figure C.5: **ProSA-web Z-scores of overall model quality for the N-terminal extracellular domain of *P. alecto* ZIP14 (P_ECD) and the hZIP14 homology models, relative to those of experimentally determined structures.** The Z-scores of P_ECD and the hZIP14 homology models fall within the range of those observed for experimentally determined structures, indicating the models are of good quality.

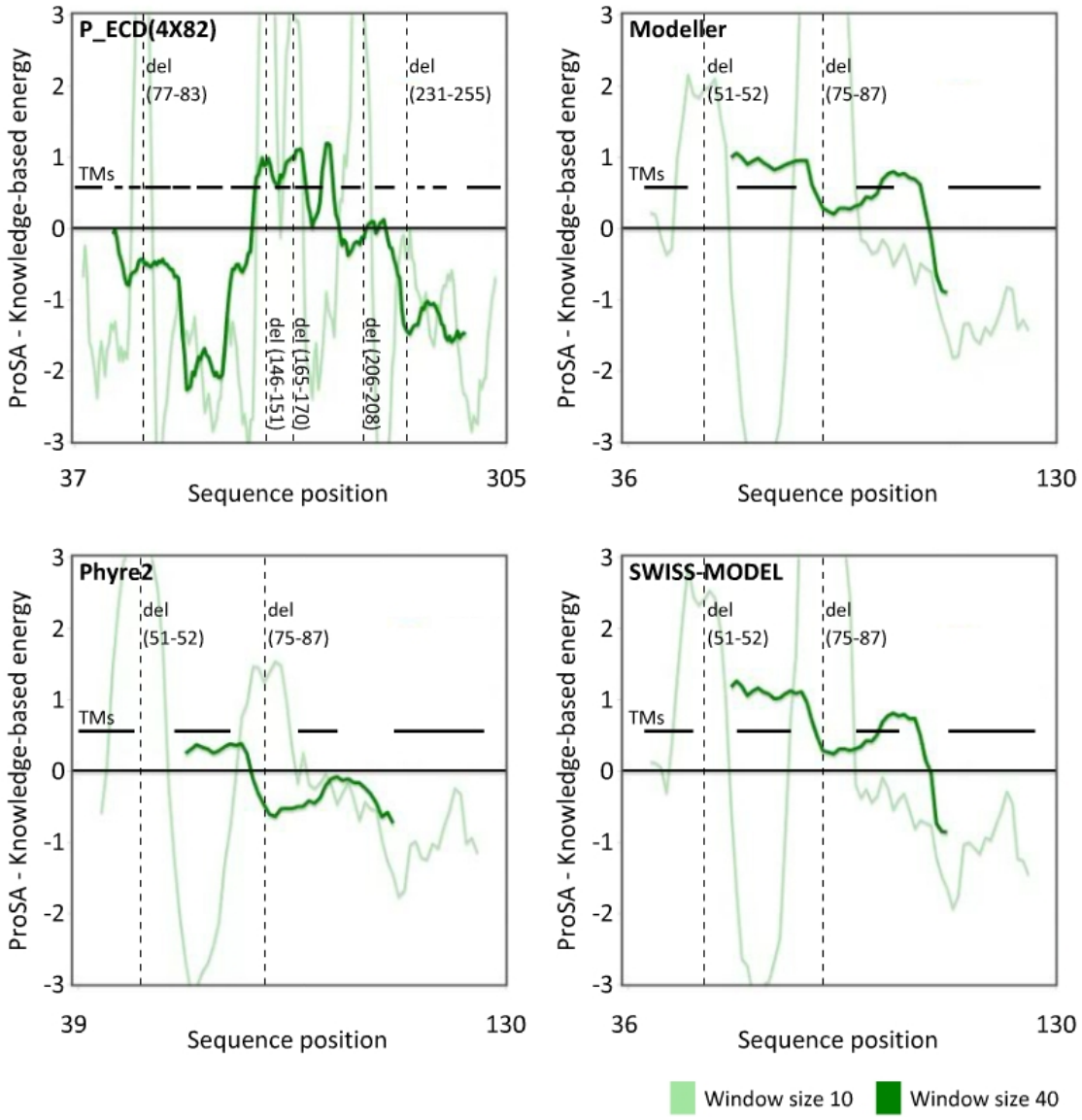


Figure C.6: **ProSA-web energy plots of ECD protein structures.** ProSA-web energy plots for P_ECD and the hZIP14 homology models, smoothed by averaging over 10 and 40 residues, as indicated. Positions of alpha-helices are marked by horizontal black lines.

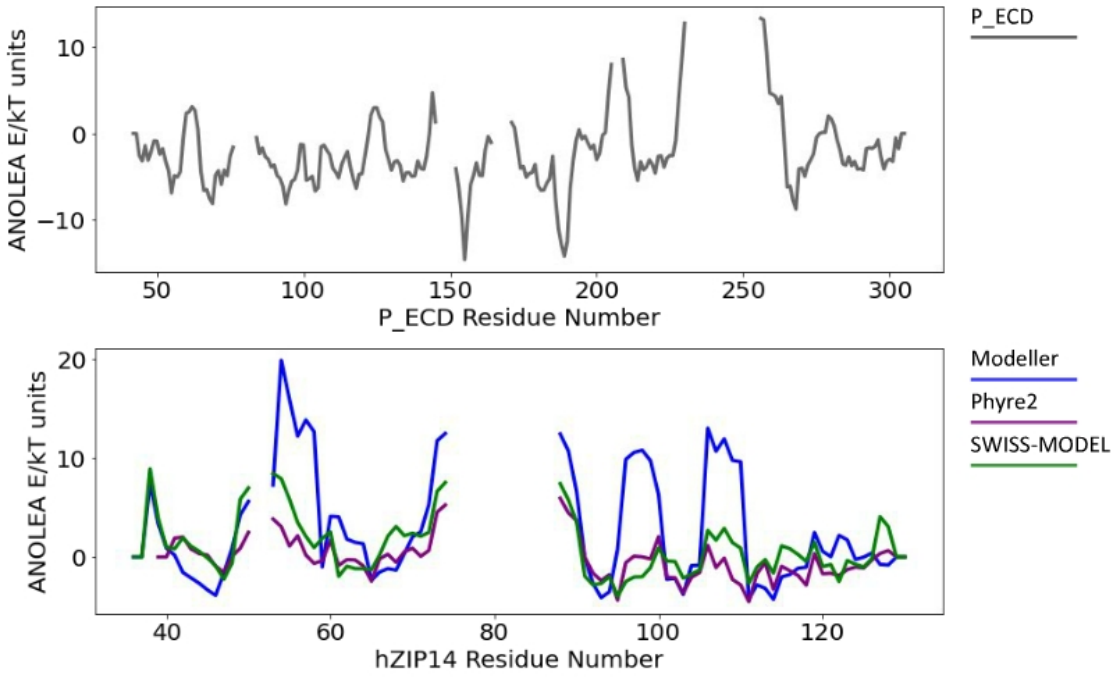


Figure C.7: **ANOLEA energy plots of ECD protein structures.** ANOLEA energy plots smoothed by averaging over 5 residues. Upper panel: ANOLEA energy plot for the P_ECD crystal structure. Lower panel: ANOLEA energy plots for each homology model.

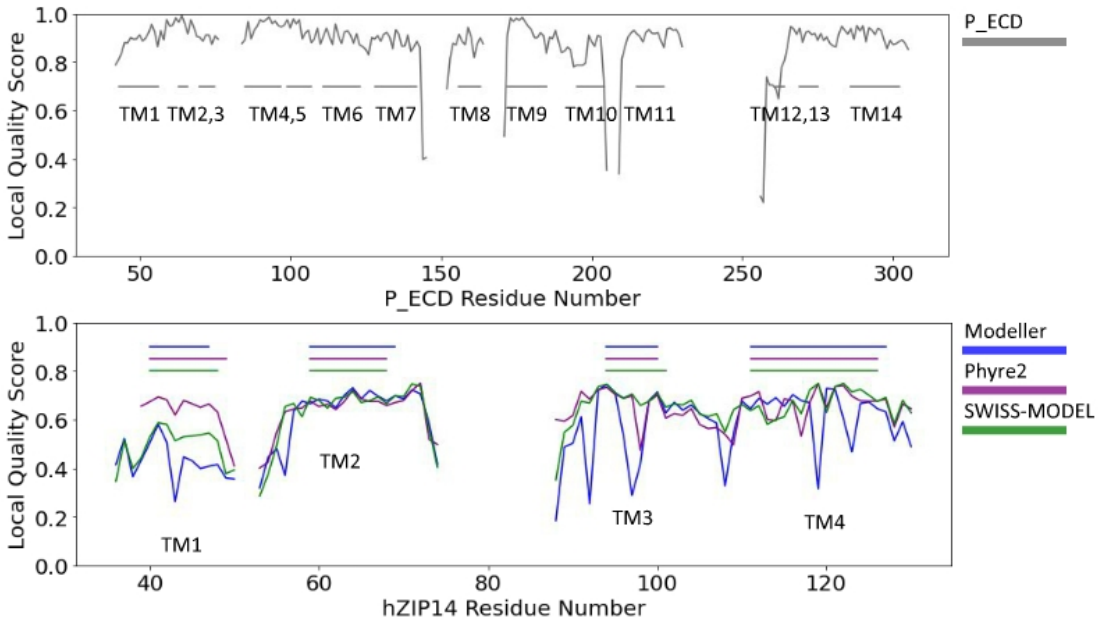


Figure C.8: **QMEANBrane local quality prediction for ECD protein structures.** Graphs of local quality predictions (predicted similarity to the correct protein structure) at each residue of the protein model. Upper panel: Local quality prediction for the P_ECD crystal structure. Lower panel: Local quality prediction for each homology model.

C.2 Scripts for Selecting Mutations in Homology Models

C.2.1 Extracting Interactions Between R-groups of Amino Acid Residues

UCSF Chimera [197] exports a text file of contacts between atoms in amino acids. Part of an example file is shown below:

```
SER 424.A OG    SER 421.A CA    -0.392  3.772
LEU 233.A C     PHE 236.A N     -0.392  3.717
LEU 346.A O     PHE 349.A CA    -0.392  3.752
ARG 185.A H     TYR 183.A C     -0.392  3.092
ASP 384.A N     GLU 381.A C     -0.393  3.718
```

The python script disregards any contacts between atoms in the polypeptide backbone, writing all other contacts to a new file.

```
f = open("...bonds.txt", "r")
f2 = open("...rbonds.txt", "w")
lns = f.readlines()

backbone = ["CA","HA","C","O","N","H"]

for i in range(len(lns)):
  strs = lns[i].split()
  if((strs[2] not in backbone) and (strs[5] not in backbone)):
    f2.write(lns[i])
```

```
f2.close()
```

C.2.2 Ordering Contacts By Conservation Scores

The file of R-group contacts was filtered to remove duplicates and contacts between residues less than 5 residues apart. Contacts were scored using the product of the normalised conservation scores. An example script is shown below:

```
import numpy as np
import matplotlib
import matplotlib.pyplot as plt

cons = np.dtype([('res','U1'),('con','f')])
dt = np.loadtxt("...score_cons.txt",dtype=cons)

dt_a = np.loadtxt("...a_rbonds.txt",dtype='str',usecols=(1,4))
dt_a = np.char.rstrip(dt_a, '.A')

dt_b = np.loadtxt("...b_rbonds.txt",dtype='str',usecols=(1,4))
dt_b = np.char.rstrip(dt_b, '.B')

bonds = []
for i in range(dt_a.shape[0]):
  bonds.append(str(min(int(dt_a[i][0]),int(dt_a[i][1])))+"."
                       +str(max(int(dt_a[i][0]),int(dt_a[i][1]))))

for i in range(dt_b.shape[0]):
```

```
  bonds.append(str(min(int(dt_b[i][0]),int(dt_b[i][1])))+"."
                       +str(max(int(dt_b[i][0]),int(dt_b[i][1]))))

bonds = list(dict.fromkeys(bonds))

dt_vals = np.zeros([len(bonds),3])
for i in range(len(bonds)):
  strs = bonds[i].split(".")
  dt_vals[i][0] = int(strs[0])
  dt_vals[i][1] = int(strs[1])
  dt_vals[i][2] = dt[int(strs[0])-1][1] * dt[int(strs[1])-1][1]

dt_vals = dt_vals[dt_vals[:,1]-dt_vals[:,0]>4]
a = dt_vals[:,2].argsort()
dt_vals = dt_vals[np.flip(a)]
print(dt_vals)
```

Appendix D

Co-evolution Results

D.1 Gremlin Co-evolution Output

For human ZIP14, residues 154-257;321-486, first five lines of data shown below as an example of the format:

i	j	i_id	j_id	r_sco	s_sco	prob
137	200	137_L	200_G	0.4244	3.111	1.000
57	214	57_A	214_A	0.4164	3.052	1.000
133	196	133_F	196_G	0.3927	2.878	1.000
36	255	36_Y	255_G	0.3715	2.723	1.000
40	262	40_L	262_I	0.3412	2.501	1.000

D.2 DeepMetaPSICOV Co-evolution Output

For human ZIP14, all residues, first five lines of data shown below as an example of the format:

```
1 3 0 8  0.7514011263847351
1 4 0 8  0.9989131689071655
1 5 0 8  0.9999839067459106
1 6 0 8  0.09303665906190872
1 7 0 8  0.02880968526005745
```

D.3 Distance Matrix Script

An abbreviated script:

```
import numpy as np
import matplotlib
import matplotlib.pyplot as plt
import math

dt = np.loadtxt("...file.txt",dtype='float',skiprows=1)
dt_less = dt[dt[:,1]-dt[:,0]>4]

dt_close = dt_less[dt_less[:,2]<=0.5]

plt.scatter(dt_less[:,0],dt_less[:,1],color='lightgray')
plt.scatter(dt_less[:,1],dt_less[:,0],color='lightgray')
plt.scatter(dt_close[:,0],dt_close[:,1],color='gray')
plt.scatter(dt_close[:,1],dt_close[:,0],color='gray')

plt.show()
```

Filenames and indices were altered accordingly. A similar script was used to

plot the co-evolution data.

D.4 Comparison of Co-evolution Data to hZIP14 Homology Model

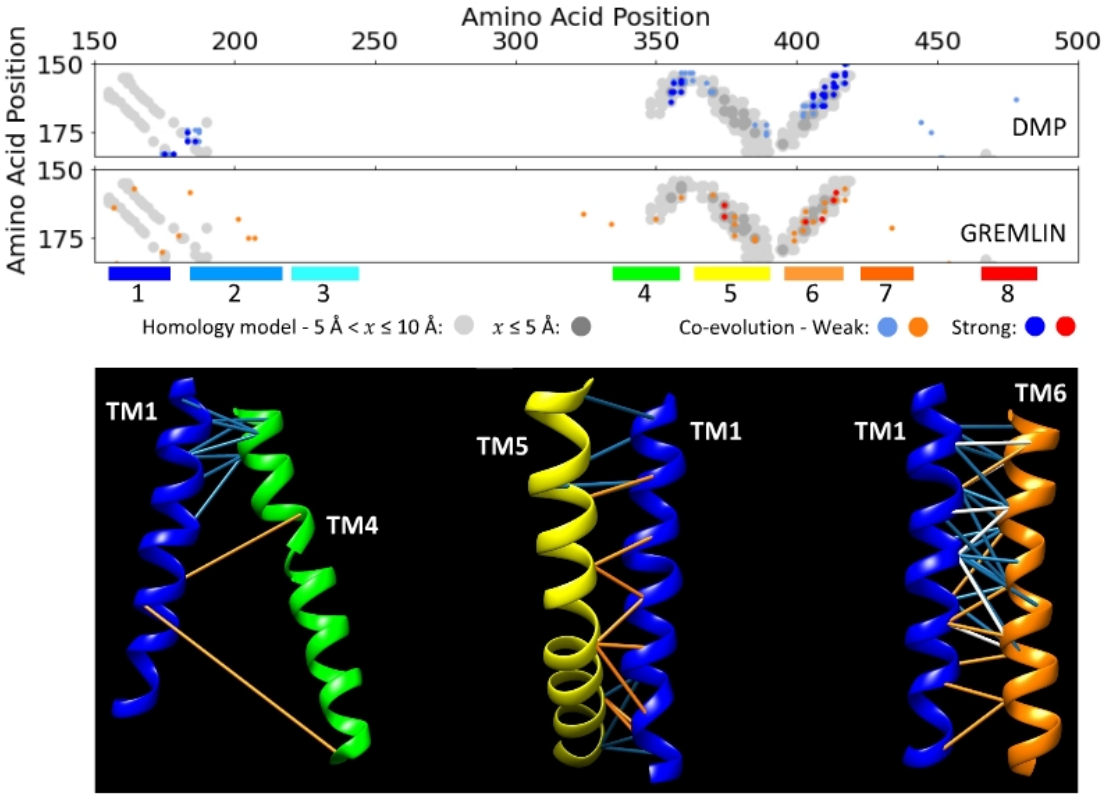


Figure D.1: **Co-evolution data support interactions of TM 1.** Upper panels: Overlay of co-evolution data (from the DeepMetaPSICOV (DMP) and GREMLIN servers) and the hZIP14 homology model distance matrix for interactions of TM 1. Lower panel: Segments of the hZIP14 homology model; lines join the α-carbons of amino acid pairs predicted to co-evolve according to the DeepMetaPSICOV analysis, the GREMLIN analysis, or both (blue lines, orange lines, or white lines, respectively).

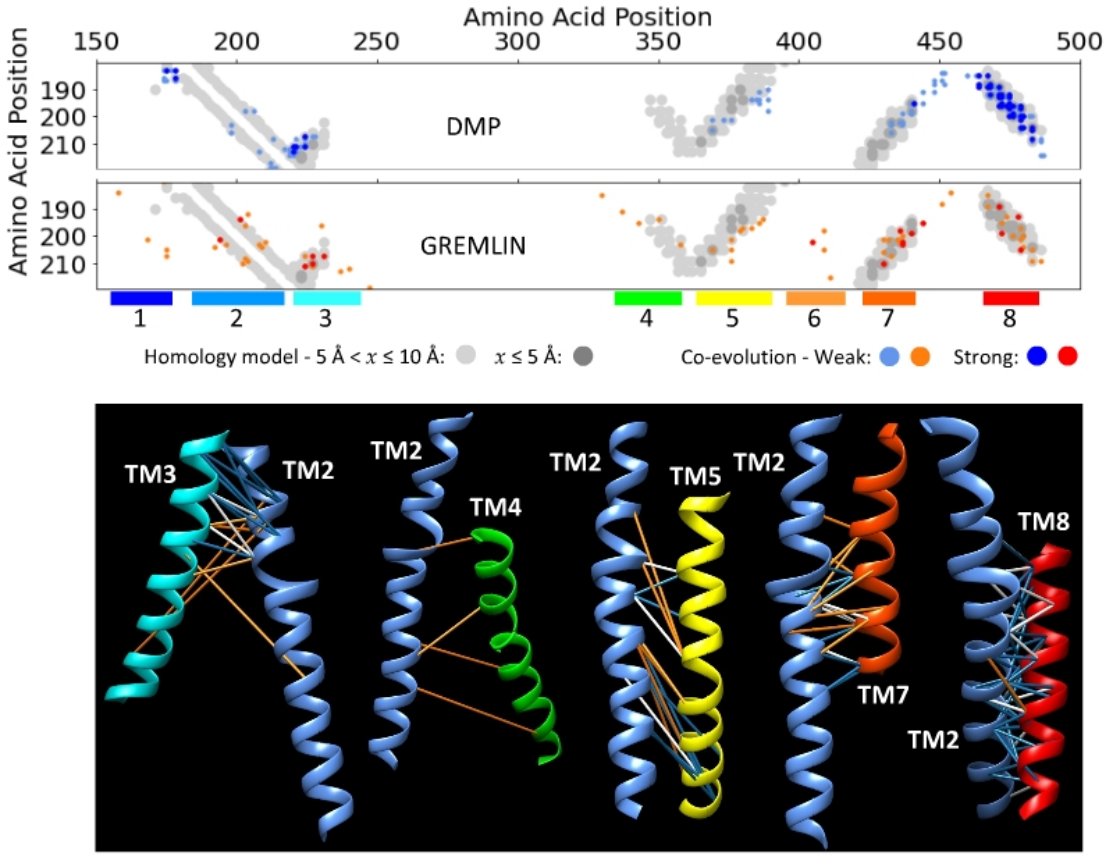


Figure D.2: **Co-evolution data support interactions of TM 2.** Upper panels: Overlay of co-evolution data (from the DeepMetaPSICOV (DMP) and GREMLIN servers) and the hZIP14 homology model distance matrix for interactions of TM 2. Lower panel: Segments of the hZIP14 homology model; lines join the α-carbons of amino acid pairs predicted to co-evolve according to the DeepMetaPSICOV analysis, the GREMLIN analysis, or both (blue lines, orange lines, or white lines, respectively).

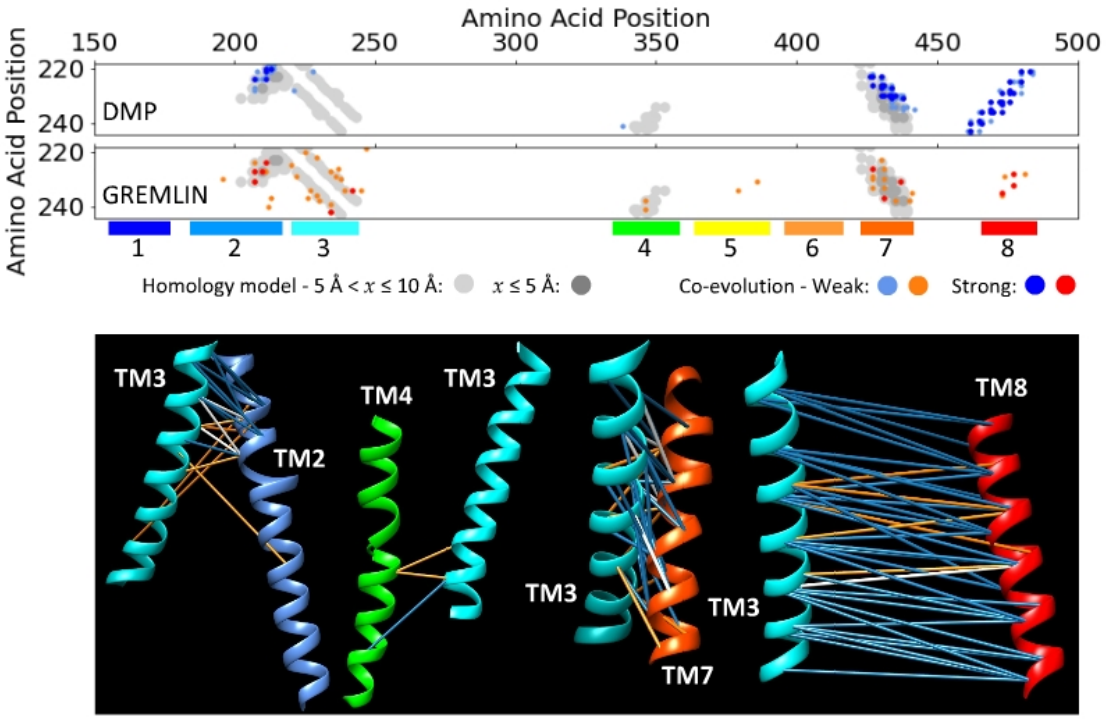


Figure D.3: **Co-evolution data support interactions of TM 3.** Upper panels: Overlay of co-evolution data (from the DeepMetaPSICOV (DMP) and GREMLIN servers) and the hZIP14 homology model distance matrix for interactions of TM 3. Lower panel: Segments of the hZIP14 homology model; lines join the α-carbons of amino acid pairs predicted to co-evolve according to the DeepMetaPSICOV analysis, the GREMLIN analysis, or both (blue lines, orange lines, or white lines, respectively).

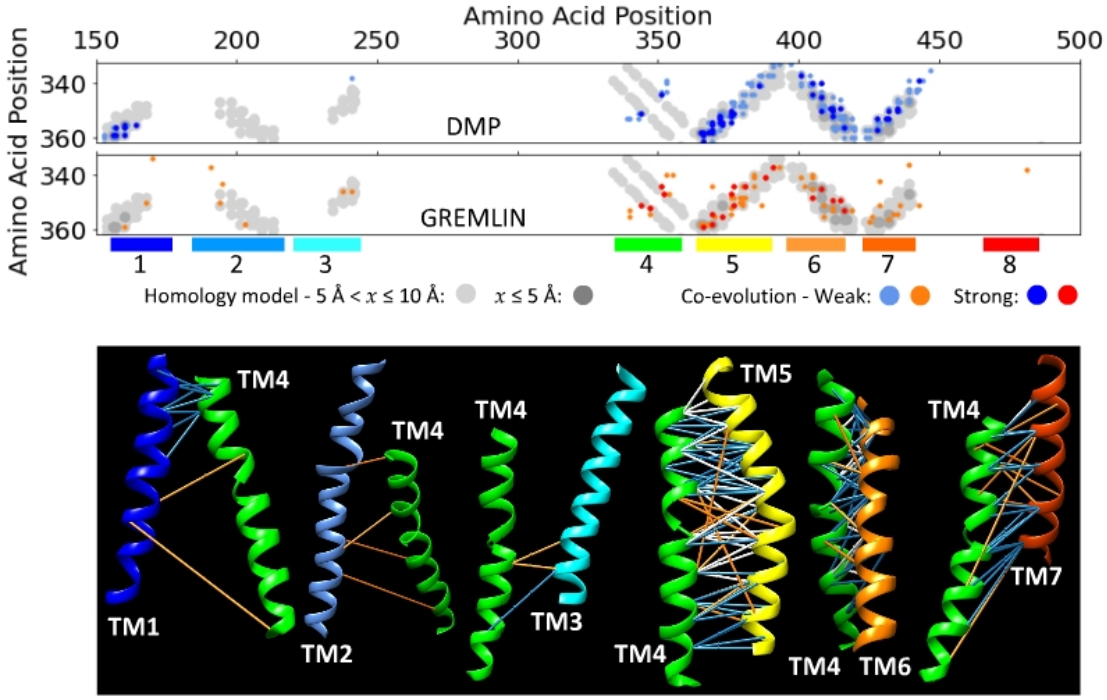


Figure D.4: **Co-evolution data support interactions of TM 4.** Upper panels: Overlay of co-evolution data (from the DeepMetaPSICOV (DMP) and GREMLIN servers) and the hZIP14 homology model distance matrix for interactions of TM 4. Lower panel: Segments of the hZIP14 homology model; lines join the α-carbons of amino acid pairs predicted to co-evolve according to the DeepMetaPSICOV analysis, the GREMLIN analysis, or both (blue lines, orange lines, or white lines, respectively).

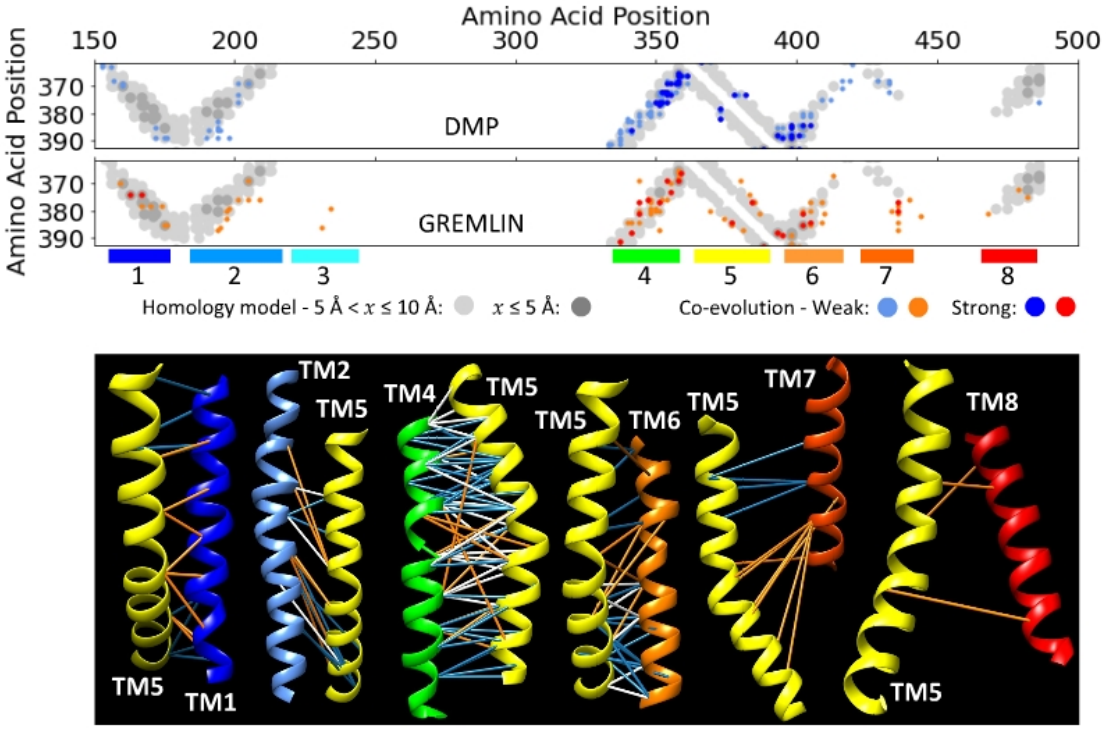


Figure D.5: **Co-evolution data support interactions of TM 5.** Upper panels: Overlay of co-evolution data (from the DeepMetaPSICOV (DMP) and GREMLIN servers) and the hZIP14 homology model distance matrix for interactions of TM 5. Lower panel: Segments of the hZIP14 homology model; lines join the α-carbons of amino acid pairs predicted to co-evolve according to the DeepMetaPSICOV analysis, the GREMLIN analysis, or both (blue lines, orange lines, or white lines, respectively).

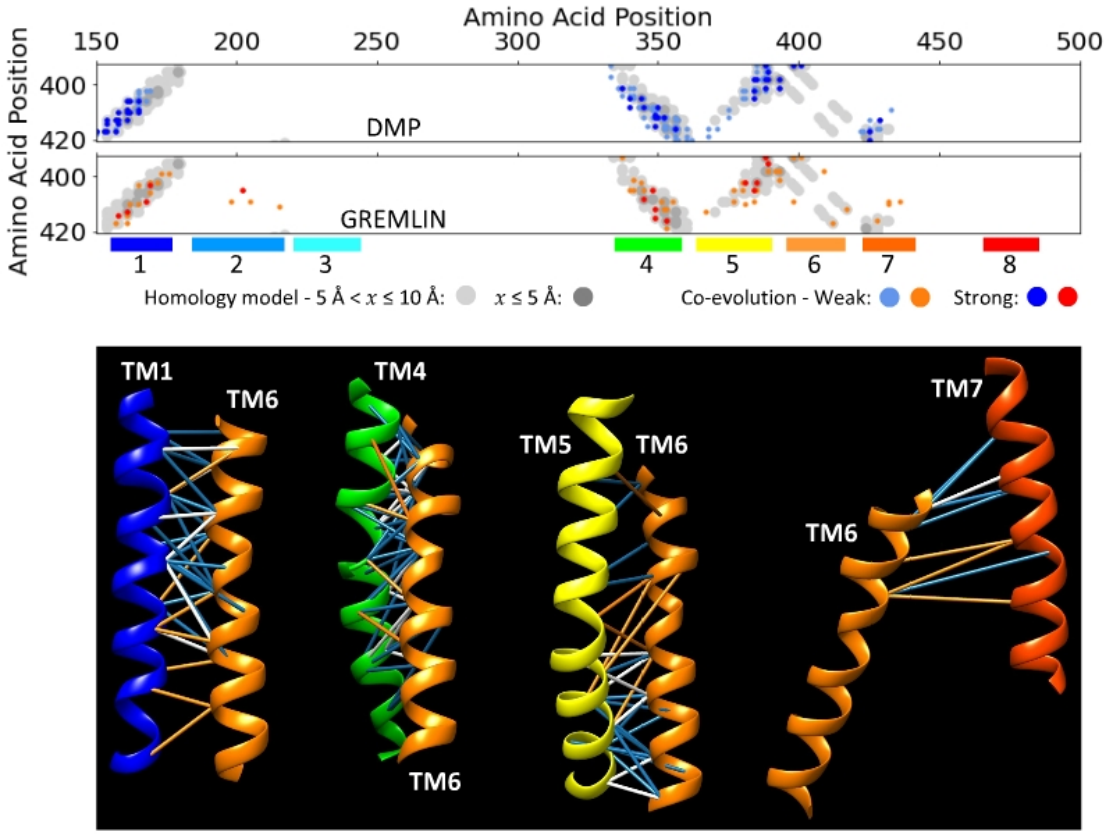


Figure D.6: **Co-evolution data support interactions of TM 6.** Upper panels: Overlay of co-evolution data (from the DeepMetaPSICOV (DMP) and GREMLIN servers) and the hZIP14 homology model distance matrix for interactions of TM 6. Lower panel: Segments of the hZIP14 homology model; lines join the α-carbons of amino acid pairs predicted to co-evolve according to the DeepMetaPSICOV analysis, the GREMLIN analysis, or both (blue lines, orange lines, or white lines, respectively).

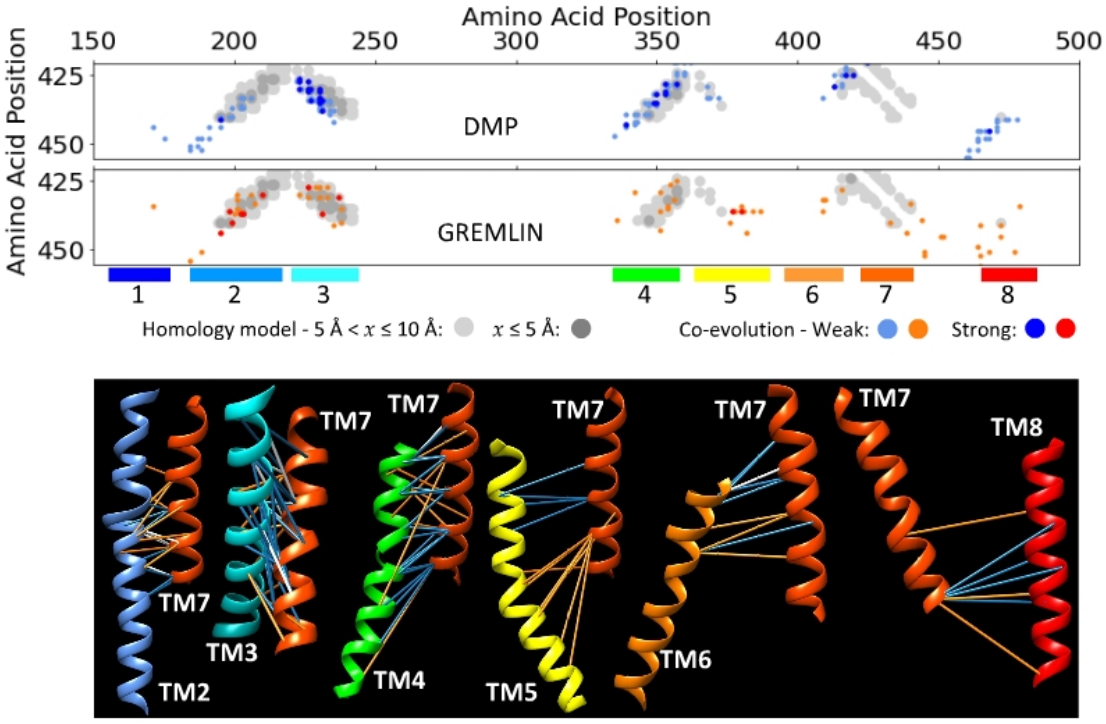


Figure D.7: **Co-evolution data support interactions of TM 7.** Upper panels: Overlay of co-evolution data (from the DeepMetaPSICOV (DMP) and GREMLIN servers) and the hZIP14 homology model distance matrix for interactions of TM 7. Lower panel: Segments of the hZIP14 homology model; lines join the α-carbons of amino acid pairs predicted to co-evolve according to the DeepMetaPSICOV analysis, the GREMLIN analysis, or both (blue lines, orange lines, or white lines, respectively).

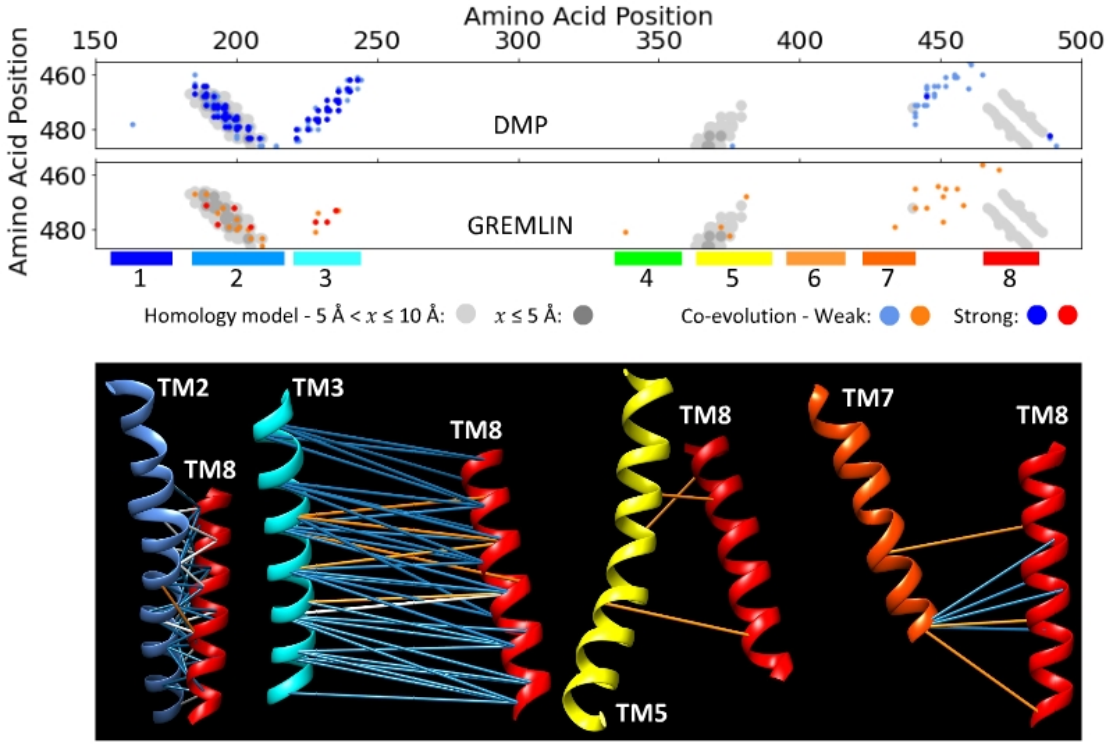


Figure D.8: **Co-evolution data support interactions of TM 8.** Upper panels: Overlay of co-evolution data (from the DeepMetaPSICOV (DMP) and GREMLIN servers) and the hZIP14 homology model distance matrix for interactions of TM 8. Lower panel: Segments of the hZIP14 homology model; lines join the α-carbons of amino acid pairs predicted to co-evolve according to the DeepMetaPSICOV analysis, the GREMLIN analysis, or both (blue lines, orange lines, or white lines, respectively).