Stress Induced Transcriptional Regulation of the Glycine Transporter Type 1A (GlyT-1A/SLC6A9) in Human Intestinal Epithelia

Livingstone K. F. Fultang (BSc, MRes)

Thesis submitted in partial fulfilment of the requirements for the degree of
Doctor of Philosophy

Institute for Cell and Molecular Biosciences
Epithelial Research Group
Faculty of Medical Sciences,
Newcastle University.
England, UK.

January 2015

ABSTRACT

There is mounting experimental evidence demonstrating protection by free glycine against stress in several cell types. The glycine transporter type 1 (GlyT-1) mediates the high affinity supply of glycine, which together with cysteine is required for the synthesis of the antioxidant glutathione. Previous work in this laboratory has established that GlyT-1 is expressed on the apical and basal membranes of intestinal epithelial cells and that its mRNA levels are regulated by stress. In the present study exactly how stress signals to transcriptional induction of GlyT-1 was investigated. Caco-2 cells transfected with reporter constructs of sequences of the GlyT-1a proximal promoter and 5'UTR cloned upstream of a β-galactosidase coding sequence, showed increased reporter activity following treatment with thapsigargin (Tg), tunicamycin (Tu), amino acid (AA) starvation, tert-Butylhydroquinone (tBHQ) or Diethyl maleate (DEM). Despite no changes in Nrf-2 mRNA levels, a significant increase in total Nrf-2 protein abundance was evident on western-blots following DEM treatment of Caco-2 cells. However, gel shift showed no protein-DNA complexes between Nrf-2 protein and a DNA probe sequence of the putative antioxidant response element (ARE) identified in the GlyT-1a 5' flank. Despite a significant siRNA mediated knock-down of Nrf-2 mRNA and protein, there was no further effect on GlyT-1a expression. Unlike Nrf-2, the knock-down of Atf-4 diminished the basal and stressed induced expression of GlyT-1a. Atf-4 was detected bound to DNA probes containing a potential amino acid response element (AARE) located in the first exon of the GlyT-1a gene by gel shift and super shift assays. QPCR assays performed on DNA isolated from Caco-2 cells by chromatin immunoprecipitation (ChIP) using antibodies against Atf-4, demonstrated 9, 5 and 2-fold enrichment of the GlyT-1a AARE following Tu, AA starvation and DEM treatment respectively. Site directed mutation of the GlyT-1a AARE showed a 75% reduction in reporter activity as well as attenuated protein-DNA interaction with a representative probe. It is evident from the data presented in this thesis that the direct interaction of Atf-4 at the proposed GlyT-1a AARE contributes to its transcriptional up-regulation following endoplasmic reticulum stress, nutrient stress and oxidative stress.

PREFACE

This PhD thesis is based on research into the transcriptional regulation of the high affinity glycine transporter (GlyT-1a) conducted by myself, L K F Fultang in the Epithelial Research Group (ERG) of the Institute for Cell and Molecular Biosciences (ICaMB), Newcastle University. Previous work conducted in this laboratory has demonstrated an important cytoprotective role of free glycine against perturbations to normal physiology in epithelial cells lining the human intestine. With an interest in cellular signalling pathways, I set out to investigate the underpinning molecular interactions which lead to the transcriptional regulation in GlyT-1 following physiological stress.

This thesis is the first written account of the research conducted in the last three years. The text is divided into four chapters. In the first chapter (Chapter 1), I review existing knowledge on the importance of the synthesis and transport of free glycine, as well as what is known about its cytoprotective role. This is followed by a review of the multi-faceted stress response pathways. This chapter lays the foundation and motivation for the work reported here. The next chapter (Chapter 2), contains a general description of all the methods used for the work carried out during this project. This includes experimental design for real time quantitative polymerase chain reaction, electrophoretic mobility shift and super-shift assays, chromatin immunoprecipitation (ChIP) assays, reporter assays and *in silico* methods. All experiments and associated protocols were approved by the appropriate departmental authorities and executed under the supervision of Dr Alison Howard and Prof Barry Hugo Hirst. The penultimate chapter (Chapter 3), provides a discussion of the key findings of this project. It begins by summarising observations on the expression of the glycine transporter (GlyT-1a) downstream of stress in Caco-2 cells. This is followed by novel evidence for the regulation of the resulting changes in expression by specific transcription factors. Some of the results described in this chapter have been presented at key international conferences (*viz*: the International Union of Physiological Sciences, IUPS Birmingham 2013, and Experimental Biology, EB San Diego 2014). In the final chapter (Chapter 4), I review and discuss the general implications of the findings from this project, whilst proposing future work that may further improve

This work is dedicated to

Dad, Mum, Mercy, Grace, Joshua and Norman.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

AAR Amino Acid Response pathways   31

AARE Amino Acid Response Element   33

AMPs Antimicrobial Peptides   3

ATCC American Type Culture Collection   44

Atf Activating Transcription Factors (e.g Atf-4, Atf-3, Atf-2, Atf-6)   26

ATP Adenosine Tri-Phosphate   7, 8, 58

BZIP Basic Leucine Zipper protein   41

CD Crohn's Disease   4

CEBP CCAAT enhancer box binding protein   passim

CGI CpG Island   79, 81, 83

ChIP Chromatin Immunoprecipitation   iii

CHOP CEBP homology protein (Also known as DDIT3, CEBPξ and GADD153)   33

C-MARE A MARE element with a CRE-like core sequence   96

CORD CO-Regulation Database   78

CRE cAMP Response Element   33

CREB CRE binding protein   25

CRM Composite cis-Regulatory Modules   50

DEM Diethyl-maleate   46

DMEM Dulbecco's modified Eagle's medium   44, 47

eIF Eukaryotic Initiaion factor   31

ENCODE Encylopedia of DNA elements Consortium   29

eQTL Expression quantitative trait loci   4

ER Endoplasmic reticulum   11

ERAD ER associated protein degradation   35

ERK Extracellular signal regulated Kinase   26

ERSE ER Stress response element   37

GCN-2 General control nonderepressible 2 protein kinase   30

GCS Glycine Cleavage System   13

GCSH Glycine Cleavage System Protein H (H-Protein)   13

GCSL Glycine Cleavage system Protein L (EC 1.8.1.4)   13

INTRODUCTION

Outline: This chapter details the theoretical framework and motivation for the study reported in this thesis. It provides a review of the importance of free glycine synthesis and transport. A discussion of the various proposed mechanisms for glycine cytoprotection is included. What is presently known about the supply and transport of glycine to cells is reviewed, detailing the various transporters involved. Facets of physiological stress in epithelial cells and the various evolutionary conserved adaptive stress response pathways are reviewed in this chapter. It concludes by defining the main hypothesis, aims and objectives of this project.

Contents:

In recent years, the metabolism and transport of glycine and one carbon units have been implicated in processes promoting epigenetic alterations, modification of redox status and genome maintenance (Amelio *et al.*, 2014, Locasale, 2013). Studies conducted in this lab on cellular models of physiological stress reveal the rapid intracellular accumulation of

glycine in gut epithelia as a function of GlyT-1: the specific glycine transporter type 1 (Howard and Hirst, 2011). GlyT-1 (also known as SLC6A9 – where SLC6 stands for solute carrier family 6 member A9) has been identified as a downstream target of the integrated stress response (ISR) in mouse embryonic fibroblast cells (Dickhout *et al.*, 2012). Its mRNA levels are elevated in intestinal epithelial cell models following induction of the unfolded protein response (UPR) and endoplasmic reticulum (ER) stress with tunicamycin and thapsigargin (Howard and Hirst, 2011). How various stresses amount to changes in GlyT-1 expression in intestinal epithelia is the subject of this thesis.

1.1   Significance of glycine transport

Glycine is a gluconeogenic, alpha-amino acetic acid with no stereoisomers. Its simplicity and size convene its function in beta-turn formation, linking the alpha helices to beta sheets in the secondary structure of various proteins. The structural implication of glycine availability for development is severe (Guay *et al.*, 2002, Wu, 2010). A glycine residue occurs at almost every third position in the collagen chain (Lodish *et al.*, 2000). Collagen, an insoluble structural protein accounts for about a quarter of all proteins in the human body. The fact that very little of maternally fed labelled glycine is transported from mother to foetus is aggravating (Lewis *et al.*, 2005). Only about 10% of maternal glycine is passed to the foetus, thus in preterm infants glycine is conditionally essential (Jackson *et al.*, 1997). Offspring of rats feed a glycine free diet during pregnancy developed several complications in visceral organs at or before attaining adulthood (El Hafidi *et al.*, 2004).

Glycine transport across the intestine is necessary for the supply of amino acids to all tissues. In posterior areas of the central nervous system, the rapid re-uptake of glycine from synaptic clefts into the axon terminal or glycinergic neurones produces inhibitory post-synaptic potentials (Betz *et al.*, 2006). Conversely, its binding to the N-methyl-D-aspartate (NMDA) receptor co-agonises the effects of aspartate or glutamate; hence, potentiating excitatory transmissions and modulating synaptic plasticity (Viu *et al.*, 1998). Glycine mediated transduction of action potentials is necessary for both sensory and motor control. Reduced plasma glycine and low glycine titre in cerebrospinal fluid has been associated with several neurological conditions (Altamura *et al.*, 1995, Maes *et al.*, 1998). For this reason, several glycine transport inhibitors have been trialled, with the main purpose being to blunt its re-uptake from synapses in the central and peripheral nervous systems (Umbricht *et al.*, 2014, Chue, 2013).

1.2 Defences of the human intestine

With its distinctive high cell proliferation rate, the highly differentiated and complex gastrointestinal tract is an energetically expensive organ requiring an increased flow of nutrients. Despite accounting for less than 5% of whole body weight, about 20% of all body protein is synthesised in the intestine (Reeds *et al.*, 2000). Although the gut wall is important for nutrient absorption and synthesis, it is precariously exposed to diverse microbes, toxins and xenobiotic compounds. In the absence of adequate defences dysbiosis in the luminal tract may result in adhesion to, and invasion of, the intestinal wall by micro-organisms. As such, a specialised two tier barrier comprising of primarily an intrinsic wall of glycocalyx coated continuous monolayer of epithelial cells with tight intracellular junctions serves as first line of defence (Figure 1.1). The intestinal epithelia are responsible for regulating the transit of micro and macromolecules between the interstitium and the underlying lamina propria. Specialised epithelial cells such as goblet cells, secrete mucus forming glycosylated proteins (including mucins) and other protective peptides for added mechanical resistance to injury. Others such as the microfold cell (M-cells) found in Peyer's patches of the ileum present pathogenic antigens to the immune cells of the lymphoid follicle for the induction of an immune response (Kucharzik *et al.*, 2000). The coat of mucus infused antimicrobial peptides (AMPs) is an important extrinsic defence protecting the epithelial barrier from changes in environmental variables such as pH, electrolyte balance and bacterial invasion (Agerberth and Gudmundsson, 2006).

The integrity of the intrinsic barrier is subject to various physiological stresses including nutrient and oxidative stress (Soderholm and Perdue, 2001). Physiological stress represents the unfavourable effect of environmental factors (stressors) on the homeostasis and normal physiology of the cell (as defined in the directory of medical subject headings: MeSH ID 68013312; tree number G07.700.830). Prolonged barrier dysfunction (as depicted in Figure 1.2) and changes in epithelial permeability have been linked to the progression and complications of various intestinal diseases (Catalioto *et al.*, 2011, Ma, 1997, McGuckin *et al.*, 2009). This includes the many diseases collectively known as inflammatory bowel disease (IBD). Generally, they result from the disproportionate and misguided response of the immune system to stressors from within the gut lumen (Geboes *et al.*, 2014).

Crohn's disease (CD; which affects any region of the gut) and Ulcerative colitis (UC; presenting only in the colon) are two of the most common forms of IBD with similar aetiology and progression but differing presentations. Genome wide analysis of expression quantitative trait loci (eQTL) for CD highlight associated gene variations in several loci such as those required for sensing pathogen or microbe associated molecular patterns (P/MAMP) and maintaining the microbial ecosystem (Barrett *et al.*, 2008). Microbial endotoxins are effective stressors of the human intestine (Walker and Porvaznik, 1978).
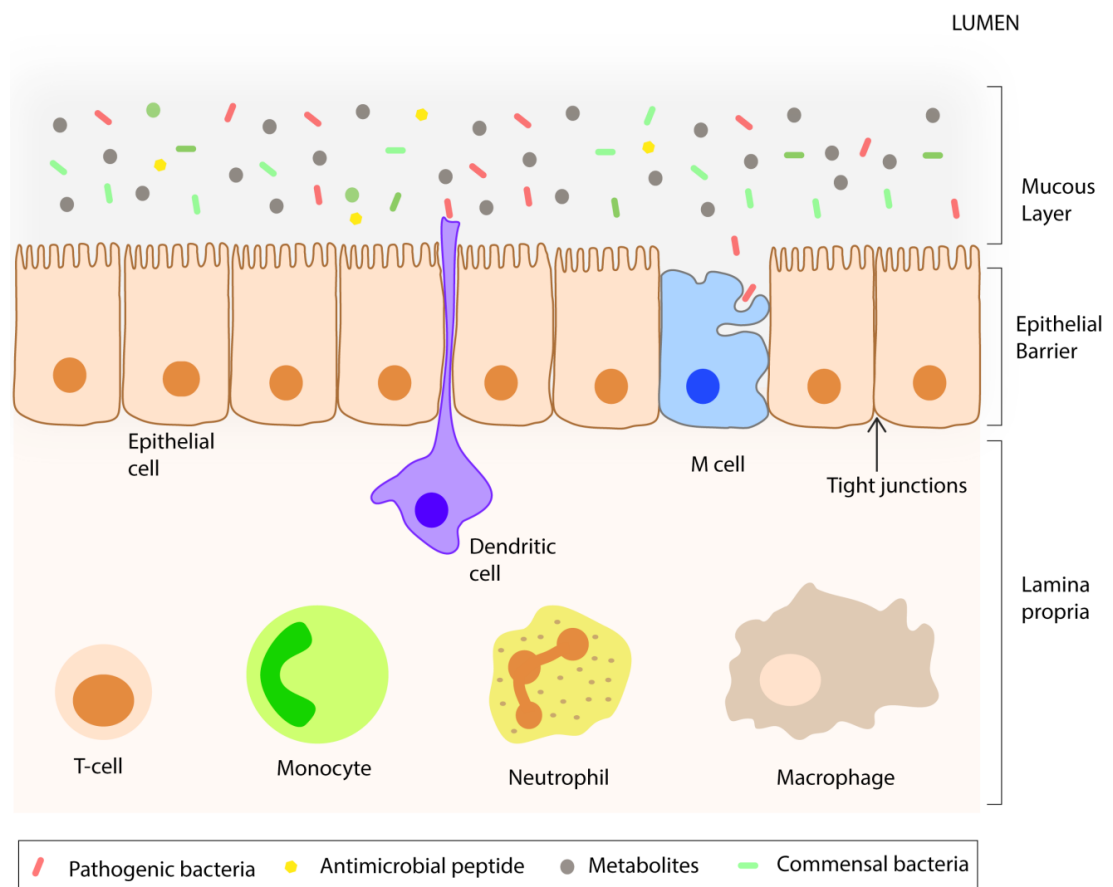
Figure 1.1: Defences within the healthy human intestine. Protection from injury is mediated by a specialised two tier barrier comprising of an extrinsic mucous layer with antimicrobial peptides, and an intrinsic network of epithelial cells forming a monolayer with tight intracellular junctions. Specialised cells such as the M-cells (or microfold cell) and dendritic cells detect pathogenic invasions and present specific antigens to lymphocytes within the lamina propria, resulting in the activation of an appropriate immune response.

Aside from IBD, several less severe forms of colitis such as Infectious Colitis are characterised by states of acute stress within the epithelial cells of the intestinal barrier. These often result from infection from parasites such as *Shigella dysenteria*. Chemical imbalance within the gut also represents a serious threat to luminal homeostasis. Stress resulting from the occurrence of free radicals may lead to chemical colitis. Brief periods of inadequate blood supply to the bowel have been linked to the occurrence of ischaemic colitis. Prolonged stress may ultimately lead to more severe CD.
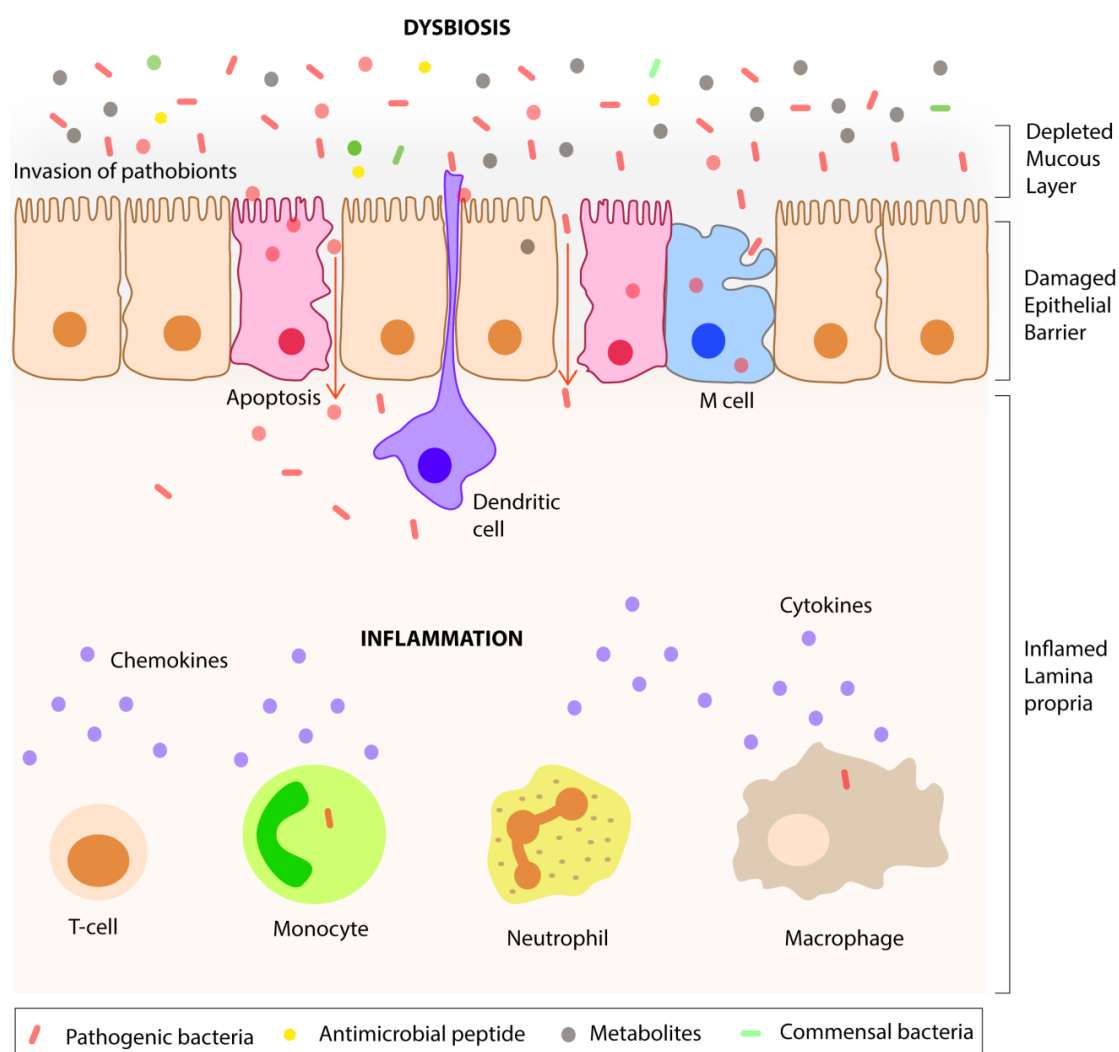
Figure 1.2: Effects of dysbiosis on gut barrier function. Toxins produced by invading pathobionts, upon accumulation in cells of the epithelial lining may lead to cell death, disruption of tight junctions and a leaky gut barrier. Dendritic and M cells present pathogen associated molecules to lymphocytes contained within the lamina propria. These lymphocytes mount an appropriate immune response against the invasion.

1.3 Cyto-protection by glycine

Over the last 25 years the role of glycine as a cyto-protective agent has been extensively researched. It is now known that it can protect a number of cell types/organs against various stresses. Glycine has been shown to have cyto-protective effects in the hepatocytes of the liver (den Butter *et al.*, 1993, den Butter *et al.*, 1994), renal tubules of the kidney (Weinberg *et al.*, 1987, Weinberg, 1990, Weinberg *et al.*, 1990, Paller and Patten, 1992), epithelia of the lung (Omasa *et al.*, 2003, Gohrbandt *et al.*, 2006), cardiocytes (fibroblasts and myoblasts) of the heart (Qi *et al.*, 2007), and epithelial cells of the intestines (Mangino *et al.*, 1996). That the accumulation of intracellular glycine retards stress induced damage in various cell types, prevents the depletion of cellular ATP levels or the accumulation of fatty acids, as well as prevents the onset of inflammation is an emerging theme.

1.3.1 Immune regulation by glycine

As glycine is one of the amino acids with the largest titre change in serum following septic shock (Vente *et al.*, 1989), the hypothesis that glycine has an immune-regulatory role has been scrutinised (Wheeler *et al.*, 1999, Zhong *et al.*, 2003). Glycine supplemented parenteral and enteral diets suppresses the immune responses resulting from intestinal ischemia/reperfusion (I/R) injury (Bower *et al.*, 1995). Likewise dietary glycine is able to prevent the onset of ulcerative colitis in the gut by inhibiting the induction of various chemokines and cytokines (Stoffels *et al.*, 2011, Tsune *et al.*, 2003).

The occurrence of glycine gated chloride channels on a variety of leukocytes, such as Kupffer cells, alveolar macrophages and neutrophils, has led to speculation on the role of glycine in blunting their activation (Wheeler *et al.*, 1999). It is now understood that the binding of glycine to these chloride channels leads to rapid membrane depolarisation of the plasma membrane and an influx of extracellular calcium ($_eCa^{2+}$) via voltage gated channels (Van den Eynden *et al.*, 2009, Wheeler *et al.*, 2000, McCarty *et al.*, 2009, Ikejima *et al.*, 1997, Schilling and Eder, 2004). Using a two hit model in rats combining an intravenous injection of a sub lethal endotoxin derived from *Escherichia coli,* together with partial hepatic I/R, Ikejima and colleagues reported that glycine feeding markedly improved survival (Ikejima *et al.*, 1996). They propose from their results that the reduced hepatic necrosis is a consequence of blunting of intracellular calcium ($_iCa^{2+}$) fluxes in

Kupffer cells preventing a cascade of eicosanoid and cytokine production which require intracellular $_iCa^{2+}$ as second messenger.

Canonical activation of leukocytes by external stimuli such as drugs, endotoxins, growth hormones or other stimuli invokes signal transduction via a G-protein coupled trans-membrane receptor (Meng and Lowell, 1997). Activation of phospholipase C and the generation of inositol tri-phosphate (IP3) culminate at the release of $_iCa^{2+}$ as secondary messenger from internal stores (Chow *et al.*, 1995). The release of $_iCa^{2+}$ from intracellular stores has for effect the rapid depolarisation of the plasma membrane and a further influx of extracellular $_eCa^{2+}$. This massive change in intracellular $_iCa^{2+}$ is believed to serve as message for the expression of key cytokines (Reilly *et al.*, 2005). Thus, current understanding suggests that upon the binding of glycine to glycine gated chloride channels, hyper-polarisation of the plasma membrane shunts most (if not all) membrane voltage operated calcium gates. This renders the cell insensitive to external stimuli such as lipopolysaccharides (LPS), growth hormone or second messengers.

1.3.2 Detoxification of xenobiotic compounds

One of the earliest identified cyto-protective roles of glycine was its conjugation to xenobiotic compounds in the 'glycination' pathway (Quick and Cooper, 1931). The enzymatic conjugation of glycine to toxins and their metabolic intermediaries generates less toxic excretory derivatives. The two step amino acid conjugation to xenobiotic compounds comprises a key detoxification pathway (Figure 1.3). The first ATP-requiring step is the formation of a xenobiotic-co-enzyme A (CoA) intermediate (Schachter and Taggart, 1953, Gatley and Sherratt, 1977). The final step is the conjugation of the mitochondrial acyl-coA to glycine in a reaction catalysed by the enzyme glycine *N* acetyltransferase (GLYAT, EC 2.3.1.13). Although glycination is believed to take place predominantly in the mitochondria of the liver and kidney, the expression of acyl-coA by several other tissues (such as corneal and intestinal epithelium) point to the possibility of pre-systemic glycination of xenobiotic compounds (Kaminsky and Zhang, 2003).

Benzoic acid is a naturally occurring aromatic acid in many plants and vegetables and widely used as common food preservative and anti-fungal agent. Additional benzoic acid is produced from microbial degradation of dietary aromatic acids such as purines and polyphenols in the gut. In the liver, benzoic acid is the end product of the catalytic conversion of Toluene - a widely used industrial solvent (van Duynhoven *et al.*, 2011).

Whilst dietary intake alone may be non-toxic, increased intracellular accumulation of benzoate can detrimentally alter cell physiology (Bas *et al.*, 2014). In man, the conjugation of intracellular glycine to benzoic acid leads to excretion of hippurate (Graber *et al.*, 2012).
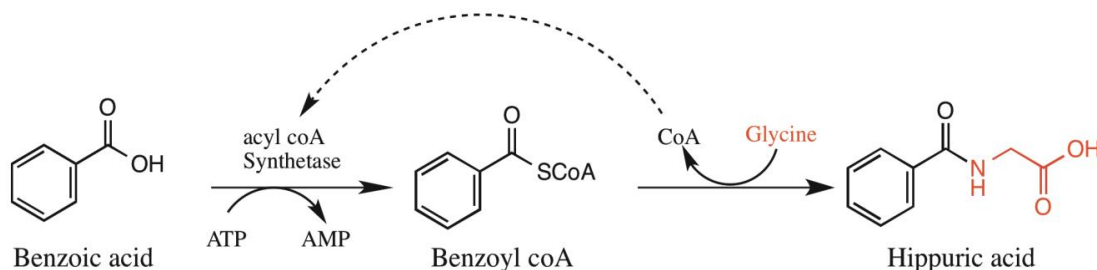
Figure 1.3: Glycination of xenobiotic compounds such as benzoic acid.

Metabolic profiling of urine samples from patients presenting with IBD has shown to have significantly diminished hippurate (Williams *et al.*, 2009), highlighting an apparent dysbiosis common in IBD patients. Whilst ruling out genetic causes for inadequate glycination, Williams and colleagues argue that it is this dysbiosis leading to excess accumulation of xenobiotic compounds in and around intestinal epithelial cells which elicits the ensuing auto-inflammatory response (Williams *et al.*, 2010). Such dependency on glycine availability for detoxification was reported by Gregus and co-workers, where the administration of exogenous glycine increased the capacity of benzo-glycine formation at a rate much higher than permitted by endogenous glycine supply in the liver of anaesthetised rats (Gregus *et al.*, 1996, Gregus *et al.*, 1992, Gregus *et al.*, 1993).

For β-oxidation of esterified fatty acids, acyl-coA hydrolases in the mitochondria maintain a recycled pool of CoA (Reddy and Sambasiva Rao, 2006). In the absence of suitable conjugates like glycine, xenobiotic compounds conjugate to and sequester CoA in an ATP-dependent reaction forming xenobiotic acyl-coA compounds. Drugs such as sodium valproate (an anti-epileptic drug), aspirin and many other salicylates or phenyl acetic acids bind to CoA in this manner; reasons why glycine treatment is recommended following an overdose of these drugs (Patel *et al.*, 1990). The depletion of CoA pools eventually disrupts oxidative phosphorylation (respiration) and β-oxidation of fatty acids (Knights *et al.*, 2007). This leads to the accumulation of long chain fatty acids, ATP depletion and an increased cascade of free radicals. The ensuing lipid-peroxidation is believed to result in necro-inflammation, fibrosis and cell death (Day and Saksena, 2002).

Such is the 'double-hit' pathogenesis in the progression from steatosis to steatohepatitis (Day and James, 1998) in the livers of obese individuals, heavy alcoholic drinkers, and patients treated with some antiviral or some non-steroidal anti-inflammatory drugs (Gaemers *et al.*, 2011). The alimentary inclusion of glycine significantly reduced the circulation of non-esterified fatty acid (NESFA) and reversed the high blood pressure in adult male Wistar rats previously fed a glycine-free diet (El Hafidi *et al.*, 2004).

1.3.3 Glutathione synthesis

Like glycination, glutathione conjugation or 'glutathionation' is a crucial detoxification mechanism employed by several cell types (Pastore *et al.*, 2003). The formation of thio-ester bonds between glutathione (GSH) and compounds with an electrophilic centre represents a key antioxidant control. GSH is a tri-peptide consisting of a peptide link between the amino groups of cysteine and glycine, and unusually to the γ-carboxyl group of the glutamate side chain (Figure 1.4). The terminal carboxyl moiety of the conjugated glycine of GSH prevents its degradation within the cell by γ-glutamyl-cylo-transferase (GGCT, EC 2.3.2.4). For the synthesis of GSH, the conjugation of glycine to γ-glutamyl-cysteine by the action of glutathione synthetase (GSS, EC 6.3.2.3) is the final step in a two-step catalytic reaction; the first step being the conjugation of cysteine to glutamate by γ-glutamyl-cysteine synthetase (γ-GCS, EC 6.3.2.2, also known as glutamate cysteine ligase). It has been suggested that microbial endotoxins such as LPS interfere with GSH synthesis by lowering the expression of the catalytic and modifier subunits of GCT (Tomasi *et al.*, 2014). Using Caco-2 cells incubated with *E. coli* as a model for the intestinal micro-environments, He *et al.* report the activation of host defence response pathways associated with the intracellular accumulation of free radicals (He *et al.*, 2013) .

Free radicals are formed by the non-enzymatic autoxidation of surrounding molecules resulting in the breakup of electron pairs. The reactivity of 'reactive' oxygen species (ROS) is known to initiate or aggravate DNA-double strand breakage, lipid-peroxidation and oxidative stress; all of which are associated with ageing and other degenerative diseases (Ray *et al.*, 2012). Whilst the toxicity of ROS is dependent on their reactivity, it is apparent that some free radicals are more potent than others. In recent years there has been an overwhelming focus on superoxides (O_2^-), peroxides (H_2O_2), hydroxyl (OH) and hyroxy-ethyl in the pathogenesis of disease and ageing (Ray *et al.*, 2012, Tiganis, 2011, Alfadda and Sallam, 2012). Be they from the prolonged inhibition to oxidative phosphorylation,

alcohol induced, pathogenic sources, or the consequence of perturbed oxygen metabolism and cellular electrolyte balance, ROS are also essential in signalling to cellular adaptive responses such as in the stress response, pathogenic defences and cell death (Tiganis, 2011).

STEP1:
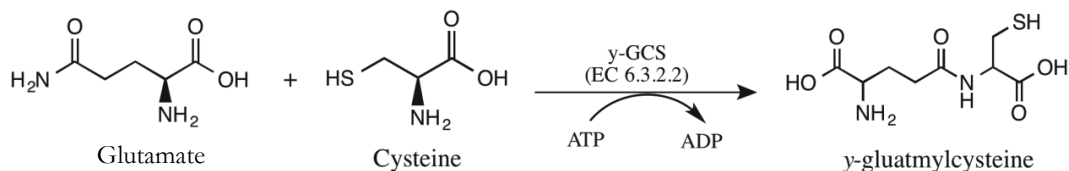
STEP2:
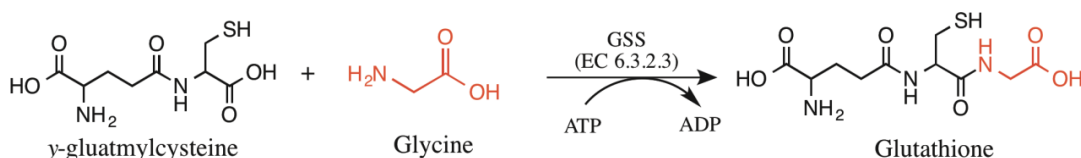
Figure 1.4: Synthesis of the tri-peptide glutathione

In extracellular space, the amino acid cysteine is often found in its oxidised form cystine (Banjac *et al.*, 2008, Iyer *et al.*, 2009). The cystine/glutamate antiporter System X_c^- mediates the high affinity import of cystine for GSH synthesis. Studies in rodents on GSH availability in states of oxidative stress highlight the cysteine/cystine redox cycle as rate limiting (Lyons *et al.*, 2000, Bitensky, 1990). However, there is an equal importance in glycine availability for synthesis of GSH (Jackson, 1986). This is evident from the increases in downstream gene targets required for the intracellular accumulation of glycine following stress amongst which are GlyT-1 and other low affinity glycine transporters (Howard and Hirst, 2011).

1.3.4 Osmo-protection by glycine

Epithelial cells lining the intestinal wall and also the proximal tubules of the kidney are particularly susceptible to severe changes in cell volume resulting from acute osmotic imbalances (Burg *et al.*, 2007, Brocker *et al.*, 2012). The regulation of solute transport across the cell membrane comprises a prescribed response to the ensuing osmotic stress

(Willem H Mager, 2000). To counter states of hyper- or hypo-tonicity, the opening of mechano-sensitive ion channels following changes in cell volume has for consequence the increased fluxes of predominantly inorganic ions across the membrane (Martinac, 2004, Inoue *et al.*, 2009). However, increased intracellular concentration of ionic compounds usually has for effect the disruption of biochemical process, build-up of free radicals, and alteration of protein structure and function. To prevent the negative cascading effects of osmotic stress to oxidative stress, and ER stress resulting from altered intracellular biochemistry, small neutral organic amino acids like glycine (or tri-methyl glycine [betaine], taurine; amongst many others) are preferred osmolytes in many animal cells (Dawson and Baltz, 1997, Yancey, 2005, Dawson *et al.*, 1998). In the kidney and intestinal cells, glycine is a potent inducer of ornithine decarboxylase (ODC; EC 4.1.1.17) required for the synthesis of polyamines from ornithine (a by-product of urea), which if accumulated may elicit osmotic stress (Chen and Canellakis, 1977, Minami *et al.*, 1985). Work done by Steeves and associates using mouse embryo shows that the intracellular accumulation of glycine in hypertonic states results from increased direct activity of GlyT-1 (Steeves *et al.*, 2003, Steeves and Baltz, 2005).

Gene expression levels of many organic osmolyte-transporters (such as members of the System A amino acid transport family), and enzymes facilitating their intracellular synthesis are known to be regulated by hyper-tonicity (Burg *et al.*, 2007, Burg *et al.*, 1997). The minimal essential osmotic response element (ORE) – a 12 bp sequence (GGAAANN(C/T)N(C/T)) first identified in the 5' flanking region of the gene encoding aldose reductase (AR, an enzyme required for the synthesis of the osmo-protectant sorbitol), has also been identified in the promoter regions of the betaine/GABA transporter (BGT-1), Sodium/myo-inositol co-transporter (SMIT), the taurine transporter (TauT) and HSP70 (Ferraris and Garcia-Perez, 2001, Ito *et al.*, 2009, Burg *et al.*, 2007, Miyakawa *et al.*, 1999). Amino acid limitation combined with hyper-tonicity has been shown to increase the binding and activity of the Tonicity responsive Enhancer Binding Protein or Osmotic Response Element Binding protein (TonEBP/OREBP) at the SNAT-2 ORE (Franchi-Gazzola *et al.*, 2004, Franchi-Gazzola *et al.*, 2001). TonEBP/ORE-BP a member of the *rel* family of transcription factors was first identified in cells of the renal medulla (Miyakawa *et al.*, 1999). Due to its very close ancestry and similar function to members of the Nuclear factor of activated T-cells (NFAT) family of transcription factors that regulate gene expression in response to osmotic stress it is has

now been reclassified as NFAT-5 or NFAT-L1 (NFAT-like protein type 1) (López-Rodríguez *et al.*, 1999, Trama *et al.*, 2000). NFAT-5, an ubiquitously expressed transcription factor, has been shown to regulate expression of the taurine transporter (TauT/SLC6A6) as well as the betaine (*N, N, N*-tri-methyl-glycine) solute exchanger (BGT-1/SLC6A12) (Hsin *et al.*, 2011, Ito *et al.*, 2005, Lopez-Rodriguez *et al.*, 2004)

As to whether NFAT-5 or other tonic response transcription factor directly regulates GlyT-1 like other members of the SLC6 family remains speculative. But with widespread acceptance on the osmolytic properties of glycine (Baltz and Zhou, 2012, Steeves and Baltz, 2005, Steeves *et al.*, 2003) and given that GlyT-1 activity increases in hypertonic states (Steeves *et al.*, 2003), it is a reasonable assumption that the transporters of glycine like those for other osmolytes such as taurine, may be regulated by osmotic stress at a transcriptional level.

1.3.5 Stimulation of Heat Shock Proteins

It has been demonstrated both *in vivo* and *in vitro* that changes in temperature lead to a well-orchestrated heat stress response (Nowak, 1990, Pelham, 1986). The transcriptional regulation of downstream heat shock proteins (HSP) occurs early following heat stress, but also following ischemia, anoxia, nutrient deprivation and other forms of stress (Morimoto, 1998, Morimoto *et al.*, 1992, Tonkiss and Calderwood, 2005). Using opossum kidney (OK) cell line, Nissim and colleagues demonstrated that when glycine or alanine was supplemented in culture media the mRNA levels of the 70, 72 and 73 kDa heat shock proteins (HSP70, HSP72 and HSP73) were significantly increased following heat stress (Nissim *et al.*, 1992). In the same experiment, treatment with glutamate, leucine, arginine or aspartate had no such effects. Similarly, supplementation of Caco-2 culture medium with 8 mmol/L glycine significantly increased the protein expression of HSP72 following heat stress (Lindemann *et al.*, 2001). In both studies the effect of glycine supplementation alone under basal conditions (no stress) had no effect on the transcription of these proteins. That glycine and not just any amino acid resulted in transcriptional regulation of HSPs points to a specific role for glycine in cellular resistance to heat stress, although a mechanism is as of yet unknown.

1.4　Glycine synthesis and Metabolism

Previously, glycine had been considered nutritionally dispensable on the foundation that its endogenous supply and metabolism provided sufficient recycling capacity for intracellular needs. Now, it is better described as conditionally essential, with periods of glycine insufficiency due to inadequate intracellular synthesis or increased requirement, for instance during periods of stress. Amino acid metabolism in the intestine accounts for over 60% of all metabolism in the portally-drained viscera or PDV (Yu *et al.*, 1992, Yu *et al.*, 1995). Studies with stable isotopic tracers reveal that glycine is important in gluconeogenesis, the biosynthesis of pophyrins such as heme, the production of creatinine and the synthesis of purines for nucleic acid synthesis (Locasale, 2013). Nucleic acid synthesis in the intestine relies predominantly on *de novo* synthesis of nucleotides from intestinal amino acids (Berthold *et al.*, 1995). An intact glycine residue is incorporated into the purines of mucosal RNA (Boza *et al.*, 1996). In addition to direct synthesis from carbon-dioxide and ammonium salts by the reverse action of the glycine synthetase complex (or glycine cleavage system GCS), glycine can also be generated from many sources (Wang *et al.*, 2013b). This includes from the amino acids threonine and serine or from glyoxylate, the ammonium salts choline, betaine (tri-methyl glycine) and sarcosine (mono-methyl glycine).

The breakdown of glycine by mitochondrial GCS serves as a source of one carbon units to many important metabolic processes. GCS comprises a concert of four proteins (T-protein, P-Protein, L-Protein and H-protein) which together catalyse the oxidative decarboxylation of glycine to carbon-dioxide. The lipoyl group of the H-protein (also known as GCSH) mediates the shuttling of moieties between the three other enzymes in the system. The methylamine moiety, resulting from glycine decarboxylation by the action of P-protein (also known as Glycine dehydrogenase, GLDC: EC 1.4.4.2) is transferred to the H-protein. T-protein (GCST: EC 2.1.2.10) an amino-methyl-transferase, catalyses the release of the methylamine from and reduction of H-protein, and transfers a methyl group to tetra-hydro folate. This yields ammonia and 5,-10-methylenetetrahydrofolate (5, 10-MTHF). The reduced H-protein is oxidised by L-protein (GCSL: EC 1.8.1.4) which uses NAD^+ as cofactor.
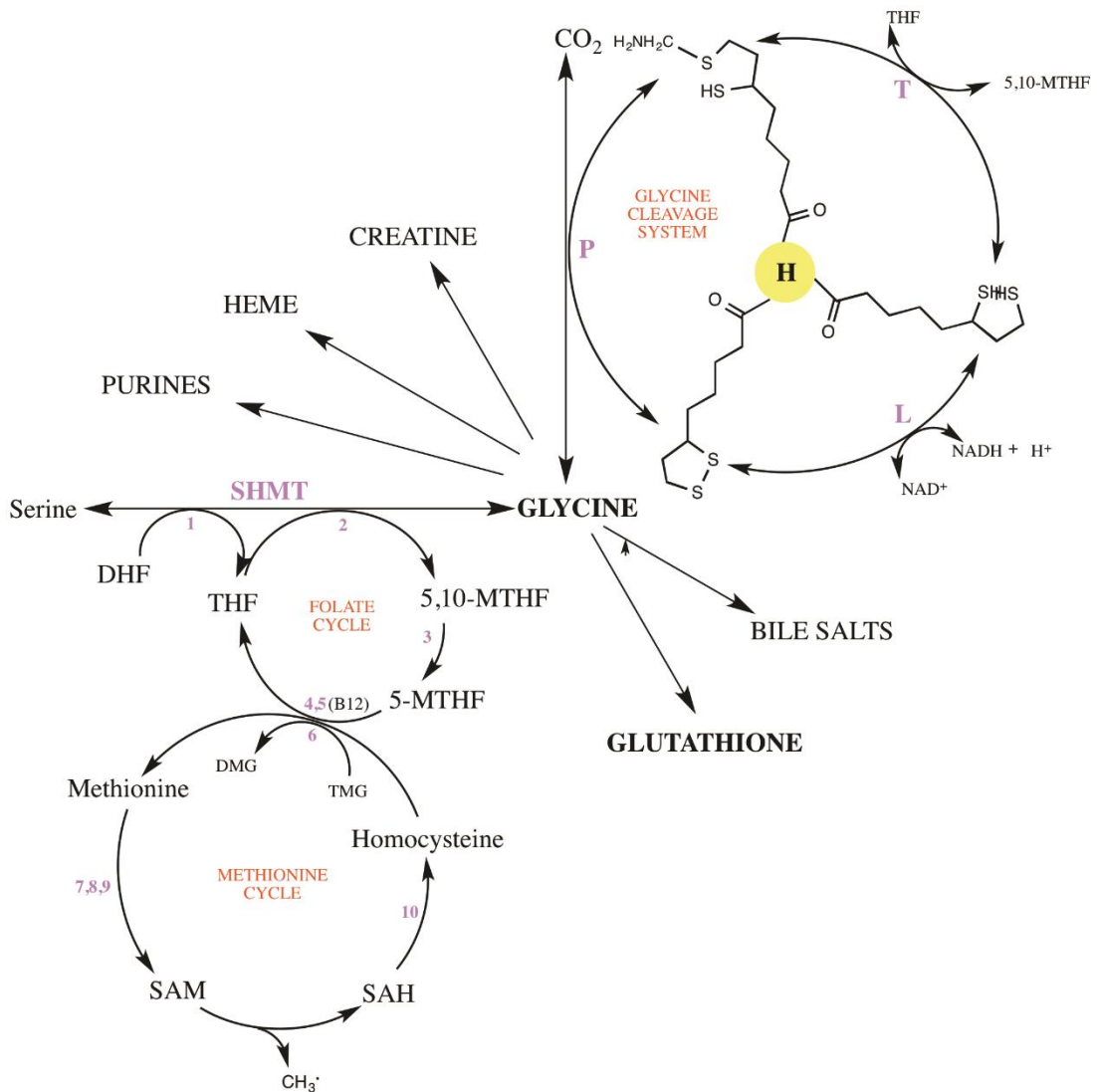
Figure 1.5: Glycine synthesis and metabolism. Enzymes are denoted in purple; 1: di-hydro folate reductase (DHFR); 2: Serine hydroxylmetyltransferase (SHMT); 3: 5, 10-methylene THF reductase (MTHFR); 4: methionine synthase (MS), 5: methionine synthase reductase (MTRR); 6: betaine-homocysteine methyltransferase (BHMT); 7,8,9: methionine adenosyl transferase 1a (MAT-1A), 2A (MAT-2A) and 2B (MAT-2B) respectively; 10: S-adenosyl-homocysteine (SAH) hydrolase (SAHH); SAM: S-adenosyl-methionine; TMG: Trimethylglycine (Betaine); DMG: Dimethyl glycine (Sarcosine);

5, 10-MTHF is an important cofactor in the synthesis of serine from glycine by the action of serine hydroxyl methyl transferase (SHMT). It is an important source of methyl groups used for DNA methylation patterns (hence gene silencing) associated with several cancers. It is reduced by the action of Methylenetetrahydrofolate reductase (MTHFR, EC 1.5.1.20) to 5-MTHF and eventually demethylated to THF or re-methylated to 5-10-

MTHF in a series of steps termed the folate cycle. The one-carbon (methyl) output from the folate cycle is donated to the methionine cycle via methylation of homocysteine by the action of methionine synthase (MS, EC 2.1.1.13) and its co-factor vitamin B_{12}.

1.5 Glycine transporters

Owing to its importance in cyto-protection as described above, during periods of physiological stress there is an increased intracellular requirement for glycine. This may result in significant shortfall between the supply of glycine from intracellular biosynthesis and metabolic demands, hence requiring added glycine transport from external sources. Most dietary amino acid and peptide absorption occurs in the early two thirds of the small intestine, predominantly in the duodenum, proximal jejunum and ileum (Freeman and Kim, 1978, Silk *et al.*, 1985). This usually involves active uptake across the apical membrane of absorptive cells, simple diffusion across the cytoplasm and transport through the basolateral membrane into the blood stream. In humans, systemic availability of amino acids represents less than half of dietary intake, with the remainder sequestered by tissues comprising the PDV (stomach, intestine, pancreas and spleen) which have a heightened energy expenditure and metabolic rate (Yu *et al.*, 1992, Yu *et al.*, 1995). In pigs, first-pass extraction accounts for up to 40% of dietary glycine and 60% of cysteine retention in absorptive cells (Reeds *et al.*, 2000). This is in contrast with the systemic availability of absorbed monosaccharide of almost all of enteral intake. In the human intestine, glycine can be transported by a multiplicity of SLC proteins summarised in Table 1.1 and described below.

1.5.1 Imino and proton driven glycine transporters

Initially identified as the lysosomal amino acid transporter (LAAT) in brain neurones, the proton assisted amino acid transporter (PAT-1; SLC36A1) is known to mediate electro genic transport of the imino acid proline, glycine, alanine, betaine (*N, N, N*-tri-methyl glycine) and sarcosine (*N*-methyl glycine) (Chen *et al.*, 2002, Boll *et al.*, 2003a, Boll *et al.*, 2003b, Boll *et al.*, 2004, Broberg *et al.*, 2012). PAT-1 activity is absent at neutral pH but highest at low pH with no enantiomer discrimination and relatively low substrate affinity (K_m 3 - 7mM) (Inigo *et al.*, 2006). The order of substrate preference for PAT-1 is highest for compounds containing an imine group within three carbon units of a carboxyl group such as proline or hydroxyproline, and lowest for amino acids such as glycine (Metzner *et al.*, 2006). Like most proton driven transporters, PAT-1 co-transports (symport) one

amino acid and one proton (H^+) sourced from the H^+/ATPase, H^+/K^+/ATPase (or V-type H^+/ATPase on lysosomes) or the immediate acidic surroundings of the cell. Evidence points to an apparent ambiguity in requirement for Na^+ by PAT-1 in rat intestine (Inigo *et al.*, 2006). This is possibly due to the Na^+ dependence of the membrane H^+/Na^+ anti-porters from whence H^+ required for its activity may be sourced.

Subcellular PAT-1, expressed on lysosomal membranes in human embryonic kidney (HEK293) cells is known to enable the amino acid dependent recruitment of the mammalian target of rapamycin (mTORC1) to the lysosome by virtue of its physical interaction with Rag/GTPases (Ogmundsdottir *et al.*, 2012). In mammalian tissue, PAT-1 mRNA is differentially expressed from the gene locus 5q33.1 in cells of the kidney, neuronal cells in the brain and cells of the spleen, but predominantly on the apical membrane of epithelial cells lining the stomach, small intestines and colon (Boll *et al.*, 2004). Due to the pH neutral environment of the kidney proximal tubular cells, PAT-1 is inefficient in the re-absorption of glycine at the apical membrane. This together with a moderate substrate specificity for imino compounds has let to doubts of its association with the morbidity of iminoglycinuria (Broer *et al.*, 2006).

Iminoglycinuria (IG; OMIM 242600) and hyperglycinuria (HG; OMIM 138500) are autosomal recessive errors of glycine and imino acid transport, resulting from complex mutations in glycine and imino acid transporters. Despite its low capacity for glycine, PAT-1 mutations have been associated with the prevalence of IG with impaired intestinal transport. It is believed that an IG phenotype is associated with the inheritance of two non-functional alleles within genes encoding for the low to medium affinity imino and glycine transport such as PAT-1 (Broer, 2008a, Broer *et al.*, 2008). An HG phenotype, whilst common in neonates with a burgeoning glycine transport, is thought to be associated with inheritance of a single non-functional allele within the genes encoding medium to high affinity glycine transporters.

Like PAT-1, PAT-2 (SLC36A2) is a proton (H^+) dependent amino acid transporter. It has thus far been identified in mouse, rat and human (Boll *et al.*, 2003a, Boll *et al.*, 2003b, Bermingham and Pennington, 2004). Although PAT-2 is less distributed in humans than PAT-1, PAT-2 mRNA expression has been detected in cells of the kidney, skeletal muscle (Bermingham and Pennington, 2004, Nishimura and Naito, 2005) and intestine (Foltz *et*

al., 2004). PAT-2 mediates the high affinity transport (K_m $100 - 200\mu M$) of glycine, proline and other L-alpha amino acids with small aliphatic side chains (Kennedy *et al.*, 2005).

1.5.2 Neutral amino acid transporters

Like PAT-1, the first sodium coupled neutral amino acid transporters (SNAT-3/SLC38A3 and SNAT-1/SLC38A1) were cloned on the basis of their sequence homology to the vesicular inhibitory or GABA amino acid transporter (VIAAT or VGAT; SLC32A1) of inhibitory neurones (Sagne *et al.*, 2001). Despite an apparent Na^+ dependence, it was subsequently suggested that there exists functional similarities between members of the SLC38, SLC36 and SLC32 families in terms of substrate specificity (Schioth *et al.*, 2013). Indeed like PAT-1, two members of the System N subfamily, SNAT-3 (SLC38A3) and SNAT-5 (SLC38A5), which share up to 22% sequence similarity to VGAT, counter-transport one H^+ ion for the co-import or export of one Na^+ with either one of glutamine, histidine, asparagine and sometimes glycine (Broer *et al.*, 2002). Amongst the System N transporters, SNAT-3 does not transport glycine, whilst SNAT-5 has very low substrate selectivity (K_m $1 - 2mM$) (Fei *et al.*, 2000, Nakanishi *et al.*, 2001). SNAT-5 substrates (in order of preference) are: asparagine, serine, histidine, glutamine, glycine and alanine (Nakanishi *et al.*, 2001). Unlike PAT-1 and despite diminished transport at low extracellular pH, SNAT-5 transporters are differentially expressed on plasma membranes in the brain, liver, lung and intestinal tract (Mackenzie and Erickson, 2004).

SNAT-1, SNAT-2/SLC38A2 and SNAT-4/SLC38A4 (members of the System A subfamily) share up to 60% sequence similarity to System N transporters. System A transporters do not require the counter transport of the H^+ ion, but are solely dependent on a Na^+ gradient for substrate co-transport (Hatanaka *et al.*, 2000, Albers *et al.*, 2001, Hatanaka *et al.*, 2001). Whilst no transport of glycine has been demonstrated for SNAT-1 – which is mostly expressed in the brain (Albers *et al.*, 2001), SNAT-2 and SNAT-4 mediate efficient low affinity glycine transport (Hatanaka *et al.*, 2001, Hatanaka *et al.*, 2000). SNAT-2 mRNA levels increase as a consequence of activation of osmo-sensitive transcription factor NFAT-5/TonEBP (See section 1.3.4) in response to osmotic stress in cultured human fibroblast cells (Franchi-Gazzola *et al.*, 2001, Franchi-Gazzola *et al.*, 2004).

Other members of the neutral amino acid subfamily of transporters such as the B^0 neutral amino acid transporter B^0AT-1 (SLC6A19) are expressed on the apical epithelial surface in the intestine (Broer, 2008a, Broer, 2008b). B^0AT-1 is expressed on epithelial cells of the duodenum, jejunum and ileum, and also mediates reabsorption in the proximal tubule of the nephron (Seow *et al.*, 2004). Like members of the System N subfamily (described above), B^0AT-1 is very dependent on the availability of a Na^+ gradient for the transport of its substrates. Although B^0AT-1 has a high preference for leucine, it is able to transport several other non-aromatic L-amino acids with low selectivity (K_m 1 – 12mM) (Bohmer *et al.*, 2005). This includes methionine, valine, asparagine, alanine, serine, threonine, glycine and proline, listed here in order of preference. Like B^0AT-1, NTT-4 (SLC6A17) is another neutral amino acid transporter which is expressed solely in the central nervous system. NTT4 is Na^+ dependent and mediates low affinity transporter of glycine (Zaia and Reimer, 2009).

B^0AT-3 (SLC6A18) transports glycine with higher affinity than B^0AT-1 (K_m 0.9 – 2.3mM)(Broer and Gether, 2012). However unlike B^0AT-1 it is expressed in the kidney but not expressed in the intestine, and is dependent on both a Na^+ and Cl^- ion gradient (Singer *et al.*, 2009). Another neutral and basic transporter ATB^{0+} (SLC6A14) has similar transporter affinities for glycine as B^0AT-3. It is expressed at its highest levels in the adult and foetal lung as well as trachea and salivary glands, although very low levels have been detected in the stomach and colon (Sloan and Mager, 1999). This transporter (ATB^{0+}) like B^0AT-3, is Na^+ and Cl^- dependent; with transport stoichiometry of 2-3 Na^+ for 1 Cl^- and 1 amino acid.

Table 1.1: List of SLC transporters with glycine affinities; assembled with data curated in the database of SLC transporters by Alexander *et al.* (2013).

System	cDNA	SLC	Affinity	Mechanism	
A	SNAT-1	SLC38A1	Medium	Symport	Na^+
	SNAT-2	SLC38A2	Medium	Symport	Na^+
	SNAT-4	SLC38A4	Medium	Symport	Na^+
N	SNAT-5	SLC38A5	Low	Symport	H^+
asc	4F2 hc/asc1	SLC3A2/SLC7A10	Low	Antiport	
B^0	B^0AT-1	SLC6A19	Low	Symport	Na^+
	B^0AT-3	SLC6A18	Medium	Symport	Na^+
$B^{0,+}$	$ATB^{0,+}$	SLC6A14	High	Symport	Na^+, Cl^-
PAT	PAT1	SLC36A1	Low	Symport	H^+
	PAT2	SLC36A2	Medium	Symport	H^+
Gly	GlyT-1	SLC6A9	High	Symport	$2\,Na^+ : 1\,Cl^- : 1\,Gly$
	GlyT-2	SLC6A5	High		$3\,Na^+ : 1\,Cl^- : 1\,Gly$
	B^0AT3 (XT2)	SLC6A18	High	Symport	Na^+
PepT	PepT1	SLC15A1	Low	Symport	H^+
	PepT2	SLC15A2	Low	Symport	H^+

1.5.3 Peptide transporters

Additional glycine supply may be achieved by the catabolism of di- and tri-peptides transported into the cell by the high-capacity, low-affinity (K_m 0.2 – 10 mM) peptide transporters PepT-1 (SLC15A1) and PepT-2 (SLC15A2) (Doring *et al.*, 1998, Daniel and Kottra, 2004). PepT-1 is a sodium-independent, proton-dependent transporter predominantly expressed on the apical membranes of the small intestine but also epithelial cells of the proximal tubules in the kidney, the epithelial cells of the bile duct and vascular smooth muscle cells. It mediates the proton coupled transport of preferentially dipeptides ($2H^+$ co transported with one zwitterionic peptide). PepT-2 is predominantly expressed in the kidney and cells of the nervous system. In RNA-seq of coding RNA from the tissue samples of 95 human individuals, two of the seven possible PepT-2 transcripts were detected in the colon and duodenum (Fagerberg *et al.*, 2014; array data visualised in ArrayExpress).

1.5.4 High affinity specific glycine transporters (GlyTs)

GlyT-1 (SLC6A9) and GlyT-2 (SLC6A5) mediate the high affinity active uptake of glycine across the plasma membrane of absorptive cells and synaptic ends of neuronal cells (Borowsky *et al.*, 1993). GlyT-1 and GlyT-2 share up to 50% sequence homology but differ in tissue distribution and function. The basolateral uptake of glycine across the human intestine was characterised as a single high affinity transport system sensitive to sarcosine (Christie *et al.*, 2001) and with transport kinetics consistent with those previously published for system Gly transporters (Hillman *et al.*, 1968, Narkewicz *et al.*, 2000). Both GlyT-1 and GlyT-2 are significantly expressed in the white and grey matter of the central nervous system of developing mice and rats (Jursky and Nelson, 2002, Cubelos *et al.*, 2005) where they facilitate inhibitory potentials at glycinergic synapses (Betz *et al.*, 2006). Whilst GlyT-2 is mostly detected in the spinal and spinothalamic white matter, expression of GlyT-1 is detectable predominantly in the grey matter in regions rich with NMDA (N-methyl-D-aspartate) receptor-containing synapses including significant subpopulations of glutamatergic neurons (Adams *et al.*, 1995, Cubelos *et al.*, 2005, Betz and Eulenburg, 2012). GlyT-1 antibodies also detected expression in the liver, intestines and kidney (Jursky and Nelson, 2002). Homozygous mice lacking GlyT-1 were reported to die within one day after birth from severe motor and respiratory complications (Gomeza *et al.*, 2003a, Gomeza *et al.*, 2003b, Tsai *et al.*, 2004). GlyT-2 deletions in mice resulted in significant weight gain and the development of a fatal hypertonia and hyperekplexia (an exaggerated

startle reflex characterised by touch induced seizures and muscle stiffness) within two postnatal weeks (Rousseau *et al.*, 2008). Correspondingly, autosomal dominant inheritance of homozygous, heterozygous or compound heterozygous mutations within the GlyT-2 gene has been associated to hyperekplexia type 3 (HKPX-3, OMIM-614618) and proposed as causative or contributory to sudden infant death syndrome (OMIM 272120).

Glycine transporters (GlyTs) are strictly dependent on extracellular Na^+ and Cl^- ion fluxes. By combining electrophysiological and radiological techniques, Roux and Supplisson (2000) confirmed a requirement for 3 Na^+ and 1 Cl^- by GlyT-2 (K_m 27 μM) and 2 Na^+ and 1 Cl^- for GlyT-1 (K_m 72 μM). It is believed that this difference in stoichiometry results from varying constraints of glycine transport (such as rate of transport required) at site of transporter localisation (Roux and Supplisson, 2000, Supplisson and Roux, 2002). From *N*-glycosylation and cysteine scanning mutagenesis assays it is understood that GlyT-1 and GlyT-2 have twelve trans-membrane domains (TMD) connected by six extra- and five intra-cellular loops functioning in nutrient sensing and selectivity (Olivares *et al.*, 1994, Olivares *et al.*, 1995). The coupling of GlyTs to their extracellular substrates leads to a conformational change favouring an inwardly open over an outwardly open topology (Zafra and Gimenez, 2008, Perez-Siles *et al.*, 2011). Whilst the exact substrate pockets or glycine binding sites of GlyTs are yet to be deciphered, substitution of the tyrosine 289 residue located in TMD3 to a tryptophan or serine residue significantly reduces transport kinetics (Ponce *et al.*, 2000). Comparative modelling to the *Aquifex aeolicus* leucine transporter (LeuT$_{aa}$) suggests the glycine 305 residue in TMD6 of GlyT-1 and serine 481 of GlyT1 may also be important to glycine selectivity (Zafra and Gimenez, 2008). The Drosophila disc large (Dlg) homologous region (DHR; or PDZ) domain located on the intracellular carboxyl (C) terminal is supposed to function in protein trafficking and regulation of protein density at the cell surface or synaptic interface (Adams *et al.*, 1995, Betz *et al.*, 2006). It has been suggested that removal or modification of the C-terminal during states of elevated calcium comprise a well-defined control in glycine homeostasis (Baliova and Jursky, 2010).

Several isoforms for both GlyT-1 and GlyT-2 have been described. GlyT-1 isoforms are expressed in the brain, kidney, pancreases, lung, placenta, liver and intestine, whilst GlyT-2 is mostly found in the brain (Lopez-Corcuera *et al.*, 2001a, Lopez-Corcuera *et al.*, 2001b). Three isoforms of GlyT-2 have been defined (GlyT-2A, GlyT-2B and GlyT-2C); which differ in their N-terminal regions (Liu *et al.*, 1993, Ponce *et al.*, 1998). By convention, the

protein isoforms a labelled with uppercase letters A, B, C etc., whilst the DNA transcripts a labelled with the corresponding lowercase alphabet i.e. a, b, c etc. Variations at both the N- and C-termini are evident amongst the isoforms of both GlyT-1 and GlyT-2; with GlyT-2A having an unusually long N-terminus (about 200aa residues) when compared to other members of SLC transporters (usually 30-45aa).

Alternative promoter usage at the same gene loci 1p33 spanning 44.1 Mb, as well as variations in splicing of the coding exons may result in up to twelve transcript variants of GlyT-1 (Borowsky *et al.*, 1993, Borowsky and Hoffman, 1998). Of the twelve 'predicted' variants, mRNA from the expression of three GlyT-1 variants has been detected in humans (Kim *et al.*, 1994, Borowsky and Hoffman, 1998). These are: variant 3 or GlyT-1a (NM_001024845, 3236bp), variant 1 or GlyT-1b (NM_006934.3, 3228pb) and variant 2 or GlyT-1c (NM_201649.3, 3390bp). GlyT-1a, b and c are identical in their 3' end and have been identified in a number of mammals including rat, mice, human and cattle. Of the three transcripts, GlyT-1a – a 14 exon (out of the possible 17 exons) transcript, has been shown in rats to be under the control of a promoter upstream of exon 1. With exon 3 out of transcriptional frame and exon 1 completely untranslated, GlyT-1a mRNA encodes the shortest protein (with only 633 amino acid residues) from a translational start site in exon 2. Variations in the large 5'-UTR region of GlyT-1a convey a distinctive N-terminus when compared to GlyT-1B and GlyT-1C.

Transcription of GlyT-1b is (in rats) under the control of an alternate promoter located between exons 2 and 3 (Borowsky and Hoffman, 1998). Despite a shorter transcript, GlyT-1b codes for an isoform 19 AA residues longer than GlyT-1a. Exons 1 and 2 are completely excluded from GlyT-1b and its translation starts from the alternate exon 3, but maintains the same subsequent exon usage as GlyT-1a. Hence, GlyT-1B differs from GlyT-1A only at the 5'-terminus in that, the first 10 amino acid residues of GlyT-1A which result from the translated region of exon 2 are swapped for 29 amino acid residues from the translated region of exon 3. Like GlyT-1b, GlyT-1c is under the control of an alternate promoter between exon 2 and 3. Whilst the N and C termini of their corresponding protein isoforms are identical, the GlyT-1C isoform is 54 amino acid residues longer than GlyT-1B. This is due to the inclusion of an additional in-frame exon (exon 4 of the gene) into the GlyT-1c sequence.

To date GlyT-1c mRNA has only been identified in human cells or tissues (Kim *et al.*, 1994, Herdon *et al.*, 2001). GlyT-1b and GlyT-1c mRNAs are expressed predominantly in neuronal tissues whereas GlyT-1a is found in the brain and peripheral organs. GlyT-1a mRNA is the prevalent transcript detectable in human intestinal cells, based on PCR evidence from work done in this lab, although low amounts of GlyT-1b and GlyT-1c mRNAs are also detected in these cells. Attempts at identifying the cellular localisation of the different isoforms of GlyT-1 proteins have remained futile, hindered in part by the lack of specific antibodies to the various isoforms. Characterisation of GlyT-1 in Caco-2 cells revealed greater localisation to the basolateral membrane (Christie *et al.*, 2001). Four times more glycine is transported by GlyT-1 at the basolateral membrane of intestinal epithelia as opposed to the apical membrane. Such transport polarity lays further emphasis on the importance of glycine transport to the epithelial barrier lining the gut.

Despite its classification as 'protein coding' in the genome assembly, no mRNA level evidence has been demonstrated for the fourth transcript variant 4 (NM_001261380.1, 3183bp) annotated by the NEDO human cDNA sequencing project (Japan). Additionally, Hanley *et al.* by yeast 2 hybrid assays identified two potential bovine GlyT-1 isoforms with C-terminal variations (GlyT-1D and GlyT-1E) which specifically interact with the rho-1 subunit of the GABA receptor and have significantly different transport kinetics (Hanley *et al.*, 2000). However these isoforms are yet to be described in human species. Of the remaining seven, two transcripts, variant 5 (NR_048584.1) and variant 6 (NR_048549.1), are long non protein coding RNAs in human. Evidence of the existence of five more transcripts are inferred based on locus specific information determined by homology analysis from Sanger's HAVANA project but have never been confirmed in humans.

In this and other labs, recent and on-going functional characterisation of GlyT-1 provide hints into the extent at which intrinsic and extrinsic factors modulate GlyT-1 availability, localisation and function in various cellular physiological states. Factors such as pH changes, metal binding, nutrient availability and calcium fluxes within the cell have all been postulated to regulate the localisation and activity of the protein. The transcriptional regulation and expression of GlyTs have been largely studied in their capacity as facilitators of neurotransmission across glycinergic synapses. Most of what is known stems from studies in neuronal and renal transport systems. Whilst it is now accepted that the intracellular accumulation of glycine is important to the viability of the cell, the

molecular events leading to the expression, localisation and activity of GlyTs following stress are yet to be understood. Amino acid availability, metabolism and transport have been demonstrated to regulate gene expression at key dogmatic points; from nucleic acid to mRNA to protein levels. It remains unknown as to whether GlyT-1 expression control involves specific coordinated regulatory networks during epigenetic, transcriptional, translational or post-translation processing.

1.6   Regulation of Glycine transporters

As with most members of the SLC6 family, transporter regulation include epigenetic, transcriptional and translational modulation, as well as post-translational effects. A complexity of multiple protein-protein and protein-DNA interactions oversee processes such as acetylation and methylation, phosphorylation, SUMOylation, glycosylation and ubiquitination, which together regulate cell surface density and activity of the transporter by 1) altering the accessibility of chromatin domains to transcription factors, 2) transcription factor and cofactor recruitment and specificity at the promoter and enhancer regions of the transporter gene, 3) promoting transcription of the transporter gene, RNA stability and translation of the protein, 4) localisation and trafficking of the protein to the cell surface, 5) fusion of the protein trafficking vesicles to the plasma membrane mediated by the numerous SNAP (Soluble *N*-ethylmaleimide sensitive factor (NSF) Attachment Protein) Receptor proteins or SNARE and 6) internationalisation and removal of unneeded transporter by phosphorylation of the N or C-terminus of the transporter, or Ca^{2+} mediated modification of the PDZ domains. Work done to understand regulation of a variety of SLC6 transporters along the above avenues is extensive and beyond the scope of this thesis Broer and Gether (2012), (Kristensen *et al.*, 2011, Pramod *et al.*, 2013). However, but for a few conclusive reports on GlyTs regulation (reviewed in this section) most of what is known is inferred or implied.

1.6.1 Epigenetic regulation

Trans-generational inheritance of an epigenome conveys environmentally induced adaptation to changes in gene expression via processes of DNA methylation, histone modifications, RNA silencing amongst many others (Jaenisch and Bird, 2003, Portela and Esteller, 2010). Research into epigenetic regulation of GlyT-1 has focused on two key epigenetic processes, histone methylation and acetylation. These epigenetic modifications have the ability to regulate the level of GlyT-1 expression by altering chromatin

accessibility to polymerases and transcription factors (Jaenisch and Bird, 2003). The nuclear localisation of the chromosomal high mobility group nucleosome binding protein HMGN3a has been reported to facilitate unravelling of the chromatin to reveal the GlyT-1 promoter (Barkess *et al.*, 2012). Unlike HMGN-3b, the bi-partite nuclear localisation and nuclear binding domain together with the regulatory domain of HMGN3a improved the ability of the CREB binding protein co-associated factor (p300/pCAF, also known as the Lysine (K) acetyl-transferase 2B or KAT-2B), a member of the GCNC-5/pCAF nuclear histone acetyl-transferases (HATs) to acetylate histone 3 (H3Ac) at lysine 14 residue (H3K14) along the body of SLC6A9 gene locus promoting chromatin unravelling (West, 2004). pCAF is a known cofactor of activating transcription factor 4 (Atf-4) and other ATF/CREB factors, for the expression of genes downstream of the amino acid starvation response (Cherasse *et al.*, 2007). Acetylation of histone 3 at lysines 9 and 14 as well as at lysine 5, 8, 12 and 16 of histone 4 constitute a primer for transcriptional activation whilst di- and tri-methylation at lysine 27 of histone 3 are markers for transcriptional repression (Kouzarides, 2007). The significance of the HMGN proteins to GlyT1 expression was emphasised in an experiment by Shan and colleagues where expression of the Jumonji domain-containing protein 3 (JMJD3; a histone 3 lysine 27 de-methylase known to be up-regulated by Atf-4 following amino acid starvation) was rescued in Atf-4 knockout HepG2 and MEF cells (Shan *et al.*, 2012).

The role of activating transcription factor 2 (Atf-2) a member of the BZIP family of activating transcription factors in the epigenetic control of genes regulated by amino acid availability has been studied (Gibney and Nolan, 2010, Feil and Fraga, 2011, Seong *et al.*, 2011). Reporter and gel shift assays demonstrate Atf-2 and Atf-4 are crucial components for immediate/early expression of target genes of the amino acid response, such as the neutral amino acid transporter ATB^0 (SLC1A5), the cell-surface antigen heavy chain 4F2 (SLC3A2; also known as CD98) which when associated with xCT/SLC7A11 comprise the cysteine/glutamate exchanger (System X_c^-) or with LAT-2/SLC7A8, transports leucine and glycine with low affinity (System Y^+) (Averous *et al.*, 2004).

Following nutrient stress, phosphorylation of Atf-2 at threonine 71 precedes acetylation of histone H4 and H2B (Waas *et al.*, 2001, Bruhat *et al.*, 2007). This results in the modification of chromatin structure to enhance transcription of multiple genes potentially including GlyT-1 by Atf-4. In the early events leading to the recruitment and activity of Atf-4, phosphorylation of pre-bound Atf-2 by mitogen activated protein kinase (MAPK),

extracellular signal regulated kinase (ERK) and the MAPK/ERK kinase (MEKK1-p38) is regarded as an absolute requirement (Livingstone *et al.*, 1995, van Dam *et al.*, 1995, Waas *et al.*, 2001, Turjanski *et al.*, 2007). In drosophila, it has been demonstrated that the non-Mendelian inheritance of disrupted heterochromatin resulting from a similar phosphorylation of the analogous Atf-2 protein (dAtf-2) conveyed resistance to heat and osmotic stress across multiple generations (Seong *et al.*, 2011).

1.6.2 Post-translational regulation

The regulation of cell surface density of amino acid transporter in lieu of modulating transporter activity provides control on how much substrate transits across the cell membrane. This is particularly true for neuro-transmission where a rapid turnover of the transporter by either endocytosis or exocytosis is required to maintain the cadence of an action potential across the synapse (Geerlings *et al.*, 2000, Geerlings *et al.*, 2001, Horton and Quick, 2001). Extracellular levels of Calcium ions ($_cCa^{2+}$) have been shown to modulate GlyT-2 and GlyT-1 numbers on the cell surface of neuronal cells in synaptic junctions. This was demonstrated by Baliova *et al.* in two separate studies first for GlyT-2 and subsequently for GlyT-1A and GlyT-1B protein turnover in rat spinal cord synaptosomes (Baliova and Jursky, 2004a, Baliova and Jursky, 2004b, Baliova and Jursky, 2005). From their work, increased synaptic Ca^{2+} mediated the proteolytic cleavage of the N-terminal of both GlyT-1 and GlyT-2 at distinct sites which could be blocked by phosphatases. Fairly recently Baliova and colleagues showed a Ca^{2+} dependent cleavage of the last 12 amino acids comprising the PDZ domain in the distal C-terminus of GlyT-1C, preventing its trafficking to the membrane (Baliova and Jursky, 2010).

1.6.2.1 Regulation of transporter trafficking

GlyT-1 function is dependent on localisation of the properly folded protein at membranes across which it transports glycine. Glycine transporter trafficking by the SNARE system is a tightly controlled process. Most of what is known of this system comes from investigations on neuronal GlyTs interactions with SNARE proteins; predominantly with the synaptobrevin or vesicle associated membrane protein (VAMP), the synapse specific Syntaxin-1A and synaptosomes associated protein 25 (SNAP-25). The conjugation of Syntaxin-1A to SNAP-25 provides a docking port for VAMP on membranes of trafficking vesicles allowing for membrane fusion at cell surface membrane. When expressed alone or in combination with Syntaxin-1A in COS cells SNAP-25 had no

additional functional effects on glycine transport (Lopez-Corcuera *et al.*, 2001a). However it was observed that the direct physical interaction between Syntaxin-1A and either both GlyT-1 or GlyT-2 decreased plasma membrane concentrations without any changes in transporter expression (Geerlings *et al.*, 2000). Whilst no such investigations on GlyTs trafficking and docking have been carried out on epithelial tissues of the intestine, investigations of the SNARE system in Caco-2 cells highlight a complex network of over 60 proteins, notably Syntaxin-3 and SNAP-23 which regulate the docking of trafficking vesicles at either the apical or basal membrane (Riento *et al.*, 1998), This could form the basis of an investigation into GlyTs trafficking and docking at membranes of intestinal epithelial cells.

1.6.3 Regulation of transporter surface density

Protein kinase C (PKC) is considered a major player in transporter trafficking to the membrane by modulating the physical interaction between the transporters and the SNARE proteins (Beckman *et al.*, 1998). Activation of PKC isoforms by phorbol esters has been reported to inhibit transport by glycine transporters in glioblastoma C6 cells (Gomeza *et al.*, 1995, Morioka *et al.*, 2008) as well as L-arginine transport in cultured pulmonary endothelial cells (Krotova *et al.*, 2003). This appears to be due to an increase in internalisation rate of the transporters from the membrane (Fornes *et al.*, 2008). Similar effects of PMA have been demonstrated for GlyT-2 in transfected COS cells (Fornes *et al.*, 2004). Several PKC sites common to all GlyTs and indeed other Na^+ and Cl^- dependent transporters have been proposed. In GlyT-2, substitution of the charged lysine 422 residue and the polar threonine 419 and serine 420 residues of the second intracellular loop with acidic residues blocks internalisation of the transporter (Fornes *et al.*, 2004). In line with the above observation, it was determined from work done in this lab that phorbol 12-myristate 13-acetate (PMA) an activator of PKC significantly reduced the basolateral glycine uptake via the basolateral GlyT-1 transporter in Caco-2 cells after 4 hours treatment (Christie *et al.*, 2001). For GlyT-1, it has been suggested that a reduction in maximal transport rate by PMA treatment might result from indirect phosphorylation mechanisms (Sato *et al.*, 2002). Work has identified the cysteine 159 residue close to the putative substrate pocket of GlyT-1C (Vargas-Medrano, 2010) and the lysine 619, 577 and 581 residues in the N-terminal of GlyT-1b as phosphorylation sensitive sites (Vargas-Medrano *et al.*, 2011, Fernandez-Sanchez *et al.*, 2009). However, unlike GlyT-2, there is as of yet no confirmatory evidence of phosphorylation, neither at the above sites or the

many studied polar threonine and serine residues of the intracellular loop regions of either GlyT-1a, GlyT-1b or GlyT-1c (Vargas-Medrano *et al.*, 2011).

1.6.4 Regulation of transporter distribution across membranes

Aside from phosphorylation, protein N-terminal glycosylation represents an important sorting mechanism for transporter localisation to either the apical or basolateral membrane of polarized epithelial cells (Vagin *et al.*, 2009, Martinez-Maza *et al.*, 2001, Poyatos *et al.*, 2000, Olivares *et al.*, 1995). There is significant membrane polarity in localisation of GlyT-1 in human intestinal epithelia; with higher surface density on the basolateral membrane when compared to the apical membrane (Christie *et al.*, 2001). Oliveras and co-workers describe four putative glycosylation sites in the N-terminal of GlyT-1 used in surface targeting of the transporter (Olivares *et al.*, 1995). Though treatment with tunicamycin (a glycosylation inhibitor) and the mutation of these sites significantly reduced transporter activity GlyT-1 in transfected COS cells, enzymatic de-glycosylation of localised GlyT-1 protein did not alter transporter activity. In contrast to GlyT-1, enzymatic de-glycosylation of GlyT-2 reduced transporter activity by up to 40% in COS cells whilst mutation of putative glycosylation sites in the large second extracellular loop (EL2) completely blocked localisation of the transporter to the apical surface of Madin-Darby canine kidney (MDCK) cells (Martinez-Maza *et al.*, 2001).

1.6.5 Modulators of transporter activity

The association of glycine signalling to several neuro-psychotic disorders has resulted in extensive research into molecular means of directly affecting transporter activity. Several extrinsic as well as intrinsic inhibitors of glycine transport have been proposed. Ju and colleagues reported on the ability of Zn^{2+} at low doses to inhibit GlyT-1b at glycinergic synapses in the central nervous system (Ju *et al.*, 2004) by binding to two histidine residues (His-410 and His-421) of the large second and fourth loop extracellular loop respectively and inhibiting conformational changes crucial to glycine transport. Amongst the many other classes of glycine inhibitors, substitute sarcosine based compounds such as 3H-(R)-NTPS or (R)-N-[3-phenyl-3-(4'-(4-toluoyl)-phenoxy)-propyl]-Sarcosine irreversibly block glycine transport. It appears Sarcosine based inhibitors block glycine in a concentration independent, non-competitive manner, whilst non-Sarcosine based glycine analogues are mostly reversible and very much dependent on glycine concentration (Mezler *et al.*, 2008).

1.6.6 Transcriptional regulation

Given that most studies on regulation of glycine transport thus far has dwelled on post-transcriptional aspects, very little is known of the transcriptional regulatory networks modulating transporter expression. Transcriptional regulation of gene expression provides a unique mechanism for cellular adaptations to specific environmental events. Gene transcription is facilitated by the presence of core promoter sequences characterised by an array of tissue-specific and/or development-specific and even event-specific cis- and trans-acting regulatory transcription factor binding sites (TFBS). An assembly of transcriptional factors (TF), scaffolding proteins, cofactors and DNA polymerases at the promoter form the pre transcriptional complex and promote expression of the gene in response to specific cytosolic events. Experimental analysis and characterisation of the promoter regions of several amino acid transporters confirm a characteristic use of specific TF combinations at multiple promoters and transcriptional start sites (TSS) (Palii *et al.*, 2004, Fafournoux *et al.*, 2000). Which factors regulate the expression of GlyTs in the human intestinal cells following stress is the subject of the work presented here.

Although the resulting transcripts from alternate promoter usage in the GlyT-1 gene are known to result in several distinct and possibly functionally diverse GlyT-1 transporter isoforms, little is known about cellular events leading to the preferential usage of one promoter over the other. An advantage of alternate promoter usage and multiple reading frames in target genes may be to improve mRNA stability by favouring different 3' post-transcriptional processing and alternative splicing (Ayoubi and Van De Ven, 1996). Such is believed to boost protein translation and mRNA turnover of tissue restricted and event-specific protein isoforms during a targeted response to specific physiological events such as stress. It is however unlikely that alternate promoter usage in GlyT-1 expression in human intestinal cell serves a similar purpose.

Aside from multiple promoter usage, the regulation of gene expression by long-range enhancers, repressors, silencers and insulators within distal promoters, intronic regions and even within 3' flanking regions is now better appreciated with evidence from projects such as the ENCODE (Skipper *et al.*, 2012). Indeed, amino acid limitation or starvation controls the expression of CCAAT enhancer box binding protein (C/EBPβ) via elements located downstream of the coding sequence (Chen *et al.*, 2005). What remains outstanding

is linking these elements to modules within specific regulatory networks in response to specific events such as physiological stress.

In eukaryotic cells, extremes of both intracellular or extracellular environmental variables and changes in cellular homeostasis can alter gene expression of proteins required for withstanding the effects of the stressor. Depending on the nature of the stress and the cell type, a manifold response ensues ranging from the activation of survival pathways to the initiation of programmed cell death. A conserved adaptive signalling pathway involving several membrane receptors, kinases and transcription factors, link the cellular stress induced events to target genes and proteins. In the next section aspects of the cellular response to shortfalls in nutrient availability, to unfolded proteins, endoplasmic reticulum stress and oxidative stress are reviewed; specifically discussing how such events may link to up-regulation in transcription of the glycine transporters.

1.7   Stress induced gene regulation

Cellular stress defines a physiological disturbance to intracellular homeostasis. These include states of protein wasting in disease or inflammation, states of uncontrolled or compromised protein synthesis, poor nutrient availability, increased metabolic toxicity, physical or mechanical injury etc. The cellular stressor can be defined as any compound or event resulting in the disruption of homeostasis. Such disruptions can alter various cellular processes, including by interfering with protein glycosylation, as well as tampering with ionic and metabolite balance. To maintain homeostasis, the epithelium must adapt at a molecular and cellular level to changes in its environment. Elaborate mechanisms for sensing and responding to specific cellular stresses have evolved to enable the cell to maintain its viability, or in cases where the former is unachievable the activation of programmed cell death.

1.7.1 The amino acid response (AAR)

Amino acid (AA) depletion can have severe consequences on cellular metabolism; from resulting in biosynthetic and metabolic alterations to epigenetic changes and up regulation of gene expression of specific transporters. All mammalian cells possess evolutionary conserved adaptive pathways to sense and efficiently respond to specific changes in amino acid availability (Berlanga *et al.*, 1999). First identified in yeast, the mammalian general control nonderepressible 2 (GCN-2) protein kinase – a member of the pancreatic

eukaryotic initiation factor-2alpha kinase (PEK) family, is capable of sensing single amino acid deficiencies during protein elongation (Austin *et al.*, 1986). Under normal conditions, the binding of GTP to the eukaryotic translation initiation factor 2α (eIF-2α) mediates the delivery of a methionine charged t-RNA to the 40S ribosome. The accumulation of uncharged t-RNA as a result of amino acid unavailability for protein elongation leads to the direct phosphorylation of eIF-2α at its serine 51 residue (Gietzen *et al.*, 2004). Phosphorylated eIF-2α effectively blocks the required catalytic conversion of GDP to GTP by eIF-2β with net effect being the general suppression of protein translation. This marks the start of the AAR.

The initiation of the AAR is greatly conserved from yeast to mammalian cells. Following eIF-2α phosphorylation, the up regulation of some members of the activating factor family (Atf) of transcription factors, notably Atf-4 is ubiquitous across all cell types. Whilst Atf-4 mediated gene regulation appears to be the central pathway for AAR (more on this later), accessory pathways are understood to confer specificity to the response. For instance, blockade of the interdependent mitogen activated protein kinase (MAPKK or MEK; a tyrosine/threonine kinase) has been shown to prevent phosphorylation of eIF-2α by GCN-2 as well as block transcriptional induction of downstream Atf-4 target genes in human hepatoma cell lines (Thiaville *et al.*, 2008a, Thiaville *et al.*, 2008b, Thiaville *et al.*, 2008c). In line with this observation, Lopez-Fontanals and colleagues demonstrated that the increase in SNAT-2/SLC38A2 mRNA abundance following amino acid limitation was dependent on a MAPKK/MEK pathway whilst transporter activation by hyper-tonicity was not (Lopez-Fontanals *et al.*, 2003). Likewise activation of MAPKK/MEK pathway by amino acid signals is believed to mediate the phosphorylation of existing Atf-2 required for epigenetic modulation of Atf-4 target genes.
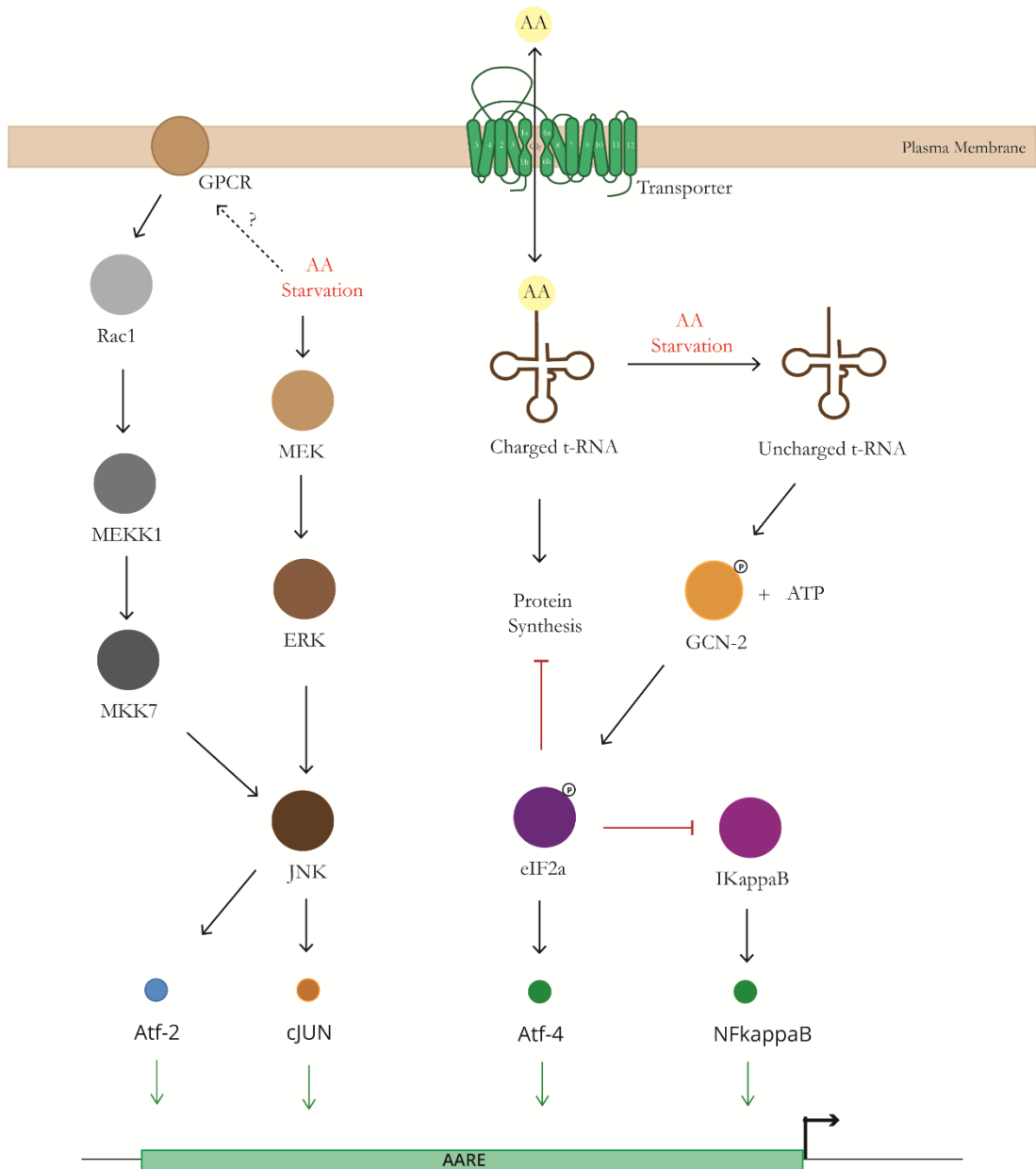
Figure 1.6: Illustration showing pathways of the amino acid response (AAR). In periods of amino acid insufficiency the mammalian general control nonderepressible 2 (GCN-2) phosphorylates the 2α subunit of the eukaryotic elongation factor (eIF-2α) a consequence of which is the down regulation of protein synthesis and activation of specific transcription factors such as Atf-4. Complementary to the AAR is activation of MAP kinase (MEKK) pathways which may activate other BZIP factors such as Atf-2 and cJUN required for transcriptional regulation of the target genes; GPRC: G-Protein coupled receptor; Rac1: Ras-related C3 botulinum toxin substrate 1, a RhoA GTPase; ERK: Extracellular signal regulated kinase. JNK: c-Jun N-terminal kinase.

A large group of amino acid transporters are regulated by Atf-4 following induction of the AAR. These include the cationic arginine and lysine amino acid transporter gene (Cat-1), SNAT-2, the excitatory amino-acid transporter (EAAT) and xCT – the light chain of the System X_C^- cystine/glutamate exchanger (Lopez *et al.*, 2007, Palii *et al.*, 2004, Sato *et al.*, 2004). Whilst an increase in Atf-4 mRNA following AAR is documented, its primary mode of activation downstream of phosphorylated eIF-2α is by translation control of existing mRNA. There are two upstream open reading frames (uORF) within Atf-4 mRNA. The 5' translation start site used in normal conditions is out of frame with the Atf-4 coding sequence negatively regulating it. Following stress, ribosomal 5'-cap scanning overlooks the first uORF in favour of a second, in-frame with the coding sequence; resulting in translation of a functional transcription factor. Activated Atf-4 promotes transcriptional induction of target genes within 30 minutes of AAR initiation. From transcription factors to transporters and more, the repertoire of Atf-4 target genes is large. Sequence analysis of the 5'-flanking regions of these genes revealed characteristic short nine base pair sequences collectively termed the amino acid response element or AARE (Averous *et al.*, 2004, Kilberg *et al.*, 2009).

First identified in the promoter region of the human C/EBP homologous protein (CHOP), AAREs are short 8-9bp composite cAMP response element and Atf-binding sites (CRE/ATF) with consensus 5'-TGATG/(N)AA(N)-3' (Bruhat *et al.*, 2002, Kilberg *et al.*, 2009). AAREs have been characterised within several amino acid transporters. In Cat-1, the binding of Atf-4 to the AARE in the 5'-UTR of the gene between nucleotides +45 and +53 relative to the transcription start site (TSS) mediates its up-regulation in response to amino acid limitation (Fernandez *et al.*, 2003). In many cell types, xCT is required for the supply of cysteine for use in glutathione synthesis. Sato *et al.* demonstrated using gel super shift and reporter assays that Atf-4 complexes with amino acid response elements in the 5' flanking region of the murine xCT gene (Sato *et al.*, 2004), and that the expression of this gene was induced by deprivation of not only cysteine, but shortages in other amino acids. Two AAREs were subsequently described in the 5'-flanking regions between nucleotides -94 and -86 relative to the TSS, one in a forward orientation on the sense strand, and the other 9bp further on the same strand in a reverse orientation (Sato *et al.*, 2004). Although cysteine availability is considered to be the rate-limiting step in the synthesis of glutathione, other experimental evidence suggests an equal importance for the supply of glycine to the final step of glutathione synthesis (Jackson *et*

al., 2004). As such, it is possible that the GlyT-1 transporter is regulated in a similar way to xCT. Interestingly, the coupling of Atf-4 to the single AARE in the SNAT-2 promoter decreases its mRNA abundance in liver cells but increases SNAT-2 mRNA following amino acid limitation in skeletal muscle or 3T3-L1 mouse adipocytes (Gaccioli *et al.*, 2006, Alfieri *et al.*, 2005, Luo *et al.*, 2013). Whilst it is not clear whether the former is a result of Atf-4 or other transcription factors, it may be that regulation of these transporters at their AARE is highly tissue specific.

Permutations in dimerisation of members of the BZIP and other superfamily of TFs such as nuclear factor kappa light chain enhancer protein (NF-κB) are important to modulating the calibre and tenure of the AAR. Phosphorylation of eIF-2α by GCN-2 is also believed to promote dissociation of the inhibitory kappa B protein (IκB), mimicking the effects of the active IκB kinase (IKK) in up-regulating NF-κB within an hour of AA starvation (Jiang *et al.*, 2003). Micro-array analysis of several NF-κB target genes has shown several to be up-regulated following amino acid starvation in HepG2 cells (Shan *et al.*, 2010); amongst which are the amino acid transporters SNAT-2/SLC38A2, the transcription factors C/EBPβ, Atf-3, Atf-5 and CHOP (also known as DNA damage-inducible transcript factor 3 [DDIT3] or GADD153) amongst many. CHOP is a negative repressor of the AAR by promoting the de-phosphorylation of eIF-2α

1.7.2 The unfolded protein response (UPR) and ER stress

The endoplasmic reticulum (ER) is a versatile organelle important to several cellular functions, including for calcium storage and providing a site for protein folding. Perturbations in ER function have been linked to the pathogenesis of several disease states, including many neuro-pathological conditions like Alzheimer's and Parkinson's diseases (Schroder and Kaufman, 2005). Inhibition of protein glycosylation or disruption in Ca^{2+} homeostasis as well as oxygen deprivation elicits a well prescribed ER stress response (Bailey and O'Hare, 2007, Takayanagi *et al.*, 2013, Vaughn *et al.*, 2014). Where the stressor is the accumulated unfolded proteins, the ensuing ER stress pathways are collectively termed the unfolded protein response (UPR). The accumulation of unfolded client protein within the ER leads to the dissociation of chaperon Immunoglobulin heavy chain binding protein (BIP) from its ER membrane dock (Sommer and Jarosch, 2002). The result is the initiation of multifaceted signalling pathways of the UPR.

Three adaptive pathways comprise the UPR. First and predominantly, the dissociation of BIP from its ER membrane-docking site activates the double stranded RNA-activated protein kinase R (PKR)-like ER kinase (PERK) (Harding *et al.*, 2003, Avivar-Valderas *et al.*, 2011). As with GCN-2 in the AAR response, PERK mediates the phosphorylation of eIF-2α. This leads to the translational activation of Atf-4 and a response similar in kind to the AAR (Harding *et al.*, 2000, Harding *et al.*, 2003). Second to and independent of the PERK/Atf-4 pathway, phosphorylation of the inositol requiring endonuclease (IRE-1) follows activation of the UPR (Xu *et al.*, 2014). In brief periods of stress IRE-1α actuates pro-survival pathways by splicing of the basic leucine zipper factor x-box binding protein type 1 (XBP-1). Upon translocation to the nucleus, spliced XBP-1 (XBP-1s) regulates gene expression of proteins of the ER degradation systems (ERAD) via cis-acting UPR response elements or UPREs (consensus: 5'-CAGCGTG-3') in their promoter regions (Wang *et al.*, 2009). In states of prolonged ER stress IRE-1 overseas the Regulated IRE-1 dependent decay (RIDD) of anti-apoptotic pre-miRNAs such as the pre-miRNA of the pro-apoptotic protease caspase-2 (CASP-2) (Sano and Reed, 2013, Maurel *et al.*, 2014). IRE-1α also mediates the activation of the Jun N-terminal kinase (JNK) pro-apoptotic pathways (Kato *et al.*, 2012, Cheng *et al.*, 2014).
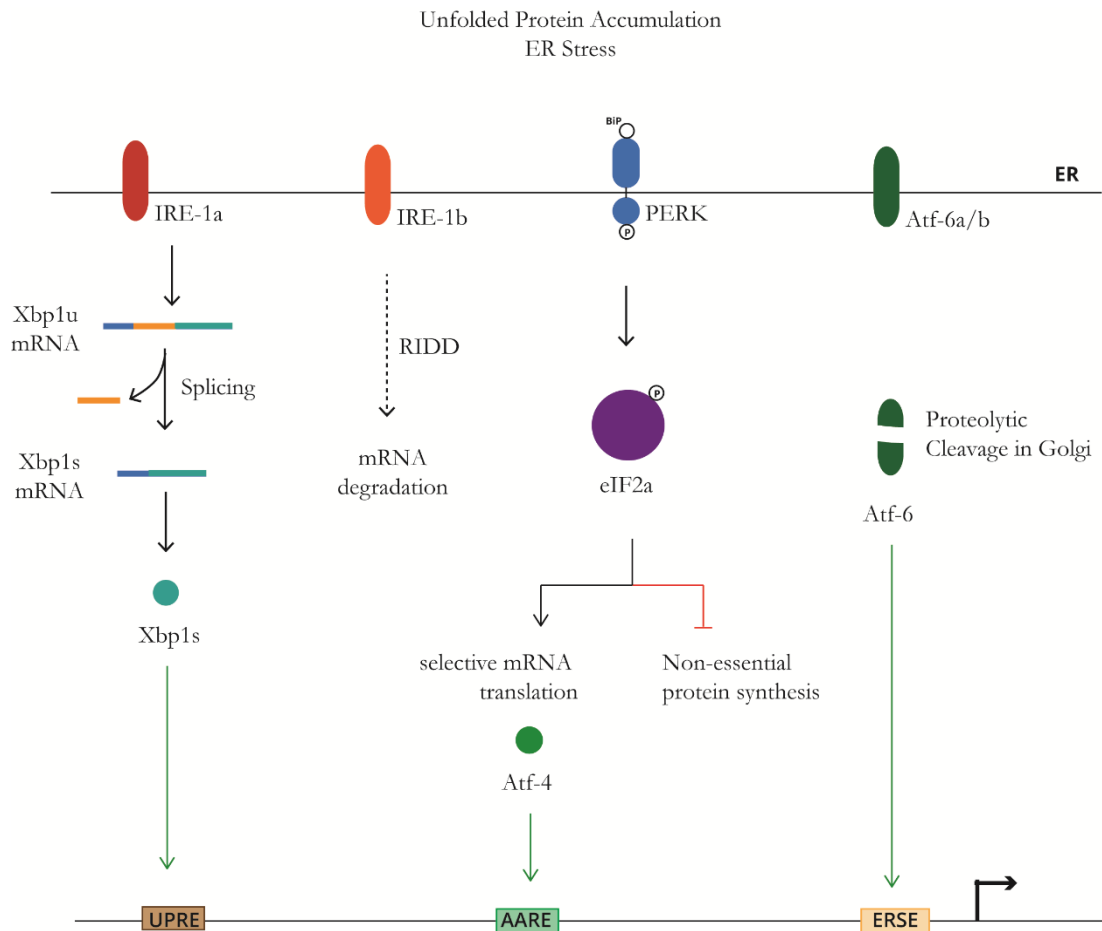
Figure 1.7: Illustration showing pathways of the UPR and ER stress response. The accumulation of unfolded proteins in the ER, and ER stress leads to the activation of several ER kinases, amongst which is the protein kinase R (PKR)-like ER kinase (PERK). As in the AAR PERK phosphorylates eIF2a and leads to the activation of Atf-4. Along with the PERK pathway is the activation of the x-box binding protein (XBP-1) by the Inositol requiring endonuclease IRE-1 which is a serine/threonine protein kinase resident at the ER membrane. Upon activation XBP-1 binds to unfolded protein response elements in the promoters of target genes. The proteolytic cleavage of Atf-6 in the Golgi following ER stress leads to the release of an active transcription factor able to regulate target genes via ER stress response elements (ESRE) in target gene promoters.

In later phases of ER stress, a negative feedback mechanism emanating from the accumulation of stable C/EBP-β LIP (liver-enriched inhibitory protein) isoforms previously minimised by proteolytic degradation and eIF-2α phosphorylation, supresses gene expression (Nakajima *et al.*, 2011). Lastly, the accumulation of unfolded proteins within the ER promotes trans-localisation of Atf-6 to the Golgi apparatus where in proteolytic cleavages activates the transcription factor (Yoshida *et al.*, 2000, Yoshida *et al.*, 2001, Sommer and Jarosch, 2002). Upon translocation to the nucleus Atf-6, alongside Atf-4, regulate the expression of response genes via a cis-acting endoplasmic reticulum stress element or ERSE with consensus 5'-CCAAN-N_9-CCACG-3'.

The overlapping nature of the three ER stress response pathways has led to ambiguity in nomenclature in the literature of the characteristic response elements. As originally identified, UPRE only describes elements to which the spliced XBP-1s binds, whilst ERSE corresponds to Atf-6 binding sites. A third response element with consensus 5'-ATTGG-N-CCACG-3' has been characterised and termed ERSE-II to which either XBP-1s or Atf-6 bind. Complicating further, is the understanding that the activation of PERK-Atf-4 pathway by ER stress can lead to gene regulation by elements identical to the consensus AAREs described above.

Recently, our understanding of the damaging effects of prolonged accumulation of unfolded proteins has highlighted additional facets of ER stress, notably disruption in oxygen metabolism and the accumulation of reactive oxygen species. Closer investigations of the interrelationship between ER stress and oxidative stress reveal highly specialised signalling pathways leading to the accumulation of ROS. ER-stress induced Ca^{2+} release: a) activates the calcium dependent protease m-Calpain which b) activates the ER procaspase 12. This in turn leads to c) the depolarisation of the inner mitochondrial membrane and activation of procaspase-9 mediated apoptosis as well as d) direct ROS release from the respiratory chain in response to Ca^{2+}.

1.7.3 Response to reactive oxygen species (ROS)

Despite advances in understanding ROS signalling in mammalian cells, puzzling questions remain on its role in several cellular processes. However it is evident that ROS accumulation from toxins, imbalanced oxygen metabolism or pre-existing ER stress leads to altered cellular biochemical processes and cell physiology.

As the primary sites for oxygen metabolism within the cell, the matrix and inner membrane of the mitochondrion are an important sources of ROS. Superoxides (O_2^-) produced in the mitochondria by the respiratory chain are impermeable to the mitochondrial membrane. O_2^- fluxes appear to be dependent on the availability of electron donors in the respiratory chain, oxygen concentration and $NADH/NAD^+$ ratios in the mitochondrial matrix. Dismutation[†] of the elevated O_2^- in the mitochondria is catalysed by the mitochondrial specific enzyme superoxide dismutase (SOD, EC 1.15.1.1). Hydrogen peroxide, the resulting by product of this reaction is membrane permeable and upon interaction with iron forms highly reactive hydroxyl specie (OH^-). Several redox homeostatic systems respond to specific threats from ROS within the cell, these include turnover of $NADH/NAD^+$ and glutathione conjugation. In fact, increased expression of anti-oxidant genes is characteristic of a target response to ROS accumulation.

For the transcriptional regulation of antioxidant genes, nuclear factor of the erythroid derived 2 like factor 2 (NFE2-L2 or Nuclear factor related factor 2, Nrf-2), another BZIP factor has garnered significant attention as an important mediator. Nrf-2 has been shown to regulate the expression of xCT via electrophile response (EpRE) cis-elements in its promoter region, as well as mediating the expression of a host of enzymes including the two electron reductase NADPH/Quinone oxido-reductase 1 (NQO1, EC 1.6.5.2) and glutathione S-transferases (GST, EC 2.5.1.18). In neuronal cells where xCT is not expressed, Nrf-2 up regulates the expression of excitatory amino acid transporter 3 (EAAT3) which transports cysteine required to facilitate glutathione synthesis in these cells. Miyamoto and colleagues describe a regulatory arm of the oxidative stress response in corneal epithelium where the phosphorylation of Nrf-2 by PERK increases Atf-4 expression following stress (Miyamoto *et al.*, 2011). This together with reports on possible Nrf-2/Atf-4 co-regulation and interactions supports a crossover in target response to ROS.

The modulation of gene expression by Nrf-2 is achieved at a post-transcriptional and possibly post-translational level. Under normal physiological states, Nrf-2 is sequestered within the cytoplasm by its association with the Kelch domain of the Kelch-like erythroid CNC-homologue (ECH) associated protein KEAP-1, pending ubiquitination and

---

[†] A disproportionation reaction in which both oxidised and reduced products are generated

degradation in a Culicin-3 (Cul3) dependent pathway (Gan *et al.*, 2010). Several models have been postulated as to how Nrf-2 escapes its association with KEAP-1 following stress. Clements *et al*, propose the involvement of DJ-1/PARK-7 (A cancer associated protein first identified in Parkinson's disease) (Clements *et al.*, 2006). Other radical explanations propose the phosphorylation of Nrf-2 by upstream activated protein kinases, such as MAPKK/MEK, phosphatidyl inositol kinase (PI3K) and even Protein Kinase C (PKC) facilitating escape from KEAP-1 and nuclear trans localisation. Zipper and Mulcahy suggest that the modification of the reactive cysteine residues of KEAP-1 enables the dissociation of Nrf-2, hence proposing KEAP-1 and not Nrf-2 as the sensor of oxidative stress (Zipper and Mulcahy, 2003). In Contrast, Li *et. al.* identified a canonical redox sensitive nuclear export signal in the transactivation domain of the Nrf-2 protein and a redox sensitive leucine zipper domain of the Nrf-2 protein, suggesting that the Nrf-2 protein itself is sensitive to oxidative stress (Li *et al.*, 2010). The observations that phosphorylation of Nrf-2 by PERK kinase facilitates its nuclear localisation and that Nrf-2 up regulates the expression of genes involved in the anti-oxidant response including Atf-4, provides insights into the possible link between phosphorylation events similar to those of both AAR, ER stress and redox homeostasis.

1.7.4 The Integrated Stress Response (ISR)

GCN-2 for AAR and PERK for UPR/ER stress response are only two of four known ER resident kinases capable of sensing changes in cellular homeostasis. Another, the Heme regulated inhibitor (HRI or EIF2αK1), is activated in response to heme deprivation as well as oxidative and heat stress in erythroid tissues. The accumulation of idiogenic double-stranded DNA (dsDNA, such as from bacteria) or severe inflammation activates the Protein kinase R (PKR or EIF2AK2). Regardless of the stressor, phosphorylation of eIF-2α converges several distinct stress sensing pathways to the termination of unneeded translation and the up-regulation of specific mRNA required in a targeted response; notably that for Atf-4. This has led to the proposal of a single central integrated stress response (ISR) pathway with Atf-4 transcription at its core and presiding over distinct and appropriate cellular responses to specific stress types. Harding and colleagues observed a rapid and sustained accumulation of Atf-4 protein and downstream genes such as GlyT-1 in wild-type mouse fibroblasts following ER stress or amino acid removal (Harding *et al.*, 2000). This observation was reversed in PERK and Atf-4 mutants.
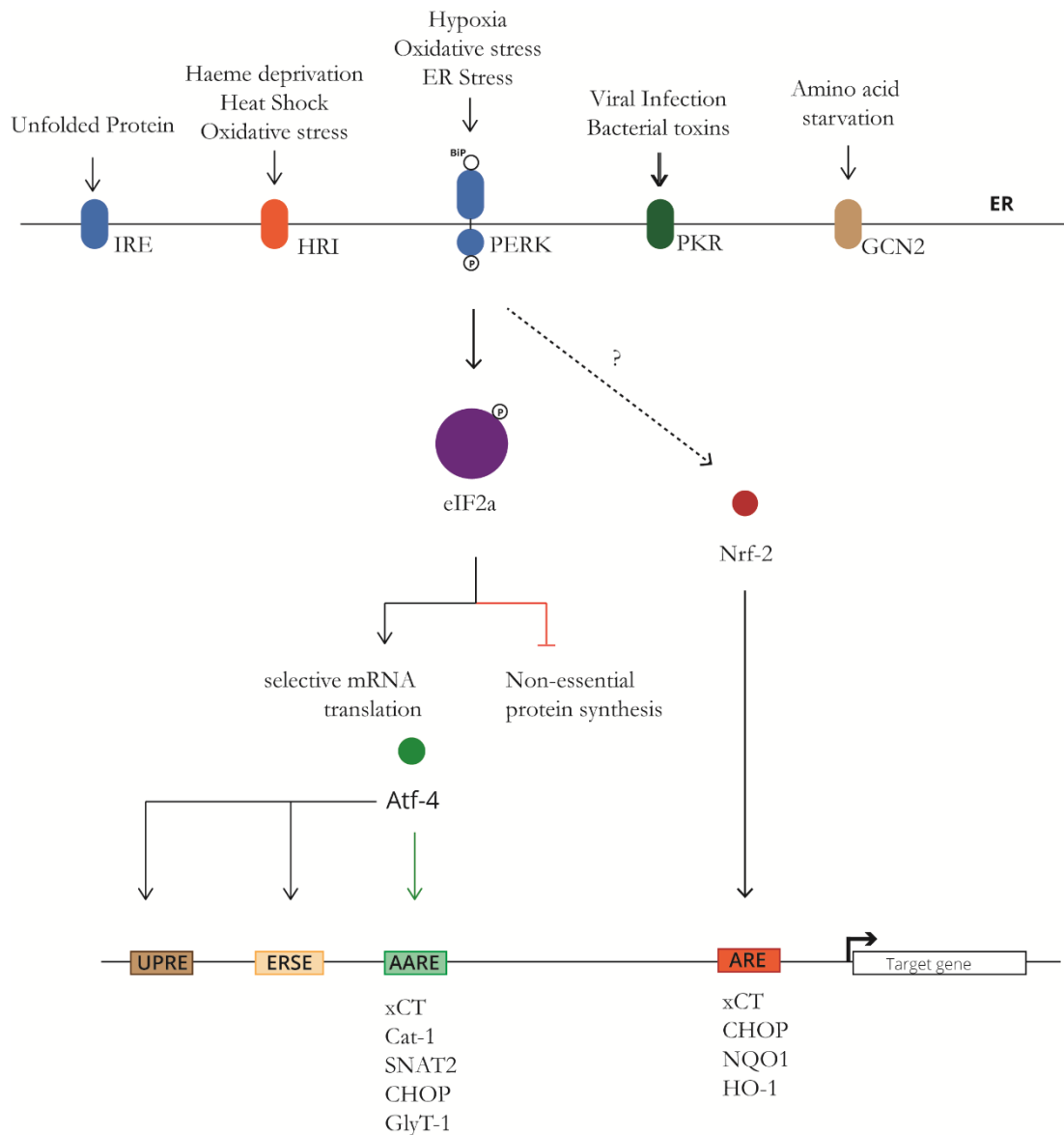
Figure 1.8: Illustration of the main pathway of the integrated stress response (ISR); A convergence point of several stress response pathways in the phosphorylation of eIF2a. That Atf-4 is activated after eIF2a has led to investigations in its role in several stress responses; this includes in response to unfolded protein, heme deprivation, viral infection, amino acid starvation and oxidative stress. It has been suggested that Nrf-2 another stress response BZIP factor may be activated by PERK although exactly how is not fully understood. Upon activation, both Atf-4 and Nrf-2 may regulate the expression of target genes via specific response elements in the promoters.

1.8 Hypothesis, Aims and Objectives

Given that stress plays an important role in the pathogenesis of several diseases, it is important to understand the gene regulatory modules and networks thereof, mediating the cellular adaptations to environmental changes. Several glycine transporters are responsible for the high affinity supply of glycine; and in the human intestine, GlyT-1a is the predominant isoform. The supply of glycine to epithelial cells of the human intestine, together with cysteine serves as an important precursor to the production of glutathione, bolstering cellular defences against oxidative stress, as well as protecting against other types of stress. Thus motivated, the molecular events involved in the transcriptional regulation of GlyT-1a, downstream of the stress response as contributory to glycine cyto-protection were investigated.

Using cellular models of the human intestinal epithelium and stress initiating agents, I aimed to explore the role of components of the integrated stress response in the regulation of the human GlyT-1a gene, examine key players involved in the stress response and amino acid transport control as potential transcriptional regulators of GlyT-1a and to propose a model network for the regulation of GlyT-1a during stress

To achieve these aims, a number of hypotheses have been developed based on a present understanding:

1) That the transcriptional up-regulation of GlyT-1a is mediated by the action of Atf-4, downstream of amino acid starvation, oxidative stress and endoplasmic reticulum stress;

2) The transcription of GlyT-1a is controlled by specific stress response elements and other regulatory motifs; and

3) Permutations of binding partners amongst the various members of the basic leucine zipper (BZIP) family of transcription factors such as Atf-4 and Nrf-2 modulate the specificity of the response to stress.

<div align="center">

CHAPTER 2.

MATERIALS AND METHODS

</div>

Outline: This chapter describes the experimental procedures used throughout this project. All protocols described here were validated and approved by respective administrative bodies and conducted under defined safety precautions.

Contents:

2.1 Intestinal cellular model

In the work described here, two intestinal cellular models were used. These are the human colonic carcinoma cell-line (Caco-2), and the human illeo-caecal carcinoma cell-line (HCT-8).

2.1.1 Caco-2, The Human colonic carcinoma cell line

Originally described by Fogh *et al.* (1977), colonic carcinoma cell lines like SW480, T-84, HT-29 and Caco-2, have extensively been used as *in vitro* models to investigate various intestinal properties (Sambuy *et al.*, 2005). These include for the study of gut physiology, barrier permeability, amino acid transport, nutrient absorption and metabolic regulation.

When grown in culture, Caco-2 cells show striking structural and biochemical similarities to human colonocytes but differentiate to small intestinal-like cells (Engle *et al.*, 1998). Their desirability as intestinal models over other human colonic carcinoma cell lines stems from the fact that, in addition to better representing several *in vivo* properties, they spontaneously differentiate in culture after confluence (Delie and Rubas, 1997, Engle *et al.*, 1998, Sambuy *et al.*, 2005).

When grown in culture, Caco-2 monolayers exhibit high cell functional polarity with well differentiated microvilli on the apical side, tight junctions between adjacent cells and characteristic expression of enzymes and membrane transporters (Hilgers *et al.*, 1990, Mariadason *et al.*, 2000). Most colonic carcinoma cells lines show varying degrees of differentiation when grown in culture (Bolte *et al.*, 1997). Some such as T-84 cell lines unlike Caco-2, exhibit uncharacteristic protein expression profiles. HT-29 on the other hand, does not differentiate in culture unless induced under specific conditions to form some enterocyte-like cells (Rousset, 1986). Subpopulations of Caco-2 cells have heterogeneity of properties, with satellites of microvilli, tight junction formation, and distinct morphologies post confluence.

The suitability of cultured Caco-2 cell line as an intestinal model has ushered its rapid propagation across many laboratories. Several groups have attempted to develop specific homogenous clones for specific applications. Culture conditions, supplements, etc. are all effective selective pressures that can significantly alter results from Caco-2 modelled studies; making the interpretation and correlation of data more difficult. For this purpose, the work described here used the parental Caco-2 cells (ATCC no. HTB-37™) which was routinely cultured under the same conditions as described in Section 2.1.3.

2.1.2 HCT-8, The Human illeo-caecal carcinoma cell line

HCT-8 is a human illeo-caecal colorectal adenocarcinoma derived cell line (Tompkins *et al.*, 1974). It is now known to be the same as the parental HRT-18 cell line (Vermeulen *et al.*, 1998, Vermeulen *et al.*, 1997). Their distinctive high proliferation rate and increased migration provides a low passage model for the study of intestinal physiology on phenotypically characteristic human colonocytes. When compared to Caco-2 cells, their characteristic invasive phenotype is associated with a smaller cell size and rounded morphology which is believed to be the result of the loss of an α-catenin gene (Vermeulen *et al.*, 1997). α-catenin is a crucial requirement for the formation of the E-cadherin-catenin

complex which when linked to the cytoskeleton enables cell-cell adhesion and retention of an epithelioid morphology (Wijnhoven *et al.*, 2000).

Like Caco-2 cell lines, several HCT-8 clones have been developed, notably the many α-catenin positive HCT-8 clones which despite their reduced invasion *in vitro*, have very similar tumorigenic behaviour as that of α-catenin negative clones when transplanted to nude mice (Van Hoorde *et al.*, 2000). Such similarities in *in vivo* characteristics have led to their interchangeable use with Caco-2 cells to study colonic parameters. Unlike Caco-2 which undergo spontaneous enterocytic differentiation and polarization (and can also de-differentiate), HCT-8 despite their ability to form a polarized monolayer, when grown on micro-porous filters do not differentiate in *in vitro* culture no matter the conditions; However, some differentiation markers such as the actin binding protein Villin, are expressed, all be it at a much lower level when compared to Caco-2 (Chantret *et al.*, 1988).

Some of the work described here used the parental HCT-8 cell line (ATCC no. CCL-224) obtained from the American Type Culture Collection (ATCC) routinely cultured under the same conditions as described in Section 2.1.3

2.1.3 Routine cell culture

To minimize contamination from infectious splashes, all tissue culture was performed using aseptic techniques in a SafeFlow® 1.2 laminar flow hood (Class II). Caco-2 cells (passage number 91-96) and/or HCT-8 cells (passage number 30-40) were routinely cultured in Dulbecco's modified Eagle's medium (DMEM) with 4.5 g/L glucose, supplemented with 10% (*v/v*) foetal calf serum, 1.2% (*w/v*) non-essential amino acids, 100 units/mL penicillin and 100 μg/mL streptomycin. Unless otherwise stated all reagents were obtained from Sigma UK. Cells were grown to 90-100% confluence under conditions of 5% CO_2, 95% air and 37°C in a CO_2 cell culture incubator (Sanyo). The culture medium was replaced every 48-72 hours. Confluent cells were 'passaged' for either experimentation or routine culture with seeding densities as shown in Table 2.1. To passage cells, the confluent monolayers were washed once with sterile phosphate-buffered-saline (PBS: pH7.3, 137 mM $NaCl_2$, 2.7mM KCL, 4.3 mM Na_2HPO_4, 1.4 mM KH_2PO_4). Cells were then incubated with 5 mL of trypsin-EDTA solution (containing 0.25% trypsin and 0.03% EDTA) for either 5 minutes (Caco-2) or 2 minutes (HCT-8) at 37°C to enzymatically disrupt and detach the monolayer. 10 mL of routine culture medium (containing 10% foetal calf serum) was added to stop trypsinisation. The

dissociated cell suspensions were then transferred to a 25 mL universal and centrifuged for 3 minutes at 1500 rpm to pellet the cells. The resulting supernatant containing neutralized trypsin was discarded and cell pellet re-suspended in 10 mL of routine culture medium. The cells were counted on an automated T4 cellometer (Nexcelcom) and seeded onto new culture vessels as shown in Table 2.1. Cells were often inspected under a light microscope and/or tested for mycoplasma infection.

Table 2.1: Cell seeding densities for varying culture vessels. The amount of culture medium required for routine culture is indicated as well as volume of trypsin required to detach the monolayer from culture vessel following 5 minute incubation at 37°C. Caco-2 or HCT-8 cells were routinely cultured in T-175 flasks. Transfections for reporter assays and gene knockdown experiments were performed on cells grown on 12-well plates. Nuclear protein extracts were from Caco-2 cells grown on 6-well plates. All culture vessels were from Corning, UK.

	Surface Area (mm2)	Caco-2 Seeding Density	HCT-8 Seeding Density	Trypsin (mL of 0.25% trypsin, 0.03% EDTA	Volume Growth Medium (mL)
12-well plate	401	0.3×10^6	0.25×10^6	1	1
6-well plate	962	0.5×10^6	0.3×10^6	2	3
T-175 flask	1750	3.0×10^6	0.5×10^6	5	30

2.1.4 Stress treatment

Regimes for stress treatments were maintained as used previously in this lab (Howard and Hirst, 2011). Prior to stress treatment, cells were grown under optimal conditions in complete medium for an extended period to ensure there was no exertion of either nutrient or other physiological stresses. For consistency, the morphology and confluence of cells were noted and maintained across all experiments. No foetal calf serum was added to stress medium

2.1.4.1 Amino acid and Glycine starvation

For amino acid starvation assays, cells were incubated for the desired length of time in freshly made amino acid free stress medium with formulation shown in Table 2.2. For glycine starvation, freshly made glycine free stress medium as per the formulation shown

in Table 2.2, supplemented with all other amino acids except glycine. Whilst antibiotics, and Glucose and Phenol Red were added to amino acid and glycine starvation media, no foetal calf serum was used in this medium. Given that the medium used for routine cell culture contained 30 mg/L of glycine, for Glycine supplementation assays twice as much glycine (60 mg/L) was used.

2.1.4.2 Oxidative Stress treatment

For the induction of oxidative stress response, either 0.2 mM diethyl-maleate (DEM) for 4 hours or 60 μM tert-Butyl Hydroquinone (tBHQ) for 16 hours was added to serum free stress medium. DEM – an electrophilic agent is widely used in models for oxidative stress. DEM effectively depletes glutathione levels within the cell, by binding to and oxidising its free sulfhydryl (thiol) groups, leading to the build-up of free radicals and other reactive oxygen species within the cell. These are able to cause oxidative modifications to proteins, leading to their fragmentation, or proteolytic degradation, or increase intracellular accumulation of oxidised protein.

Table 2.2: Formulation of base medium for nutrient availability studies.

Component	For Amino acid Starvation	For Glycine Starvation	For Glycine Supplement.
Calcium Chloride (CaCl2)			265 mg/L
Ferric Nitrate (FeNO3)			0.100 mg/L
Potassium Chloride			400 mg/L
Magnesium Sulphate			97.72 mg/L
Sodium Bicarbonate			3700 mg/L
Sodium Chloride			6400 mg/L
Sodium Phosphate Dibasic			141 mg/L
Penicillin			100 U/mL
Streptomycin			100 μg/mL
Glycine‡	0 mg/L	0 mg/L	60 mg/L
L-Arginine	0 mg/L		84 mg/L
L-Alanine	0 mg/L		66 mg/L
L-Histidine	0 mg/L		42 mg/L
L-Glutamine	0 mg/L		584.4 mg/L
L-Leucine	0 mg/L		105 mg/L
L-Cystine	0 mg/L		62.6 mg/L

‡ According to Documenta Giegy Scientific tables, the physiological concentrations of glycine in blood serum is 17.7mg/L.

L-Lysine	0 mg/L	146 mg/L
L-Methionine	0 mg/L	30 mg/L
L-Serine	0 mg/L	42 mg/L
L-Threonine	0 mg/L	95 mg/L
L-Tryptophan	0 mg/L	16mg/L
L-Tyrosine disodium Salt	0 mg/L	103 mg/L
L-Valine	0 mg/L	94 mg/L
Glucose		4500 mg/L
Phenol Red		15.9 mg/L
MEM Vitamin Solution§		1 % v/v of 100x Stock. (See footnotes)

Evidence points to an ER-independent redox regulatory mechanism with many cell types. tBHQ promotes the oxidative stress response by preventing the degradation of key antioxidant transcription factors such as Nrf-2. Peroxide formation in the mitochondria may be a consequence of mitochondrial calcium overload induced which is induced by tBHQ (Peng and Jou, 2010).

2.1.4.3 Induction of UPR and ER stress

For comparison between an UPR specific response and ER stress, thapsigargin was used to elicit a Ca^{2+} ER stress response whilst Tunicamycin was used to mimic an UPR response. Cultured Caco-2 cells were incubated in serum free DMEM medium containing 50nM thapsigargin for 16 hours or alternatively in serum free DMEM medium containing 1µM tunicamycin for 16 hours.

Tunicamycin and Thapsigargin, two stress agents known to elicit the UPR and ER stress in several cell types were used. Derived from Streptomyces, the glucosamine-containing antibiotic Tunicamycin is a potent inducer of the UPR. By specifically inhibiting, dolichol pyro phosphatase mediated N-linked glycosylation of asparaginyl residues, it facilitates the formation of N-glycosidic protein-carbohydrates, inducing ER stress. Toxico-genomic analysis of tunicamycin stress in the cells of liver and small intestinal tract of C57BL/6J mice have revealed activation of a host of several downstream genes; these include chaperones, proteins involved in ubiquitination and proteolysis, electron and amino acid transporters and detoxification enzymes and other proteins required for cell

§ MEM Vitamin solution from Gibcco® (Catalogue Number 11120). 100x Stock Contains: D-Ca panthothenate (100 mg/L), Choline chloride (100 mg/L), Folic acid (100 mg/L), i-Inositiol (200 mg/L), Nicotinamide (100 mg/L), Pyridoxal HCl (100 mg/L), Ribovflavin (10 mg/L), and Thiamine HCl (100 mg/L).

growth, apoptosis and the cell cycle (Nair *et al.*). Thapsigargin is a *Thapsia garganica* (also known as 'deadly carrots') plant derived lactone built from three isoprene units. It binds tightly to (and only to) the unique carboxyl tail of the intracellular sarcoplasmic or ER Ca^{2+}-dependent ATPase (SERCA) pump of the ER membrane. This leads to a rapid release of Ca^{2+} from these stores and ER stress. It is believed that the ensuing multifaceted stress response following Ca^{2+} signalling results from changes in ionic balance within the ER. The uncontrolled calcium influx to the mitochondrion is believed to play an important role in inducing the apoptotic machinery (See section 1.7.2).

2.2 *In Silico* Identification of GlyT-1 transcriptional regulatory network

Modern computer algorithms used in the identification of gene regulatory networks can be classified into three main categories. This includes 1) those which perform a comparative promoter analysis of co-regulated genes from micro-array expression or other expression profiling data sets, to identify regulatory sequences of known transcription factors, and/or 2) where there is limited expression data for the given genome or genes of interest, comparative genomic (mostly phylogenetic foot-printing) analysis using orthologous promoter sequences from multiple species allows for identification of both putative TFBS and their corresponding TFs. 3) Lastly, a combinatorial approach as shown in Figure 2.1, consisting of approaches 1) and 2) to build a refined promoter dataset to which detailed composite element analysis is performed for identification of known TFBS. A combination of all of the above was used for the identification of conserved regulatory motifs within the GlyT-1a promoter as described below.

1

Analysis of the GlyT-1a sequence

- Search for CGI islands
- Identify Basal regulatory elements (BRE)
- Determine conservation (PhastCons)
- Histone acetylation marks (ENCODE)
- TFBS search (TRANSFAC MATCH)

2

Promoter Collection

- Identificaiton of GlyT-1a co-regulated genes
- Factor dependent expression profile analysis using CORD .
- Only genes with more than 80% correlation to expression of GlyT-1a are used

4

TFBS/Composite Model Analysis

- Use TRANSFAC Composite Model
. Analyst Tool to find putative Composite regulatory modules (CRMs)

3

Search Overrepresented TFBS in promoters

- Use TRANSFAC F-MATCH to detect
. frequently occuring TFBS in co-regulated promoter collection

5

Phylogenetic Analysis

- Use MAFFT alignment of GlyT-1a
. orthologs to determine conservation score of putative CRMs

6

Experimental Validation

- Use EMSA, ChIP and Reporter assays to
. validate Transcription factor binding to predicted sites

Figure 2.1: Strategy for the identification of GlyT-1a regulatory modules

2.2.1 Identification of GlyT-1a co-regulated promoters

In recent years, the use of high throughput expression profiling to identify gene expression changes following specific cellular stresses simplifies the task of identifying probable gene regulatory modules (regulatory and co-regulated protein networks). Computational clustering algorithms on the available genome wide expression profiles from microarray datasets in public databases such as the Gene expression omnibus NCBI-GEO(Edgar *et al.*, 2002) and GEO2R – an R-based web application for analysing GEO data (Barrett *et al.*, 2013) are widely used to this effect. There is presently no comprehensive expression dataset for the integrated stress response (ISR) in epithelial cells of the human intestine. However, using the CO-Regulation database (CORD), genes with similar factor dependent expression profiles to GlyT-1a/SLC6A9 were catalogued from public micro-array datasets. Analysis was performed on the promoters of the co-regulated genes using the TRANSFAC F-MATCH algorithm to identify enriched TFBS. From the list of enriched TFBS, composite models were built using the TRANSFAC composite model analysis tool to detect the arrangement of TFBS within putative Composite regulatory modules (CRMs)

2.2.2 Conservation analysis of the putative GlyT-1a CRM

Combining phylogenetic analysis with detailed analysis of promoters of co-regulated and co-expressed genes, to ascertain the occurrence of similar motifs provides added evidence to possible biological function of such putative sites (Wang and Stormo, 2003). For phylogenetic analysis of the GlyT1-a promoter, sequence spanning 2000 bases upstream of the transcription start site (TSS) of 56 orthologous transcripts (Appendix A.3), catalogued in the database of orthologous groups (OrthoDB) were obtained from corresponding species-specific ENSEMBL databases using the BioMart Perl API. All Perl scripts used are documented in Appendix A. Sequences were aligned using the MAFFT plugin of the Geneious Bioinformatics suite version R7 (Kearse *et al.*, 2012). Arbitrary conservation scores for putative TFBS (regulatory motifs) were calculated using a similar method as described by (Hestand *et al.*, 2008); briefly the percentile conservation score is calculated as the sum of the percentage alignment at each position of the TFBS motif, divided by the motif length.

2.3   Endpoint Polymerase Chain Reaction (PCR)

2.3.1 Genomic DNA extraction for endpoint-PCR

Genomic DNA from $2 \times 10^6 - 3 \times 10^6$ Caco-2 or HCT-8 cells was extracted using the GenElute® Mammalian genomic DNA Miniprep kit (Sigma) according to the manufacturer's instructions. DNA content of the final 50 μL eluent was quantified on a Biomate™ 3 spectrophotometer from Thermofisher. 2 μL extracted genomic DNA was also electrophoresed for 1 hour with 5V per cm on a 0.5x TBE agarose horizontal gel and visualized under UV in a Bio-Rad Gel Documentation System. Genomic DNA was stored at 4°C and used as template for amplification of the GlyT-1a 5'flanking region by endpoint PCR.

2.3.2 Primer design for endpoint-PCR

Primer pairs for endpoint-PCR were designed using the NCBI Primer3 portal, and synthesised by Integrated DNA Technologies (IDT, Belgium). Corresponding sense and antisense oligonucleotide primers for endpoint-PCR were designed to be at least 19 bases in length, GC content within a 45-60% range. In addition to these, care was taken to avoid the tendency of secondary structure (hairpins, self-dimers, cross-dimers) formation. Where possible GC clamps were included at the 3' primer end. Primer melting temperature (T_m) of between 55°C and 65°C were preferred. All primer T_m were re-calculated using the base-stacking nearest neighbour thermodynamic algorithms (SantaLucia, 1998). These calculations allow for salt correction of melting temperatures as shown in Equation 1 below.

$$T_{m_{\text{primer}}} = \frac{\Delta H}{\Delta S + R\ln([primer])} - 273.15$$

Equation 1: Calculation of PCR primer melting temperature (T_m). Where ΔH (kcal/mole): Change in enthalpy obtained by adding all dinucleotide enthalpy values of each nearest neighbour base pair adjusted for helix initiation factors (Peyret *et al.*, 1999, SantaLucia, 1998); ΔS (kcal/mole): Change in entropy adjusted for the contribution of Mg^{2+} salts used to the entropy of the system; R is the universal gas constant (1.987 cal/°C Mol); [primer] is the final concentration of the primer. Where primers have a high tendency for self-complementarity [primer]/4 is used (SantaLucia, 1998).

T_m values calculated using this formula did not vary significantly from values calculated using the GC content formula (Equation 2) for sequences longer than 14 bases. To estimate the optimal annealing temperatures (T_a) for effective DNA-primer hybrid stability during amplification used T_m values of both the primers and the anticipated product were considered as shown in Equation 3 above as originally described by (Rychlik *et al.*, 1990).

$$T_{m_{target}} = 81.5^{o}C + 0.41^{o}C \; \acute{} \; (\%G + \%C) + 16.6^{o}C \; \acute{} \; (\log_{10}[Na^{+}] + [K^{+}]) - 675/N$$

Equation 2: Salt adjusted formula for calculation of melting temperatures (T_m) For DNA sequences of length N longer than 14 bases and assumes the presence of monovalent ions in the PCR reaction.

$$T_a = 0.3^{o}C \; \acute{} \; T_{m_{primer}} + 0.7^{o}C \; \acute{} \; T_{m_{target}} - 14.9$$

Equation 3: Annealing temperature ($T_{a)}$ calculation for endpoint PCR. Where, $T_{m_{primer}}$ is the melting temperature of the primers and $T_{m_{target}}$ is the melting temperatures of the target sequence or amplicon. As the nearest neighbour thermodynamic model is only applicable to short oligonucleotides $T_{m_{target}}$ was calculated based on GC content as shown in Equation 2.

2.3.3 Endpoint-PCR using a Thermo Scientific Px2 Thermal Cycler

Amplification of genomic DNA target sequences by PCR was performed on a Px2 thermal cycler (Thermo Scientific) using 1.25 U of GoTaq® DNA polymerase (Promega) per reaction. Briefly for each reaction, a 50 μL volume was used. This consisted of 10 μL of 5x GoTaq® Green PCR buffer (pH 8.5, 1X contains 1.5 mM Mg²⁺) from Promega, 5 μL each of 10 μM forward and reverse primers (final concentration of each primer 0.5 μM), 5 μL of a 10 x dNTP mix (final concentration of each dATP, dTTP, dCTP, dGTP in reaction mix was 0.2 mM), 2.5 μL of the template DNA (less than 0.5μg final

concentration), 1.25 U of GoTaq® polymerase enzyme and nuclease free water. A standard temperature program was used for amplification which included a single initial 2 minute activation step at 95°C followed by 35 cycles of denaturing at 95°C for 1 minute, annealing at the calculated T_a (as described in Section 2.3.2 for each primer set) for 90 seconds and a final extension step at 72°C for S seconds depending on the size of the expected product, where S is equal to 60 seconds per each kilo base of the expected product length. A final extension cycle of 8 minutes was added to the program before final incubation at 4°C.

2.3.4 Electrophoresis, Imaging and Purification of endpoint-PCR products

10 µL of endpoint-PCR products were analysed on a 1% agarose/1x Tris-borate EDTA (TBE) horizontal gel by electrophoresis. Samples were electrophoresed for 1 hour with 5V per cm of gel length and visualized under UV and imaged on a Bio-Rad™ Gel Doc XR system. Endpoint-PCR products were extracted from the excised gel bands using either the GenElute™ Gel extraction Kit (Sigma) or the Montage™ gel extraction kit (Millipore) according to their respective manufacturer's specifications. PCR products were assayed for contaminants and quantified on a Biomate 3 spectrophotometer. Samples were stored at -20°C until required for cloning or use as a secondary template in ligation-mediated PCR (see Section 2.5.1).

2.4 Quantification of gene expression by real time PCR (QPCR) assays

Real time quantitative PCR (QPCR) was used to quantify gene expression or gene signal enrichment. All QPCR procedures were conducted and data obtained for presentation in accordance with the revised MIQE (Minimum Information for the publication of Quantitative Real-Time PCR Experiments) guidelines (Taylor *et al.*, 2010).

2.4.1 Transfection of mammalian cells with short interfering RNA (siRNA)

Caco-2 or HCT-8 cells were seeded onto 12-well plates (Corning) and grown for 24 hours in complete medium at optimal conditions as described in Section 2.1.3. Transfection was only performed on cell monolayers with confluency between 60-80%. For gene knockdown analysis, mammalian cells were transfected with at least two siRNAs targeting the transcription factors Atf-4 (s1702 and s1704 (Silencer Select, Life Technologies); sc35112 (Santa Cruz)), ATF6 (s223543 and s223544 (Life Technologies)), XBP-1 (s14913 and s14915 (Life Technologies)) and Nrf2 (s9491 and s9492 (Life Technologies)). 5 ρmol of each siRNA, diluted in 200 μL of OPTIMEM® (Life Technologies) was complexed to 2 μL of Lipofectamine RNAi MAX transfection reagent (Life Technologies) for 20 minutes at room temperature. The transfection complex was then added to cells pre-incubated in 1 mL of antibiotic free medium for 24 or 48 hours. This was then followed by stress treatment with tunicamycin, thapsigargin, amino acid starvation, or DEM as described in Section 2.1.4

2.4.2 Total RNA isolation

Following transfections, total RNA was extracted from cells using the SV Total™ RNA extraction system (Promega, Southampton UK) according to the manufacturers specification. Adherent cells were lysed in a 10 mM Tris-HCl (pH 7.5) based lysis buffer containing 4 M guanidine thiocyanate (GTC), and 0.97% β-mercaptoethanol to inactivate the ribonucleases present in the cell extracts. Cell lysates were aspirated through a 20-gauge syringe needle to shear genomic DNA and diluted further in a dilution buffer containing excess GTC and incubated at 70ºC for 3 minutes to selectively precipitate proteins from solution. Samples were then filtered by vacuum through SV Total spin baskets to allow for binding of RNA to the silica surface of the component glass fibres. RNA samples were then treated for 15 minutes with DNAseI enzyme to cleave any contaminating DNA and the reaction stopped with a DNAseI stop solution (final concentrations: 2M GTC, 4mM Tris HCl [pH7.5], 57% ethanol). A single filtration pass

of wash solution (final concentrations: 60 mM Potassium acetate, 10mM Tris HCl [pH 7.5, RTP], 60% ethanol) through the RNA-bound glass fibres of the spin baskets under vacuum was used to precipitate out the resulting RNA. Isolated RNA was then eluted in 50 µL of nuclease free water. The SV Total™ offers purer DNA-free RNA due to an additional DNAseI-I treatment step. The concentration of the extracted RNA was quantified by measuring the absorbance at 260 nm on a Biomate 3 spectrophotometer. RNA integrity was determined on an Agilent™ Bio Analyser (Agilent). A RNA integrity Number (RiN) greater than 8 was required for use of the sample. RNA samples were stored at -80°C until required for reverse transcription as described in the next section (Section 2.4.3)

2.4.3 Reverse transcription reaction

First strand complimentary DNA (cDNA) synthesis from isolated total RNA was performed by incubating 0.5 µg of extracted total RNA, 500 ng of random hexamers (GE Biosciences), 0.5 mM of each dNTP (Promega, Southampton UK), 1x reaction buffer (50 nM Tris HCl [pH 8.3, RTP], 75 mM KCl, 3 mM $MgCl_2$, 10 mM DTT), 20 U RNAsin (Promega, Southampton UK) and 100 U of M-MLV RNAse H^+ Reverse Transcriptase (Promega, Southampton UK) in a final volume of 20 µL for 120 minutes at 42°C. Two controls were included with each sample batch; these included a no RNA control and a no reverse transcriptase enzyme control. The samples where then incubated at 70°C for a further 1 minute to inactivate the enzymes. The resulting cDNA was diluted by a factor of one in five. Both undiluted and diluted cDNA samples were stored at -20°C until required for QPCR.

2.4.4 Primer design for QPCR

Primers for QPCR shown in Table 2.3 were designed with considerations as described in Section 2.3.2. Additionally to dissuade amplification from possible contaminating genomic DNA primers were designed – where possible, to span an exon-exon boundary, and preferentially for both the upstream and downstream primers to be at least 1000 bases apart on different exons on the same genome. Primers amplifying products not greater than 250 bp were preferred.

Table 2.3: Primer pairs used for quantification by QPCR. Annealing temperatures (Tₐ) used for QPCR amplification from genomic DNA are indicated. Primers for QPCR were synthesised and HPLC purified by Integrated DNA technologies (IDT, Belgium). Sequences are listed in the 5' to 3' orientation.

Primer Pair	Target template accession number	Primer Sequences	Temp. (T_a)	Product Size
Atf-4[**]	NM_001675	CTTCAAACCTCATGGGTTCTCCAG	60	171
		AACAGGGCATCCAAGTCGAACTC		
Atf-6	NM_007348.3	TCTCGTCTCCTGCCTCAGTG	60	151
		TTATCTTCCTTCAGTGGCTCC		
XBP-1[**]	NM_005080	GACAGCGCTTGGGGATGGATG	60	149
		GAGGGGTGACAACTGGGCCTG		
Nrf-2[**]	NM_006164	CAGCCAGCCCAGCACATCCA	60	120
		TGTCTGCGCCAAAAGCTGCA		
Transporters				
xCT	NM_014331	GTTTTGCACCCTTTGACAATGA	55	120
		GTTCATCCCAGCTTTGTTTTCC		
GlyT-1	NM_001024845	TCGGGAGGCTGATGCAACTTTC	55	220
		GGCACAGCACCATTCAGCATC		
PepT-1	NM_005073	TTCCACAATCATCACACC	50	192
		GGTTACGAGGGTGATGTT		

2.4.5 Preparation of QPCR standards and normalisation

For QPCR assays a relative standard curve method was used (Larionov *et al.*, 2005). Endpoint-PCR (as described in Section 2.3) was used for the amplification of corresponding sequences from cDNA templates. Products from endpoint PCR were electrophoresed on a 1x TBE agarose gel and visualised under UV light. Gel bands corresponding to the expected product size were excised and purified as described in Section 2.3.4 . Purified PCR products were cloned into the TA insertion site of a pGEM-T-Easy plasmid vector (Promega, Southampton UK, plasmid map illustrated in Appendix A.4). *E. coli* strain JM109 (Promega, Southampton UK) were transformed with the recombinant plasmid. 20 hours post transformation and incubation on ampicillin infused Luria Bertani (LB) agar plates at 37ºC, a single positive colony was exponentially grown for a further 18hours in LB broth. The cloned plasmids were extracted from JM109 using

---

[**] Primers used detect all transcript variants (Atf-4 NM_001675 or NM_182810; Nrf-2 NM_001145412, NM_001145413 or NM_006164; XBP-1: NM_005080.3 and [NM_001079539.1 with product size 123bp])

the GenElute™ Plasmid Miniprep kit (Sigma). Plasmids were then sequenced using the common M13 20-mer forward sequencing primer (5'-GTAAAACGACGGCCAGT-3') by GeneVision (Newcastle, UK). A BLAST search of the sequencing result was also carried out to confirm accuracy of amplicon and sequence integrity. Plasmid DNA samples containing identical full length target sequence for each gene of interest were stored at -20ºC, and a 1/100 or 1/1000 dilution series to cover appropriate cycle threshold ranges, was used as QPCR standards as described below.

2.4.6 QPCR using Roche Light Cycler 480

Relative quantification of expressed mRNA message for the genes of interest was completed on a Roche Light Cycler 480 using the SyBR-Green I technique. Briefly, sample duplicates of 10 µL QPCR reaction volume were used. Each consisted of 5 µL of 2X SyBR Green mix (Roche UK), 0.5 µL of a 10 µM primer pair mix, 2.0 µL of the template DNA and 2.5 µL of nuclease free water. The thermo cycling program consisted of an initial denaturing step at 95ºC for 5 minutes followed by 45 cycles of a further denaturing step at 95ºC for 5 seconds, annealing at the optimal T_a for each primer set as shown in Table 2.3 for 10 seconds, and an extension step at 72ºC for a further 10 seconds. This was followed by melt curve analysis and a cooling step to 4ºC. The No RNA control and No Reverse transcriptase controls generated during cDNA synthesis (Section 2.4.3) together with a no-template-control (NTC) were used as negative controls of PCR amplification.

2.4.7 Data Normalisation and Analysis of QPCR assay

In QPCR, the crossover point (C_P also known as the cycle threshold C_T value) represents the cycle at which the measured fluorescence from the SyBR-Green I dye intercalated in the grooves of a nascent DNA amplicon exceeds a predefined threshold. C_T values are proportional to the accumulating concentration of amplicon from the PCR reaction. Owing to the logarithmic kinetics of the PCR amplification process, C_T values are calculated in the exponential phase of the reaction. Intra sample duplicate variations in C_T values of more than 1% were flanked as pipetting errors and the sample omitted from relative quantification calculations. Arbitrary concentrations of amplicon are determined from the standard curve of known concentration. The QPCR efficiency and coefficient of variability was calculated from the standard curve and recorded for an assay of intra-experimental consistency. Data are represented as the relative arbitrary concentration of

experimental group $\pm$ standard error of the mean (SEM). All statistical analysis was performed using the commercially available software package GraphPad (InStat, USA). Analytical tests used and inferences are highlighted in the data figure legends in the results chapter (Chapter 3).

For data normalisation against intra-sample expression variations, arbitrary concentrations calculated from a standard curve of C_T values from QPCR of three housekeeping genes: GAPDH, ATP-5B and TOP-1 are used. The suitability for these genes to be used as internal control was determined as per the GENorm algorithm (Guenin *et al.*, 2009). All data were normalised to a geometric mean (normalisation factor) derived from the arbitrary concentrations of these three housekeeping genes. For stress treated samples, relative messenger RNA (mRNA) abundance was reported as a percentage of the unstressed control sample average. A One-way analysis of variance (ANOVA) was used for comparisons between experimental groups.

2.5 Reporter gene Assays

To identify potential regulatory motifs, *in silico* analysis of the sequence upstream from the TSS to the translational start codon in exon 2 of the GlyT-1a nucleotide sequence was performed. Analysis of the first intronic sequence between exons 1 and 2 was carried out (as outlined in Section 2.2). Five β-galactosidase reporter plasmid constructs were designed containing sequences of the GlyT-1a corresponding to the promoter and the entirety of exon 1 (pG1PromE1); the promoter and 5-UTR (pG1Prom5U, excluding intronic sequence); the promoter, exon 1, intron 1, and the UTR of Exon 2 (pG1PromInt15U), the UTR sequence of exon 2 (pG1E2) and last a construct with just the GlyT-1a 5UTR (pG15U). In this section I describe how these constructs were generated and used (Sequence maps are included in the results section).

Table 2.4: Primer for reporter plasmid construction, All primers were synthesised and desalted by Integrated DNA technologies (IDT, Belgium)

Product	Target	Primer Sequences
Primer 1	GlyT1-Prom S	5'-TGCAGCCCTCCCAGGAAATAGC-3'
Primer 2	GlyT1-Exon 1 AS	5'-GCGGCGGTGGGTTGGGGCTC-3'
Primer 3	GlyT1-Exon 1 S	*5'-GAAGGAATTGGACTCCATCC-3'*
Primer 4	GlyT1-Exon 2 AS	5'-GCGGCGGTGGGTTGGGGCTC-3'
Primer 5	GlyT1-Exon 2 S	*5'-GAAGGAATTGGACTCCATCC-3'*
Primer 5b	GlyT1-Exon 2 S (ext)	*5'-aggccccc*GAAGGAATTGGACTCCATCC-3'
Primer 6	GlyT1-Intron 1 AS	5'-TTGCGGCGAGACAGACATCC-3'
Primer 7	GlyT1-Intron 1 S (ext)	*5'-tcgccgcaa*GGGCACCGGTTTCCTCCCAC-3'
Primer 8	GlyT1-Intron 1 AS	5'-GGGGGCCTGGCCTTGTGTAAG-3'

2.5.1 Preparation of reporter plasmids constructs.

End point PCR as described in Section 2.3 was used to amplify a 1346bp amplicon from Caco-2 genomic DNA template, corresponding to the promoter and exon 1 of GlyT-1a (G1PromE1) between nucleotides -1159 and +187 (relative to transcription start site TSS of NM_001024845). For this amplification, Primer 1 (a forward primer with annealing site in the promoter region of GlyT-1a) and Primer 2 (a reverse primer in exon 1 of GlyT-1a) was used. Primer sequences are indicated in Table 2.4.

Amplification by endpoint PCR (using Primer 3 and Primer 4) of the full 272bp GlyT-1a 5'UTR (G15U) was carried out using Caco-2 cDNA as template to exclude the large 6.9kb intronic sequence between exon 1 and exon 2 (Figure 2.2). To obtain a full length 1431bp promoter and 5'UTR sequence (G1Prom5U), ligation mediated PCR (LM-PCR) was used to ligate the G1PromE1 and Gl5U amplicons via their overlapping ends of exon 1 sequence.

The resulting endpoint PCR and LM-PCR products were purified as described in Section 2.3.4. To confirm the sequence accuracy of amplification and sequence integrity, the G1PromE1, G15U and G1Prom5U products were cloned into the TA insertion site of a pGEM-T-Easy plasmid vector (Promega, Southampton UK, plasmid map illustrated Appendix A.4). The cloned plasmids were replicated in *E. Coli* strain JM109 (Promega, Southampton UK) and sequenced using the common M13 20-mer forward sequencing primer (5'-GTAAAACGACGGCCAGT-3') by GeneVision (Newcastle, UK). Exact sequences were then sub cloned into the TA sites of a pBlueTOPO-TA vector (Life Technologies) or between the EcoR1 endonuclease sites, upstream of the β-galactosidase (LacZ) reporter gene sequence of pSF-PromMCS-BetaGal reporter vector from Oxford Genetics.
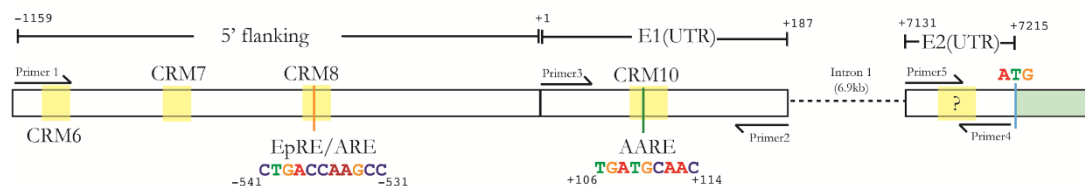
Figure 2.2: GlyT-1a Sequence map for the generation of reporter constructs. G1PromE1 was generated using Primer 1 and Primer 2. The G1PromE1 amplicon was ligated to an amplicon representing the GlyT-1a 5UTR (E1 UTR + E2 UTR) amplified from cDNA using Primer 3 and Primer 4. The locations of the putative composite regulatory modules (CRM) are shown in yellow. Numbering is relative to the Transcription start site (TSS) in Exon 1. All Primer sequences are indicated in Table 2.4.

All plasmid inserts were designed such that the transcription start sites within the proximal promoters were in the same frame with the reporter gene. The resulting β-galactosidase constructs were named correspondingly, pG1PromE1, pG15U and pG1Prom5U.

Reconstituted reporter plasmids were replicated in *E. coli* strain JM109, grown and selected on an antibiotic infused LB agar plates. Following a further 18 hour incubation of antibiotic selected positive colony in LB broth, amplified reporter plasmids were extracted using the QIAGEN Plasmid Maxi Kit (Qiagen) and sequenced to confirm sequence integrity. All reporter plasmids were stored at -20ºC until required for transfection as described below.

2.5.2 Transient transfection of reporter plasmids and stress treatment

Prior to transfection cells were grown under optimal conditions as described above (Section 2.1.3). Reporter plasmid constructs were transfected into Caco-2 cells grown on a 12 well plate. 2µg of reporter plasmid diluted in 200 µL JetPrime™ Buffer (Polypus) was complexed to 2µL JetPrime™ reagent (PolyPlus) for 20 minutes at RTP. 75 µL of the DNA/Reagent complex was added together with 1mL of complete DMEM growth medium to each sample well of Caco-2 cells (40-50% confluent). The "no-transfection" control cells were incubated with only JetPrime™ buffer. Cells were then incubated under conditions of 5% CO_2, 95% air and 37**°C** in a CO_2 cell culture incubator (Sanyo). The transfection medium was replaced 4 hours post transfection to minimise any adverse stress effects of the transfection reagents. After a further 20 hour incubation and growth phase, cells were treated with tunicamycin, thapsigargin, amino acid Starvation, tBHQ or DEM as described in Section 2.1.4. To investigate the influence of polyphenols on the response activity at the transfected GlyT-1a promoter, cells transfected as described above were treated with either 50 µM Resveratrol for 24 hours, 10 µM Genistein for 24 hours or a polyphenol soup from alcoholic brown seaweed extract (generously donated by Dr Peter Chater) for 30 minutes before stress treatment.

2.5.3 Extraction and quantification of total protein

Following stress-treatment, Caco-2 cells were washed with 1 mL PBS per well. Cells were then lysed by a single freeze-thaw cycle in the presence of 100 µL/well of lysis buffer (250 mM Tris [pH6.4], 0.25% (*v/v*) Nonindent P-40, 2.5 mM EDTA). Cell lysates were transferred to a 1.5 mL micro-centrifuge tube and centrifuged at 13000 G for 5 minutes at 4ºC. The cleared supernatant, containing total protein was transferred to a new micro-centrifuge tube, quantified, and sampled for β-galactosidase activity or stored at -80ºC until required. The protein concentrations of the cell lysates were determined using the Thermo Scientific Pierce 660 nm Protein assay in micro plate format. A standard curve

generated using bovine serum albumin (BSA) samples of known amounts was used to quantify sample protein concentration. Duplicate absorbance readings at 660 nm wavelengths were obtained for each sample 10 minutes after addition of Pierce reagent using a FLUOstar Omega spectrophotometer (BMG LabTech).

2.5.4 Beta-Galactosidase activity assay

Reporter gene activity was determined using the galactosidase analogue chlorophenol red-β-D-galactopyranoisde (CPRG) (Sigma). β-galactosidase enzyme catalyses the hydrolysis of the galactosidase analogue converting it from a yellowish-orange substrate to chlorophenol red chromophore. To quantify specific enzymatic activity, 20 μL of the cleared cell lysate from each sample (Section 2.5.3) was incubated for 60 minutes at 37ºC with 130 μL of 1 mg/mL of CPRG diluted in buffer containing 25 mM MOPs, 100 mM NaCl, 10 mM $MgCl_2$ at pH 7.5. The enzymatic reaction was stopped by the addition of 80 μL of 0.5M Na_2CO_3. Absorbance was measured at 575 nm using a FLUOstar Omega spectrophotometer (BMG LabTech).

2.5.5 Data Analysis

β-galactosidase specific activity (U/min/mg) was calculated as the amount of chlorophenol red formed per minute of incubation time, and corrected to the total protein in each individual cell lysate sample. Repeated experimental data was expressed as mean ± standard error of the mean (SEM) and relative to the background β-galactosidase activity of the un-transfected control cells. Statistical differences between experimental groups were calculated using the commercially available software package GraphPad (InStat, USA) and outlined in the corresponding data figure legends.

2.6   Electrophoretic gel mobility shift assays (EMSA)

To investigate specific TF interaction at putative regulatory motifs within the GlyT-1a proximal promoter identified by *in silico* analysis (Section 2.2), gel shift assays of crude nuclear extract and representative synthetic labelled GlyT-1a DNA probes were used as described below.

2.6.1 Preparation of crude nuclear extracts

Following stress treatment of routinely cultured Caco-2 cells on 6 well plates (as described in Section 2.1.3 and 2.1.4), adherent cells were briefly washed with 1ml of ice cold PBS, scraped and transferred to a micro-centrifuge tube. Stress treated, as well as untreated control cells were then pelleted by centrifuging at 4°C at 1500 G for 5 minutes. The cell pellet was re-suspended in ice cold cell lysis buffer (10 mM HEPES at pH 7.9, 1.5mM MgCl$_2$, 10 mM KCl, 0.5 mM DTT, 25% *v/v* glycerol, 0.1% *v/v* NP-40) for 15 minutes at 4°C and centrifuged for a further 5 minutes at 4°C and 1500 G. The supernatant containing the cytosolic components was discarded. The remaining pellet was re-suspended in nuclei lysis buffer (20mM HEPES at pH 7.9, 1.5 mM MgCl$_2$, 400mM KCl, 0.5 mM DTT, 25% *v/v* glycerol, 1 complete mini EDTA-free protease inhibitor cocktail table [Roche]) and incubated for 30 minutes on ice. The crude nuclear extract was centrifuged at 4°C and 1500 G for 15 minutes and the supernatant collected. The protein concentration from each sample was determined using the Thermo Scientific Pierce 660 nm Protein assay in micro plate format. Protein samples were stored in aliquots at -80°C until required for binding assays. Buffers and protein extracts were stored in ice at all times and sample aliquots subject to more than one freeze thaw cycle were discarded.

2.6.2 Preparation of Infrared labelled DNA Probes and competitors

For investigations into Atf-4 or Nrf2 and GlyT-1a interactions two Infrared dye (IRD) labelled DNA probes were designed. /5IRD700/G1Exon1, an 88bp DNA probe spanning +97 to +185 downstream of the GlyT-1a TSS of the GlyT-1a gene transcript was designed to investigate Atf-4/GlyT-1a interactions at predicted regulatory elements. For investigations into Nrf-2/GlyT-1a interactions, /5IRD800/G1Prom, a 50bp probe spanning between 559 and 509 bases upstream of the GlyT1a TSS was synthetically generated (IdtDNA, Belgium). Previous attempts to generate these probes by endpoint PCR using IRD-700 5'end-labelled PCR primers proved time consuming and impractical. As such, the full length single stranded oligonucleotide was eventually synthetically

generated by IdtDNA to include an IRD-700 or IRD-800 dye at the 5' end. Synthetically generated oligonucleotides were HPLC purified to separate truncated or incomplete sequences from the full length labelled probes. Double-stranded DNA (dsDNA) probes were generated by incubating equimolar amounts of the single stranded probes at 95°C and allowed to cool to room temperature. A third probe – /IRD700/G1Exon1/Mut, was designed and synthesised by IdtDNA as above, to include a double mutation in the identified AARE and was used to investigate the specificity of Atf-4 interactions at the putative AARE.

To demonstrate affinity of the Atf-4/GlyT-1a at the putative AARE, competition assays using an unlabelled competitor probe were performed. A 23bp oligonucleotide probe containing the specific amino acid response element (AARE) of the CHOP (also known as C/EBP-ζ) promoter was used as binding competitor to the /5IRD700/G1Exon1 and /5IRD700/G1Exon1/Mut probes. Excess amounts of the unlabelled representative GlyT-1a probe was used to 'out-compete' positively identified complexes with either of these probes. All single stranded competitor oligos were synthetically generated by IdtDNA, HPLC purified and annealed as described above. All labelled oligonucleotide probes were kept under minimal light and stored at -20°C until required for binding reactions as described below.

2.6.3 Binding reactions and gel shift assay

Binding reactions were performed using the Odyssey EMSA buffer kit (Licor, Cambridge UK) according to the guidelines. For each binding reaction, 1 ρmol of the labelled probe was incubated with either 5 μg of crude nuclear protein extract, competitor sequence or both, in a 20 μL binding reaction. Final binding reaction solution in addition to the probes and DNA, consisted of 200mM EDTA, 1M KCl, 1% NP-40, 1 μg/μL DNA polymer dI.dC, 2.5% Tween-20, 1x Binding buffer. To ascertain observed complexes resulted from specific protein-DNA interactions, incubations with BSA as opposed to crude nuclear extracts or no protein were used in binding reactions. The reaction was incubated at 4°C for at least 60 minutes and at most 16 hours in the dark. Following the addition of 2 μL of 10x orange loading dye (Licor, Cambridge UK) binding reaction samples were electrophoresed on a mini-PROTEAN TGX precast 7.5% poly-acrylamide gel (PAGE) (Bio-Rad Laboratories, Hemel Hempstead UK) in 0.5% TBE at 5 mV per cm of gel for 60 minutes in the dark.

Gels were imaged using a Licor® Odyssey™ Infra-red Scanner. Scanning resolution was set at 169 µM with a focus offset of 1.5 mm (i.e. the thickness of the plate plus half the thickness of the PAGE gel). Scanning intensities were tweaked per gel to obtain the best possible complex image for each channel (700nm or 800nm) scanned.

2.6.4 Gel super-shift assays

For super-shift assays, an Atf-4 specific antibody (#sc200x, Santa Cruz), was added to the binding reactions described above, prior to the addition of the IRD probe. Reaction conditions for super-shift assays were identical to those described above. A smaller binding reaction volume of 10 µL was used. Super-shift samples were electrophoresed on a 26-well 7.5% Criterion™ TGX Precast PAGE gel (Bio-Rad Laboratories, UK) and imaged on a Licor® Odyssey™ Infra-red Scanner.

2.6.5 Data Analysis

The resulting intensities (I) of the complex bands and their relative migration (M) down the PAGE provide two exact variables for characterising the protein-DNA interaction forming the said complex. With the molecular weight of the IRD probe used being constant, the mobility of the observed complex is proportional to the molecular weight of the interacting protein, whilst the intensity of the complex band quantifies the interaction to provide an estimate of protein affinity to the DNA probe. Comparative lane by lane analysis was performed between lanes using the Licor® Odyssey™ Software suite (version 3.0) to determine the relative mobility factor (RF) – defined as a percentage migration of the complex band from the top of the lane (0%) to the bottom of the lane (100%); Lane profiles from which an automatically defined background intensity threshold had been deducted were used to identify fainter band complex non-evident by visual inspection of the gel image.

2.7 Chromatin Immunoprecipitation (ChIP) Assays

ChIP assays were performed using a modified protocol originally described by (Carey *et al.*, 2009) and adapted in (Johnson *et al.*, 2007). The ChIP work-flow described here is outlined in Figure 2.3. Assays were performed on a total of 100 million Caco-2 cells per stress treated sample cultured as described above in Section 2.1.3.

2.7.1 Cross-linking, nuclei isolation and Sonication

Routinely cultured Caco-2 cells in a T-175 flask were stressed as described in section 2.1.4. Cells were transferred to 10 mL of serum free medium in a universal macro-centrifuge tube. Total protein bound DNA was cross-linked by adding formaldehyde to a final concentration of 1% (v/v) then stopped 10 minutes later by the addition of glycine to a final concentration of 0.125M. The cells were pelleted at 15000g for 5 minutes at 4°C and washed in ice cold PBS and incubated in RIPA buffer (1% NP-40, 0.1% SDS, 50 mM Tris-HCl [pH 7.4], 150 mM NaCl, 0.5% Sodium Deoxycholate, 1 mM EDTA) for 20 minutes at 4°C. Pelleting at 15000g for 10 minutes retrieved the intact nuclei.

The intact Caco-2 nuclei containing cross-linked chromatin were incubated in 5 mL of freshly prepared nuclei lysis buffer (50 mM NaCl, 10 mM Tris [pH 7], 20% Sucrose, 10 mM EDTA, 1 complete Roche protease inhibitor table) and sonicated using a U300H ultrasonic bench top cell sonicator (Ultrawave) for 15s at full power with 1-min cooling intervals on ice between each burst. Sonicated DNA was electrophoresed on a 1% agarose/TBE horizontal gel, to confirm the obtained DNA fragments were of length less than 500bp.

2.7.2 Immunoprecipitation (IP) and DNA Extraction

Complexed protein-DNA aliquots, representing nuclei extracts from about 1×10^6 cells were incubated overnight at 4°C with 0.5µg/mL of Atf-4 antibody (SC-200x, Santa Cruz). The antibody-bound complex was then precipitated using protein-A magnetic Dynabeads (Life Technologies) according to the manufacturers' specification. Dynabeads bound protein-DNA complexes were washed five times with ice cold IP wash buffer (100 mM Tris-HCl [pH 7.5], 500 mM LiCl, 1% NP-40, 1% sodium deoxycholate, 4°C) followed by a single wash in 1 x Tris/EDTA (TE) buffer (10 mM Tris-HCl pH 7.5 + 0.1 mM EDTA). The immune-precipitate was then eluted in 200 µL of IP elution buffer (1% SDS, 0.1 M NaHCO$_3$). The DNA fragments bound to the antibody complexed protein in the

immuno-precipitate was released by reversing the cross-linking overnight at 65° C and purified using a phenol/chloroform/isoamyl extraction. To determine total DNA input, control total DNA was purified from sonicated chromatin. Total DNA concentration of each sample was measured on a Biomate3 Nano cell (Thermofisher scientific) prior quantification by real-time QPCR.

2.7.3 Quantification of ChIP signal enrichment by quantitative QPCR

The relative abundance of DNA sequence containing the putative regulatory motif was measured by real-time QPCR. Products from preliminary PCR amplifications for the corresponding sequences were characterised by cloning and sequencing as described elsewhere. A sense primer 5'-CTGTGGAAGAGCTGCGAGC-3' and reverse primer 5'-CCAGGTGGCTTTCTTAAAGGG-3' were used to quantify the enrichment of a 100 bp sequence spanning the putative GlyT-1a AARE in input and no-input DNA samples. All reactions were optimised to generate a single PCR product from the region of interest and to avoid primer-dimer formation. Real-time quantitative PCR reaction mixes consisted of 0.5 μL of 10mM each ChIP primer set, 2 μL purified ChIP DNA fragments, 5.0μL of 2x Light Cycler 480 SYBR Green 1 Master Mix in a total reaction volume of 10 μL. PCR was performed over 45 cycles and included an initial hot start activation at 95°C for 5 minutes, denaturation at 95°C for 5 seconds, annealing at 55°C for 15 seconds and extension at 72°C for 20 seconds. Fluorescence was measured at the end of the extension step in each cycle. After PCR, melting curves were acquired by stepwise increases in the temperature from 65 to 90°C to ensure that a single product was amplified in the reaction. PCR analyses were performed in duplicate for each sample. The immunoprecipitated material was quantified relative to a standard curve of genomic DNA.

2.7.4 Data Analysis

The results are expressed as the relative amounts of quantified AARE in the "input" (precipitated chromatin) of stress treated Caco-2 cells versus the control un-stressed group.

1 Caco-2 cell culture

- 100 million Caco-2 cells cultured under unstressed conditions, 16hr of Tunicamycin 8 hours of AA starvation, or 4 hours DEM

2 Crosslinking with Formaldehyde

- Following stress treatment cells are treated with 1% Formaldeyhide for 10 minutes
- Cross-linking is stopped with 0.125M Glycine

3 Chromatin Sonication

- Wash and resuspend nuclei in Nuclear lysis buffer
- Sonicate using U300H Ultrawave

4 Isolation of Intact Caco-2 nuclei

- Cells are incubated in RIPA buffer for 20 minutes on Ice
- Wash and pellet to collect nuclei

5 Immunoprecipitation Sample Prep

- Check Sonicated Chromatin is less than 300 bp long by running on agarose gel
- Store an aliquote of Sonicated sample for "No-Input" control

6 Immunoprecipitation

- Immunoprecipitation using Dynabeads complexed to Atf-4 antibody (SC200x)
-

7 Quantifiy sequence enrichment by PCR

- Using primers to known region of the GlyT-1a Sequence

8 Reverse Crosslinking

- Reverse crosslink by incubating at 65 deg.
- Reverse crosslink both IP "input" sample and "No-Input" control
- DNA extraction by phenol/chloroform

Figure 2.3: ChIP experimental workflow

RESULTS & DISCUSSIONS

Outline: In this chapter I discuss the project results. This chapter begins by providing RT-QPCR evidence that the expression of GlyT-1a is increased in response to stress, and that similar changes in expression are observed with other potentially co-regulated SLC transporters. Outcomes of a detailed bioinformatics analysis of the GlyT-1a flanking sequence are reviewed. In this chapter I review the interactions of two potential regulatory transcription factors (Nrf-2 and Atf-4) at their respective identified putative regulatory elements as experimentally ascertained by electrophoretic mobility shift (EMSA), chromatin immune-precipitation (ChIP-PCR), and reporter assays. Where indicated, n represents sample replicates while N represents experimental replicates, such that the total number of samples included in each data graph is N * n.

Contents:

3.1 GlyT-1a mRNA abundance is increased by stress

It had previously been established in this lab that the intracellular accumulation of glycine resulting from the activity of GlyT-1, confers protection from various insults to the cell. To determine if Caco-2 cells increased expression of the GlyT-1a transporter to mitigate an added requirement for glycine during periods of stress, mRNA level of the specific

GlyT-1a variant was quantified in total RNA extracted from stress treated confluent Caco-2 monolayers (as described in Chapter 2). As shown in Figure 3.1, GlyT-1a mRNA levels were significantly increased by approximately 6-fold following treatment with 1 μM tunicamycin for 16 hours. Consistent with these findings, work by Lecca and colleagues show that primary fibroblast cells from patients with type-1 congenital disorders of glycosylation (CDG-1) treated with tunicamycin or grown in glucose-deprived medium resulted in an 8.7 and 2.6 fold increase in GlyT-1a expression respectively (See Table 2 in Lecca *et al.*, 2005).

Likewise a significant 3-fold increase in GlyT-1a mRNA was observed with 16 hour treatment of 50 nM thapsigargin. Increases in GlyT-1a mRNA were observed following 4 hours of 0.2 mM DEM and also following 8 hours of amino acid starvation (Figure 3.1). To establish whether these changes in expression of the high affinity GlyT-1 transporter were indeed a preferred specific response to the added requirement for the intracellular accumulation of glycine during periods of stress, the mRNA level of the peptide transporter PepT-1 was quantified. In intestinal epithelia, PepT-1, expressed solely on the apical membrane mediates the absorption of short di and tri- dietary peptides (See Section 1.5.3). As shown in Figure 3.2, no statistically significant change in PepT-1 expression between stress treated and unstressed Caco-2 cells was observed following amino acid starvation, tunicamycin, thapsigargin or DEM treatment.

Our initial hypothesis predicted that increased glycine transport to intestinal epithelium during states of physiological stress contributed to glycine supply for the synthesis of the intracellular antioxidant glutathione (GSH), hence cyto-protection from antioxidants. Another required component for the two step synthesis of GSH is the amino acid cysteine. Cysteine is unstable in the extracellular environment and is rapidly oxidised to form cystine. Cellular cystine fluxes are generally considered rate-limiting to the intracellular provision of cystine required for the production of glutamyl-cystine, which is subsequently conjugated to glycine in the final step of the reaction to produce GSH (Section 1.3.3). In epithelial cells of the peripherally drained viscera (PDV), cystine is transported across cellular plasma membranes by the dimeric cystine-glutamate anti-porter (System X_c^-). In cells of the central nervous system cysteine transport is by the excitatory amino acid transporter 3 (EAAT-3/SLC1A3). There is extensive work in the literature detailing the importance of system X_c^- or EAAT-3 in oxidative protection, as well as chemo sensitivity and chemo resistance. The expression of xCT/SLC7A11, the

light-chain of System X_c^-, has been shown to be increased downstream of amino acid starvation in mouse fibroblast cells (Sato *et al.*, 2004) and downstream of DEM treatment (Hosoya *et al.*, 2002) or oxidative stress. Indeed, littermates of C57BL/6 mice deficient in xCT had decreased plasma concentrations of GSH and a characteristic redox imbalance, whilst fibroblast cells derived from these mice required addition of the antioxidant 2-mercaptoethanol to the culture medium for survival in routine culture (Sato *et al.*, 2005).

To determine if the expression of GlyT-1 was regulated by similar regulatory modules to that of xCT (discussed here and also in Section 1.7 above), xCT mRNA abundance was quantified in total RNA extracted from stress treated Caco-2. As shown in Figure 3.3, like GlyT-1a, xCT mRNA expression was significantly increased by approximately 4-fold following treatment with 1 µM tunicamycin for 16 hours. Also a significant 3-fold increase in xCT mRNA was observed with 8 hours of amino acid starvation whilst a doubling was observed following either a 16 hour treatment with 50nM thapsigargin or 4 hours treatment with 0.2 mM DEM. This data is consistent with observations made in this lab using the same stress conditions on HCT-8 cells (data not included)
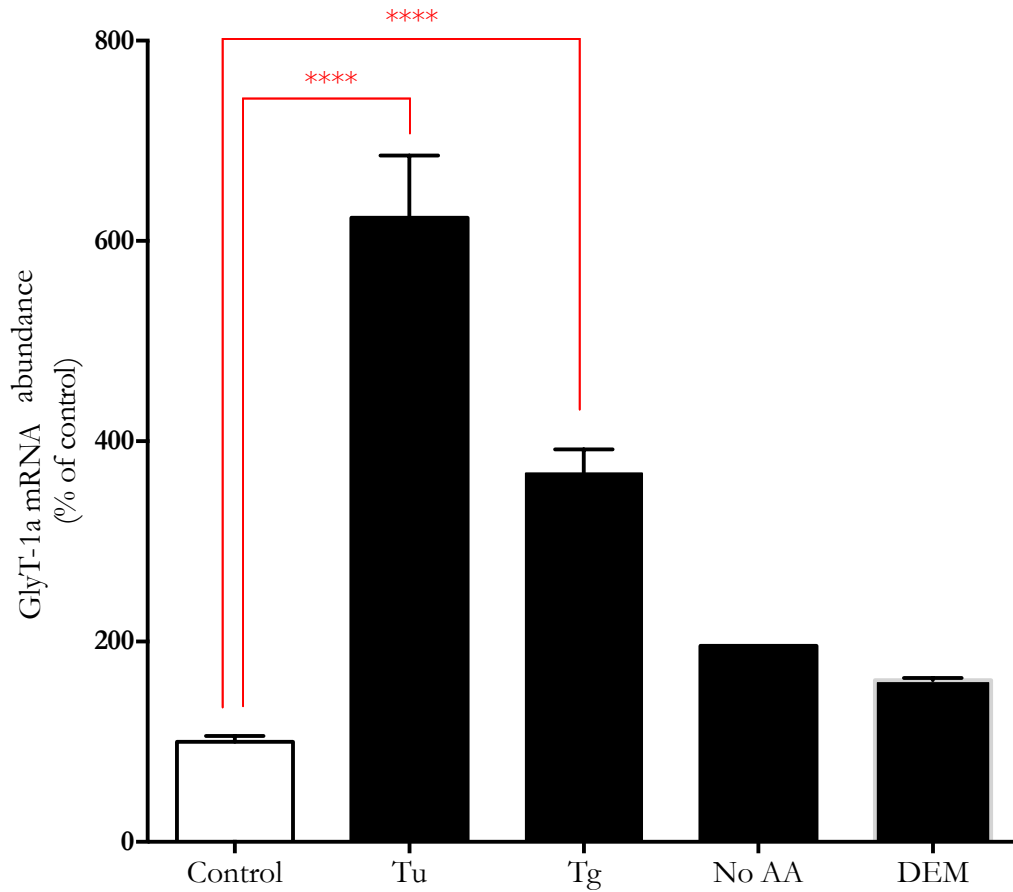
Figure 3.1: GlyT1a mRNA abundance is increased following stress in Caco-2 cells with either tunicamycin (Tu) or thapsigargin (Tg) for 16 hours, 8 hours of amino acid starvation (No AA) or 4 hours of diethyl maleate (DEM). Values are pooled from multiple experiments and represent a percentage mean of the untreated control sample displayed as mean ± SEM, n=2-3 (N = 2 for each treatment). Sample data is consistent with further experiments performed in this laboratory. One-way (non-parametric) analysis of variance (ANOVA) was performed. For comparison of each treatment to the control Dunnett's multiple comparison tests with pooled variance for each treatment was used. **** represents a P-value of less than 0.001.
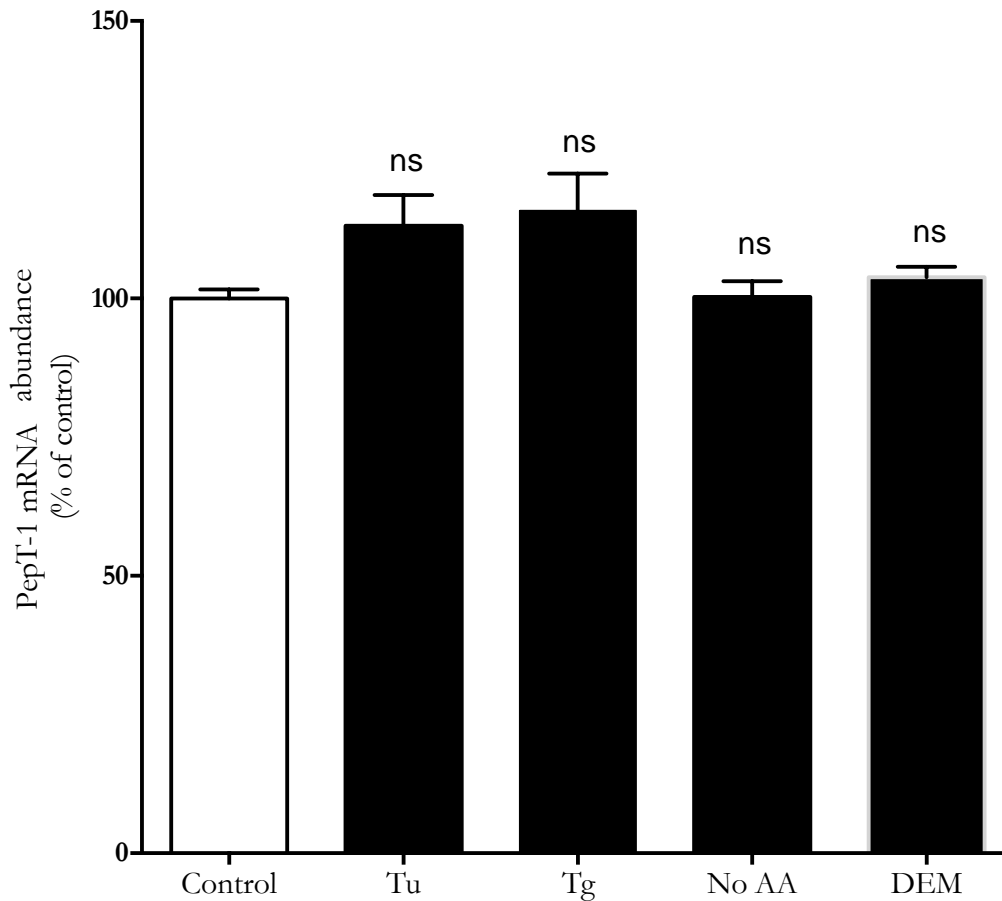
Figure 3.2: PepT-1 mRNA abundance is unchanged following stress in Caco-2 cells. Cells were treated with either tunicamycin (Tu) or thapsigargin (Tg) for 16 hours, 8 hours of amino acid starvation (No AA) or 4 hours of DEM. Values are pooled from multiple stress treatment experiments and represent a percentage mean of the untreated control sample displayed as mean ± SEM, n=2-3 (N = 3 for each treatment). One-way (non-parametric) analysis of variance (ANOVA) was performed. Dunnett's multiple comparison tests with pooled variance for each treatment was used for comparison of each treatment to the control. *ns*: not significant
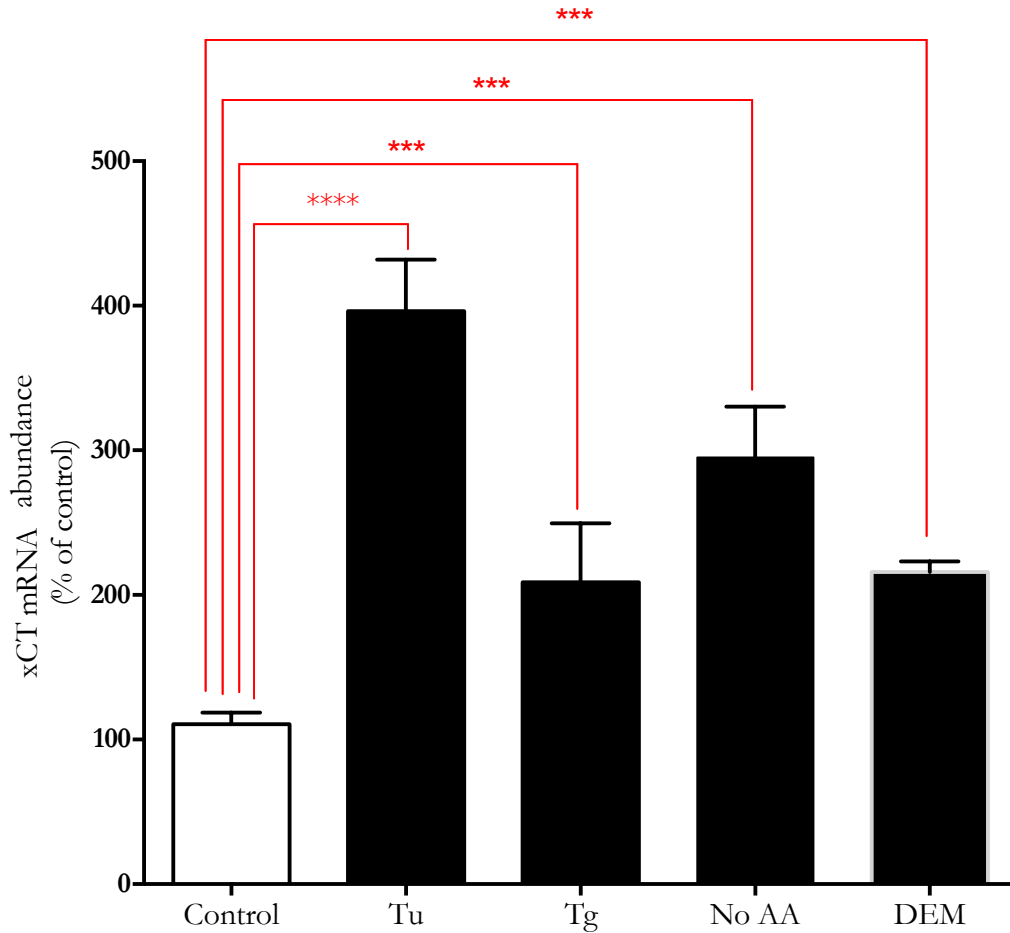
Figure 3.3: xCT mRNA abundance is increased following stress in Caco-2 cells with either tunicamycin (Tu) or thapsigargin (Tg) for 16 hours, 8 hours of amino acid starvation (No AA) or 4 hours of diethyl maleate (DEM). Values are pooled from multiple stress treatment experiments and represent a percentage mean of the untreated control sample displayed as mean ± SEM, n=2-3 (N = 2 for each treatment). One-way (non-parametric) analysis of variance (ANOVA) was performed. For comparison of each treatment to the control, Dunnett's multiple comparison tests with pooled variance for each treatment was used. *** and **** represents a *P*-value of less than 0.005 and 0.001 respectively

Taken together, the fact that mRNA levels of both xCT and GlyT-1 are increased in a similar pattern downstream of stress unlike PepT-1 suggest that these adaptive changes in expression are part of a specific response to stress and not a result of a global increase in requirements for amino acids for protein synthesis or energy production. Whilst our data shows no change in PepT-1 expression following stress, impairment in transport (measured as the uptake of [^{14}C]glycine-sarcosine) by the PepT-1 transporter in Caco-2 cells pre-treated with hydrogen peroxide for 24 hours has previously been reported and associated with decreased expression of the transporter; as well as reduced V$_{max}$ of the transporter (Alteheld *et al.*, 2005). Such a decrease in expression of PepT-1 has also recently been reported in mouse models of Crohn's disease-like ileitis or colitis as well as in tissue samples collected from IBD patients during endoscopy (Wuensch *et al.*, 2014). Alteheld *et al.* showed that uptake by PepT-1 in stressed Caco-2 cells, as well as intracellular GSH level and redox potential can be maintained by simultaneously treating cells with growth hormone (GH) or the dipeptide alanine-glutamine (Alteheld *et al.*, 2005). Recently the regulation of PepT-1 by the transcription factor Nrf-2 has been demonstrated following Sulforaphane treatment of Caco-2 (Geillinger *et al.*, 2014).

Most cellular functions and adaptive processes are characterised by modular, interactive networks of regulator proteins (including transcription factors, co-factors, modifying enzymes) and regulated genes. These modular components acting in concert produce a characteristic functional activating or repressing response by means of protein expression, silencing or modification. The work described here set out on the assumption that specific gene regulatory modules, largely comprised of key members of the integrated stress response (ISR), regulate the transcriptional expression of the GlyT-1a and xCT downstream of stress stimuli such as oxidative stress.

Canonical methods for the identification of probabilistic regulatory networks begin with large expression array datasets and rely on specific clustering algorithms to elucidate any functional meaning. From existing literature and published data, it is possible to catalogue potential regulator proteins active in states of oxidative stress, amino acid starvation, the unfolded protein response, or ER stress. Our approach, entailed detailed computational analysis of the GlyT-1 flanking sequence in search for candidate binding sites and motifs for the catalogued potential regulatory proteins, followed by experimental validation of their DNA/Protein interaction following stress treatment in Caco-2 cells.

3.2   Identification of regulatory motifs of GlyT-1a

Identification of the GlyT-1a Theoretical Promoter
The GlyT-1a 5'-flanking sequence was analysed to identify a CGI promoter and other conserved regulatory element (Section 3.2.1)

Identification of Promoters of GlyT-1a Co-regulated genes
The CORD database was searched to identify 185 genes with a factor dependent correlation in expression to GlyT-1a greater than 84%

Identification of TFBS in the GlyT-1a Promoter
The GlyT-1a CGI promoter was analysed using TRANSFAC MATCH. Over **28500** TFBS sites were identified (Section 3.2.2)

Identification of frequently occurring TFBS in co-regulated genes
From the list of 28500 TFBS a list of **131** TFBS enriched in the promoters of GlyT-1a and co-regulated genes was identified using F-MATCH

Analysis of CRM conservation
Phylogenetic analysis of the 10 postulated CRMs show extensive conservation across orthologous GlyT-1a sequences in 56 species for CRM1, CRM8 and CRM10. The binding of the Atf-4 to CRM 10 was subsequently studied by Gel Shift assays, reporter assays and ChIP assays.

Identification of composite regulatory modules
The arrangements of the 131 TFBS identified was studied to identify TFBS suitable for the binding of Atf-4, Nrf-2 and other TF known to mediate stress response. 10 CRMs were proposed (Section 3.2.3)

Figure 3.4: Flow chart showing the workflow and outcomes of bioinformatic analysis of the GlyT-1a Promoter as discussed in Section 3.2

Be they transcription factors (TFs) responsible for driving gene expression or cofactors required for epigenetic modifications of the chromatin at specific sites, the binding of proteins to nucleic materials provides an interface for the regulation of several cellular processes. Understanding the site specificity of protein-DNA interactions is an important challenge to deciphering gene regulatory modules and networks. The short sequence-specific transcription factor binding sites (TFBS) within flanking regions of stress responding genes confers varying degrees of trans-repressive or trans-activating activity upon binding of a TF. Despite the absence of ambiguity in a well-prescribed stress response, the degeneracy of TFBS poses a significant problem to the accurate *in silico* identification of factors involved in the regulation of a gene.

In-lieu of time-consuming and often expensive wet-lab methods, combinatorial and enumerative methods are used on previously identified sequences of probabilistic protein occupancy to describe TFBS consensus for applicability to the remainder of the genome. Most of what is known of TFBS is based on a limited amount of experimental evidence from assays such as ChIP on 'chip' followed by sequencing (ChIP-seq), DNase I foot-printing, electrophoretic mobility gel-shift assays (EMSA), south-western blots and promoter reporter assays (Neph *et al.*, 2012). For determining the function of specific TF at a gene promoter, newer methodologies such as protein-binding microarrays (PBM) provide comprehensive and high throughput datasets for the characterisation of binding sites (Bulyk, 2007). Databases such as UniPROBE catalogues PBM array data for a limited number of transcription factors (Newburger and Bulyk, 2009). Algorithms combining protein-specific binding data to functional expression data are emerging as a suitable computational strategy for identifying gene modules within regulatory networks (Bar-Joseph *et al.*, 2003).

Several of such consensus TFBS for eukaryotic TFs are catalogued as positional weight matrices (PWMs) or position specific score matrices (PSSM) in databases such as TRANSFAC® (Matys *et al.*, 2003) and JASPAR (Sandelin *et al.*, 2004). PWMs represent a mathematical representation of multiple sequences identified from the wet-lab methods described above and to which a given transcription factor is believed to bind. In effect, they are an L by four matrix (where L is the sequence length of the postulated TFBS, usually between 5 and 20) which describe the frequency of occurrence any of the four nucleotide bases (A, C, G, T) at any one point along L. PWMs, together with their

graphical representations known as Tom Schneider sequence logos depict the DNA binding preferences of a TF at a TFBS.

Importantly, does the occurrence of a probabilistic TFBS within the promoter region of a gene imply its specific activity at that site? Despite the many approaches and algorithms used to identify TFBS *in silico*, minimising false positives and negatives from computational prediction remains an important challenge. The use of TFBS consensus matrices in identifying putative regulatory motifs at other genomic sites ignores the issues of TF redundancy, co-binding (transcription factor tethering), tissue specificity of TF/DNA interactions as well as the regulatory context (Kuntz *et al.*, 2012). The understanding that several proteins or heterodimers of transcription factors may interact at a particular TFBS further complicates the notion of one TF to one TFBS.

For the identification of the transcriptional modules involved in the regulation of human GlyT-1a a combination of several methods and workflows were used as described in (Wasserman and Fahl, 1997, Wasserman and Sandelin, 2004). To begin, the GlyT-1a promoter sequence was characterised to identify regions of important regulatory significance. Next, using the database of co-ordinately regulated genes (CORD), several promoters from genes with significant regulatory correlation to GlyT-1a were catalogued. This was followed by detailed analysis of the promoter collection to reveal significantly enriched (frequently occurring) TFBS using the TRANSFAC Explain tool. To minimize false positives from TFBS identification, the TRANSFAC Composite Model Analyst tool was used to identify composite sites. In the following section, the results from this bioinformatic analysis of Glyt-1a are discussed.

3.2.1 Characterisation of the GlyT-1a promoter

The human GlyT-1a gene transcript (Accession numbers RefSeq: NM_001024845, Ensembl: ENST00000372310) is encoded from 16 out of a possible 17 exons (Figure 3.5), which span a 44.1 kb sequence on the reverse strand of chromosome 1 between nucleotides 44462155 and 44497164 (unless otherwise stated, all genomic coordinates noted in this section are with respect to the GlyT-1a NM_001024845 transcription of GRCh37/hg19 assembly obtained from NCBI). Exon 1 and the first 86 bases of exon 2 are completely untranslated and form the 5'UTR of the mRNA. On the genome the first and second exons are separated by a 6943bp genomic intron. The last 1062 bases of exon 16 are untranslated and form the 3'UTR.

The human GlyT-1a gene transcript sequence was studied to determine a characteristic 'core' promoter and basal regulatory elements (BRE). The core promoter is defined as the minimum essential neighbouring DNA sequence required for accurate binding and initiation of gene transcription by RNA polymerase II (RNAPII) and other co-factors (Butler and Kadonaga, 2002). Typically the core promoter extends ~50 bases either upstream or downstream of the transcription start site (TSS or A_{+1}) and may contain specific sequence motifs such as a TATA-box required for anchoring RNAPII (as illustrated in Figure 3.6), an initiator motif (INR) functionally similar to TATA-box in facilitating binding of the TATA-box binding protein (TBP, TFIID) complex, CpG islands (CGI), and sometimes downstream core promoter elements (DCE or DPE) (Kadonaga, 2002).

Analysis of the GC content of the GlyT-1a flanking sequence (chr1:44499164-44489948) suggest that its core lies within an extensive CGI as indicated in (Figure 3.7; page84). CGIs were calculated as regions where over an average of 10 sliding windows of size at least 100 nucleotides (or a region spanning no less than 1000 nucleotides, whichever is greater) the percentage G+C content is over 50 and the calculated observed/expected ratio for the occurrence of a CpG (i.e. a cytosine base immediately linked by a phosphate backbone to a guanine base) is over a threshold of 60%.

In mammalian genomes, about 70% of gene core promoters reside within CGIs. CGIs have a strong correlation with gene regulatory sites and are pronouncedly GC-rich, CpG rich and non-methylated when they occur in active promoters (Deaton and Bird, 2011). Whilst it is impossible to computationally determine the methylated status of CGIs, biochemical analysis reveals an association between methylation of the cytosine of the CpGs and gene silencing. The methylated cytosine can spontaneously de-aminate to a thymine to form TpG sites. TpG are up to 50 times more frequent than other types of transitional mutations in cancer cells. In humans, the enzyme thymine-DNA glycosylase (TDG, EC 3.2.2.29) upon interaction with several TFs and co-factors such as the pCAF/p300, is tethered to the DNA and catalyses the removal of the C/T mismatch.
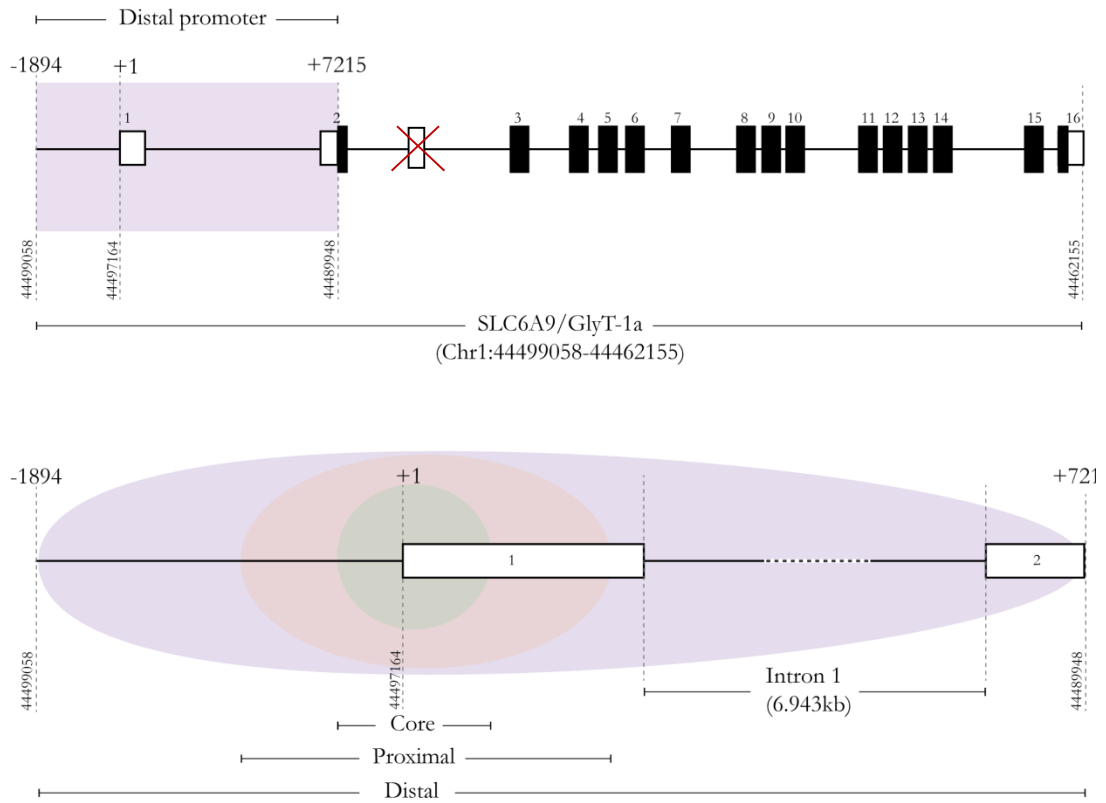
Figure 3.5: Arrangement of the human GlyT-1a/SLC6A9 gene. The human GlyT-1a transcript (NM_001024845 or ENST00000372310) is encoded from 16 out of a possible 17 exons (numbered rectangles) spanning a 44kb region of chromosome 1 between nucleotides 44497164 and 4446215 – as numbered in the latest release of the human genome assembly (GRCh37/hg19, Released 1st Feb 2009). The exon excluded from GlyT-1a lies between exons 2 and 3 (all translated exons are indicated by dark boxes). Exon 1 and the first 86 bases of exon 2 are completely un-translated (forming the 5'UTR) and on the genome are separated by a 6943 base intron (Intron 1). Likewise, the last 1062 bases of exon 16 are un-translated and define the 3'UTR. In work reported here, bioinformatic analysis was performed on key promoter regions (core, proximal and distal promoters) to identify regulatory motifs important to the transcriptional regulation of the gene. All identified motifs within these regions are reported and discussed in the text.

In addition to the methylated status of the cytosine in CGIs histone modifications are believed to be involved in controlling gene expression or repression. Whilst direct methylation of the cytosine may lead to permanent or long term gene silencing, reversible methylation of some histones such as di-methylation of histone 3 at lysine 9 (H3k9me2) may provide short term gene inactivation. The specific methylation status of histone 3 at lysine 4 is generally considered a characteristic signature of CGIs, and has previously been used to experimentally determine cell type-specific gene regions rich in regulatory enhancers. Work by Creyghton and colleagues on H3k4me1 histone marks identified by ChIP-Seq using murine ES cell revealed that, when co-associated with H3k27ac histone acetylation marks, they highlight important gene regulatory enhancer sites (Creyghton *et al.*, 2010). These sites actively promote gene expression by negating the silencing effect of methylation. Another conclusion from the later study is that enhancers associated with only the H3k4me1 mark often require additional external stimulus for gene activation. Analysis of the GlyT-1a sequence in h1-hESC, HUVEC and GM128787 cells from data outputs of the ENCODE project show that it correlates to significant H3k27ac marks (Figure 3.7), positing that it lies in a transcriptionally permissive region. Results of this analysis, depicted in Figure 3.7 demonstrate an important overlap of the Glyt-1a core promoter with the predicted CGI, high PhastCons[††] conservation scores, high GC content and high H3k27ac marks. Based on these observations, this sequence region was searched further for the occurrence of specific regulatory sites.

In a study by Suzuki *et al.,* analysis of core promoters of 1031 human genes revealed only 32% contained a TATA box, whilst 80% contained an INR motif (Suzuki *et al.*, 2001). A detailed understanding of the occurrence of general transcriptional regulatory elements around and afar from a given TSS (Figure 3.6) provides insights to the combinatorial relationship between these elements in regulating gene transcription (Smale, 2001). For instance it has been suggested that some TFs such as SP-1 exhibit differences in activity at TATA-containing versus TATA-less promoters, with TATA-less promoters requiring an additional INR motif for activity of upstream enhancers such as SP-1 (Lo and Smale, 1996).

---

†† Scores representing the predicted base-by-base conservation in multi-aligned sequences determined using the PhastCons algorithm developed by Siepel, A., Bejerano, G., Pedersen, J. S., Hinrichs, A. S., Hou, M., Rosenbloom, K., Clawson, H., Spieth, J., Hillier, L. W., Richards, S., Weinstock, G. M., Wilson, R. K., Gibbs, R. A., Kent, W. J., Miller, W. & Haussler, D. 2005. Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. Genome Res, 15, 1034-50.
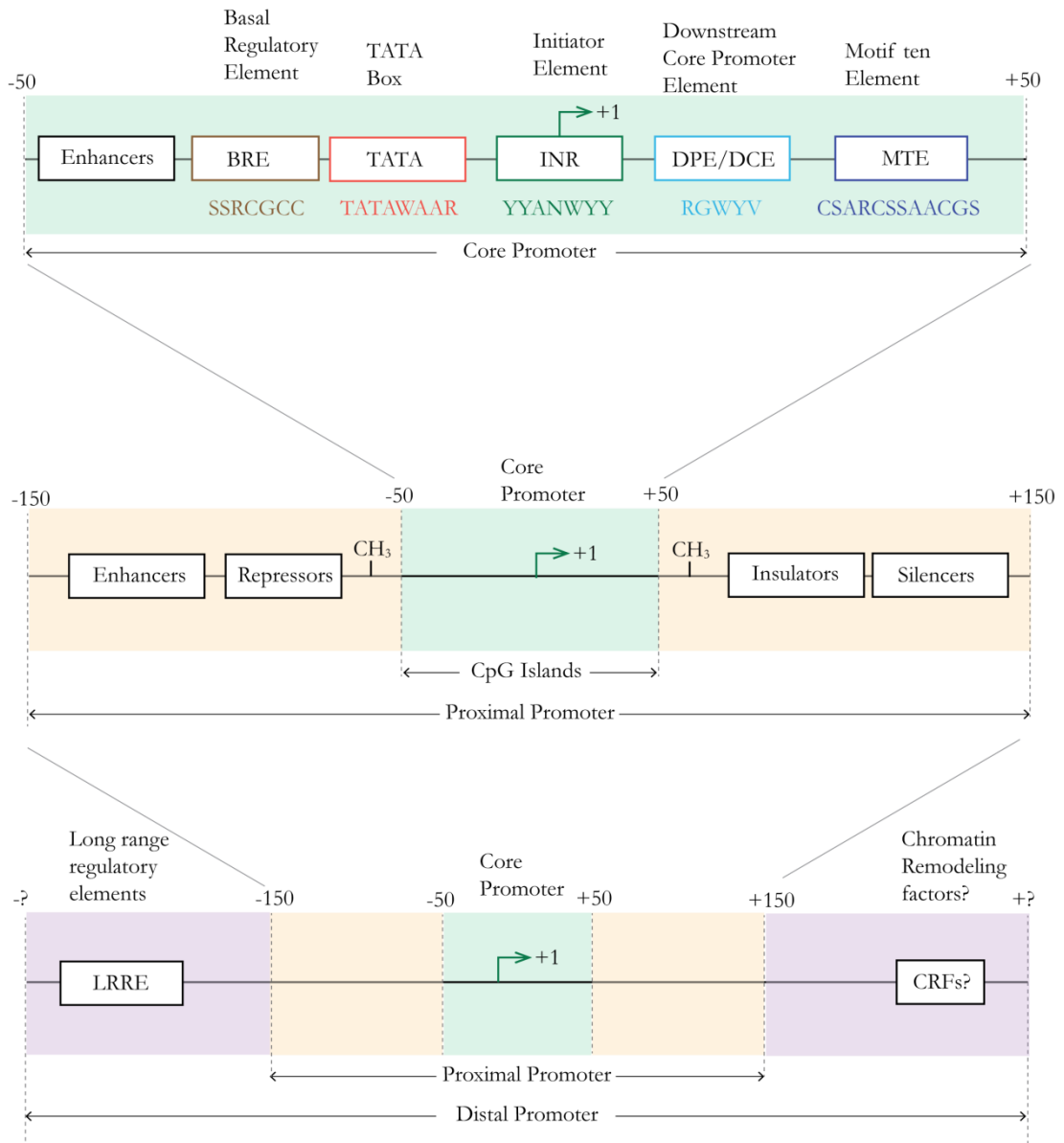
Figure 3.6: Hypothetical TFBS architecture at gene promoters The core promoter (shaded green), defines the minimum essential sequence required for assembly of the pre-transcription complex. The occurrence and arrangements of BRE, TATA boxes with respect to the INR, as well as downstream DPE and MTEs modulate RNA Polymerase II and cofactor assembly. Core promoters are characterised by CpG islands and flanked by extensive methylation sites in the proximal promoter. The proximal promoter (shaded orange) contains additional enhancer or repressor motifs capable of modulating event specific transcription. Additional insulators or silencers modulate access to the core promoter by elements at long range regulatory sites (LRREs).

In humans the occurrence of SP-1 and CCAAT-enhancer boxes (C/EB) within TATA-less promoters has been postulated to replace the requirement for TATA-boxes in recruitment of transcriptional initiator proteins and interaction with conventional TBP-associated factors such as transcription factor IID (Emami *et al.*, 1998, Butler and Kadonaga, 2002). CGI promoters often lack characteristic TATA boxes, BRE, DPE, or DCE elements, show diverse initiation patterns and can have multiple TSS (Juven-Gershon and Kadonaga, 2010). It may be for this reason that genome build algorithms (such as Sanger's Eponine algorithm) used by various online assembly databases have resulted in differing TSS annotated for the same GlyT-1a transcript (the GlyT1a TSS on Ensembl (ENST00000372310) is 100 bases upstream of that for the same transcript on NCBI (NM_001024845)). In the post genomics era, the identification of gene TSSs or INRs is generally achieved by mapping sequenced RNA or cDNA ends to genomic DNA and noted on several genome assemblies (Cullen and Barton, 2007, Pelechano *et al.*, 2014).

The GlyT-1a core promoter which resides in an extensive CGI, is TATA-less, but contains a characteristic INR motif (5'- $TGA_{+1}GAATC$-3') at its TSS (in this case that of the NCBI entry). The GlyT-1a INR motif is similar to the originally defined INR consensus sequence for vertebrates $(C/T)(C/T)A_{+1}N(A/T)(C/T)(C/T)$ (Smale and Baltimore, 1989), but differs in that the bases preceding the A_{+1} and at +3 are non-pyrimidine (Javahery *et al.*, 1994).

It has been suggested that TATA-less core promoters which have an INR domain usually contain an additional DPE element with the suggested consensus $(A/G)G(A/T)(C/T)(G/A/C)$ (Burke and Kadonaga, 1996) within the first 30 bases downstream of A_{+1} (Yang *et al.*, 2007); however, no such DPE was identified within +50 bases from A_{+1} of GlyT-1a. The absence of a characteristic TBP motif and other BRE elements within the GlyT-1a core promoter heightens the significance of additional enhancers and other regulatory elements for its transcription.
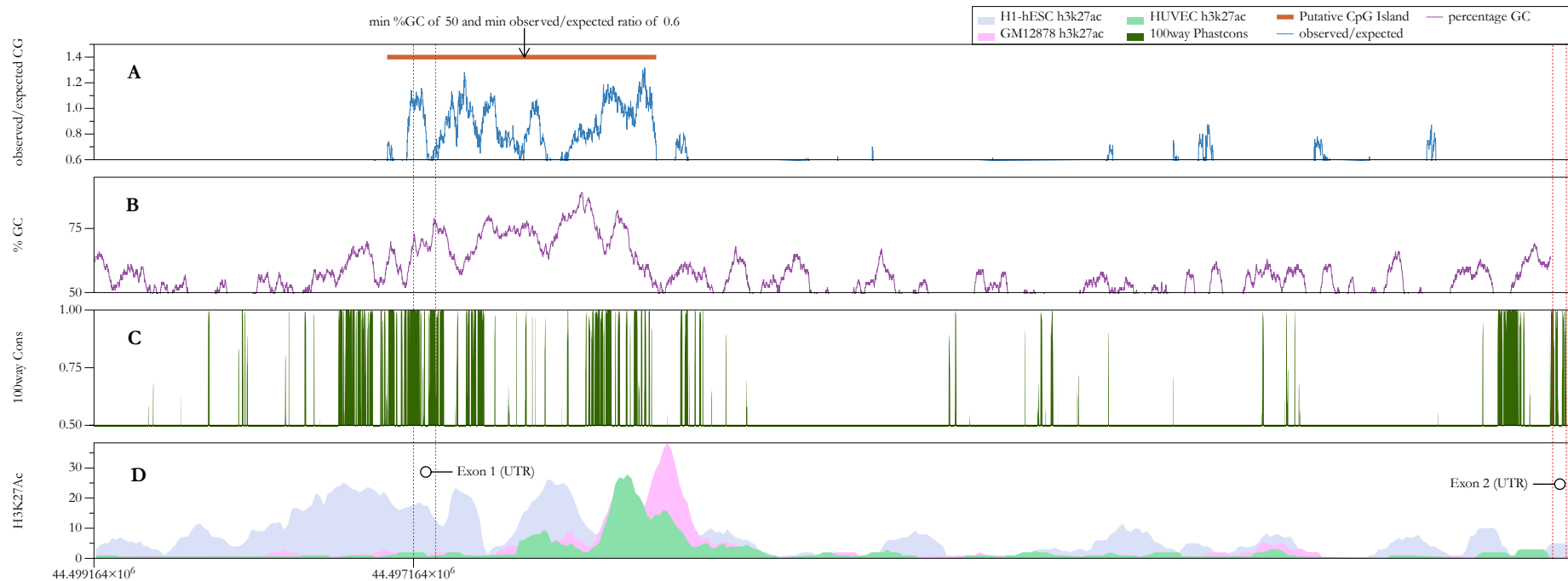
Figure 3.7: Co-occurrence in the human GlyT-1a promoter of CpG islands (CGIs), and correlation to a highly conserved region and one of high levels of H3k27ac. CGIs were predicted in a region spanning -2000 bases upstream the TSS of GlyT-1a to include the full 6.9kb of intron 1 and the 85 un-translated bases of exon 2 (Panels A and B). For comparative analysis of conservation, PhastCons scores from a Multi-z alignment of the genome regions of 100 vertebrate species (multi-100-way, Panel C) was downloaded from UCSC browser table and charted. Based on the assumption that non-coding sequences of functional significance are under heightened selective pressure, regions of high conservation represent potential sites for gene regulatory elements. Whilst CGIs are characteristic of transcriptionally permissive promoters, histone acetylation marks are representative of active promoter regions and usually coincide with enhancer elements and other regulatory motifs. H3K27ac marks (Panel D) are from three cells lines (h1-HESC, HUVEC and GM12878) studied by the ENCODE project.

3.2.2 Identification of putative GlyT-1 regulatory TFBS

A 2.2kb flanking sequence of the GlyT-1a gene transcript was analysed for the occurrence of TFBS. This sequence consisted of 2000 bases upstream of the GlyT1a TSS, the first exon (187 bases) and the first 85 un-translated (UTR) bases of the second exon. Using the MATCH algorithm with cut-off scores set to minimise false positives (minFP), 2481 sites were highlighted as potential binding sites using high quality PWMs from the TRANSFAC Matrix Table (data not included). Cut-off scores define the minimal similarity to the PWM for a given TFBS to be assumed biologically 'real' (minFP) or 'functional' (minFN). An important disadvantage of cut-off scores and in fact a limitation to the MATCH algorithm is that higher similarity to the PWM does not always imply a given site is biologically relevant. For instance TFs such as C/EBP, Sp-1 and Ap-1 all bind to functional biological sites with considerable sequence variation, as such a significantly lower similarity score (minFN of as low as 40% similarity to the core matrix) are required to computationally detect all of such sites in an input sequence. This could eventually result in numerous false negatives. On the flip-side, with high cut-off scores to minimise false positives (minFP), it is possible to miss biologically 'functional' sites which may differ significantly from the information content (IC[‡]) of a given PWM. The authors of the MATCH algorithm suggest using cut-off scores which minimise the sum (minSUM) of both error rates i.e. false-positives and false-negatives. Using minSUM cut-off scores, 28510 sites were identified in the GlyT-1a flanking sequence under investigation (data not included). Given that the biological specificity of any one TF is attained by a combinatorial relationship between several factors and co-factors (modules), determining the biological significance from a very large list of TFBS returned from computer algorithms such as this is best achieved by contextual analysis of the occurrence of these sites within specified sequence windows. One such approach is to determine which TFBS are overrepresented in co-regulated genes, and then study their arrangements within composite cis-regulatory modules (CRMs)

As originally pioneered by Spellman *et al*, the meta-analysis of large scale micro-array data to identify differentially co-expressed genes followed by promoter analysis to identify

---

‡ The information content (IC) or a PWM is the sum of the expected self-information. Self-information is a numerical measure of the information content associated with the unique outcome of a random occurrence. It is the negative log of the probability of observing a particular base at a particular position as defined in a PWM.

common (or highly represented) TFBS is a useful and popular strategy for the identification of CRMs (Spellman *et al.*, 1998). The hypothesis that genes with similar expression patterns are indicative of co-regulation overly simplifies the multi-faceted nature of gene regulation, which often may include other epigenetic and post-transcriptional, and protein modification modules. However, work done by Alloco and colleagues using 611 microarray datasets to quantify the relationship between co-expression, co-regulation and co-functionality, showed that for two genes to have a greater than 50% chance of being regulated by the same TF they must have a greater than 84% correlation in their factor-specific expression profiles (Allocco *et al.*, 2004). When the correlation in expression is condition or factor dependent it is more often than not consequential of similar regulatory modules. Hence, meaningful prediction of regulatory modules can be achieved from analysis of the promoter regions of genes sharing a strong correlation in expression under similar conditions. Based on these assumptions, an analysis of the promoter region of genes potentially co-regulated with GlyT-1a (identified by high correlation in their factor specific expression profiles) was performed to identify frequently occurring TFBS.

The Co-regulation Database (CORD) was searched to identify genes with high factor dependent co-relation in expression to GlyT-1a. Using 9490 public microarray datasets from the Gene Expression Omnibus (GEO) database and ArrayExpress database, the curators of CORD analysed over 120000 group by group comparisons of factor/condition dependent correlation in gene expression. CORD query parameters where set to allow a minimum fold change of greater than 2 in more than 3 experiments and a threshold p-value (correlation coefficient significance) of less than 0.01. Only genes with correlation of expression greater than or equal to 84% were retained for subsequent analysis. To identify co-regulated genes, CORD uses a gene ranking algorithm. Briefly, all microarray datasets in which the target gene (GlyT-1a) is differentially expressed are collected and grouped using one of two methods *viz:* 'grouped factor method' or 'individual factor method'. The Pearson correlation coefficient of the log2 fold change in expression between the target gene and each gene in the above datasets are then calculated and compared across each group. 185 potentially co-regulated genes were returned from the above query (data not shown).

Using the TRANSFAC Explain tool, the returned gene list was classified based on gene ontology (GO) of terms describing the biological processes they are involved in

(Appendix A.1). The list of co-regulated genes returned from CORD includes those involved in similar biological processes as is GlyT-1; such as in the transport of amino acids, nitrogenous and other organic compounds, neuro-transmission, cell cycle and development, anti-apoptotic genes, immune response and stress response genes. As most of these functions overlap postulated GlyT-1 function, the above list of co-regulated gene was deemed informative enough to proceed with promoter analysis for the identification of overrepresented TFBS.

Using the TRANSFAC F-MATCH algorithm, analysis of the promoter regions (spanning 2000 bases upstream to 1000 bases downstream of the TSS) of co-regulated genes identified by CORD was performed. F-MATCH uses an algorithm similar to MATCH but can simultaneously scan multiple promoters to identify over-represented TFBS as potential 'causes' for differential gene expressions. F-MATCH compares the occurrence of these over-represented TFBS in the 'test' or 'YES' list of differentially expressed gene promoters to a 'control' or 'NO' list of housekeeping genes to correct for multiplicity (errors in inference, *i.e.* where the frequency of a given TFBS is higher in the NO set than the YES set it is omitted from the result). For genes in the YES set of co-ordinately expressed genes with multiple promoters, only the promoters annotated in the TRANSPRO database with the best support were used. *P*-value cut-offs of 0.01 were used for F-MATCH. Only TFBS with significant enrichment in the promoter regions of the YES genes were studied. From this list, only TFBS present in the GlyT-1a promoter were retained for subsequent analysis. A complete list of the 131 enriched TFBS, *P*-values, and the TF which bind them is included in Appendix A.2.

Amongst the 131 TFBS identified as enriched were sites for some TFs known to be involved in the regulation of the stress response. These included, amongst many, sites for the activator protein 1 and 2 transcription factors (AP-1, 2); sites for the C/EBP proteins (C/EBP-α, β, δ, and γ); several E-box binding motifs, CREB/ATF; small MAF proteins (MAF-A, G and Z; which are known to interact with Nrf-2 in the regulation of oxidative stress), sites for NFAT (involved in the response to osmotic stress), the POU class of transcription factors involved in the regulation of cell cycle and sites for hypoxia inducible factor (HIF). Although such an analysis may provide an exhaustive list of TFBS enriched within a given set of promoters, very little can be garnered as to which are biologically relevant. Moreover, the above analysis does not discriminate between which of the promoters in the collection are actually involved downstream of the stress response

pathways under investigation here. As such, the significant enrichment of a given TFBS in the promoter set only provides a possible indication to its significance in the regulation of GlyT-1a downstream of stress. Thus, the arrangement of each of these TFBS on the GlyT-1a sequence was studied to identify possible composite regulatory modules (CRMs).

3.2.3 Composite regulatory modules (CRMs) of the GlyT-1a promoter

Simply defined, CRMs are a combination of cis- or trans-regulatory elements whose combinatorial effect on neighbouring gene expression is dependent on the individual factors bound to each regulatory site in the module. As several TFs and co-factors may be involved in the control of the spatial and temporal expression of any one gene, it is axiomatic that the outcome of any gene transcriptional regulatory network is dependent on the functioning of its constituent CRMs. CRMs offer important logic-like switches to molecular transcriptional events downstream of intracellular and extracellular stimuli such as stress (discussed later). They exert their transcriptional control on gene expression by exactly computing the titres of active nuclear localised transcription factors at any one time to satisfy maximum occupancy and function of individual TFBS in the module.

CRMs can be classified into two main groups based on the architecture and functioning of their constituent TFBS (Jeziorska *et al.*, 2009). Where the rigid arrangement of the individual TFBS within a given CRM allows for coordinated information processing of all the bound factors to generate a unified coordinated output, it is termed an 'enhanceosome'. Enhanceosomes are particularly sensitive to mutations or changes in any one of their constituent TFBS. Unlike enhanceosomes, there is a much greater flexibility in the arrangements of TFBS and occupancy of these sites by individual TF within 'billboard' CRMs. The transcriptional output of billboard CRMs is generally the sum of the competing effects of every factor in the module (Arnosti and Kulkarni, 2005).

Detecting CRMs in the promoters of responsive genes usually results in fewer false positives and better biological meaning over analysis of isolated individual TFBS (Guturu *et al.*, 2013). Whilst several computer algorithms exist for the detection of CRMs, the TRANSFAC Composite Model Analysis tool was used as it allows for the ability to define custom composite models. A composite model defines pairs of PWMs used for searching DNA sequences for TFBS located within a specified sequence window. In the absence of effective models for combinatorial interactions of TFs downstream of the stress response, composite models were designed to capture potential structural arrangement of

sites conducive to previously described protein-protein interactions of several stress response transcription factors (Table 3.1). For instance, the direct physical interaction between E-twenty-six (Ets-1 a transcription factor with several roles in cell development and tumour progression) and activator protein type 1 (Ap-1) has been previously described and associated with regulation of gene transcription in stress and cancer progression (Yordy and Muise-Helmericks, 2000). Using a composite Ap-1/Ets model of the vertebrate Ap-1 TRANSFAC matrix V$AP1_Q6_01 and Ets matrix V$ETS_Q6 (Figure 3.8), 28 Ap-1/Ets motifs were identified in the GlyT-1a flanking sequences (data not shown) where the two sites either overlap or are not separated by more than 30 nucleotides and have at least 80% similarity to their respective core PWM. Closer inspection of these motifs to identify which overlap with other TFBS for other proteins known to interact with either Ets-1 or Ap-1 can significantly improve the quality of such *in silico* predictions.

As discussed in Section 1.7, a plethora of BZIP factors control the temporal and differential transcriptional response to various stresses. BZIPs are characterised by a dimerisation 'zipper' domain of heptad repeats of hydrophobic and leucine residues as well as a basic region required for sequence specific DNA binding (Figure 3.9). Based on their originally characterised DNA binding specificity, BZIP transcription factors are classified into three subfamilies. This includes the classic AP-1 sub-family comprising factors that bind to a characteristic 12-o-tetradeconyphorbol-13-acetate (TPA) response element (also known as TRE; consensus TGACTCA; (Mitchell and Tjian, 1989)), the CREB/ATF subfamily which bind to a composite and often palindromic cAMP response element (CRE; consensus TGACGTCA; (Montminy *et al.*, 1986)) and the CCAAT/enhancer box binding protein (C/EBP) subfamily. Inter and intra family dimerisation of these BZIP factors with other protein super families is reported to regulate specific functional gene expression. All of the information required for dimerisation specificity to external stimuli is contained within the BZIP dimerisation domain, although other determinants such as the nature and calibre of external stimuli may favour one dimerisation partner over another.

Figure 3.8: Creating custom TFBS composite models using the TRANSFAC Composite Model tool. Composite models define two PWM used to search a DNA sequence to identify contextual occurrence of TFBS conducive to the interaction of the two factors. Using the Model definition form of the tool it is possible to define the similarity cut-off scores and orientation of each site, as well as the minimum overlap distance between the components and the maximum permissible distance of separation.
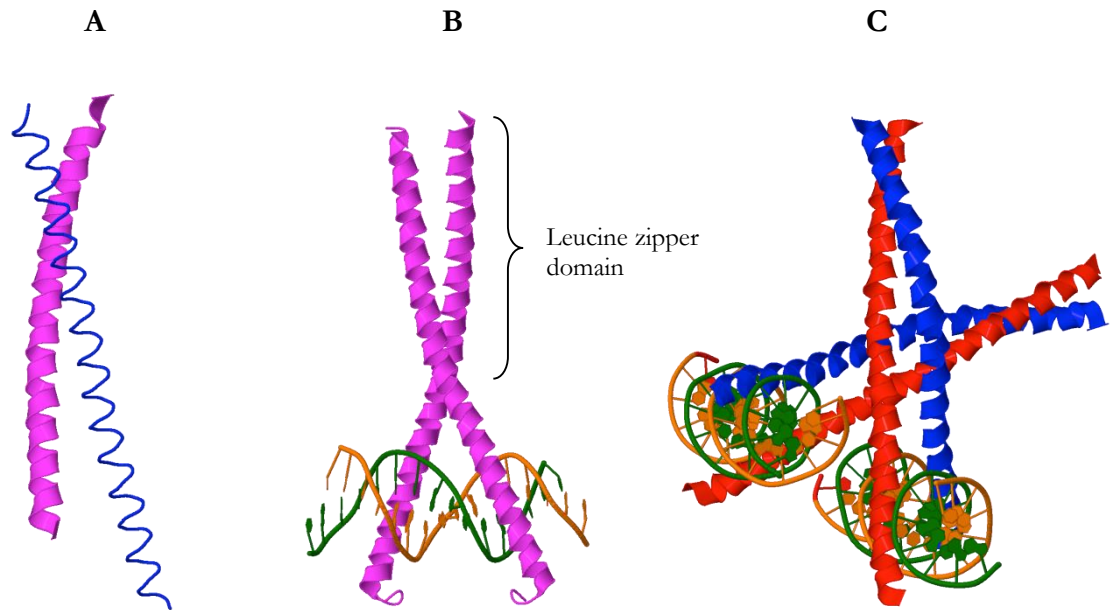
Figure 3.9: 3D Structure and dimerization of BZIP factors **(A).** Schematic of the BZIP heterodimer of mouse activating transcription factor 4 (ATF4) and CCAT/enhancer box binding protein beta (C/EBPβ**)** (MMDB ID: 15196, (Podust *et al.*, 2000)) demonstrating a characteristic BZIP coil-coil structure with homodimers of parallel alpha helices forming an inter-helix coiled-coil region via the leucine zipper and the two N-terminal basic regions (**B**)**.** Schematic of human C/EBP-β homodimers showing ternary interaction of the basic domain of the protein with the major grooves of the half sites on opposite sides of a DNA helix (MMDB ID: 61310, Tahirov *et. al.* 2006). (**C).** A BZIP tetrad of two human c-Fos/c-Jun heterodimers tethered together by additional residues within the zipper domain (MMDB ID: 72084, (Glover and Harrison, 1995, Glover *et al.*, 1995)).
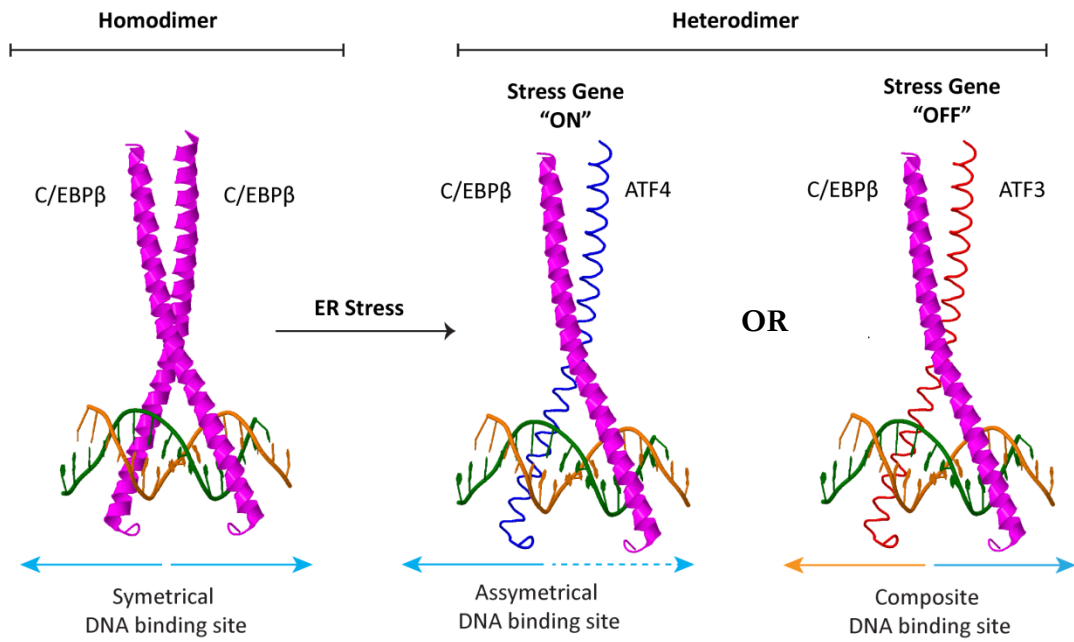
Figure 3.10: Illustration showing possible BZIP functional dimerisation patterns and the resulting DNA binding composite sites. Dimerisation is crucial for BZIP transcription factor activity and regulates the state of inducible genes. The AARE identified in GlyT-1a exon 1 is a composite ATF/CRE site with a characteristic C/EBP half site. Under normal physiological conditions homotypic dimers of two C/EBP proteins functionally interact at this site either as a repressor or an activator of gene expression. Following stress and the early induction of Atf-4, dimers of C/EBP-β/Atf-4 maintain the gene in an "ON" state. A feedback loop described at least in the transcriptional regulation of Cat-1 gene expression involves a monomer exchange at the same or similar binding site where in Atf-3 (a downstream target of ATF4) replaces Atf-4, turning "OFF" gene expression (Adapted from Tsukada *et al.*, 2011).

Of the 56 known human BZIPs, only a few self dimerise, with the rest forming and functioning as heterodimers with proteins from other families. The subset of BZIP factors investigated in the work described here was limited to those known to be active during stress. Careful consideration was given to the physical interactions of these factors downstream of stress as reported in the literature in the creation of composite models (Table 3.1).

Evidence of protein-protein interaction of these TFs was obtained from BioGrid – a database of experimentally validated protein-protein interactions (Stark *et al.*, 2006). BioGrid was preferred over other interaction databases such as StringDB (Franceschini *et al.*, 2013), as it appears to be the most actively curated with most of the interactions recorded in StringDB better documented in BioGrid. Composite models of factor pairs of Atf-4, C/EBPα, C/EBPβ, C/EBPε, C/EBPγ, c-Fos, Fos-L1, Jun and Nrf-2 were built in this manner and used to search the GlyT-1a sequence for putative CRMs. Sites not being part of a paired module, i.e. isolated TFBS for single BZIP with good matrix similarity scores, were still considered as it is understood that whilst dimerisation may be crucial to the activity of the resulting complexes, individual factors may bind to their respective sites and recruit or swap partners depending on the nature and duration of the stimuli (Figure 3.10).

As there are several PWMs of varying qualities and informational content in the TRANSFAC matrix table, deciding on which matrices to use was on a case by case basis. Assumptions were made on the theoretical permissible distance between each TFBS core in the composite model. A maximum separating distance of 9 was used as it's the longest know distance between the two halves in the ER stress response element (ERSE-I) consensus. Close or overlapping core motifs are not a stringent requirement for CRMs (Leung *et al.*, 2010). The stepwise binding of each subunit of the BZIP dimer complex to DNA could result in bending of the DNA helix to allow for co-binding of the dimerisation partner at near or distant regulatory motifs (Coulombe and Burton, 1999, Gustafson *et al.*, 1989, Kim *et al.*, 1989, Lorenz *et al.*, 1999). Nonetheless, as a key hypothesis in the computational detection of CRMs is the nearness of the core domains, a minimum overlap of five nucleotides and maximum separating distance of the individual TFBS of nine nucleotides was used.

Table 3.1: Stress response BZIP factors and their dimerisation partners involved in the stress response. Composite models of BZIP factors and their interacting partners were built to detect stress response motifs in the GlyT-1a sequence. Interacting partners were determined using the BioGrid database. (+) or (-) denote activating or repressive transcriptional activity. Typical stress response CRMs include the Amino Acid Response Element (AARE), the Nutrient Stress Response Element (NSRE-1), the ER stress response element (ESRE), the MAF recognition element (MARE), the antioxidant response element (ARE) and the Unfolded protein response element (UPRE). Shown in bold are factor pairs used to build composite models (Table adapted from Schroder and Kaufman, 2005).

Transcription factor	Cis-regulatory motif to which factor binds	Confirmed interacting partners at these sites
Atf-3	AARE	Atf-2 (-/+), Atf-3 (-), **Atf-4**, Atf-7, C/EBPξ, C/EBPγ, C/EBPα, c-JUN (-/+), NF-ϰBp50, CREBA
Atf-4	AARE, ARE, UPRE, ERSE, NSRE	**Atf-3** (-/+), Atf-7, B-Atf, **C/EBP-β**, **C/EBP-γ**, **C/EBP-α**, C/EBP-δ, **C/EBP-ε**, C/EBP-ξ, **c-Fos**, **Fos-L1**, FosB, HLF, **Jun**, JunD, **c-Maf**, Nrf-1, **Nrf-2** (+ on ARE, - on AARE)
Atf-6α	ATF/CRE, ERSE-I, ERSE-II	Atf-6α (+), Atf-6β, XBP-1(+), NF-Y(+), SRF
Atf-6β	ERSE-I	Atf-6α (-), Atf-6β
XBP-1	CRE, ERSE	Atf-6
XBP-1S	CRE, ERSE	Atf-6 (+)
Nrf-2	ARE, MARE	**Atf-4** (+), c-Jun (+), JunB(+), **cMaf**, **Maf-G**, **Maf-K** (+)

3.2.3.1 CRMs with binding sites for Atf-4 in the GlyT-1a Promoter

BZIP dimers with at least one Atf-4 subunit have been shown to interact at several stress response CRMs. These includes the amino acid response element (AARE), the antioxidant response element (ARE; consensus: G/C<u>TGAC</u>/T-N$_3$-GC (A/G), the unfolded protein response element (UPRE), the ER stress response elements (ERSE-I, consensus: CCAAT-N$_9$-CCACG and ERSE-II, consensus: AATTGG-N-CCACG) and the nutrient stress response elements. Hence, Atf-4 is widely accepted as an early regulator of downstream targets of the integrated stress response (ISR). As discussed earlier (Section 1.7), it is known to regulate the expression of several amino acid transporters and other genes downstream of amino acid limitation, unfolded protein accumulation, nutrient stress and ER stress.

At the time of analysis 94 experimental observations of 66 unique physical interactions of human Atf-4 with other proteins were documented in the BioGrid database of protein interactions. Reported interactions include with zinc finger proteins, various enzymes, other transcription factors, including the subunit 2α of the eukaryotic elongation factor (eIF-2α), subunit 3 of RNA polymerase II (RNAPII-B3), pCAF/KAT-2B, polyubiquitin-C and SUMO-3 all of which can facilitate Atf-4 activation and activity or its repression. Of the 66 unique interactions, 21 (~30%) are interactions between Atf-4 and other proteins containing a BZIP domain. For the analysis described here, only Atf-4 interactions documented in the literature to be involved in stress responses were considered (Table 3.1).

Ten putative CRMs including at least one site for Atf-4 were identified in the GlyT-1a sequence spanning 2000 bases upstream and 200 bases downstream of the TSS. (Figure 3.11a CRMs are numbered 1 to 10 for convenience). Furthest from the GlyT-1a TSS, CRM 1 spans a 21 base pair region (-1975 to -1955), but includes a sense site for Atf-4 (agCCTCAga) with 82.9% similarity to the V$ATF4_Q6 matrix. Overlapping the Atf-4 TFBS in this module are an antisense site for C/EBPα, antisense sites for small Maf proteins and Nrf-2 and sense sites for c-FOS and JUN which together form the classic Ap-1 complex (See Figure 3.11a for matrix sequences and similarity scores).

The significance if any, of the orientation of these TFBS within the putative CRMs is unknown. The approach used here relies solely on the assumption that the probability of occurrence of a TFBS in a given module relates to its probable biological function, i.e. in

the regulation of GlyT-1a. Crystal structures of the interaction of TF at symmetrical, asymmetrical or other composite DNA motifs reveal that interacting TFs within a complex may bind to opposite strands of the DNA (Guturu *et al.*, 2013). This may explain the possible functioning of modules with significantly overlapping binding sites.

A common feature across the ten identified CRMs is the sequence TGAC which is common to both the TRE (TGACTCA) and CRE (TGACgTCA) motifs. The only difference between the TRE and CRE consensus is the addition of a single guanine nucleotide between the first TGCA and the last TCA. Within CRM1 is the extended CRE/TRE like 12bp sequence <u>TGACG</u>gagcc<u>TCA</u>, which differs from the originally defined CRE consensus by the intercalation of the pentamer GAGCC between the first TGACG and the last TCA halves. The ability for BZIP factors to resolve these differences and bind in response to external stimulus appears to be conserved throughout evolution (Jones and Jones, 1989). Near exact CRE/TRE-like elements like the one identified in this CRM are known to bind several BZIP factor configurations. These include binding sites for the proto-oncogenic factors FOS and JUN, and members of the CREB/ATF subfamily of BZIP factors (Busch and Sassonecorsi, 1990).

Derivatives of the CRE and TRE have also been characterised as sites for the binding of Maf factors downstream of oxidative stress (Motohashi *et al.*, 2004, Motohashi *et al.*, 2002). The extended antioxidant response element consensus (G/C)TGA(C/T)-N_3-GC(A/G) is believed to be a mash-up of several Maf recognition sites (Erickson *et al.*, 2002). The palindromic Maf recognition element or MARE is further divided into two kinds: the C-MARE (CRE-like) of consensus TGC<u>TGA</u>(<u>GC</u>/CG)<u>TCA</u>GCA and the T-MARE (TRE-like) which has the consensus TGC<u>TGAC</u>(<u>G</u>/C)<u>TCA</u>GCA (Kurokawa *et al.*, 2009). The core of the consensus MARE sequences bear striking similarity to the CRE/TRE-like element identified in CRM1 of the GlyT-1a sequence (<u>TGACG</u>GAGCC<u>TCA</u>). As this CRM contains binding sites for both Maf (cTGACGgagcc) with 75.6% similarity to the matrix V$MAF_Q6_01 and Nrf-2 (CTGACggagcc) with 90.4% similarity to the matrix V$NFE2L2_01, we hypothesise that CRM1 may be important in the regulation of GlyT-1a downstream of oxidative stress.

Between nucleotide positions 1900 and 1100 upstream from the GlyT1a TSS, five CRMS (CRM2, 3, 4, 5, and 6; Figure 3.11a, b) contain TFBS that may favour the binding of BZIP dimers between Atf-4 and members of the C/EBP subfamily (C/EBPα, C/EBPβ,

C/EBPε and C/EBPγ). Whilst none of the respective sites overlap in either of these CRMs, the individual sites for either Atf-4 or C/EBP are all within a window not spanning more than seven bases. Closer inspection of the sequences within these modules highlights the occurrence of several TRE-like, CRE-like and CRE/Atf-like half sites (TGAGCCAT in CRM2; TCAGGAAG and TGATGCCTG in CRM6). Moreover the sequence similarity scores of the individual TFBS for the C/EBP proteins are greater than 90% in most of these modules.

Sequences identified in CRM4 (CCAGC-N$_7$-CCAAG), CRM5 (GGAAT-GGACC) and CRM6 (GGAAT-G-CCTGG) bear some resemblance to the ERSE-I consensus (CCAAT-N$_9$-CCACG; (Roy and Lee, 1999)). Repeats of the sequence GGAAN as seen in CRM5 (GGAAGgTTgtGGAAT) and CRM6 have been proposed to be required for the regulation of genes downstream of heat stress (Amin *et al.*, 1988). The consensus CNGGAAN$_2$TTCN$_2$G defines the heat shock response element (HSRE; (Xiao and Lis, 1988)). Such HSRE have been identified in the human and rat human oxygenase 1 (HO-1) gene, as well as heat shock protein 70 (HSP70) and have been shown to regulate HO-1 expression downstream of thermal stress (Okinaga *et al.*, 1996). Hence, these CRMs may warrant further experimental attention for validation of their possible involvement in the regulation of the transcription of GlyT-1a.

As illustrated in Figure 3.11b, CRM7 (ggTGAGGACTCAgg) is a short palindromic sequence of two TGAG halves, separated by an extra adenine (A) and guanine (G) base and flanked on either 3' and 5' ends by double guanine bases. Within these sequences is a reversed Atf-4 TFBS (ggTGAGGac) with 82.9% similarity to the V$ATF4_Q6 PWM as well as sites for cJUN and cFOS (see illustration for matrix similarity scores). Indeed this sequence resembles an extended TRE/Ap-1 consensus (TGA[G/C]TCA) and has been linked to the regulation of gene transcripts downstream of prolonged stress. Ap-1 TFBS often feature in the gene promoter region of several genes activated in response to external stimuli.

While studying the frequency distribution of Ap-1 TFBS in the human genome, Zhou and colleagues documented that although multiple genes had multiple sites for Ap-1 in their promoters there exist significant variations from the consensus (Zhou *et al.*, 2005). Two such variations include the A-TRE (TGAATCA), which differs from the TRE in that the fourth cytosine in the consensus is substituted for an adenine and the AA-TRE

(<u>TGACTA</u><u>A</u>). Frilling *et al.,* reported on the occurrence of two adjacent TRE/Ap-1-like sites (TGACATTGC and TGACAAAGC) in the promoter region of glutathione s-transferase (Friling *et al.*, 1992). In the later study, these Ap-1 like sites, when cloned upstream of the chloramphenicol acetyltransferase (CAT) coding sequence and transfected into HepG2 cells, generated significant reporter activity following 24 hours of either 30μM tert-Butylhydroquinone (tBHQ) or 50μM β-napthoflavone treatment. The authors concluded that these two sites may constitute a CRM required in the response to oxidative stress and labelled these sites as electrophile or antioxidant response elements.

An AA-TRE/Ap-1-like site of sequence (<u>TGACC</u><u>AA</u>) was identified in the putative CRM8 of the GlyT-1a promoter (Figure 3.11c). This site overlaps a sense Nrf-2 binding site of sequence C<u>TGAC</u>caagcc with 91% similarity to the V$NFE2L2_01 matrix. Likewise within CRM8 are sites for Maf, cFOS and cJUN. Seven bases upstream from the Nrf-2 consensus is a sense binding site for Atf-4 (caCCTCAaa) which has 90.09% similarity to the Atf-4 matrix used.

Two additional modules CRM9 (between -21 and +2 from the TSS) and CRM10 (in exon 1 between +104 and +134) within the Glyt-1a promoter contain sites for the C/EBP proteins (C/EBP-α, C/EBP-β, C/EBP-ε and C/EBP-γ) all with more than 90% similarity to the consensus sequence. Data from the ENCODE project suggest that the C/EBP-β site identified here in CRM10 by computational means is indeed a true binding site. Using a specific antibody to C/EBP-β (sc-150; Santa Cruz) ChIP-seq peaks revealed significant enrichment of the transcription factor at the sequence AAGAAAGCCACC of GlyT-1a from human embryonic stem cells (H1-hESC). There is no data to suggest that Atf-4 may interact at the other half site as at the time of analysis only 116 TF not including Atf-4 had been analysed in the ENCODE project. Nonetheless such validation of the C/EBP-β site identified here by computational means offers additional confidence to the possibility of biological significance of the other CRMs. Within CRM 10 is an antisense Atf-4 binding (gcTGATCa) site with 89.5% similarity to V$ATF4_Q6.

KEY

ATF4_Q6
CEBPA_Q6
CEBPB_Q6
CEBPE_Q6
CEBPG_Q6_01
NFE2L2_01
CFOS_Q6
FOSL1_01
CJUN_Q6
MAF_Q6_01

CRM 1
0.904 (CTGACGgagcc)   0.829 (agCCTCAga)
0.756 (cTGACGgagcc)   0.829 (agCCTCAga)
0.934 (TGACGg)   0.829 (agCCTCAga)
0.896 (TGACGgag)   0.829 (agCCTCAga)
0.919 (tcTTTCTgacgga)   0.829 (agCCTCAga)

```
-2000 CCCTAGGTTA GTCCCCAAGC CTTCCTCTTT CTGACGGAGC CTCAGAGGGC
-1950 TGGCCCCTGA GCACCCATCC TCCTCTCTTC TAACAGCCCA GGCCCCAAGC
```

0.904 (gCTTGCagcacc)   +ATF4 (+)

```
-1900 CCCTAGGCCT GAGGGCATCC TGGAAGTCTG GCTTGCAGCA CCCTACACCT
```

0.828 (cTGCCTccctt)
0.816 (cccTTCAta)
+CEBPG (-) 0.816 (cccTTCAta)
0.816 (cccTTCAta)   0.880 (ccctgAGCCAt)
0.816 (cccTTCAta)   0.926 (gATTCCctgagc)
0.816 (cccTTCAta)   0.926 (gATTCCctga)
CRM 2

```
-1850 GCCTCCCTTC ATATGGATTC CCTGAGCCAT GAAAACTCTC GGAATCTTCA
-1800 CTGAGTGTTC TTAGCGCTCA GTGCCCTCTT CTAGGCCTGT CCCCACCCCA
```

0.816 (cccTTCAta)   +CEBPE (+)
0.816 (cccTTCAta)   +CEBPb (+)

```
-1750 GCCTACTGCT GGAAGGGAGA AGGAGGGAGG CCTTTTCCCC TTCATACTCT
```

+Atf-4 (+)   0.853(ccgcCTTCCtctac)
+Atf-4 (+)   0.924 (cCTTCCtcta)   CRM 3

```
-1700 TCCTTGTCCG CCTTCCTCTA CTTGGGGGCC TCTTCGACAT CCTCCATCTG
```

CRM 4
0.822 (ttTGATGag)   0.863 (cgaggAGGCAa)
0.822 (ttTGATGag)   0.866 (ggaGGCAAGg)
0.822 (ttTGATGag)   0.925 (cgaggaGGCAAgg)

```
-1650 CCTTCCCTGG GGTGGTATCC CTTTTGATGA GCCCCAGCGA GGAGGCAAGG
-1600 AGGAAGATCC TAAAAAAAAC TTATGGAGGA AGTTACGTCC TGGCAGGAGA
-1550 GATATAGAGG GCTTGGACCC CATCTGCCCA GTCCCAGGAA TTGAGCACAC
-1500 ACATAACTAC ACAGACCTGC TGCAGTGCAC ACAATCAGTC TTCCACAGAT
-1450 GTACCTACAC AAATACACAA TCGGGGCTCA GGTGTCTGGG GGCACAGACA
-1400 CAACAGGTTG CCTAGATCCA GGTGTGTGTG CACACAACTC ACGCAAGAAA
-1350 CACCTGCACT TGGCTTGTAG GAGACTAAAA ATCCCAAACA TGCTATACCT
```

Figure 3.11a: Predicted CRMs with Atf-4 TFBS between 1300 and 2000 bases upstream of the GlyT-1a transcriptional start site (TSS). In this region of the GlyT-1a flanking sequence, four putative CRMs were identified using the TRANSFAC Composite Module Analyst Tool. Composite models were designed to capture sequences with at least 80% similarity to PWMs for Atf-4 and one other known interacting partner. Matrix-similarity scores are shown on the annotations together with the matching sequences included in brackets. Matrix core sequences are capitalised. Numbering is with respect to the transcriptional start site of GlyT-1a.

```
-1300  ACTAGGTTTG  GGGACAAAGT  AAAAGTGAAA  ATAATCTCGC  TGATAAGACT
-1250  AAGGTGTCTT  CCCAGATCTG  ATCAGGAGTC  CAATGTGAAT  AGGAGACAAC
```

KEY

ATF4_Q6
CEBPA_Q6
CEBPB_Q6
CEBPE_Q6
CEBPG_Q6_01
NFE2L2_01
CFOS_Q6
FOSL1_01
CJUN_Q6
MAF_Q6_01

CRM 5

0.836 (tcTGAGGaa) 0.913 (gtttgtGGAATg)
0.836 (tcTGAGGaa) 0.885 (tttgtGGAATggac)
0.836 (tcTGAGGaa) 0.957 (ttgtGGAATg)
0.836 (tcTGAGGaa) 0.977 (ggtttgTGGAATg)

```
-1200  AGAGAGCTGT  CTTTGTACTT  TGGCTCTGAG  GAAGGTTTGT  GGAATGGACC
```

CRM 6

0.744 (aTGCCTgggct) 0.879 (ctTGATGcc)
0.881 (tttcagGGAATg) 0.879 (ctTGATGcc)
0.871 (ttcaGGGAAtg) 0.879 (ctTGATGcc)
0.925 (tcagGGAATg) 0.879 (ctTGATGcc)
0.935 (atttcaGGGAAtg) 0.879 (ctTGATGcc)

```
-1150  GCGTGTGGCC  TGAGAGAGCT  CGCATTTCAG  GGAATGCCTG  GGCTTGATGC
-1100  CTGTTATCTT  GTTCAGTGCC  TCTGATAAAT  TAATAAACTG  GAATCTCACT
-1050  CAACAATTGG  TTTCACGTGA  GTTTGGCATG  TGTGCACTCG  CTTCTTGGAG
-1000  GCTGACTGTC  CCAAGGAAAA  AAATTAAACT  GTTTCCTGGC  CAGAAAATGG
```

0.896 (gACTCA)
0.829 (ggTGAGGac)
0.818 (aggACTCA) **CRM 7**
0.829 (ggTGAGGac)

```
 -950  TGAGGACTCA  GGGTCCAAAG  TAGAAATTGC  CTCTGTGGGC  TGGGCCCCTT
 -900  CCCTGAGCAG  ATGCCCTACT  CAGGTTGGGG  CAAAAACCAG  TTCCTCTGAC
 -850  AGAACTAACT  GGCCACTCTC  CATGAAAGGT  GGTGGCTCTC  ACAAGGTATG
 -800  CAAATGTGGG  AGGCTCATTC  CAAATGCCAG  CCACCACTCC  CCAGAGGCAG
 -750  CCTGGAGCAA  TGGGCTTTAT  ATACACACAC  ATGCATTTCC  TTAGCCCCTT
 -700  CTCTCTGTTT  CACAGAGGAG  GAAACCGAGG  CTCACCATGC  TCCTTTTGGC
 -650  AAAGACTCTT  TTTTAGCTGA  GATTGTAACC  CAGGTCTGTC  TCACCATGGA
```

Figure 3.11b: Predicted CRMs with Atf-4 TFBS between 1300 and 650 bases upstream from the transcriptional start site (TSS). In this region of the GlyT-1a flanking sequence, four putative CRMs were identified using the TRANSFAC Composite Module Analyst Tool. Composite models were designed to capture sequences with at least 80% similarity to PWMs for Atf-4 and one other known interacting partner. Matrix-similarity scores are shown on the annotations together with the matching sequences. PWM core sequences are capitalised. Numbering is with respect to the transcriptional start site of GlyT-1a.

```
ATF4_Q6
CEBPA_Q6
CEBPB_Q6
CEBPE_Q6
CEBPG_Q6_01
NFE2L2_01
CFOS_Q6
FOSL1_01
CJUN_Q6
MAF_Q6_01
```

```
                                                                    0.909 (caCCTCAaa)
                                                                    0.909 (caCCTCAaa)
                                                                    0.909 (caCCTCAaa)
                                                                    0.909 (caCCTCAaa)

-600  AGGAGGAGGG ATCCCAGGTT GGTGTCTGCA GCCCCTGTAA CACCACCTCA
+Atf-4(+)      0.910 (CTGACcaagcc)
+Atf-4(+)      0.761 (cTGACCaagcc)                  CRM 8
+Atf-4(+)      0.929 (TGACCa)
+Atf-4(+)      0.886 (TGACCaag)

-550  AAAGTAGAGC TGACCAAGCC GGTGGTCCAG GAGGGAAATT TGAGAAAGGG
-500  CTACCCCGAG CCCTCCAGTT CCAAGGAGTA GGACCTAGGC CTTACCCAGC
-450  AATGCCCACA GTCTCTAGCG GCTCTGCTCC AAAGGGAATG GAATTTGCTC
-400  TGACTTCTCT CTAGGGGAAA CGCAGGCGCA GAAGGCCAGG GGCCTGGAGG
                              0.794 (gTGCCcaggc)           0.864 (ggCTTCAtt)
-350  GCTAACTCCA GCCTCAGCCC CCCATAGTGC CCCAGGCCTT CTGGCTTCAT
-300  TCCAGCCTAG CCCGAAGGGG CCCACATCTG CTTGCTCAAG GGCCCCAGCT
-250  GGAGGGCAGC GAGGCAGGGG CGGAGAAGGG GACTCCTTTG TGGCGCTCAC
-200  TCCCAGGGGT TCGTGGGGAT CAATCATCTG AAAGCTGGAG AGACTTGTAA
-150  TTTTATTCCC TGGGAGGAGG AGCCTGAGAA GTCGGCACGG GAGGAGGGGA
-100  TGGGATGGAG AGGGGAGGAC GGCCAGGGGA GCCGTTCGGG TAGGGGAGGG
              0.887 (gatctcCCAAGc)
                                    0.807 ctCTTCAaa
                                    0.807 ctCTTCAaa      0.915 (aCTTTCaccaaa)
                                    0.807 ctCTTCAaa      0.913 (caaaCTTTCaccaa)   CRM 9
                                    0.807 ctCTTCAaa      0.959 (aCTTTCacca)
                                    0.807 ctCTTCAaa      0.952 (acTTTCAccaaac)

 -50  ACTTGGCTAG ATCTCCCAAG CCTTCTCCAC TCTTCAAACT TTCACCAAAC
  +1  TCTTAGCCCC CGCCTCCCCG AAACCAAAAC ACAACAGGCT GTGGAAGAGC
 +51  TGCGAGCGGC GCGGCGGGGC GGGCGAACTC GGCGCGGCAC AGAGCCTCGG
       0.895 (gcTGATGca)                0.931 (ctttaaGAAAGc)
       0.895 (gcTGATGca)                0.902 (tttaaGAAAGccac)
       0.895 (gcTGATGca)                0.965 (ttaaAGAAAGc)          CRM 10
       0.895 (gcTGATGca)                0.966 (cctttaAGAAAgc)

+101  GAGGCTGATG CAACTTTCCC TTTAAGAAAG CCACCTGGGC GCACCGCGGT
+151  GCGGACCCAG CACGCCTGGG CCGGGGGCTG CAGCATGGTA AGGGGGCTGG
```

Figure 3.11c: Predicted CRMs with Atf-4 TFBS between 600 bases upstream to 200 bases downstream from the GlyT-1a TSS. In this region of the GlyT-1a flanking sequence, four putative CRMs were identified using the TRANSFAC Composite Module Analyst Tool. Composite models were designed to capture sequences with at least 80% similarity to PWMs for Atf-4 and one other known interacting partner. Matrix-similarity scores are shown on the annotations together with the matching sequences included in brackets. PWM core sequences are capitalised. Numbering is with respect to the transcriptional start site of GlyT-1a.

3.2.3.2 CRMs with binding sites for Nrf-2 in the GlyT-1a Promoter

At the time of analysis 189 experimental observations of 56 unique direct physical interactions of human Nrf-2 (NFE2L2) with other proteins were documented in the BioGrid database. These include, as expected, interactions with KEAP-1 and Cullin-3 which are known to sequester the protein in the cytoplasm and mediate its degradation via ubiquitination. Interactions with Atf-4, Maf-G and Maf-F amongst many other BZIPs are catalogued. Physical interactions between Nrf-2 and several kinases have also been reported leading to the speculation that Nrf-2 may directly be activated by the same kinases which result in the phosphorylation of eIF2 and the activation of Atf-4. In this analysis, only interactions documented in the literature to be involved in the stress response were considered (Table 3.1).

In the computational search for CRMs involving Atf-4, two sites for possible Atf-4 and Nrf-2 interaction were identified (CRM1 and CRM8; Section 3.2.3.1). To determine if additional modules with binding sites for Nrf-2 and other dimerisation partners existed in the sequence, composite models of Nrf-2 and six other BZIP factors (Table 3.1) were used to scan the GlyT-1a flanking sequence under investigation. Aside from CRM1 and CRM8 which contain binding sites for both Nrf-2, Atf-4 and small Maf proteins (discussed above), no additional CRMs with binding sites for Nrf-2 were identified either in the 2000 bases upstream of the TSS, the first exon, or the untranslated region of the second exon of GlyT-1a.

3.2.4 Phylogenetic foot-printing of putative GlyT-1a CRMs

Although occurring within relatively conserved regions, the putative GlyT-1a CRMs 2-7 and 9 showed significant variations within the primate phylogenetic scope (Data not included). These variations suggest a faster evolution of these sequences when compared to other putative functional regions such as CRM1, CRM8 and CRM10. The observation that functional regulatory motifs within the promoter regions evolve at much slower rate than background sequences, elevates the usefulness of cross-species comparative genomic analysis (Wasserman and Sandelin, 2004). Given that highly conserved non-coding sequences within homologous genes represent areas of significant selective pressure, analysis beginning or ending with phylogenetic foot-printing of homologous promoter sequences by means of local and global alignment algorithms can lead to fewer false positive when identifying regulatory motifs solely from the use of PWMs. Several

multiple sequence alignment (MSA) programs like the classic command line (CLI) library CLUSTAL software family including CLUSTAL X, CLUSTAL W and more recent CLUSTAL Ω, MUSCLE (Multiple Sequence Comparison by Log-Expectation), MAFFT (Multiple Sequence Alignment based on Fast Fourier Transform) and T-COFFEE (Tree-based Consistency Objective Function for alignment Evaluation), enable the discovery of conserved motifs based on the evolutionary biological relationships between multi-aligned input sequences. Practicality, for instance in terms of speed of execution or biological accuracy determines the suitability of any one of these alignment programs, each of which employs distinct progressive or iterative statistical evaluations to determine the most biologically relevant relationship with maximum parsimony (maximum likelihood or simplest assumption) between two or more sequences. Biological accuracy, whilst unrelated to speed of execution is dependent on several factors, including the number of sequences being aligned, sequence lengths and method specific parametric differences. Given that phylogenetic footprinting analysis are based on the assumption of similarities in conserved sequences across assumed orthologous species, any overlooked differences in base-wise conservation do not necessarily mean a non-functional regulatory motif. In a clustering analysis of such highly degenerate yet experimentally validated TFBSs in the mammalian genome, Zhang *et al* observed a non-random distribution of mismatches, conserved across orthologs in a manner that was significantly more than expected by chance (Zhang *et al.*, 2006). Based on the very statistical nature of phylogenetic analysis it is possible that regulatory motifs go unidentified where there is no sufficient ancestrally relevant input data or over-reliance on base-wise conservation. False negatives from phylogenetic footprinting analyses could result from a large evolutionary separation (across multiple speciation events) between input sequences.

Several attempts have been made to qualify and quantify the relationships between sequence conservation and function (Duret *et al.*, 1993, Roy *et al.*, 2013); Reviewed in (Cooper and Brown, 2008). To date, there exists no exact measure of correlation between these two variables (a measure of the strength and efficacy of natural selection). In fact one of the highlights of the ENCODE project was the discovery of several regulatory modules not within highly conserved regions. Cooper *et al* argue that the false discovery rate (FDR) from comparative genomic approaches may be down to several technical challenges as well as the phylogenetic scope used in such analysis (Cooper and Brown, 2008). In this study, orthologous GlyT-1a sequences from 56 vertebrate species (as listed

in appendix A.3) were aligned using the MAFFT algorithm to assay the conservation of the CRMs identified as described in section 3.2.3.

3.2.4.1 Conservation of putative GlyT-1a CRM1 across vertebrate species

As illustrated in Figure 3.12, the CRM1 sequence is highly conserved within the primate subset of the phylogenetic scope analysed. Using a similar method as described by (Hestand *et al.*, 2008) the calculated conservation score of primate CRM1 amounts to 95.21%. Fourteen of the twenty-one bases of CRM1 have 100% conservation across the 12 primates analysed. The least conserved base in the primate subgroup is the guanine at position 11 of the CRM sequence, where it is substituted for an adenine in four out of the 12 species (i.e. 66.6% conservation). The preceding cytosine at position 10 of the CRM is also substituted for a thymine in three of the 12 primate species analysed (i.e. 75% conservation score). The Atf-4 binding half site within this module of sequence (agCCTCTGAga) is extensively conserved, suggesting it may be an important functional region in primates. Including other divergent species to the phylogenetic analysis suggest that this composite module might have evolved under relaxed structural and evolutionary constraints in a manner that is consistent with billboard CRMs (See Figure 1 in Meireles-Filho and Stark (2009)). The flexibility in function of billboard CRMs often coincides with characteristic flexibility in structural conservation of the individual TFBS. In CRM1 despite the absence of the full motif in the guinea pig sequence, some of the marsupials and Sarcopteryggi species, a CRE/ATF-like binding sequence (in the second half of the module) has been maintained throughout a larger evolutionary timespan. As argued by Miereles-Filho *et al*, billboard motifs such as CRM1 gain other TFBS which often are sites for interacting partners to proteins which bind at the maintained site (Meireles-Filho and Stark, 2009). Such CRMs may appear 'shuffled' over large evolutionary timespans. Such shuffling of TFBS for BZIP factors coincides with the fact that most metazoan BZIP factors are able to recognise composite sites with either CRE-like, TRE-like, MARE-like sequences and CCAAT half sites whilst ancestral BZIPs only recognise either CRE and TRE motifs (Deppmann *et al.*, 2006, Amoutzias *et al.*, 2007). Reinke *et al.* (2013) show that interactions between Atf-4, Mafs, and cap and collar (CNC) factors like NFE2 have only 'recently' been acquired.

```
                         Bits 2
Sequence Logo                 1   TCTTTCTGA    TAGAGC        C         TCAGA
                              0

REV Human                  TCTTTCTGA--CGGAGC------C---------TCAGA
REV Chimp                  TCTTTCTGA--CGGAGC------C---------TCAGA
REV Gorilla                TCTTTCTGA--CAGAGC------C---------TCAGA
REV Orangutan              TCTTTCTGA--TGGGGC------C---------TCAGA
REV Gibbon                 TCTTTCTGA--CGGAGC------C---------TCAGA
REV Rhesus                 TCTTTCTGA--CGGAGC------C---------TCAGA
REV Crab-eating-macaque    TCTTTCTGA--CGGAGC------C---------TCAGA
REV Baboon                 TCTTTCTGA--CGGAGC------C---------TCAGA
REV Green-monkey           TCATTCTGA--CGGAGC------C---------TCAGA
REV Marmoset               TCTTTCTGA--TAAAGC------C---------TCAGA
REV Squirrel-monkey        TCTTTCTGA--TAGAGC------C---------TCAGA
REV Bushbaby               TCTTCCTGA--CAGAGC------C---------TCTGA
REV Chinese-tree-shrew     TCTTTCCGA--TACCAC------C---------TCCGA
REV Squirrel                 TTAGGGTTAGAAC------C---------TCTGA
REV Lesser-Egyptian-jerboa   TCTGA--TAGGGC---------------TGA
REV Prairie-vole             TCTGG--T-GTGT-----AC---------CCTGA
REV Chinese-hamster          TCTGG--TAGTTC-----AT---------TCTAC
REV Golden-hamster           CCTGG--TAGCAC-----AC---------TCTGC
REV Mouse                    TCTAG--TAATAC---CAAT---------TCTGC
REV Rat                      TCTGG--TAATAC---TAAT---------TCTGC
REV Naked-mole-rat           TCTGC--TAGAGC------C---------CCTAA
REV Guinea-pig
REV Chinchilla               TCCGA--TAGAGC------T---------CCTAA
REV Brush-tailed-rat         TCTGC--TAGAGT------C---------CCTGA
REV Pika                     TCAGA--GGGTTAGCCAAGC---------CCACA
REV Alpaca                 TCTTCCTAA--TAGAGG------C---------TCAGA
REV Dolphin                TCTCCCTGA--TAGATC------C---------TCAGG
REV Killer-whale           TCTCCCTGA--TAGAGC------C---------TCAGG
REV Cow                    TCTCCTTGA---------------C---------TCAGG
REV Sheep                  TCTCCCTGA---------------C---------TCAGG
REV Horse                  TCTTCCTGA--TGGGGC------T---------TCAGA
REV White-rhinoceros       TCTTCTTGA--TGGGGC------C---------TCAGA
REV Cat                    TCTTCATAA--TGAGTC------C---------TCAGA
REV Dog                    TCTTCCTGA--TAAGGC------C---------TCAGA
REV Ferret                 TCCTCCCAG--TAAGGC------C---------TCAGG
REV Panda                  TCCTCCCG--TAAGGC------C---------TGGGA
REV Pacific-walrus         TCCTCC-GA--TAAGGC------C---------TCAGA
REV Weddell-seal           TCCTCCCAA--TAAGGC------C---------TCAGA
REV Black-flying-fox       TCTTCCTGG--TAGGGC------C---------TGAG
REV Megabat                TCTTCCTGG--TAGGGC------C---------TGAG
REV Microbat               TCCTCCTGG--TAGGGC------C---------TCAGA
REV Big-brown-bat          T--TCCTGG--TAGGGC------C---------TCAGA
REV Hedgehog               ATTTCCTGT--AGAGCA------C---------CAGAA
REV Shrew                  TCCTCCGGC--GAGGCC------C---------TCAAA
REV Star-nosed-mole        TCTTCCTGG--TAGGAA------C---------CCTGA
REV Elephant               TCCT------CAGAGC------AAGGG----TCAGA
REV Cape-elephant-shrew    TCTT------TGAGAT------GAGATCTTCTCTGA
REV Manatee                TC------------------------TCAGA
REV Cape-golden-mole       CCCT------CAAGAT------GAAGT
REV Aardvark               TCCT------CAGGAT------AAAG-----TCTGA
REV Armadillo              TTGT------TGGAGC------C---------TCTGA
REV Opossum
REV Wallaby
REV Platypus
REV American-alligator       TCTGG--ACAAGG------G---------TCAGA
REV Lamprey
```

Primates

Figure 3.12: Conservation of the putative CRM1 across 56 vertebrate species. Spaces in the sequence represent gaps whilst adjacent bases are indicated with a dash (-). See text for discussion (section 3.2.4.1 page 104)

3.2.4.2 Conservation of putative GlyT-1a CRM8 across vertebrate species

As illustrated in Figure 3.13, there is extensive conservation of the CRM8 sequence over a much larger phylogenetic scope, when compared to CRM1. The 28bp GlyT-1a CRM8 sequence has a conservation score of 93.72% across the twelve primate subgroups studied. When just the individual TFBS in the module are considered, the identified Atf-4 binding sequence in this putative module (caCCTCAaa; which has a 90.9% similarity to the V$ATF4_Q6 PWM) has 95.41% conservation, whilst the Nrf-2 binding sequence (CTGACcaagcc; which has a 91.01% similarity to the V$NFE2L2_01 PWM) has 89.34% conservation rate amongst primates. The heptamer AGTAGAG which separates the putative Atf-4 binding site from the Nrf-2 half site in this module has a 97.6% conservation rate across the twelve primate species studied.

Extending the conservation analysis to include 31 species diverging from the supraprimate (euarchontoglires) and placental mammal (laurasiatheria) clades, reveals that the TFBS for the Nrf-2 sequence is maintained over these phylogenetic scopes (77.77% conservation). Conservation for the Atf-4 site in this sub-scope is only 27.24%. The significance of the seven base pair region separating the Atf-4 site from the Nrf-2 site in this module is unknown given that no factor which binds it was found in the CRM analysis. This heptamer sequence has a conservation score of 67.28% across the scope of these 31 species (59.69% across all 56 species analysed).

The human GlyT-1a putative CRM8 sequence was not detected in rat, wallaby, platypus, American alligator or lamprey. As is the case with CRM1, the conservation pattern of CRM8 across vertebrate speciation events suggests that the binding site for Nrf-2 may have been maintained whilst a binding site for Atf-4 was gained only late in evolution and found only in primates. Three ways by which regulatory modules might gain TFBS include via the insertion of transposable elements, chromosomal rearrangement and *de novo* mutations (Rubinstein and de Souza, 2013). Rubinstein *et al* argue that whilst these events may lead to TFBS shuffling within the module without altering ancestral expression patterns, the acquisition of novel TFBS to a module might provide an additional spatio-temporal domain for gene regulation. For instance the gaining of a silencer site for the binding of a repressor protein by a CRM might enable a feedback opportunity to switch-off the expression of the regulated gene from an already present activating TFBS (Ochi *et al.*, 2012).

| Sequence Logo | Bits 2 1 0 |

```
                          C                    A C        C  AGA    GTAGGGCTGACCAAGCC

REV Human                 C----------CACC--TCAAAA----GTAGAGCTGACCAAGCC
REV Chimp                 C----------CACC--TCAAAA----GTAGAGCTGACCAAGCC
REV Gorilla               C----------CACC--TCAAAA----GTAGAGCTGACCAAGCC
REV Orangutan             C----------CACC--TCAAAA----GTAGAGCTGACCAAGCC
REV Gibbon                C----------CACC--TCAAAA----GTAGAGCTGTCCAAGCC
REV Rhesus                C----------CACC--TCAAAA----GTAGAGCTGACCAAGCC
REV Crab-eating-macaque   C----------CACC--TCAAAA----GTAGAGCTGACCAAGCC
REV Baboon                C----------CACC--TCAAAA----GTAGAGCTGACCAAGCC
REV Green-monkey          C----------CACC--TCAAAA----GTAGAGCTGACCAAGCC
REV Marmoset              C----------CATC--TCAAAA----GTAGAGCTGAGCAAGTC
REV Squirrel-monkey       C----------CACC--TCAAAA----GTAGAGCTGAGCAAGCC
REV Bushbaby              C----------CAAC--ACAAGA----CTAGGGCT

REV Chinese-tree-shrew    C----------CACC--ACCAGA----GTAGGGCCAAACAAACC
REV Squirrel              C----------CACC--GCCAGA----GCAGGGCTGACCAAACC
REV Lesser-Egyptian-jerboa T----------GACC--ATCAGG----GTAGGGCTAAG-AAACC
REV Prairie-vole          C----------TACC--ACCTGG----GGAGGGCTGATCAAATC
REV Chinese-hamster       C----------TACC--ACCTGG----GTAGGGCTGATCAAATC
REV Golden-hamster        C----------TACC--ACCTGG----GTAGGGCTGATCAAATC
REV Mouse                 C----------AACT--AACCAG----GTAGGGCTGATCAAATC
REV Rat
REV Naked-mole-rat        C----------CACT--GGTGGG----GTTGGGCTGAACAAGAC
REV Guinea-pig            C----------TATC--ACCAGG----TTAGGGCTGAATAAGCC
REV Chinchilla            C----------TATC--GACAGG----GTGGGGCTGAATAAGCC
REV Brush-tailed-rat      G----------TACC--GCCAGG----GTAGGGTTGAATCAGCC
REV Pika                                              GCCGACCGAGCC

REV Alpaca                C----------AACT--GC--GA----GTAGGGCTGACCAAATC
REV Dolphin               C----------AATG--GC--AA----GTAGGGTTGACCAAACC
REV Killer-whale          C----------AATG--GC--AA----GTAGGGTTGACCAAACC
REV Cow                   C-------------------------GTAGGTCTGACCAAACC
REV Sheep                 C----------AACC
REV Horse                 C----------AACT--GC--GA----GTAGAGCTGACCGAGCC
REV White-rhinoceros      C-------------T--GC--GC----GTAGGGCTGACCAAGCT
REV Cat                   G----------AAGC--AG--GA----ATAGGGGTGACCAAACC
REV Dog                   C----------A-GC--AG--GA----GTAGGGCTGCCCAAACC
REV Ferret                C----------AAGC--AG--AA----GTAGGGCTGACCAAACC
REV Panda                 C----------AAGC--AG--GA----GTAGGACTGACCAAACC
REV Pacific-walrus        C----------AAGC--AG--GA----GTAGGGCTGACCAAACC
REV Weddell-seal          C----------AAGC--AG--GA----GTAGGGCTGACCAAACC
REV Black-flying-fox      C----------AATG--GT--GA----GTAGGGTTGACCAAGCC
REV Megabat               C----------AATG--GT--GA----GTAGGGTTGACCAAGCC
REV Microbat              C----------AATG--GC--GA----GTAGGGTTGACCAAGCC
REV Big-brown-bat         C----------AATG--GC--GA----GTAGGGTTGACCAAGCC
REV Hedgehog                          GAGG--GC--GA----GT-AGGGTAACCAAGCT
REV Shrew                             CAGG--GC--AC----GTAAGGCTCACCA
REV Star-nosed-mole                   AAGG--GC--GA----GTCAGGCTGACCAGACC

REV Elephant              C-----------------CAGA----ATGGGGCTGCCCAGGCC
REV Cape-elephant-shrew   C-------------A-----GTCAGA----ATGGGGCTGCCCAGGCC
REV Manatee               A----------C-----GCTAGA----ATGGGGCTGCCCAGGCC
REV Cape-golden-mole      A----------A-----ACGAGA----GTGGAGCTATCTAAGTC
REV Aardvark              A----------G-----GCTAGA----GTGGGGCTGCCTAGGCC

REV Armadillo             A----------G-----GCCAGA----GTGGGGCTGACCAGGCC
REV Opossum               CCGTCCACCCT--CCGCTCCAGGTCAAGTGGGG-TGAAGCTGCC
REV Wallby
REV Platypus
REV American
REV Lamprey
```

Figure 3.13: Conservation of the putative CRM8 across 56 vertebrate species. MAFFT alignment of 56 orthologous species reveals extensive conservation from small mammals to primates. Spaces in the sequence represent gaps whilst adjacent bases are indicated with a dash (-). Note the complete absence of this module in rat GlyT-1a, as well as the absence in Sarcopteryggi and small mammals. See text for discussion

3.2.4.3 Conservation of putative GlyT-1a CRM-10 across vertebrate species

The 32 bases of the putative GlyT-1a CRM10 are within the first non-coding exon of GlyT-1a and as shown above contains binding sites for Atf-4 and several C/EBP BZIP factors. Conservation of this CRM was analysed across 56 orthologous GlyT-1a species. Amongst the primate phylogenetic scope it has a conservation score of 98.44% (Figure 3.14). With the exception of the American lizard and the lamprey, the CRM10 sequence is extensively conserved in the remaining 54 GlyT-1a orthologs analysed (96.37% conservation). The Atf-4 half site of the sequence (gcTCATGca) is unchanged in 51 of the 56 species. The first guanine of the Atf-4 site is substituted for an adenine in sheep and a cytosine in the GlyT-1a sequence from American alligator and lamprey. It is completely missing in the gibbon (primate) and platypus. A similar conservation pattern is noted in the C/EBP (AAGAAA--GCCACC) sites.

Sequence Logo

Bits: 2, 1, 0

Sequence Logo (top): GCTGATGCAACTTTCC   CTTTAAGAAA GCCACC

Species	Sequence
REV Human	GCTGATGCAACTTTCC------CTTTAAGAAA--GCCACC
REV Chimp	GCTGATGCAACTTTCC------CTTTAAGAAA--GCCACC
REV Gorilla	GCTGATGCAACTTTCC------CTTTAAGAAA--GCCACC
REV Orangutan	GCTGATGCAACTTTCC------CTTTAAGAAA--GCCACC
REV Gibbon	CTGATGCAACTTTCC------CTTTAAGAAA--GCCACC
REV Rhesus	GCTGATGCAACTTTCA------CTTTAAGAAA--GCCACC
REV Crab-eating-macaque	GCTGATGCAACTTTCA------CTTTAAGAAA--ACCACC
REV Baboon	GCTGATGCAACTTTCA------CTTTAAGAAA--GCCACC
REV Green-monkey	GCTGATGCAACTTTCA------CTTTAAGAAA--GCCACC
REV Marmoset	GCTGATGCAACTTTCC------CTTTAAGAAA--GCCACC
REV Squirrel-monkey	GCTGATGCAACTTTCC------CTTTAAGAAA--GCCACC
REV Bushbaby	GCTGATGCAACTTTCC------CTTTAAGAAA--GCCACC
REV Chinese-tree-shrew	GCTGATGCAACTTTCC------CTTTAAGAAA--GCCACC
REV Squirrel	GCTGATGCAACTTTCC------CTTTAAGAAA--GCCACC
REV Lesser-Egyptian-jerboa	GCTGATGCAACTTTCC------CTTTAAGAAA--GCCACC
REV Prairie-vole	GCTGATGCAACTTTCC------CTTTAAGAAA--GCCACC
REV Chinese-hamster	GCTGATGCAACTTTCC------CTTTAAGAAA--GCCACC
REV Golden-hamster	GCTGATGCAACTTTCC------CTTTAAGAAA--GCCACC
REV Mouse	GCTGATGCAACTTTCC------CTTTAAGAAA--GCCACC
REV Rat	GCTGATGCAACTTTCC------CTTTAAGAAA--GCCACC
REV Naked-mole-rat	GCTGATGCAACTCCC------CTTTAAGAAA--GCCACC
REV Guinea-pig	GCTGATGCAACTTTCC------CTTTAAGAAA--GCCACC
REV Chinchilla	GCTGATGCAACTTTCC------CTTTAAGAAA--GCCACC
REV Brush-tailed-rat	GCTGATGCAACTTTCC------CTTTAAGAAA--GACACC
REV Pika	GCTGATGCAACTTTCC------CTTTAAGAAA--GCCACC
REV Alpaca	GCTGATGCAACTTTCC------CTTTAAGAAA--ACCACC
REV Dolphin	GCTGATGCAACTTTCC------CTTTAAGAAA--GCCACC
REV Killer-whale	GCTGATGCAACTTTCC------CTTTAAGAAA--GCCACC
REV Cow	GCTGATGCAACTTTCC------CTTTAAGAAA--GCCACC
REV Sheep	ACTGATGCAACTTTCC------CTTTAAGAAA--GCCACC
REV Horse	GCTGATGCAACTTTCC------CTTTAAGAAA--GCCACC
REV White-rhinoceros	GCTGATGCAACTTTCC------CTTTAAGAAA--GCCACC
REV Cat	GCTGATGCAACTTTCC------CTTTAAGAAA--GCCACC
REV Dog	GCTGATGCAACTTTCC------CTTTAAGAAA--GCCACC
REV Ferret	GCTGATGCAACTTTCC------CTTTAAGAAA--GCCACC
REV Panda	GCTGATGCAACTTTCC------CTTTAAGAAA--GCCACC
REV Pacific-walrus	GCTGATGCAACTTTCC------CTTTAAGAAA--GCCACC
REV Weddell-seal	GCTGATGCAACTTTCC------CTTTAAGAAA--GCCACC
REV Black-flying-fox	GCTGATGCAACTTTCC------CTTTAAGAAA--GCCACC
REV Megabat	GCTGATGCAACTTTCC------CTTTAAGAAA--GCCACC
REV Microbat	GCTGATGCAACTTTCC------CTTTAAGAAA--GCCACC
REV Big-brown-bat	GCTGATGCAACTTTCC------CTTTAAGAAA--GCCACC
REV Hedgehog	GCTGATGCAACTTTCC------CTTTAAGAAAAAGCCACC
REV Shrew	GCTGATGCAACTTTCC------CTTTAAGAAA--GCCACC
REV Star-nosed-mole	GCTGATGCAACCTTCC------CTTTAAGAAA--GCCACC
REV Elephant	GCTGATGCAACTTTCC------CTTTAAGAAA--GCCACC
REV Cape-elephant-shrew	GCTGATGCAACTTTCC------CTTTAAGAAA--ACCACC
REV Manatee	GCTGATGCAACTTTCC------CTTTAAGAAA--GCCACC
REV Cape-golden-mole	GCTGATGCAACTTTCC------CTTTAAGAAA--ACCACC
REV Aardvark	GCTGATGCAACTTTCC------CTTTAAGAAA--GCCACC
REV Armadillo	GCTGATGCAACTTTCC------CTTTAAGAAA--GCCACC
REV Opossum	GCTGATGCAACTTTCC------CTTTAAGAAA--TCCACC
REV Wallaby	GCTGATGCAACTCTCC------CTTTAAGAAA--TCCACC
REV Platypus	CTGATGCAATTTATC------CTTTAAGATA--TCTACC
REV American-alligator	CCCATTCCCACTCTCCTGCCCGCCTGGAAAGA
REV Lamprey	CGTCGCACGACGGCTC------CCT---------GCCGTC

Figure 3.14: Conservation of the putative CRM10 across 56 vertebrate species. MAFFT alignment of 56 orthologous species reveals extensive conservation from small mammals to primates. Spaces in the sequence represent gaps whilst adjacent bases are indicated with a dash (-). Note the American alligator TGCCCG sequence which separates the first two halves and the insert "AA" in the hedgehog sequence.

Primates (bracket grouping Human through Bushbaby)

3.3 Stress Induced Regulation of GlyT-1a gene

Ten putative composite regulatory modules (CRMs) have been identified in the GlyT-1a flanking sequence and 5'UTR by bioinformatics analysis (Section 3.2.3). These include a potential amino acid response element (AARE; of sequence TGACTGCAAC) in CRM10 of the first exon and an electrophile or antioxidant response element (EpRE/ARE; of sequence CTGACCAAGCC) in CRM8. To determine if these and other CRMs of the GlyT-1a promoter respond to various stress treatments *in vitro*, reporter constructs of representative GlyT-1a sequences cloned upstream of the β-galactosidase coding sequence in a reporter plasmid were generated as discussed in Section 2.5.1 (Page 59).

A reporter construct (pG1PromE1) that contained 1159 bases of GlyT-1a genomic sequence upstream of the TSS and the whole of the 187 base first exon of GlyT-1a linked to a β-galactosidase reporter gene was generated and validated by sequencing (Figure 3.15). As the first exon of GlyT-1a contains the translation start codon (ATG) of an upstream open reading frame (uORF), the sequences were cloned in frame so as to allow for the expression of β-galactosidase only from its start codon. Despite the presence of the predicted AARE in this construct, no significant difference in β-galactosidase activity was noted between Caco-2 cells transfected with this construct (pG1PromE1), those transfected with a promoter-less, 'empty' β-galactosidase vector (pNoPromLacZ) or in un-transfected control cells. Treatment of transfected Caco-2 cells for 16 hours with 1μM tunicamycin had no additional potentiating effect on the β-galactosidase activity (Figure 3.16).

A possible explanation for the lack of activity of this construct is that the AARE functions as a 'billboard' type CRM requiring a collection of other TFBS working in concert to produce the desired transcriptional outcome. In other words, additional sequences not included in the promoter construct may be required for effective functioning of the AARE of exon 1. Indeed, such an observation has been previously reported for the AARE elements of Cat-1 (Fernandez *et al.*, 2003). In the latter study, Fernandez and colleagues showed that although the Cat-1 gene contains an AARE downstream of its TSS (precisely, in its first exon) when C6 glioma cells were transfected with a Cat-1/luciferase reporter construct spanning 1.4kb upstream of the TSS and including exon 1, no difference in activity was observed between stressed and unstressed states. Results from the same study showed that replacing the exon 1 sequence with the complete Cat-1

5'-UTR which consists of exons 1, 2 and part of exon 3 (which contains the Cat-1 translational start codon) resulted in significant increases in expression from this construct in starved over fed states.
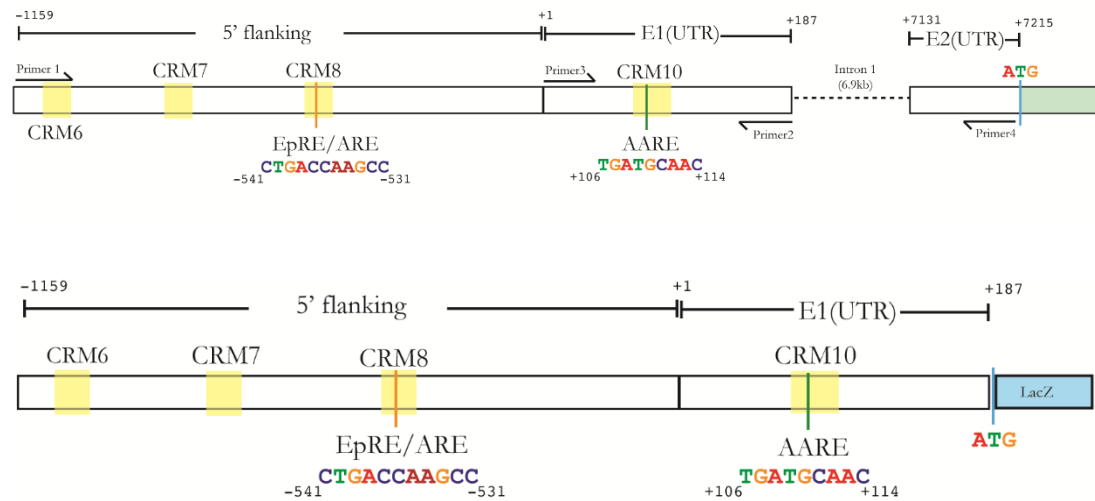
Figure 3.15: Illustration of the flanking region of the GlyT-1a gene(upper), showing the locations of the identified composite regulatory modules (CRMs). The GlyT-1a first and second exons are interrupted by a 6.9kb intron. The derived reporter construct pG1PromE1 (lower) was obtained from cloning the GlyT-1a flanking sequence amplified from genomic DNA using a sense primer (Primer 1) and reverse primer (Primer 2). Features are numbered relative to the transcription start site (TSS) and the position of a putative amino acid response element (AARE; TGATGCAAC) and electrophile or antioxidant response elements (EpRE/ARE) are given. Primers 3 and 4 were used for generation of a second construct as discussed below.
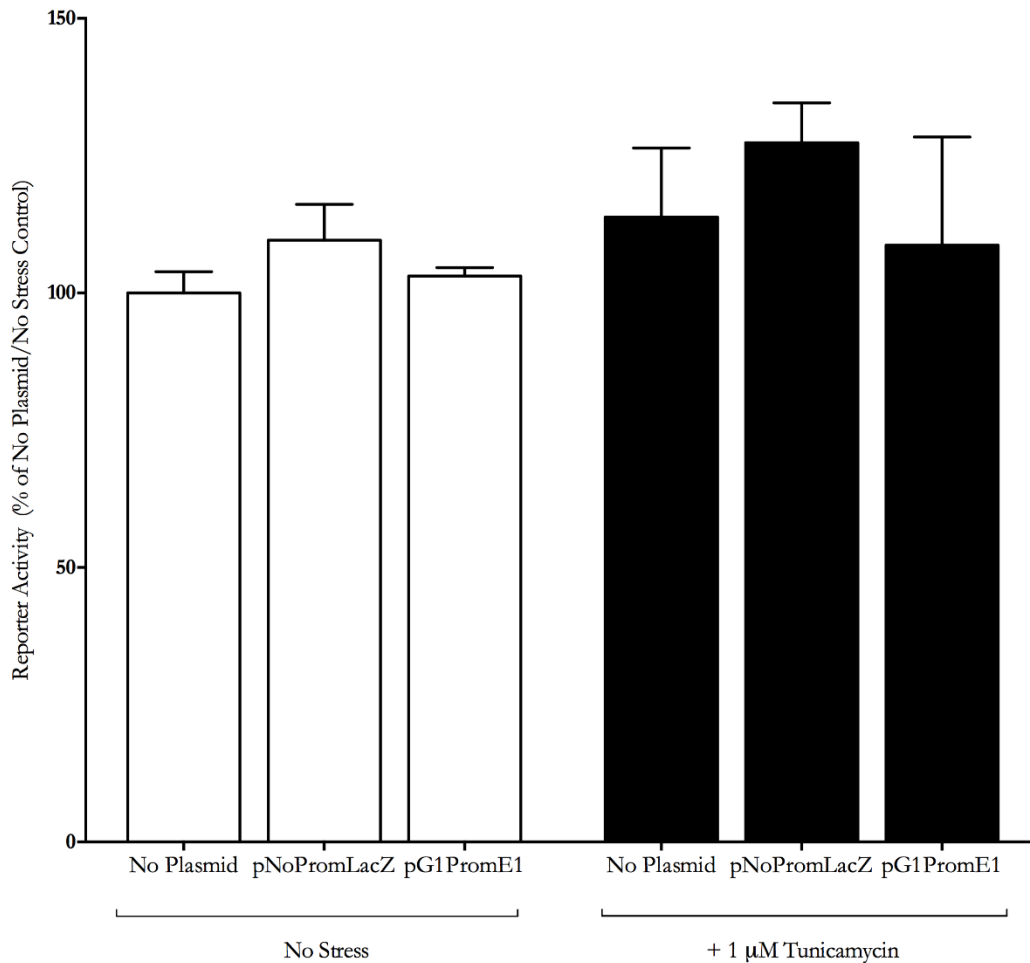
Figure 3.16: No activity is observed with construct pG1PromE1. Bars show the relative basal and stressed induced β-galactosidase activity from Caco-2 cells transfected with the pG1PromE1 that contains a sequence representative of the GlyT-1a promoter (-1159 bases upstream of the transcription start site and exon 1 only). In the absence of stress, cells transfected with the pG1PromE1 construct had comparable β-galactosidase activity to the untransfected cells this was not increased following the induction of stress with tunicamycin. No difference in β-galactosidase activity was observed between pG1PromE1 and the promoter-less 'empty' plasmid (pNoPromLacZ). Data are pooled from two experiments each of three experimental replicates (N=2, n=6) and shown here as percentage mean ± SEM relative to the average activity from the un-transfected unstressed control (No Plasmid). Reporter activity was determined by CPRG assay as described in the Section 2.5.4

The stochastic model for the arrangement of sites in a given CRM postulates that the occurrence of TFBS only increases the probability that the gene may be regulated by factors which bind to them (Walters *et al.*, 1995). This model also suggests there is no correlation between the binding of a TF to its binding site and the calibre (rate or intensity) of expression. The rheostatic response model on the other hand postulates that TFBS in a given CRM work in concert to produce an outcome with calibre that is dependent on TF occupancy of all the required sites. Indeed Sato and colleagues demonstrated that mutation of one of the two AAREs (which are separated by just nine bases) in the flanking sequence of the mouse xCT gene, resulted in a reduction in reporter activity from luciferase constructs (Sato *et al.*, 2004).

Motivated by the Cat-1 observations described above, and on the assumption that additional TFBS, possibly located in the UTR of the second exon of GlyT-1a might be required for optimal functioning of the predicted AARE in the GlyT-1a first exon, a second reporter construct (pG1Prom5U) was generated. As described in section 2.5.1, the insert cloned to pG1Prom5U was generated using ligated mediated PCR (LMPCR) to join two PCR products. The first product, containing a sequence spanning 1159 bases upstream of the GlyT-1a TSS plus the entirety of exon 1 (187 bases) was generated from genomic DNA (Figure 3.17). This was ligated to a second PCR product amplified from cDNA to include the GlyT-1a exon 1 and first 86 bases of exon 2 (Figure 3.18). The resulting 1431 base pair product was cloned upstream and in frame of the β-galactosidase coding sequence and the resulting pG1Prom5U validated by sequencing.

Transfecting Caco-2 cells with the pG1Prom5U reporter construct resulted in β-galactosidase activity that was significantly increased when cells were treated with 1μM tunicamycin for 16 hours when compared to the transfected unstressed control, but also to the activity measured when cells were transfected with the construct pG1PromE1 (Figure 3.19). Whilst it is impossible to ascertain from this experiment if the additional 86 bases of the UTR of exon 2 of GlyT-1a contain all the TFBS required for effective functioning of CRM10 these data suggest that an element or elements in this sequence are required for its activity. Taken together, this suggests that downstream of ER and UPR stress TF occupancy possibly at both the AARE of exon 1 and at least one other TFBS in the UTR of exon 2 are sufficient to enhance transcription of GlyT-1a.

Figure 3.17: Illustration showing the features of the pG1Prom5U reporter (lower) amplified from the GlyT-1a flanking sequence (upper). Primers 1 and 2 were used to amplify the promoter region from genomic DNA while primers 3 and 4 were used to amplify the UTR of exons 1 and 2 from cDNA. The resulting PCR products were joined by ligation mediated PCR to generate an insert which was cloned upstream of the β-galactosidase coding sequence (lacZ). Features are numbered relative to the TSS and the position of a putative amino acid response element (AARE) and electrophile or antioxidant response elements (EpRE/ARE) are given.

Figure 3.18: Gel image showing individual PCR products and the pG1Prom5U insert Band sizes were determined using the 1kB O'Gene Ruler ladder in lane 1. A 272 bp band (lanes 3-6) represents the uninterrupted UTR of the first and second exons amplified from GlyT-1a cDNA. Lanes 8, 9 and 10 show a 1346bp product representing the GlyT-1a flanking sequence amplified from genomic DNA. The two PCR products were ligated by LMPCR and cloned upstream of the β-galactosidase coding sequence. Lanes 12 and 14 show the resulting recombinant pG1Prom5U plasmid. Lanes 13 and 15 show the insert cut from the pG1Prom5U using endonucleases flanking each side of the cloned insert on the plasmid.

Figure 3.19: The full GlyT-1 5'UTR is required for stress induced reporter activity. Bars show the relative unstressed (open bars) and stressed (closed bars) induced activity of GlyT-1a promoter reporter constructs when transfected into Caco-2 cells. A reporter construct containing the GlyT-1a flanking promoter sequence (1159 bases upstream of the TSS) and the full 5'UTR (i.e. the whole of exon 1 and the first 86 bases of exon 2), but excluding intron 1 (pG1Prom5U) shows a significant increase in relative β-galactosidase activity in tunicamycin stressed Caco-2 cells; this when compared to the unstressed and un-transfected control samples but also when compared to a construct containing just the flanking promoter and exon 1 (pG1PromE1). Data are pooled from two experiments each with three experimental replicates (N=2, n=6) and shown here as percentage mean ± SEM relative to the average activity from the un-transfected and unstressed control (no plasmid). Reporter activity was measured by CPRG assay as described in Section 2.5.4. One-way ANOVA was used for multiple comparisons. *** indicates a P-value of < 0.005.

It is important to note that the sequence arrangement of the pG1Prom5U excludes the large 6.9kb intron which interrupts the first and second exon of GlyT-1a on the genome. Whilst introns have previously been considered to be part of 'junk' DNA their significance as important sites for gene regulation are recently being elucidated. From work by Li *et al.* (2012), it was concluded that the first introns of eukaryotic genes are generally larger, richer in CGIs and contain more basal regulatory elements (BREs) such as TATA boxes and CCAAT boxes than other intron locations. The functionality of additional AARE elements identified in the GlyT-1a first intron is assayed and discussed later (Section 3.4, page147). To determine if the arrangement of TFBS contained in the pG1Prom5U construct were sufficient for transcriptional induction of GlyT-1a following other stress types including oxidative stress and nutrient stress, Caco-2 cells were transfected with this reporter construct and subjected to DEM, tBHQ, and thapsigargin challenge and amino acid and or glycine starvation.

3.3.1 Glyt-1a Promoter activity in response to oxidative stress

The doubling of GlyT-1a mRNA abundance in Caco-2 cells following DEM treatment is evident from QPCR analysis of RNA extracted from stressed Caco-2 cells (Figure 3.1, page72). The pattern of changes in GlyT-1a mRNA expression following DEM treatment is consistent with that for xCT. A founding hypothesis for the work discussed here proposed that the transcription of GlyT-1a downstream of oxidative stress, like that for xCT is mediated by Nrf-2. Therefore we sought to determine if downstream of oxidative stress GlyT-1a mRNA expression is regulated via elements contained in the CRMs included in the reporter construct pG1Prom5U.

3.3.1.1 Regulation by Di-ethylene Maleate (DEM)

Treating Caco-2 cells transfected with the pG1Prom5U reporter construct with 0.2mM DEM for four hours resulted in a doubling of reporter activity when compared to the transfected and unstressed control, but also to cells transfected with the promoter-less construct pNoPromLacZ (Figure 3.20). DEM has been shown to effectively deplete glutathione levels resulting in heightened predisposition to oxidative stress (Mitchell *et al.*, 1983). As discussed elsewhere in this text, a knock on effect of increased oxidative stress is the expression of typically antioxidant transcription factors such as Nrf-2. Whilst the subsets of target genes for both Nrf-2 and Atf-4 may overlap, it is possible that as to whether these transcription factors result in activation or repression of the target depends

on the nature of the stressor. For instance, CHOP, a gene known to be regulated by amino acid availability via multiple AAREs in its promoter is known to be activated by Atf-4 but also potentially repressed by Nrf-2 activation (Zong *et al.*, 2012).

Nrf-2/Atf-4 dimers have been described in the literature (He *et al.*, 2001, Reinke *et al.*, 2013). As to whether the interaction of these BZIP dimers at distinct TFBS mediates activation or repression of the given gene transcript is arguable. That Nrf-2, upon activation, binds to and sequesters Atf-4 from interaction at functional AARE or other functional sites of a given CRM is a possible explanation to the 'competitive' repressive properties of Nrf-2. Downstream of oxidative stress, Nrf-2 is most frequently known to dimerise with small Maf proteins for the transcriptional activation of genes via ARE (Li *et al.*, 2008). However, He and colleagues, showed that hepatoma cells derived from dominant negative mouse mutants of either Nrf-2 or Atf-4, and transfected with a luciferase fusion construct of ARE containing sequence from the haem oxygenase (HO-1) gene, had reduced reporter activity following $CdCl_2$ treatment (He *et al.*, 2001). This suggests that there may be an equally important requirement for both Atf-4 and Nrf-2 in regulating HO-1 expression from this ARE. As to whether the two interact at the same ARE as a heterodimer cannot be determined from the above study. There is no published evidence that Nrf-2 and Atf-4 dimers interact at AARE.

Although the reporter activity observed from pG1Prom5U is significantly increased with DEM treatment, the magnitude of increase is less than that observed in tunicamycin treated cells. This is possibly due to the differing mode of actions of these stress inducing compounds. Given that DEM and not tunicamycin induces Nrf-2, could this be a specific effect of Nrf-2 at particular antioxidant response elements in the GlyT-1a sequence?
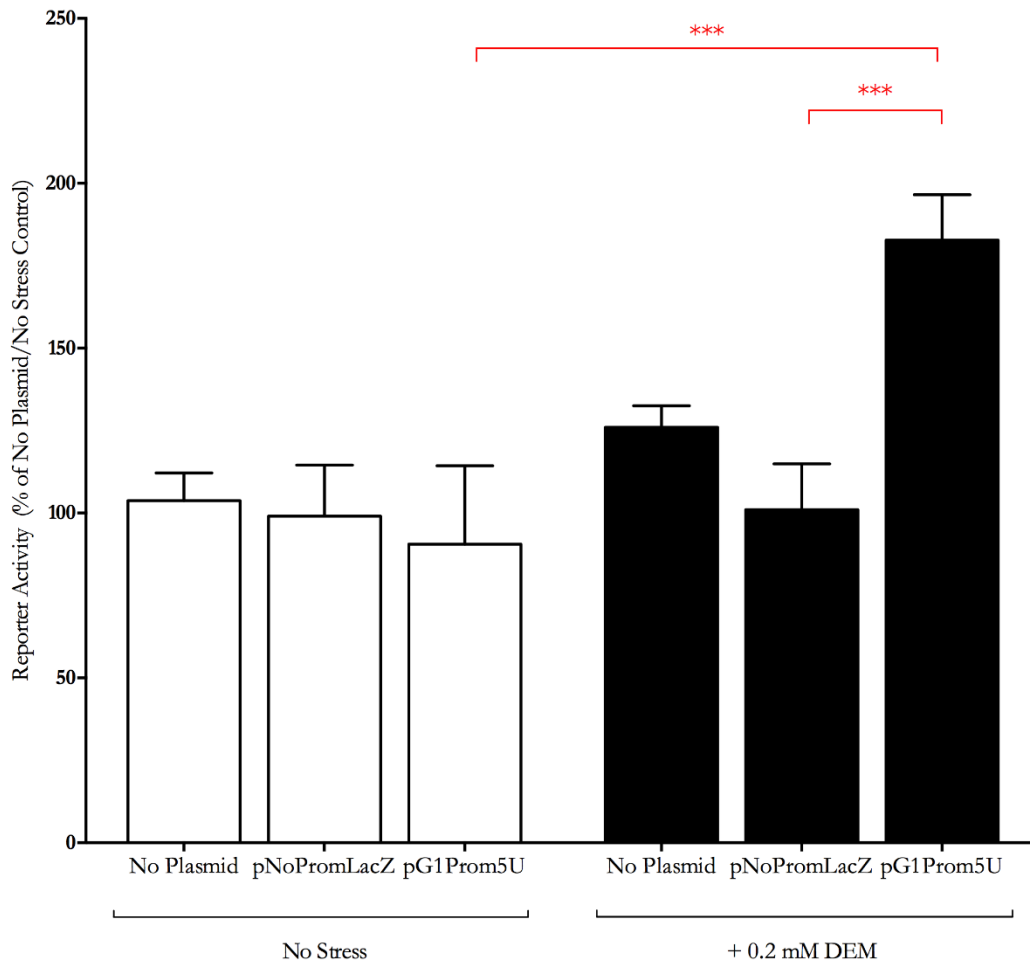
Figure 3.20: GlyT-1a promoter activity after DEM treatment in Caco-2 cells Confluent monolayers of cells were exposed to 0.2 mM DEM for 4 hours, 24 hours after transfection with the appropriate construct and assayed for β-galactosidase activity. Bars show the relative β-galactosidase activity in unstressed (open bars) and stressed (closed bars) Caco-2 cells. pG1Prom5U shows a relative β-galactosidase reporter activity in DEM stressed Caco-2 cells that is 55% higher than that observed in the stressed cells transfected with the promoter-less 'empty' plasmid (pNoPromLacZ). However when compared to activity from unstressed cells transfected with pG1Prom5U shows 100% more activity. Data are pooled from three experiments each with three experimental replicates (N=3, n=9) and shown here as percentage mean ± SEM relative to the average activity of the untransfected and unstressed control. One-way ANOVA was used for multiple comparisons. *** indicates a *P*-value of < 0.005.

3.3.1.2 Regulation by Polyphenols

Despite the controversy surrounding the efficacy or modes of action of dietary antioxidants such as vitamin C and carotenoids, epidemiological studies in populations consuming large amounts of plant derived polyphenols show a significantly lower prevalence of degenerative diseases (Arts and Hollman, 2005, Vauzour *et al.*, 2010). In recent years, dietary phenolic compounds have extensively been studied and generally accepted to play important roles in the prevention of cancers and other degenerative diseases such as Alzheimer's and cardiovascular disease (Vita, 2005, Choi *et al.*, 2012). By their direct and indirect regulation of TFs and consequential modulation of expression of cytokines, growth factors and nutrient transporter genes, polyphenols can have profound effects on several biological processes. It is believed that the antioxidant scavenging capacity of a phenolic compound is dependent on its molecular structure, particularly the positions of hydroxyl groups in relation to the other chemical moieties in its structure (Pandey and Rizvi, 2009, Kurek-Gorecka *et al.*, 2013, Malik and Mukherjee, 2014). This is particularly true for flavonoids, a ubiquitous group of polyphenols found in fruits and vegetables and often added to other food substances and drinks. Several thousand flavonoids (previously referred to as Vitamin P) have now been identified and are classified based on their structure as flavonols, flavonones, isoflavones, catechins, anthocyanidins and chalcones, which have varying anti-oxidative or pro-oxidative activity (Cos *et al.*, 1998, Nijveldt *et al.*, 2001, Heim *et al.*, 2002).

It has been shown that the treatment of Caco-2 cells with anti-oxidative flavonoids such as genistein, an isoflavone found in soy, attenuated H_2O_2-induced peroxide formation by boosting the expression of phase II detoxification enzymes such as haem oxygenase 1 (HO-1) and the glutamate-L-cysteine ligase catalytic subunit (GCLC) (Zhai *et al.*, 2013). Pre-treating Caco-2 cells with specific siRNA to Nrf-2 specifically blocked the effects of genistein. Zhai and colleagues proposed that genistein activates ERK and PKC which in turn facilitate the dissociation of Nrf-2 from the cytosolic KEAP-1 and encourages its nuclear translocation (Zhai *et al.*, 2013). In another study by Maher, pre-treatment of rat phaechromocytoma cells (PC12) with the isoflavone fisetin not only increased glutathione (GSH) levels but maintained GSH under conditions of oxidative stress via an ERK dependent pathway (Maher, 2006). Work following on from this showed that the effects of fisetin in maintaining GSH levels downstream of oxidative stress were reversed by transfecting the cells with siRNA to either Nrf-2 and/or Atf-4 (Ehren and Maher, 2013).

A conclusion from the latter study was that the concurrent regulation of both TFs (Atf-4 and Nrf-2) by fisetin is required for maintaining GSH levels. Like genistein, resveratrol, a polyphenol found in red wine has been shown to up-regulate the expression and activation of Nrf-2 by similar pathways downstream of oxidative stress (Cheng *et al.*, 2012).

In the work described here, a possible potentiating effect of genistein and other polyphenols on the transcription of GlyT-1a in Caco-2 cells was investigated. No change in reporter activity was observed in Caco-2 cells transfected with the reporter construct pG1Prom5U and treated with 10 μM genistein or 50 μM Resveratrol for 24 hours (Figure 3.21 and Figure 3.22). Given that the rationale in this case is that treatment with genistein or resveratrol may promote expression of GlyT-1a as part of an anti-oxidative response, this observation is in line with that by Maher and Ehren whose data showed an underlying oxidative state is required for the potentiating effect of the flavonoid fisetin on anti-oxidative genes, (Ehren and Maher, 2013, Maher and Hanneken, 2005).

To test if an underlying oxidative state may be required for flavonoid regulation of GlyT-1a expression, Caco-2 cells transfected with the pG1Prom5U construct were pre-treated with either resveratrol or genistein as above before the induction of oxidative stress. No differences in promoter activity were noticed following 4 hours of 0.2mM DEM treatment of transfected Caco-2 cells either with or without genistein (Figure 3.23) or resveratrol (Figure 3.24) pre-treatment.

Extracts from several seaweeds are an important source of anti-oxidative phenolic compounds (Chojnacka *et al.*, 2012). These extracts, often used as dietary supplements, have been shown to protect against oxidative stress in several cell models (Linares *et al.*, 2004, Raghavendran *et al.*, 2004, Matanjun *et al.*, 2008). They regulate the expression of detoxification enzymes via an Nrf-2 ARE dependent pathway (Wang *et al.*, 2013a). To determine if these may result in changes in expression of GlyT-1a, Caco-2 cells transfected with the pG1Prom5U reporter construct were subjected to a 30 minute pre-treatment with brown seaweed alcoholic extract prior to DEM treatment. Although it was impossible to determine the total phenolic content of the extract or what components may be involved, it had been shown that treating 3T3-L1 adipocytes in this way resulted in the attenuation of oxidative damage via an Nrf-2 dependent pathway (Lee *et al.*, 2011). The incubation of pG1Prom5U transfected Caco-2 cells, with seaweed alcoholic extract

for 30 minutes did not change promoter activity either without (Figure 3.25) or with oxidative stress induced by 4hours of 0.2mM DEM treatment (Figure 3.26).
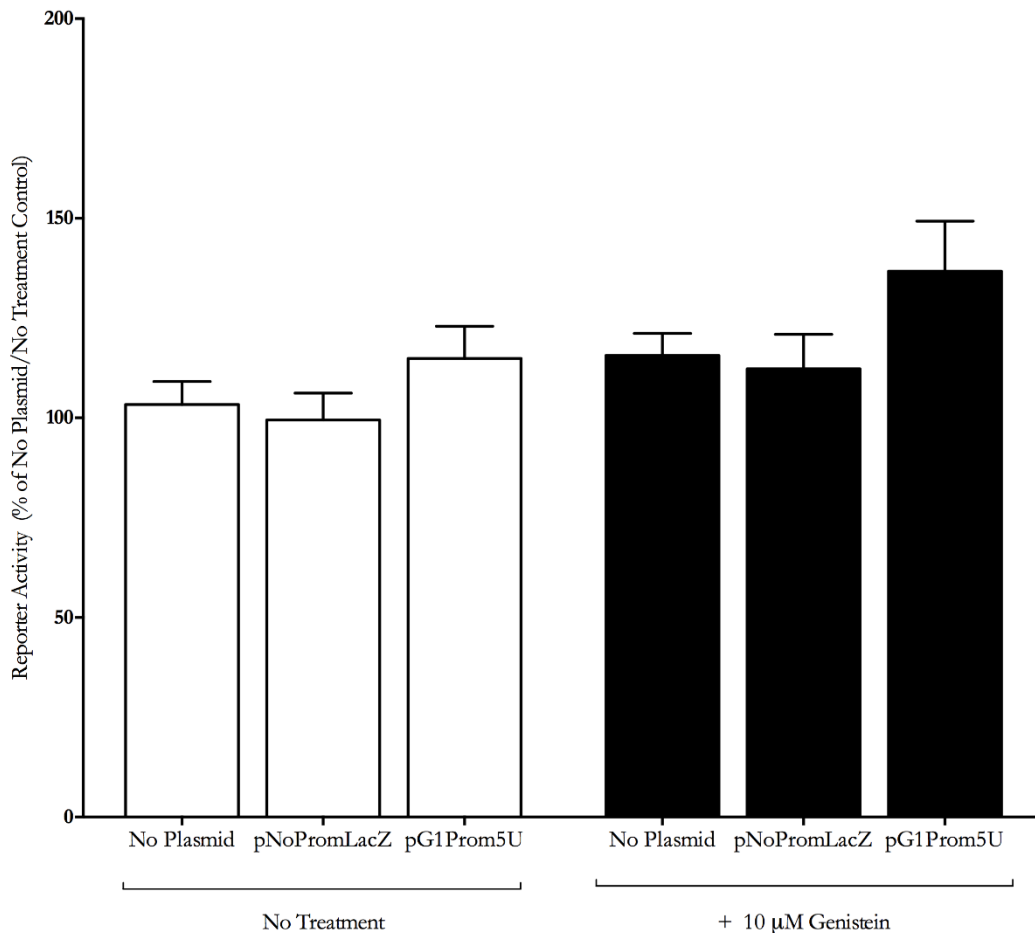
Figure 3.21: Genistein has no effect on GlyT-1a promoter activity. Genistein, a dietary polyphenol is known to potentiate the activity of the anti-oxidative factors Atf-4 and Nrf-2 as described in the text. Bars show the relative β-galactosidase activity measured in total protein extract from unstressed (open bars) and stressed (closed bars) Caco-2 cells by CPRG assay as described in the Section 2.5.4. Cells transfected with the construct pG1Prom5U showed no increase in β-galactosidase when compared to those transfected with pNoPromLacZ or untransfected cells. Stimulating the cells with genistein for 24 hours had no additional effect. Data are pooled from two experiments (N=2, n=6) and values represent the percentage mean ± SEM relative to the average activity of the untransfected and unstressed control (No Plasmid).
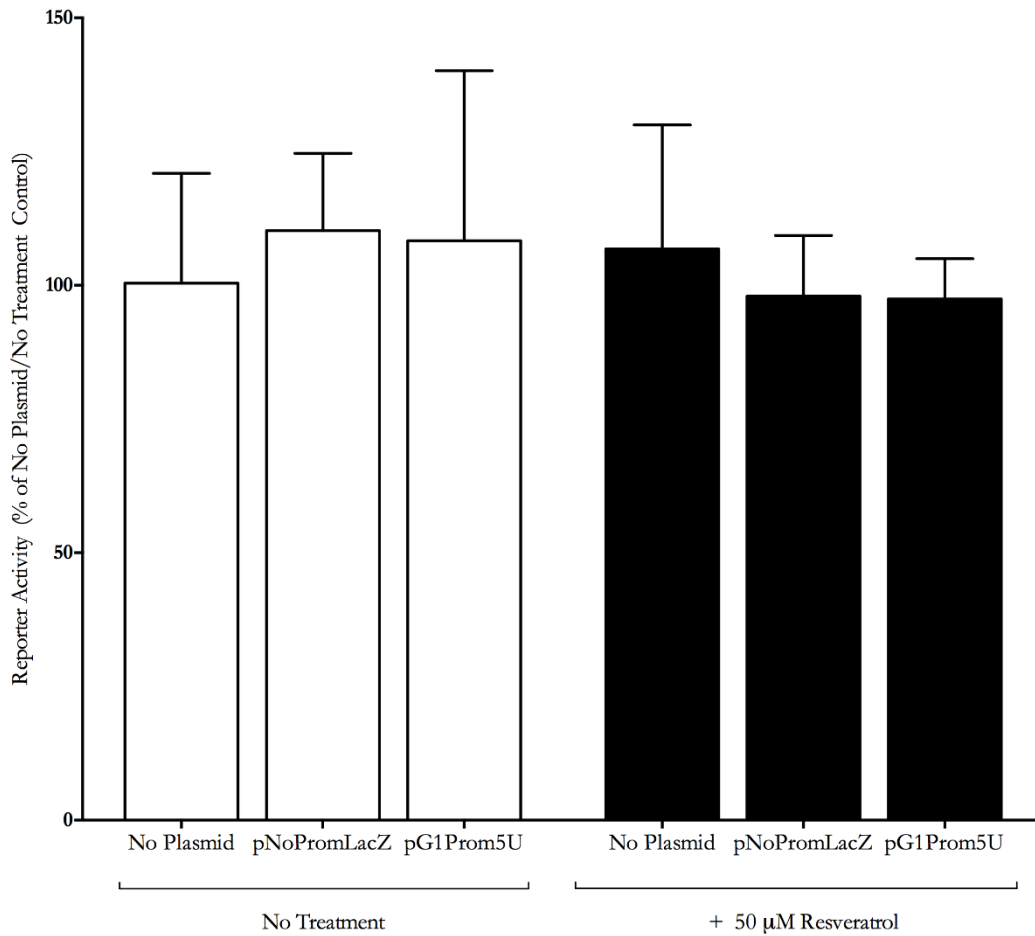
Figure 3.22: Resveratrol treatment has no effect on GlyT-1a promoter activity. Resveratrol, a dietary polyphenol found in red wine is known to potentiate the activity of the anti-oxidative transcription factors Atf-4 and Nrf-2 as described in the text. Bars show the relative β-galactosidase activity measured in total protein extract from untreated (open bars) and treated (closed bars) Caco-2 cells by CPRG assay as described in the Section 2.5.4. Cells transfected with the construct pG1Prom5U showed no increase in β-galactosidase when compared to those transfected with pNoPromLacZ or untransfected cells. Stimulating the cells with resveratrol for 24 hours had no additional effect. Data are pooled from two experiments (N=2, n=5-6) and values represent the percentage mean ± SEM relative to the average activity of the untransfected and unstressed control (No Plasmid).
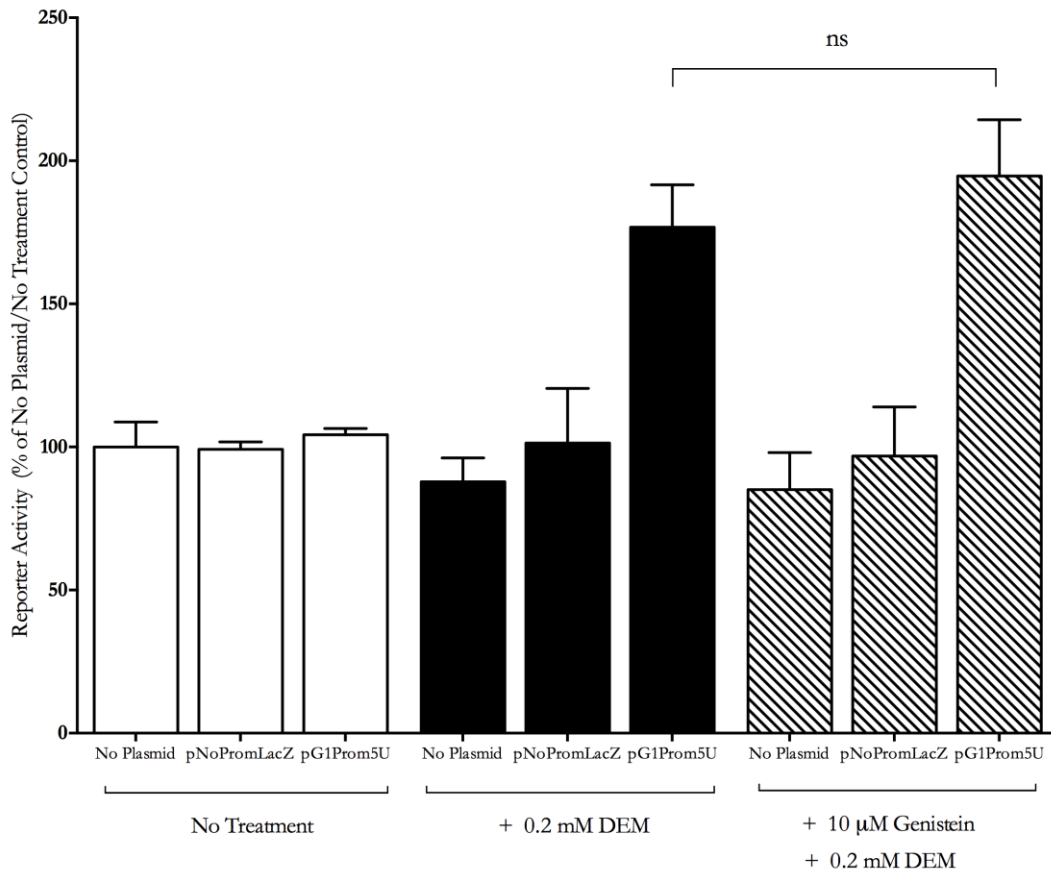
Figure 3.23: Genistein pre-treatment has no effect on DEM induced activity. pG1Prom5U, when transfected into Caco-2 cells results in β-galactosidase reporter activity that is not affected by pre-incubating the cells with genistein for 24 hours prior to 4h treatment with DEM. Bars show the relative β-galactosidase activity measured in total protein extract of Caco-2 cells relative to the untransfected and untreated control (open bars). DEM only treated samples are shown as closed bars, whilst samples pre-treated with genistein prior to DEM treatment are shown as hatched bars. Data are pooled from two experiments (N=2, n=6). Bar values are the percentage mean ± SEM for each sample group relative to the average activity of the untransfected and unstressed control (No plasmid). One-way ANOVA was used for comparisons.
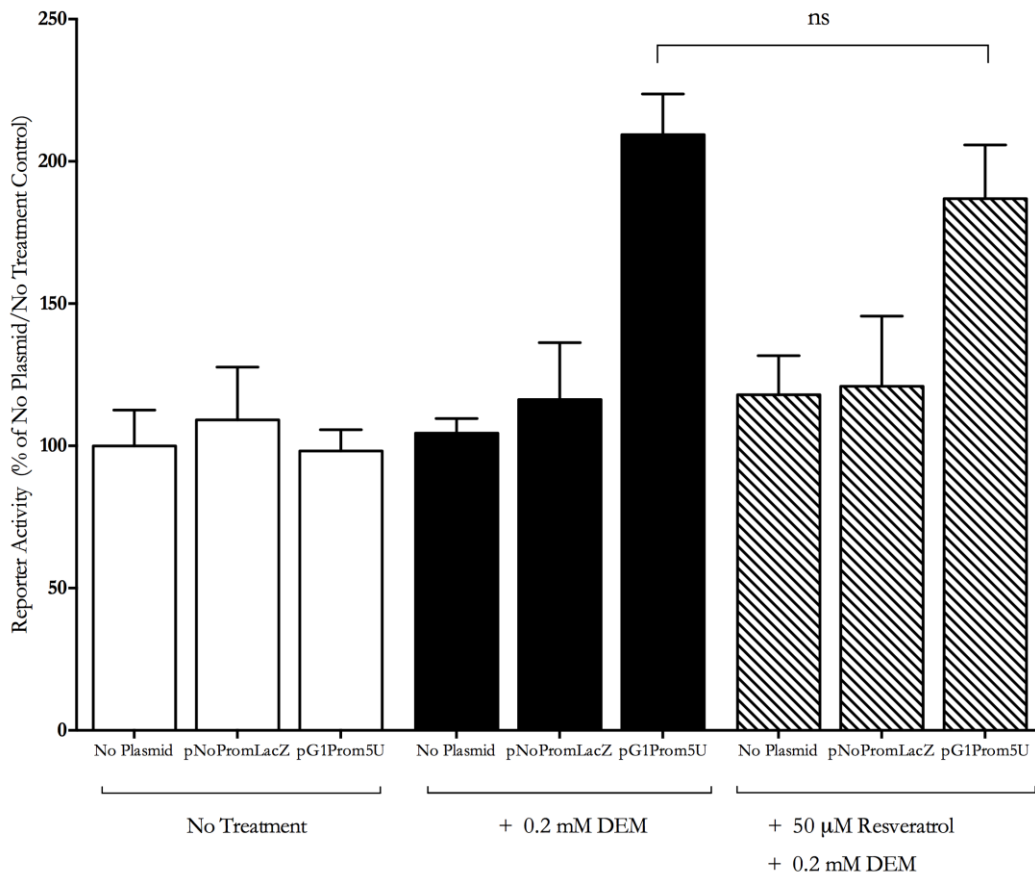
Figure 3.24: Resveratrol pre-treatment has no effect on DEM induced activity. β-galactosidase reporter activity from Caco-2 cells transfected with the construct pG1Prom5U was not affected by pre-incubating these cells with resveratrol for 24 hours prior to 4 hour treatment with DEM. Bars show the beta-galactosidase activity measured in total protein extract of Caco-2 cells relative to the untreated and untransfected control (no plasmid). Untreated samples are shown as open bars; DEM only treated samples are shown as closed bars, whilst samples pre-treated with resveratrol prior to DEM treatment are shown as hatched bars. Data are pooled from two experiments (N=2, n=6). Bar values are the percentage mean ± SEM for each sample group relative to the average activity of the un-transfected and unstressed control. One-way ANOVA was used for comparisons.
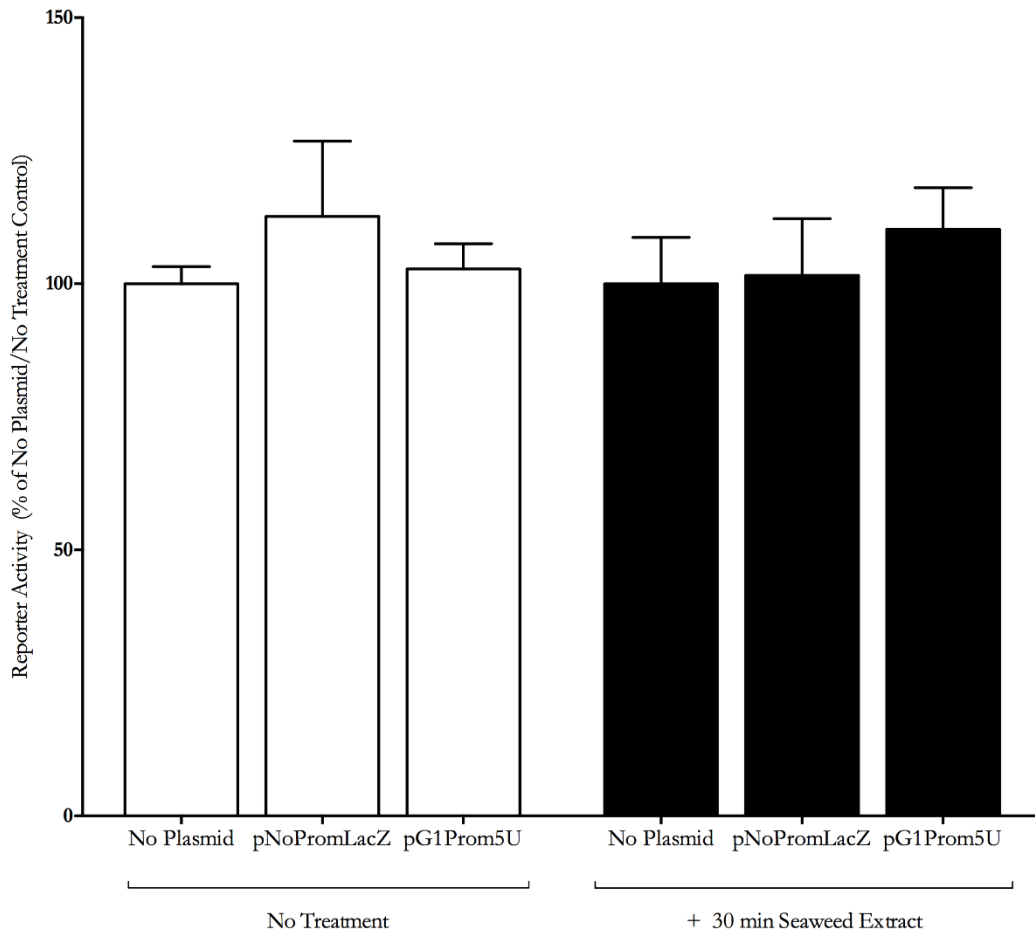
Figure 3.25: Seaweed extract has no effect on GlyT-1a promoter activity. Alcoholic brown seaweed extract containing un-quantified and un-determined polyphenols was added to the Caco-2 cell culture medium for 30 minutes. This had no effect on β-galactosidase activity from cells transfected with the construct pG1Prom5U. Bars show the β-galactosidase activity in total protein extract for untreated (open bars) and treated (closed bars) Caco-2 cells relative to the untransfected controls (No Plasmid). Bar values are the percentage mean ± SEM for each sample group relative to the average activity of the un-transfected and unstressed control (N=2, n=6).
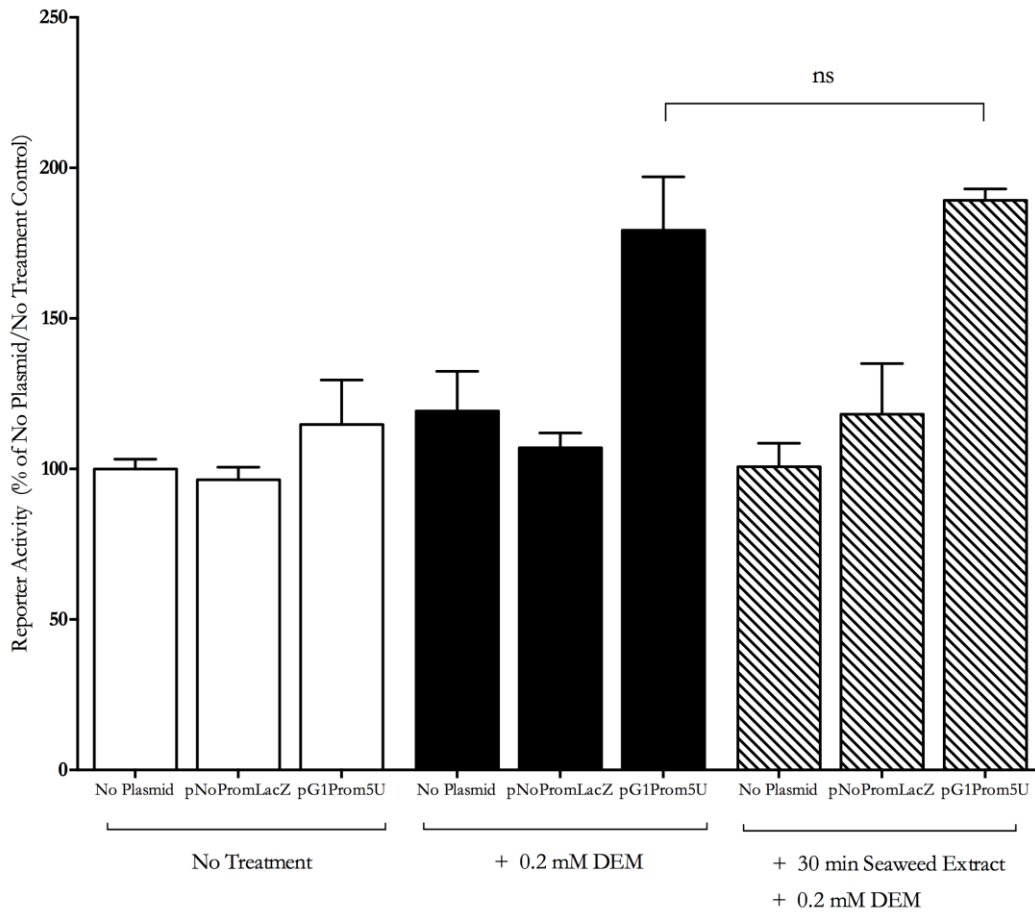
Figure 3.26: Seaweed extract pre-treatment has no effect on DEM induced activity. β-galactosidase reporter activity from Caco-2 cells transfected with the construct pG1Prom5U was not affected by pre-incubating these cells with alcoholic brown seaweed extracts for 30 minutes prior to 4h treatment with 0.2mM DEM. Bars show the β-galactosidase activity measured in total protein extract of Caco-2 cells relative to the unstressed and untransfected control (No plasmid). Untreated samples are shown as open bars; DEM only treated samples are shown as closed bars, whilst samples pre-treated with seaweed extract prior to DEM treatment are shown as hatched bars. Data are pooled from two experiments (N=2, n=6). Bar values are the percentage mean ± SEM for each sample group. One-way ANOVA was used for comparisons.

3.3.1.3 Regulation by tert-Butylhydroquinone (tBHQ)

2, 5-di-(tert-butyl)-1, 4-hydroquinone (tBHQ) is used as a preservative in several packaged foods. As it is often used in combination with the antioxidant butylated hydroxyanisole (BHA), which is itself metabolised to tBHQ, some have argued that tBHQ possesses some "antioxidant" properties. Imhoff *et al* demonstrated that in HeLa cells, the induction of Nrf-2 by tBHQ is a result of increased oxidation of the mitochondrial thioredoxin-2 (Trx-2) and unlike DEM, has no effect on glutathione (GSH) levels (Imhoff and Hansen, 2010).

It appears that expression of Nrf-2 target genes following tBHQ treatment is largely dependent on calcium fluxes (Cheung *et al.*, 2011). Pre-treating human platelets with a calcium chelating agent such as ethylene glycol tetra-acetic acid (EGTA) significantly attenuated tBHQ induced expression of HO-1 by Nrf-2 (Redondo *et al.*, 2008). Like thapsigargin, tBHQ is known to effectively inhibit endomembrane Ca^{2+}/ATPase and can alter the intracellular chemical balance by causing the rapid depletion of intracellular calcium stores, normally sensitive to inositol triphosphate (IP_3) (Mason *et al.*, 1991). Unlike thapsigargin, changes in intracellular calcium fluxes induced by tBHQ do not lead to the rapid intracellular influx of extracellular calcium. This is because tBHQ may also block L-type channel activity and calcium exchange at the cellular plasma membrane (Nelson *et al.*, 1994). Such L-type voltage activated channel activity typical in skeletal, smooth and cardiac muscles, as well as neuronal cells, has also been characterised in epithelial cells (Zhang and O'Neil, 1996). The relationship between intracellular and extracellular calcium fluxes and oxidative stress is still being elucidated. It is understood that the additional metabolic requirement for the ATP-dependent extrusion of the accumulated cytosolic calcium may significantly alter cellular biochemical properties leading to the formation of free radicals (Guzman *et al.*, 2010). The added oxidative phosphorylation required for the production of ATP may lead to a significant increase in the formation of superoxide and reactive oxygen species (ROS) by-products in the mitochondrion. Moreover, most of the calcium handling proteins such as calmodulin that contain disulphide linkages of cysteine residues are particularly susceptible to oxidative damage, further exacerbating the emergency.

With the above understanding of tBHQ mode of action, it is possible to navigate the ambiguity in existing literature. It appears that in studies where tBHQ is administered at

dosage low enough to illicit only mitochondrial oxidative stress and the induction of Nrf-2 (or other TFs), tBHQ is considered 'cytoprotective' or as an anti-oxidant. For instance, incubating the neuronal NT2N cell line with 10 μM tBHQ effectively protected the cells from H_2O_2 induced oxidative damage and the formation of amyloid proteins (Eftekharzadeh *et al.*, 2010). The same is true for human neural stem cells (hNSCs) where 20 μM tBHQ treatment for 24 hours protected against oxidative damage (Li *et al.*, 2005). On the contrary, doses of tBHQ in the millimolar range resulted in significant cell death (Okubo *et al.*, 2003). In this study, Caco-2 cells transfected with the pG1Prom5U reporter construct were treated with 60 μM tBHQ for 16 hours to determine if this may lead to regulation of the GlyT-1a promoter. As shown in (Figure 3.27), a ~3 fold increase in promoter activity was observed in Caco-2 cells transfected with the pG1Prom5U promoter and treated with tBHQ.

Reports of Atf-4 induction by tBHQ exist in the literature (Lewerenz and Maher, 2009, Li and Johnson, 2002). Western blot analysis of proteins from tBHQ treated urinary bladder T24 carcinoma cells, suggest an increase of both Atf-4 and Nrf-2 protein titres, which are required for the up regulation of xCT (Ye *et al.*, 2014). It is therefore possible that the effect seen here with tBHQ treatment of pG1Prom5U transfected Caco-2 cells is a result of action of Atf-4 activity at the AARE contained within this construct.

The effects of polyphenols such as genistein, or resveratrol on the promoter activity induced by 60 μM tBHQ were also studied. If the increases in promoter activity induced by tBHQ were via an Nrf-2 independent mechanism, it was expected that potentiating nuclear translocation of Nrf-2 with polyphenols may result in sequestration of Atf-4 and a reduction in the amounts available to regulate GlyT-1a via the AARE in exon 1. No statistically significant reduction in promoter reporter activity was observed in a single experiment with either genistein (Figure 3.28) or resveratrol (Figure 3.29). Genistein pre-treatment slightly increased tBHQ induced promoter activity. In contrast to the observations with genistein and resveratrol pre-treatment, Caco-2 cells transfected with the construct pG1Prom5U and treated with seaweed extracts for 30 minutes prior to stress resulted in a 25% reduction in tBHQ induced promoter activity (Figure 3.30).
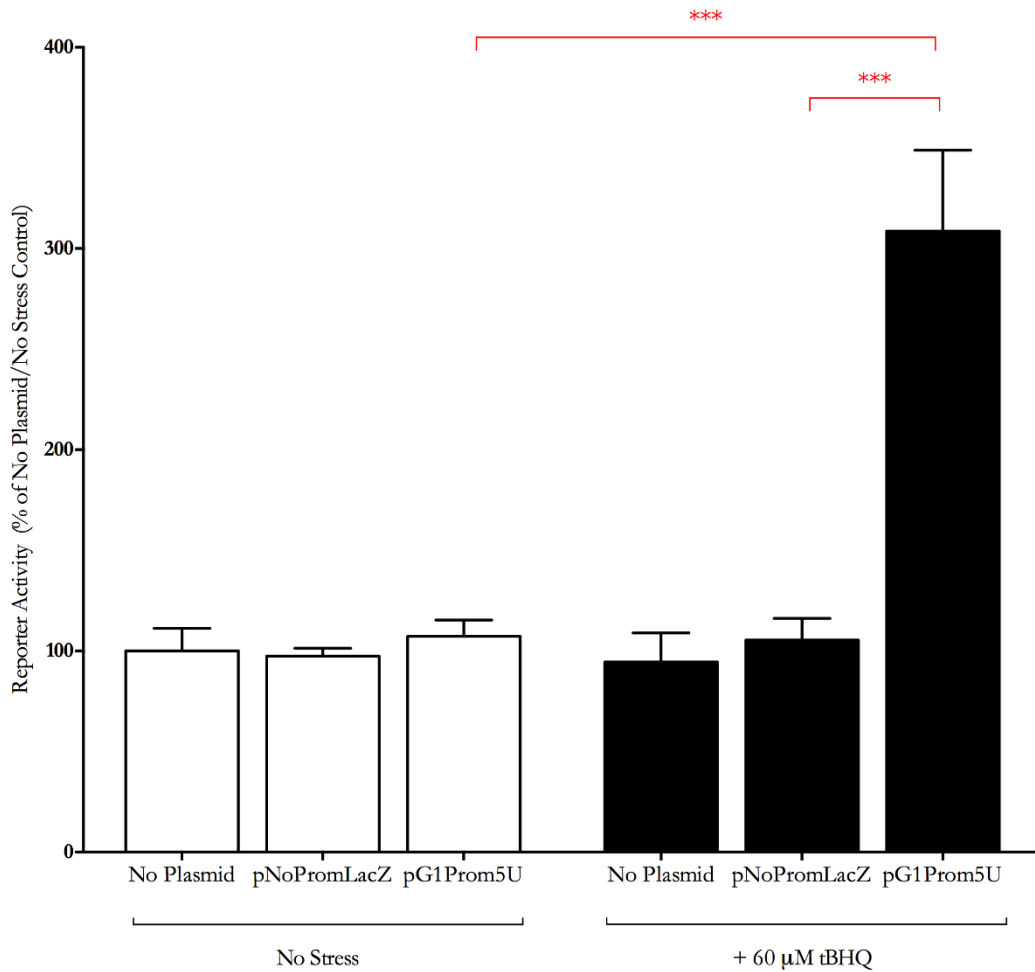
Figure 3.27: GlyT-1a promoter activity after tBHQ treatment of Caco-2 cells. Bars show the relative β-galactosidase activity in unstressed (open bars) and stressed (closed bars) Caco-2 cells. Transfecting cells with the construct pG1Prom5U followed by 16 hour treatment with 60μM tBHQ showed a relative β-galactosidase reporter activity 200% higher than that observed in cells transfected with the promoter-less 'empty' plasmid (pNoPromLacZ) and stressed with tBHQ. Data are pooled from three experiments each with three experimental replicates (N=3, n=9) and shown here as percentage mean ± SEM relative to the average activity of the un-transfected and unstressed control. One-way ANOVA was used for multiple comparisons. *** indicates a P-value of < 0.001.
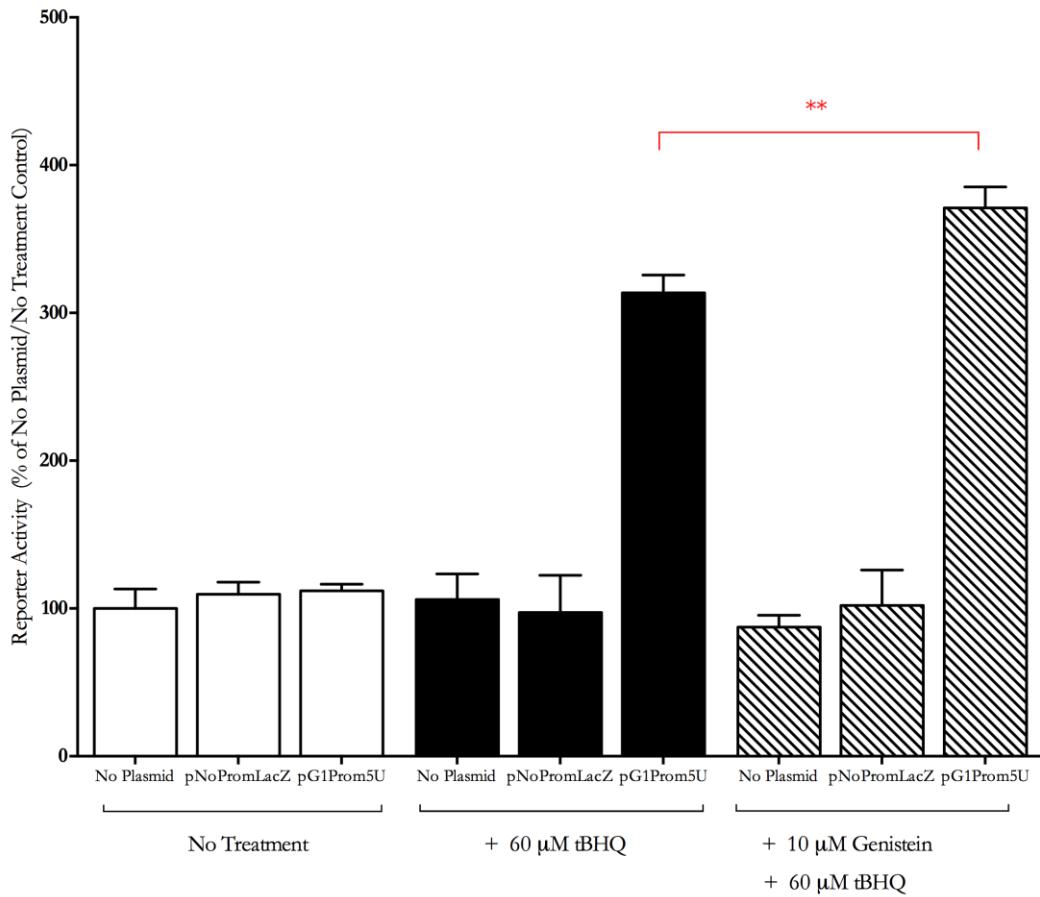
Figure 3.28: Genistein may potentiate tBHQ induced pG1Prom5U activity. In a single experiment, pG1Prom5U, when transfected into Caco-2 cells resulted in β-galactosidase activity that was greater when the cells were pre-incubated with genistein for 24 hours prior to a 16 hour incubation with tBHQ. Data are from a single experiment with three experimental replicates (N=1, n=3) and shown here as percentage mean ± SEM relative to the average activity of the un-transfected and unstressed control (No Plasmid). One-way ANOVA was used for multiple comparisons. ** indicates a *P*-value of < 0.05.
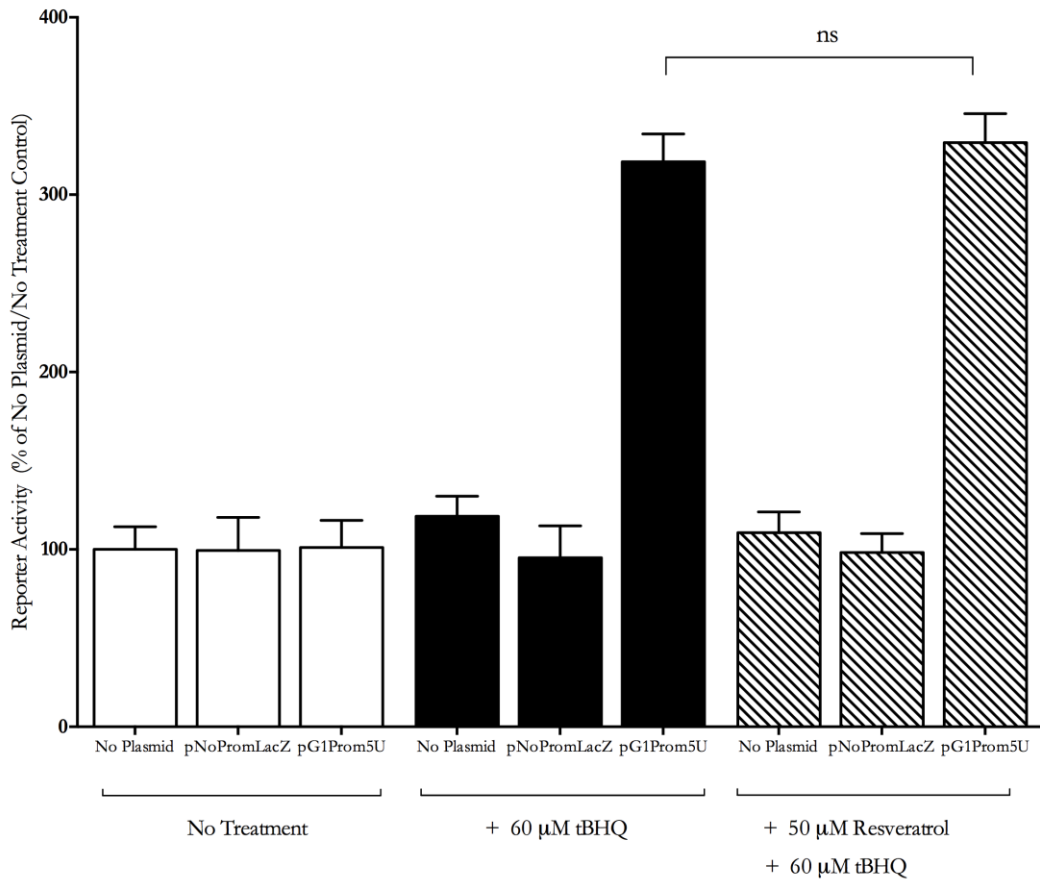
Figure 3.29: Resveratrol has no effect on tBHQ induced GlyT-1a promoter activity. The pG1Prom5U construct when transfected into Caco-2 cells resulted in β-galactosidase activity that was no different either with or without resveratrol treatment prior to incubation for 16 hours with tBHQ. Bars show the β-galactosidase activity measured in total protein extracts from transfected Caco-2 cells relative to the un-transfected and untreated control (No Plasmid). Untreated samples are shown as open bars; samples treated with tBHQ only are shown as closed bars, whilst samples pre-treated with resveratrol prior to tBHQ incubation are shown as hatched bars. Data are from a single experiment with three experimental replicates for each sample group (N=1, n=3). Bars represent the percentage mean ± SEM for each sample group relative to the average activity of the un-transfected and unstressed control. One-way ANOVA was used for multiple comparisons
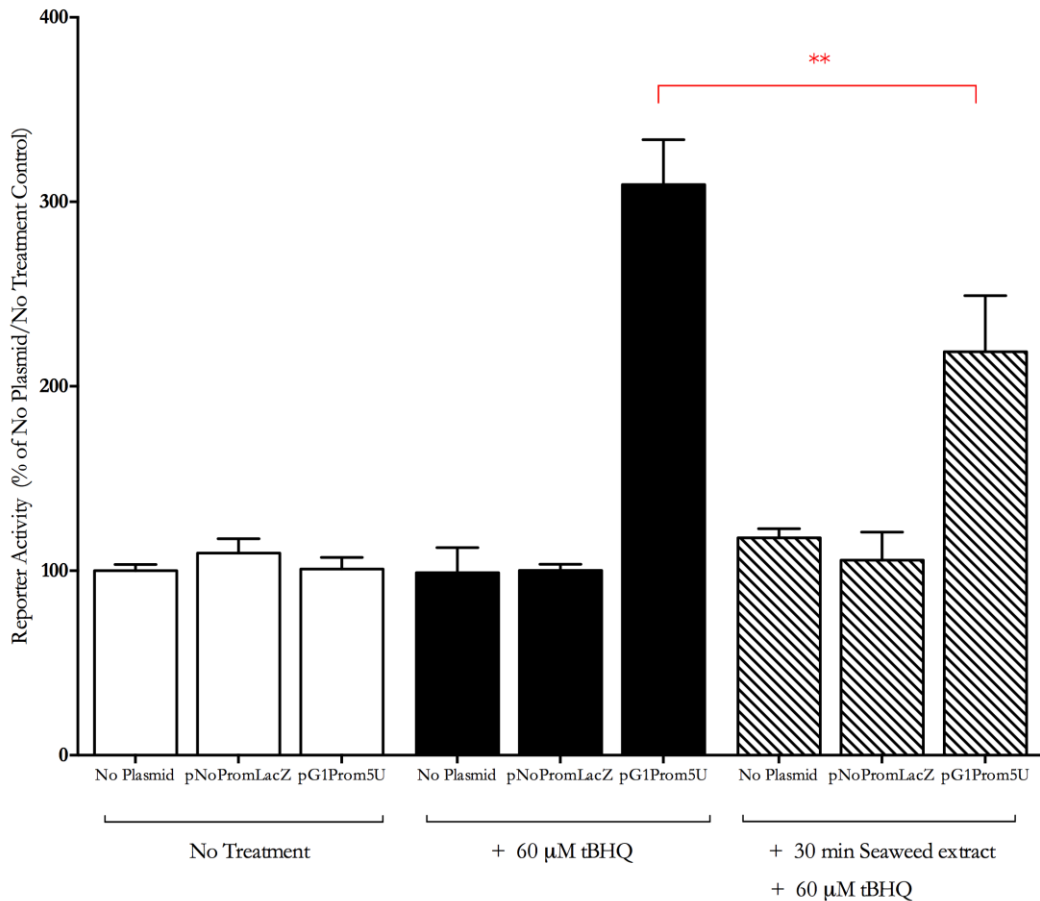
Figure 3.30: Seaweed extract pre-treatment reduced tBHQ induced promoter activity. Caco-2 cells transfected with the GlyT-1a promoter reporter construct pG1Prom5U resulted in β-galactosidase activity that was significantly lower in samples pre-treated with brown seaweed alcoholic extract for 30 minutes prior to tBHQ treatment when compared to samples not pre-treated with seaweed extract. Bars show the β-galactosidase activity measured in total protein extracts from transfected Caco-2 cells relative to the un-transfected and untreated control (No Plasmid). Untreated samples are shown as open bars; samples treated with tBHQ only are shown as closed bars, whilst samples pre-treated with seaweed extracts prior to tBHQ treatment are shown as hatched bars. Data are from a single experiment with three experiment replicates for each sample group (N=1, n=6). Bars represent the percentage mean ± SEM for each sample group relative to the average activity of the un-transfected and unstressed control. One-way ANOVA was used for multiple comparisons. ** indicates a P-value of < 0.05.

To understand this result, closer consideration was given to the phenolic content of alcoholic extracts from the brown algae *Fucus vesiculosus* used here. Characterisation of these seaweeds in the lab of Prof Jeff Pearson (ICAMB, Newcastle University) suggested that their total polyphenol content, steadily released by an *in vitro* digestion method, demonstrated significant dose dependent antioxidant capability as determined by the ferrous ion-chelating assay (data not shown). Whilst flavonoids have thus far not been found in any brown algae, the phenolic content of *Fucus vesiculosus* is composed predominantly of phlorotannins (Wang *et al.*, 2012), which are fucol-type tannin with a heavily hydroxylated and polymerised phenolic structure. The treatment of lung fibroblast V79-4 cell lines with a phlorotannin compound increased the expression and activity of Nrf-2 and subsequently HO-1 enzyme in a dose dependent manner (Kim *et al.*, 2010). In the latter study, the authors also demonstrated that phlorotannins lead to the direct activation of Nrf-2 following the phosphorylation of ERK and Akt. Phlorotannins from *Fucus vesiculosus* extracted using 70% acetone have been shown to prevent the accumulation of advanced protein glycated end (AGEs) products by scavenging reactive carbonyl species (Liu and Gu, 2012). AGEs are non-enzymatically glycated and oxidized proteins formed from the reaction between reducing sugars and the amine residues on proteins, lipids and nucleic acids. AGEs are implicated in the pathogenesis of several degenerative diseases including Alzheimer's, ischemic heart disease and inflammatory bowel disease (Ciccocioppo *et al.*, 2013, Takeuchi *et al.*, 2004, Sasaki *et al.*, 1998). The accumulation of endogenous AGEs in the ER, during hyperglycaemic states or oxidative or hypoxic stress induces ER stress and injury and lead to an ER stress (Adamopoulos *et al.*, 2014, Yamabe *et al.*, 2013). AGE induced ER stress is very dependent on an underlying oxidative state and in prolonged stress is believed to result in the activation of apoptotic mechanisms via MAPK/ERK dependent pathways (Alikhani *et al.*, 2007).

3.3.2 GlyT-1a Promoter activity in response to the ER Stress

As the ER is the major site for protein folding within cells, perturbations in environmental conditions conducive to protein folding have profound effects on cell viability. The accumulation of unfolded or misfolded proteins within the ER leads to ER stress and activates the unfolded protein response (UPR) as discussed in Section 1.7.2. Downstream of the UPR are genes that are ultimately required for effective protein folding, trafficking and degradation. The relationship between ER and oxidative stress is extensively researched. Current understanding suggests unresolved ER stress or prolonged UPR often leads to further accumulation of ROS through oxidative protein folding by eukaryotic protein disulphide isomerase (PDI; EC 5.3.4.1) an ER resident enzyme (Tu and Weissman, 2004, Simmen *et al.*, 2010).
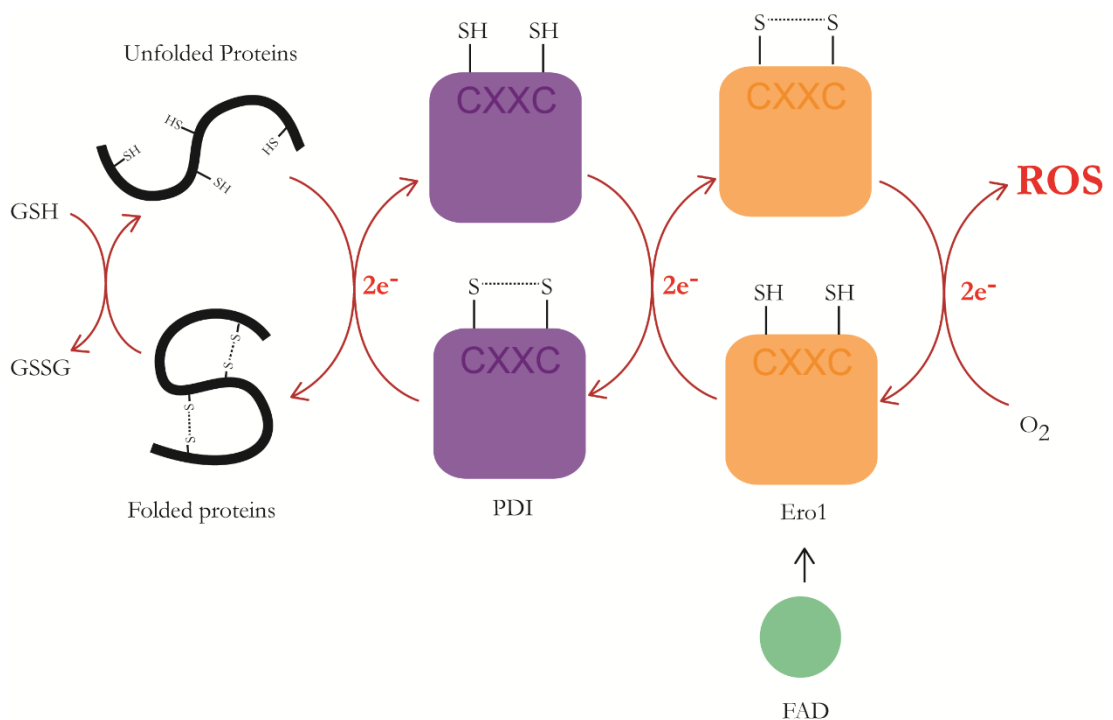
Figure 3.31: The relationship between ER stress and oxidative stress. Prolonged ER stress caused by the accumulation of unfolded proteins in the ER leads to the onset of oxidative stress via the shuttling of electrons between protein disulphide isomerase (PDI), the cysteine residues of unfolded proteins or peptides and the flavoenzyme ER oxidoreductin 1 (Ero-1), which are important controls for protein folding in the ER. See main text (Section 3.3.2 above) for description.

The relay of electrons between Protein disulphide isomerase (PDI), the cysteine residues of unfolded proteins or peptides and the ER oxidoreductin 1 (Ero-1; EC 1.8.3.2) flavoenzyme mediates oxidative protein folding in the ER. Reduced PDI mediates the formation of disulphide linkages between intramolecular cysteine residues of unfolded proteins (Wang and Tsou, 1993). In its oxidized form, PDI mediates the rearrangement of mispaired disulphide bridges. PDI is able to perform both oxidation and reduction by virtue of two distinct redox active thioreductin domains. Free electrons resulting from such oxidative folding by PDI are relayed to Ero-1 (Figure 3.31). Together with the cofactor Flavin adenine dinucleotide (FAD), Ero-1 transfers the relayed electron from PDI to molecular oxygen resulting in the formation of superoxide species. With time, unfolded protein accumulation leads to the build-up of ROS within the ER. ROS accumulation in the ER may signal to key antioxidant pathways or apoptotic pathways resulting in controlled cell death.

3.3.2.1 Regulation by Tunicamycin

One of the hypothesis for this project posited that for the supply of glutathione to scavenge ROS, glycine supply via the specific transporter GlyT-1 is required. To test if prolonged ER stress may lead to the transcriptional induction of GlyT-1, Caco-2 cells transfected with the GlyT-1a promoter reporter construct pG1Prom5U were treated with 1 μM tunicamycin for 16 hours. As shown in Figure 3.32, in stressed cells promoter activity was significantly increased by approximate fivefold when compared to the promoter-less vector control. When compared to the transfected unstressed control, the difference in activity was ~8 fold. This observation is consistent with the change in GlyT-1a mRNA levels downstream of tunicamycin treatment in these cells.
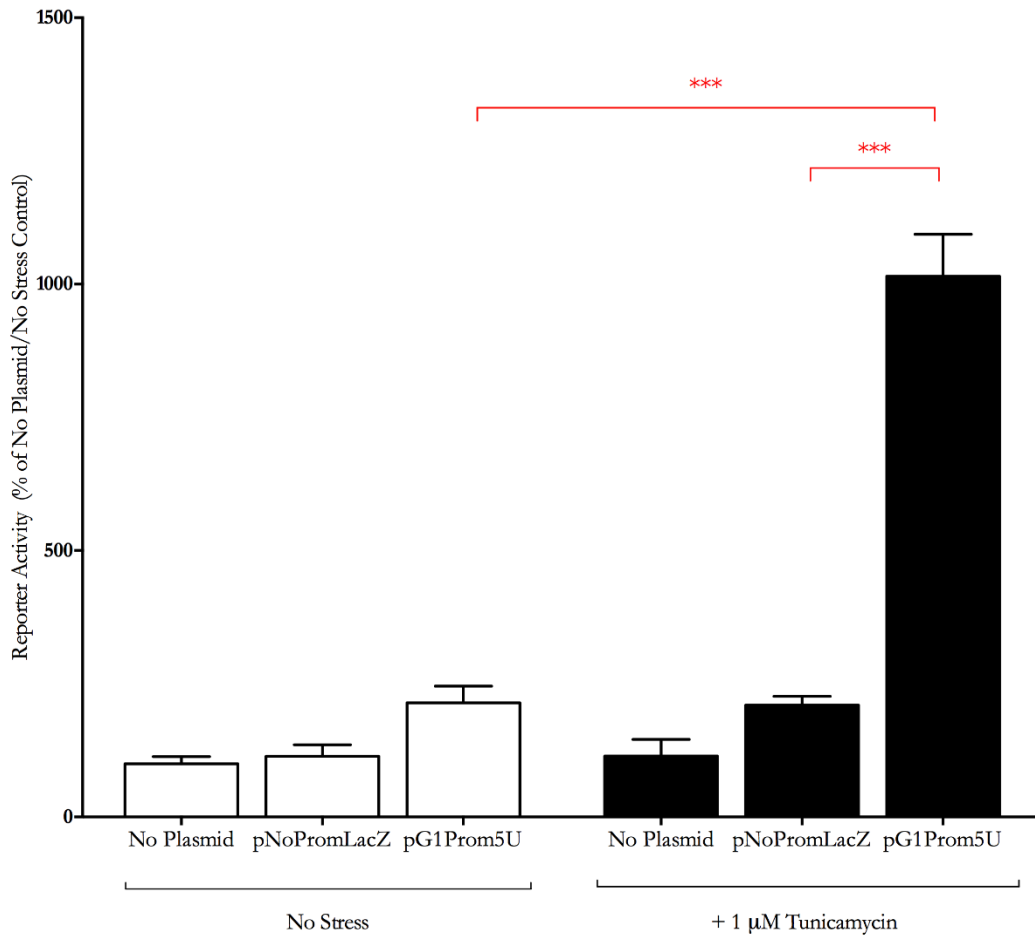
Figure 3.32: GlyT-1a promoter activity is increased by tunicamycin treatment. Bars show the relative β-galactosidase activity in unstressed (open bars) and stressed (closed bars) Caco-2 cells. Transfecting cells with the construct pG1Prom5U followed by 16 hour treatment with 1 μM tunicamycin treatment shows a relative β-galactosidase reporter activity that is ~5 fold higher than that observed in cells transfected with the promoter-less 'empty' plasmid (pNoPromLacZ) and stressed with tunicamycin. When compared to the transfected and unstressed control the difference in pG1Prom5U activity is ~8 fold. Data are pooled from multiple experiments (N=6, n=2-3 per experiment) and shown here as percentage mean ± SEM relative to the average activity of the un-transfected and unstressed control (No Plasmid). One-way ANOVA was used for multiple comparisons. *** indicates a *P*-value of < 0.001.

3.3.2.2 Regulation by Thapsigargin

ER stress caused by the accumulation of unfolded proteins may lead to the production of ROS and the formation of ER-localised peroxides and/or superoxide. ROS mediated inactivation of the sarco/endoplasmic reticulum Ca^{2+}/ATPase (or SERCA pumps) by cysteine S-glutathionylation and activation of the inositol triphosphate receptor (IP_3R) leads to the rapid flooding of the cytosol with calcium (Lock *et al.*, 2011). Such increases in intracellular calcium lead to its rapid uptake to the mitochondrial matrix via mitochondrial calcium uniporters (MCUs) and H^+/Ca^{2+} exchanges on the ER mitochondrial associated membranes (MAMs). The consequences of elevated calcium in the mitochondrial matrix are diverse and generally end in more oxidative stress from impaired function and the activation of cell death.

Thapsigargin selectively blocks the SERCA pumps leading to calcium overload in both the cytosol and mitochondria as described above. In addition to increased ROS accumulation, thapsigargin treatment also induces a characteristic ER stress response (Samali *et al.*, 2010). To determine if the induction of ER stress via alternative means (other than *N*-glycosylation induction by tunicamycin) resulted in similar effects on GlyT-1a promoter activity, Caco-2 cells were transfected with the pG1Prom5U construct and subject to 50μM thapsigargin challenge for 16 hours. As shown in Figure 3.33, Glyt-1 promoter activity was ~3 times greater following thapsigargin treatment and consistent with findings of GlyT-1 mRNA expression downstream of thapsigargin treatment (see Figure 3.1, page72).

That both tunicamycin and thapsigargin resulted in significant transcriptional induction of the GlyT-1a promoter suggest that the up-regulation of the glycine transporter downstream of the ER stress is part of a targeted response. Interestingly, the fold change in activity is consistent with that observed with tBHQ. Although both thapsigargin and tBHQ result in increased cytosolic calcium and ER stress, there are significant differences in their downstream effects owing to the differing extent at which they deplete intracellular calcium stores (Scamps *et al.*, 2000).
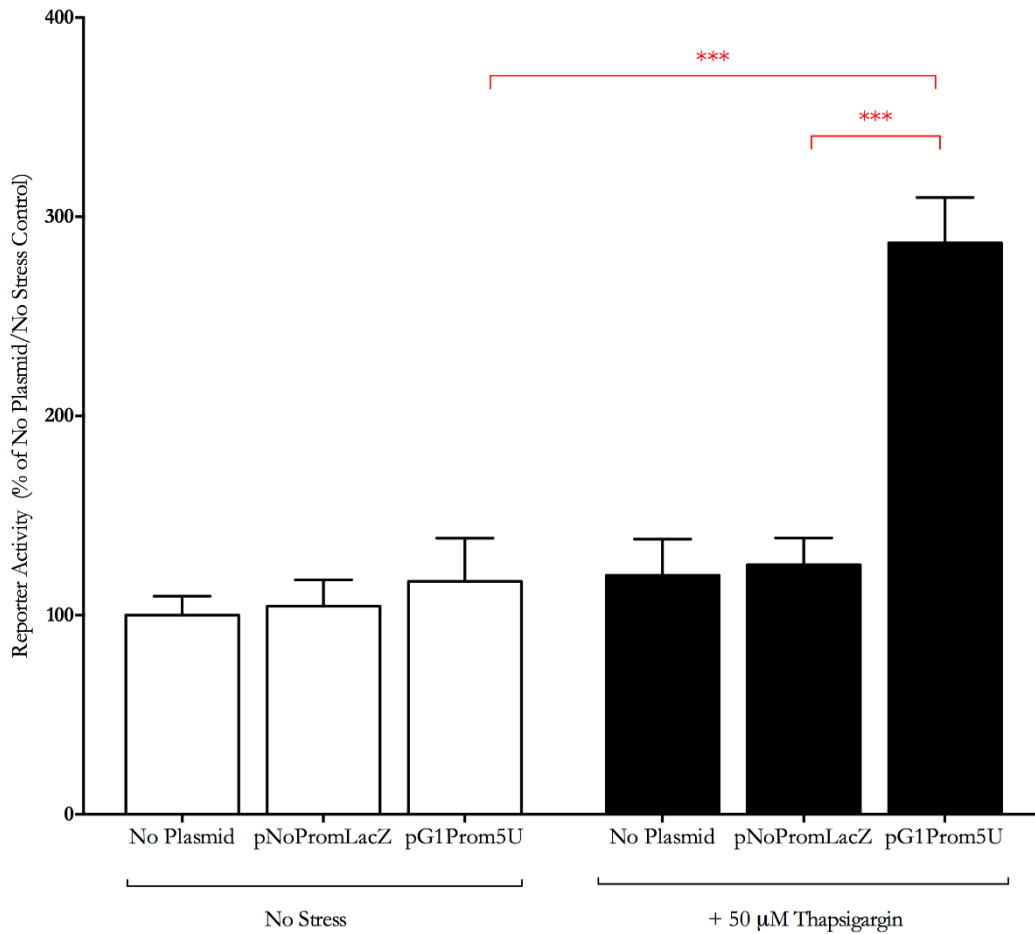
Figure 3.33: GlyT-1a promoter activity is increased by thapsigargin treatment. Bars show the relative β-galactosidase activity in unstressed (open bars) and stressed (closed bars) Caco-2 cells. Transfecting cells with the construct pG1Prom5U followed by 16 hour treatment with 50μM thapsigargin shows a relative β-galactosidase reporter activity that is ~2.5 fold higher than that observed in cells transfected with the promoter-less 'empty' plasmid (pNoPromLacZ) and stressed with tunicamycin. Data are pooled from multiple experiments (N=3, n=9) and shown here as percentage mean ± SEM relative to the average activity of the un-transfected and unstressed control (No Plasmid). One-way ANOVA was used for multiple comparisons. *** indicates a P-value of < 0.001.

3.3.3 GlyT-1a Promoter activity in response to nutrient availability

In previous work carried out in this laboratory, removing amino acids from culture medium had a time dependent effect on the expression of GlyT-1 mRNA in both Caco-2 and HCT-8 cells (Data not included). Amino acid starvation may have a cascading effect on several cellular functions ultimately leading to the induction of core stress regulatory pathways. To determine if amino acid availability affected Glyt-1a promoter reporter activity, Caco-2 cells were transfected with the pG1Prom5U and subject to states of either complete amino acid starvation or amino acid supplementation.

3.3.3.1 Regulation by amino acid starvation

Whilst no significant changes in promoter activity were observed in Caco-2 after 4 hours amino acid deprivation (Figure 3.34), there was a doubling of promoter activity after 8 hours (Figure 3.35) when compared to the control. Such a timed program in the expression of human genes downstream of nutrient availability has been reported previously (Kilberg *et al.*, 2009, Lopez *et al.*, 2007); this is believed to be consequential to the temporal changes in regulatory transcription factors (Thiaville *et al.*, 2008a). There appears to be a temporal correlation between the titres of Atf-4, Atf-3-FL (the longest protein isoform of Atf-3, believed to be a repressor of transcription) and C/EBP-β and the expression of the human asparagine synthetase gene (ASNS) downstream of nutrient limitation (Chen *et al.*, 2004). In that study, Chen and colleagues noted a peak of Atf-4 protein (a 20-fold increase) in total cell extract after 2 hours of histidine limitation which coincided with the point at which the protein isoforms of ASNS, Atf-3-FL and the LAP isoform of C/EBP-β are first detected by Western blots. With time, as the protein titres of Atf-3 and C/EBP-β, and the mRNA abundance of ASNS steadily rose to a peak at between 8 and 12 hours after the onset of amino acid limitation there was a corresponding decrease in the expression of ASNS as well the protein titre of Atf-4. The authors concluded that a regulatory module involving the interaction of Atf-4 at nutrient stress response elements (NSRE) in the ASNS promoter leads to its expression shortly after nutrient stress. With time, as the stress burden is reduced interactions at the same NSRE by Atf-3-FL or dimers of Atf-3/C/EBP-β function as a negative feedback control loop. It is also possible that in cases of prolonged stress, a similar mechanism is used to switch the ensuing response from pro-survival to pro-apoptosis.

Figure 3.34: Effect of amino acid starvation on GlyT-1a promoter activity. Bars show the relative β-galactosidase activity in unstressed (open bars) and stressed (closed bars) Caco-2 cells. Transfecting cells with the construct pG1Prom5U followed by 4 hours of amino acid starvation resulted in β-galactosidase activity that was not different from that seen in cells transfected with the promoter-less 'empty' plasmid (pNoPromLacZ) and subject to amino acid starvation in a similar way. Data are pooled from two experiments (N=2, n=6) and shown here as percentage mean ± SEM relative to the average activity of the un-transfected and unstressed control (No Plasmid). One-way ANOVA was used for multiple comparisons.

Figure 3.35: Effect of prolonged amino acid starvation on GlyT-1a promoter activity. Bars show the relative β-galactosidase activity in unstressed (open bars) and stressed (closed bars) Caco-2 cells. Transfecting cells with the construct pG1Prom5U followed by 8 hours of amino acid starvation resulted in β-galactosidase activity that was higher than that seen in cells transfected with the promoter-less 'empty' plasmid (pNoPromLacZ) and subject to amino acid starvation in a similar way. This significant increase in activity was also apparent when comparing unstressed cells transfected with the pG1Prom5U construct. Data are pooled from two experiments (N=2, n=6) and shown here as percentage mean ± SEM relative to the average activity of the un-transfected and unstressed control (No Plasmid). One-way ANOVA was used for multiple comparisons. *** indicates a P-value of < 0.001.

A similar temporal program in expression from the arginine/lysine transporter (Cat-1) was observed by Lopez and colleagues (Fernandez *et al.*, 2003, Lopez *et al.*, 2007). Atf-4 protein titres peaked 1 hour after the removal of cysteine, methionine and glutamine from the culture medium of C6 glioma cells. The subsequent increase in both Atf-3 and C/EBP-β protein corresponded to a steady decrease in Cat-1 mRNA expression. As proposed for GlyT-1, Cat-1 has an AARE in the first exon of the gene to which Atf-4 has been shown to bind (Fernandez *et al.*, 2003). There is therefore a possibility for the temporal regulation of GlyT-1 by similar TFs.

3.3.3.2 Regulation by glycine availability

GlyT-1a promoter activity in response to glycine starvation or supplementation was determined. No significant difference in promoter activity between test and control samples was observed after incubating transfected Caco-2 cells in glycine free culture medium for four hours (data not included). Eight hours after the removal of glycine from culture medium a 75% increase in promoter activity from pG1Prom5U was observed (Figure 3.36). Although the increase in promoter activity following glycine starvation is four times less than that observed when all amino acids are removed from the culture medium the resulting activity at the GlyT-1a promoter is evidence for a specific response to glycine shortages and not just part of an overall response to the added requirement for amino acids. Certainly, intracellular shortages in amino acids may have profound effects on the synthesis of crucial proteins required by the cell. However, it is likely that regulation of GlyT-1a by TF modules involving Atf-4 interaction at the AARE is not just a random consequence of the underlying stresses for two main reasons. Firstly, an important consequence of stress is the overall reduction in protein synthesis via phosphorylation of the eukaryotic eIF-2a implying an overall decrease in amino acids required for protein synthesis. Secondly, that some amino acids like glycine have added functions beyond being mere components for protein synthesis points to the requirement for adaptive regulation.
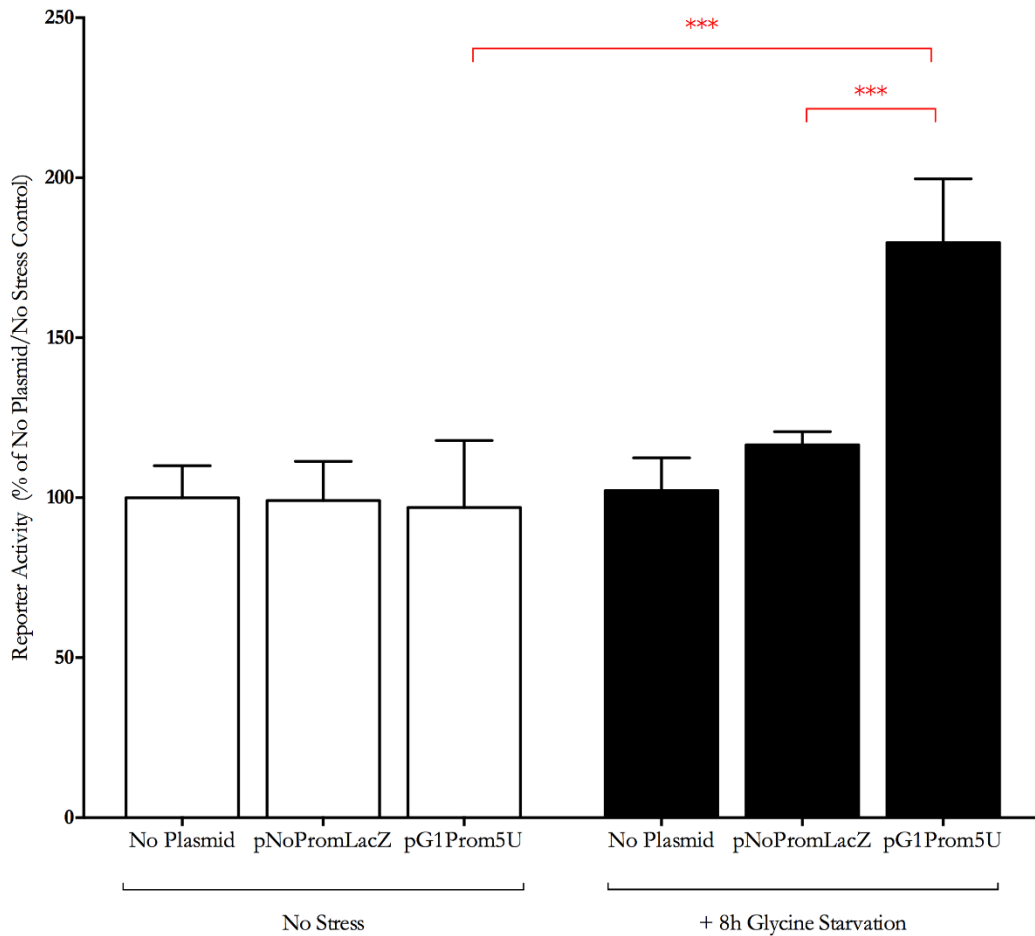
Figure 3.36: Effect of glycine starvation on pG1Prom5U. Bars shown here represent the relative unstressed (open bars) and glycine starvation (closed bars) induced activity of GlyT-1a promoter reporter construct pG1Prom5U when transfected into Caco-2 cells. Compared to the transfected unstressed control, 8hours of amino acid starvation resulted in a 75% increase in reporter activity from the pG1Prom5U construct. When compared to activity from cells transfected with the pNoPromLacZ constructed and starved of Glycine, the pG1Prom5U generates 50% more activity. Data shown as percentage mean $\pm$ SEM relative to the average activity of the un-transfected and unstressed control (N=2, n=6). One-way ANOVA was used for comparisons. *** represents P-values of less than 0.001

In all human cells, signalling to the mammalian target of rapamycin (mTOR) is an important regulatory pathway downstream of nutrient availability. The mTOR protein is a highly conserved serine-threonine kinase present in two distinct complexes, mTOR-C1 and mTOR-C2 which have very differing functions. Whilst mTOR-C2 signals to the structural dynamics and survival of the cell via the activation of protein kinase B (PKB, also known as Akt), mTOR-C1 predominantly regulates cell growth and proliferation. mTOR-C1 is essential for the phosphorylation and activation of the ribosomal protein S6 kinase (S6K) which in turn regulates protein translation via eIF-3. mTOR-C1 also directly controls mRNA translation by its ability to phosphorylate eIF-4E. It is unclear how mTOR-C1 senses amino acid sufficiency. From current understanding, in states of increased amino acid availability, the physical interaction between mTOR-C1 with the Ras related GTPases (or Rag) tethers it to the lysosome. At the lysosome mTOR-C1 is activated by the Ras homolog enriched in the brain (or Rheb) G protein. The observation that growth factors had no effect on mTOR-C1 signalling in the absence of amino acids suggested the requirement for an upstream gated mechanism to control mTOR-C1 interaction with Rag and Rheb (Jewell *et al.*, 2013). Bar-Peled *et al* suggest two GTPase activators of mTOR-C1 (GATOR 1 and 2) which specifically block the interaction between mTOR-C1 and Rag downstream of amino acid starvation (Bar-Peled *et al.*, 2013). Likewise the tuberous sclerosis protein complexes TSC 1 and 2, inactive in states of amino acid sufficiency, block mTOR-C1 interaction with Rheb in amino acid starvation (Carracedo *et al.*, 2008).

There is extensive evidence for the activation of Atf-4 following amino acid starvation (some of which are discussed earlier). Although mTOR-C1 activation is known to regulate protein translation via phosphorylation of various eIF proteins, there is no evidence to suggest that mTOR-C1 is responsible for Atf-4 activation following amino acid starvation. Double knockout TSC mutants of mouse embryonic fibroblast (MEF) cells were associated with increased mTOR-C1 and decreased Atf-4, Atf-6 and CHOP despite the extensive eIF-2α phosphorylation (Kang *et al.*, 2011). No promoter activity was observed either after four or eight hour incubation of Caco-2 cells transfected with pG1Prom5U in glycine supplemented culture medium (Figure 3.37). Atf-4 activation itself is believed to block mTOR by up regulating TSC via the Regulated in development and DNA damage response 1 gene (REDD-1 or DDIT-4) – an Atf-4 target gene (Ait Ghezala *et al.*, 2012).
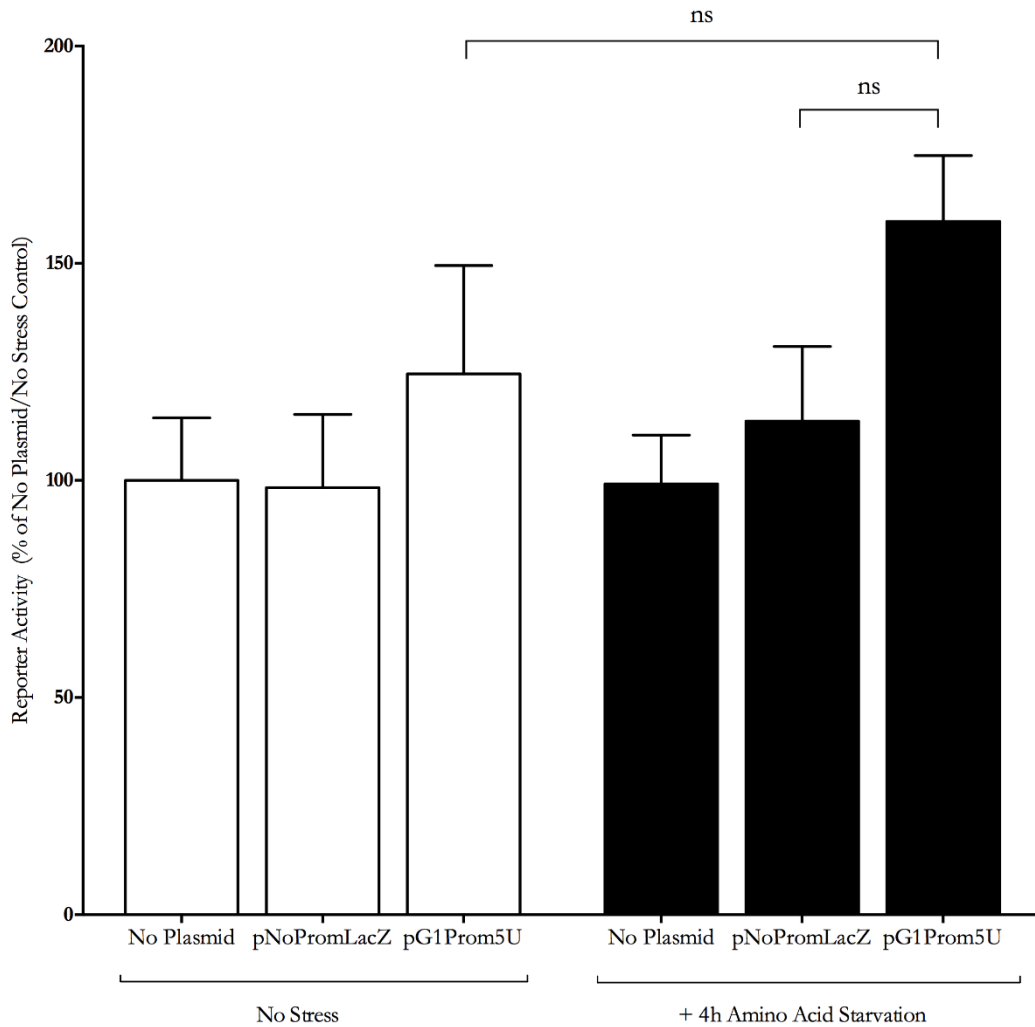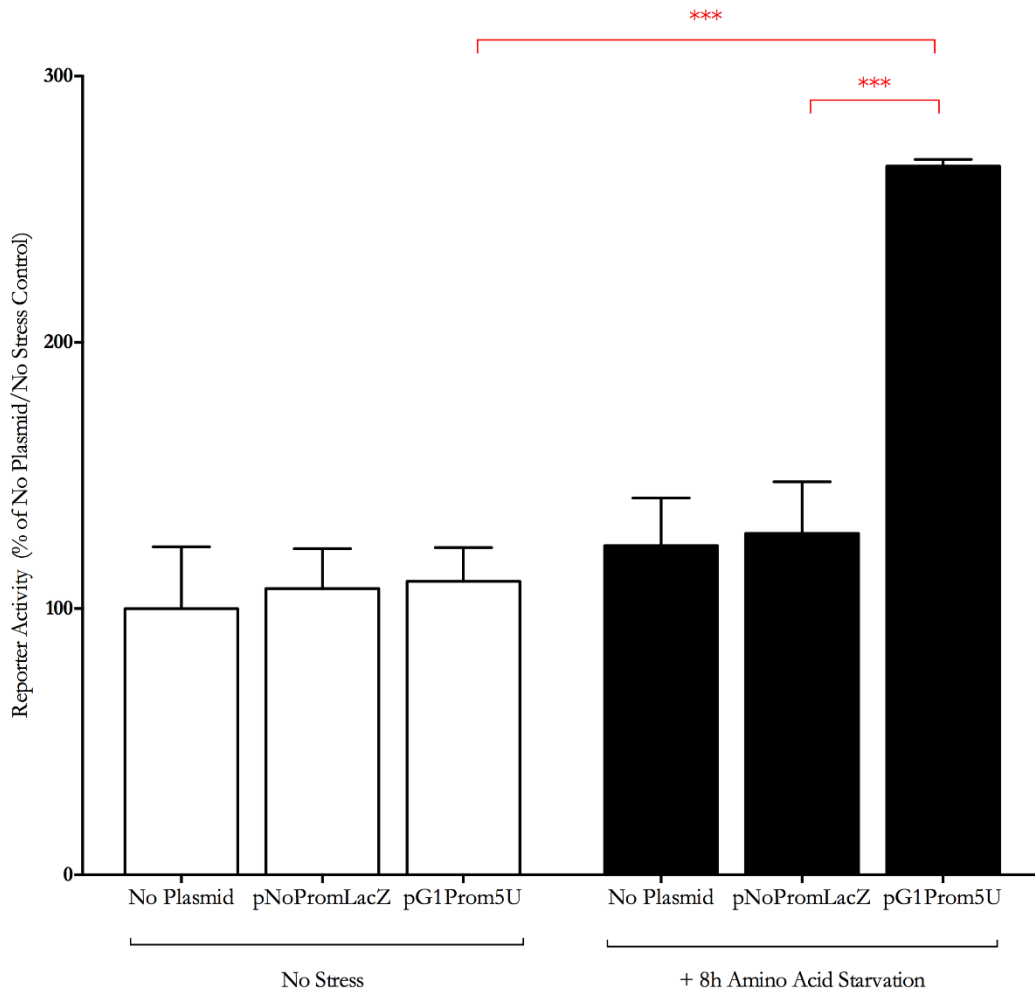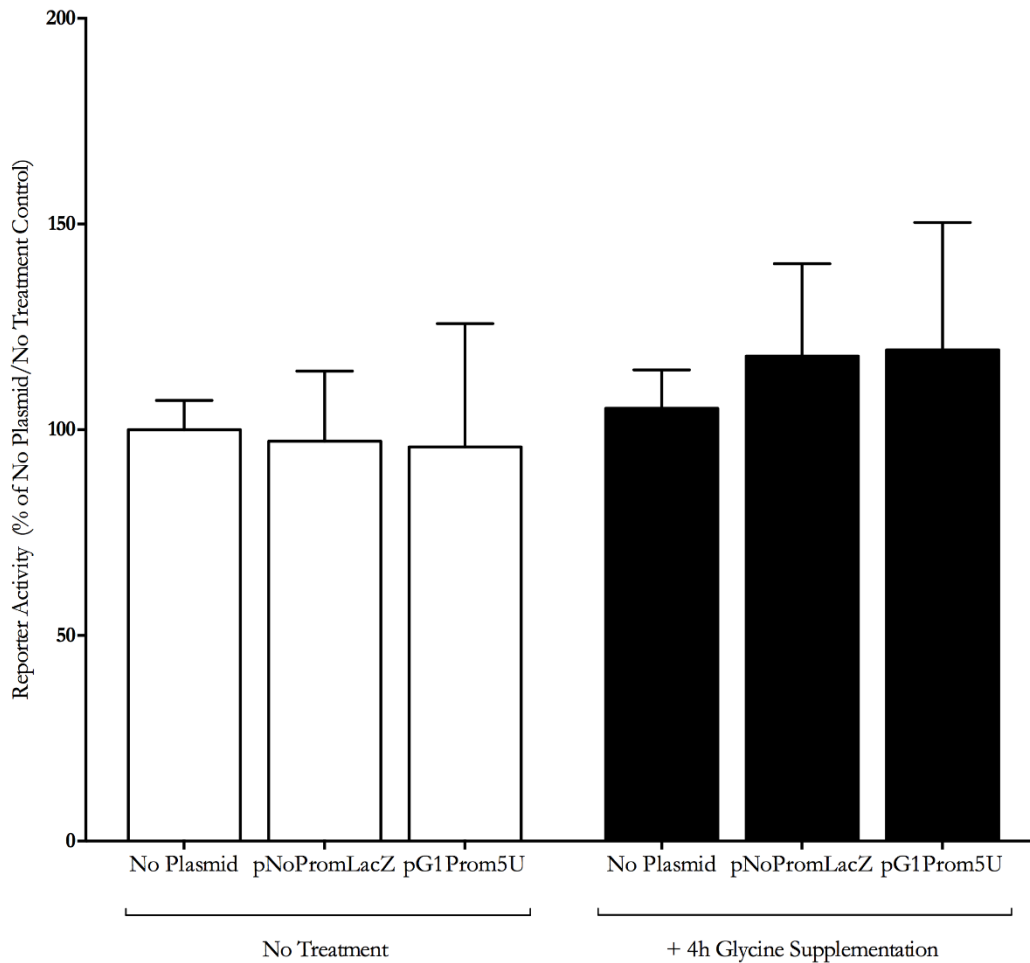
Figure 3.37: Effect of glycine supplementation on GlyT-1a promoter activity. Bars show the relative β-galactosidase activity in unstressed (open bars) and stressed (closed bars) Caco-2 cells. Transfecting cells with the construct pG1Prom5U followed by incubation for 4 hours in glycine supplemented medium resulted in β-galactosidase activity that was not different from that seen in cells transfected with the promoter-less 'empty' plasmid (pNoPromLacZ) and incubated in glycine supplemented medium. Data are pooled from three experiments (N=3, n=9) and shown as percentage mean ± SEM relative to the average activity of the un-transfected and unstressed control (No Plasmid). One-way ANOVA was used for multiple comparisons.
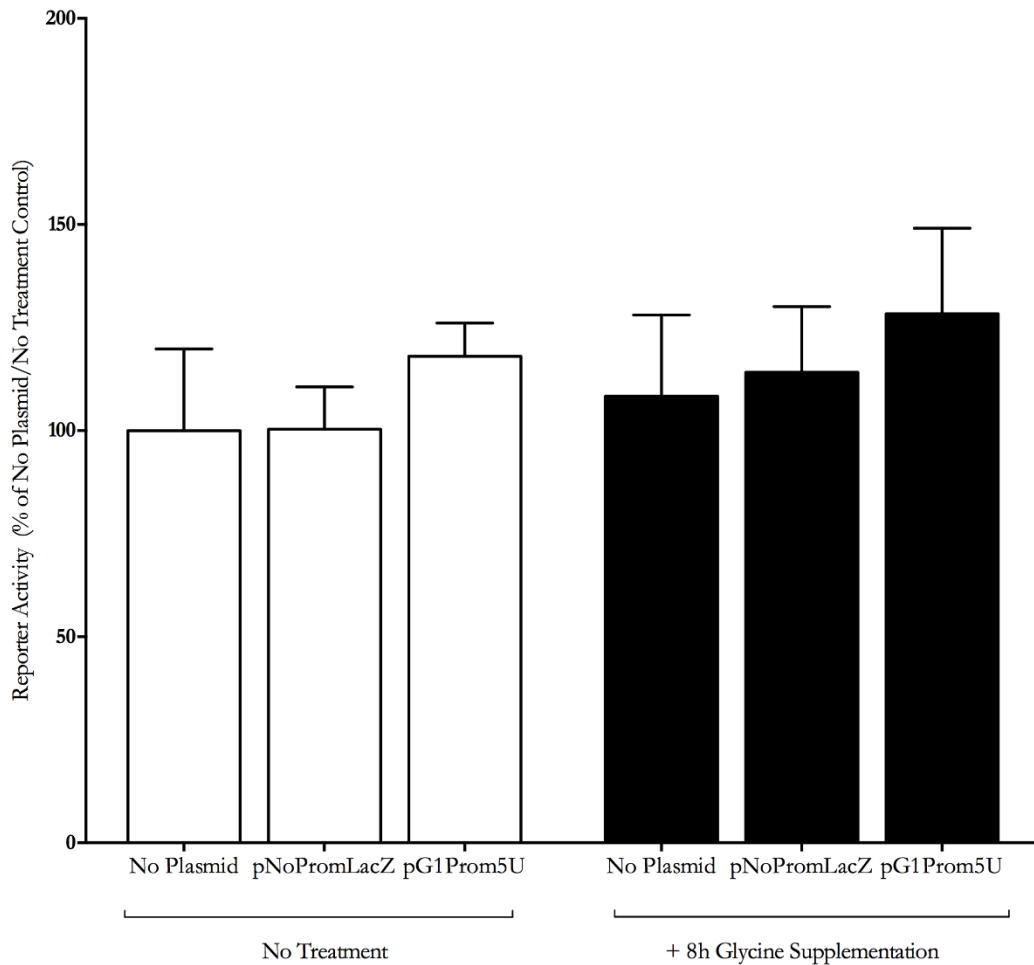
Figure 3.38: Effect of prolonged glycine supplementation on GlyT-1a promoter activity. Bars show the relative β-galactosidase activity in unstressed (open bars) and stressed (closed bars) Caco-2 cells. Transfecting cells with the construct pG1Prom5U followed by incubation for 8 hours in glycine supplemented medium resulted in β-galactosidase activity that was not different from that seen in cells transfected with the promoter-less 'empty' plasmid (pNoPromLacZ) and incubated in glycine supplemented medium. Data are pooled from three experiments (N=3, n=9) and shown here as percentage mean ± SEM relative to the average activity of the un-transfected and unstressed control (No Plasmid). One-way ANOVA was used for multiple comparisons.

3.4   Additional motifs may be necessary for GlyT-1a transcription

Data discussed in section 3.3 showed that a reporter construct (pG1Prom5U) inclusive of the predicted CRMs 6, 7, 8 and 10 contains elements sufficient for transcriptional induction of GlyT-1a by various stress treatments. Previously, Caco-2 cells transfected with another construct pG1PromE1 (which also contains all the aforementioned CRMS but unlike pG1Prom5U excludes the UTR of the second exon) did not generate significant reporter activity when stressed (Figure 3.16, page112). This observation was puzzling given that both constructs pG1PromE1 and pG1Prom5U contained a predicted ARE of the GlyT-1a flanking sequence (CRM8) and a predicted AARE in the GlyT-1a first exon (CRM10). Therefore a new hypothesis was proposed that although these sequence motifs may indeed be necessary for transcription downstream of stress other as of yet unidentified motifs, probably contained in the exon 2 UTR, are also required for effective or enhanced transcription.

3.4.1 Mutation of the GlyT-1a AARE sequence reduces reporter activity

To begin testing this hypothesis the importance of the AARE in pG1Prom5U was investigated. The first and last adenines in this AARE were mutated by site directed mutagenesis to thymine and guanine respectively (TG**A**TGCA**A**C reverse complement AC**T**ACGT**T**G to TG**T**TGCA**G**C reverse complement AC**A**ACGT**C**G). The mutated bases were selected on the basis that they had the highest base-wise conservation scores in the Atf-4 binding consensus described in the TRANSFAC database. The resulting mutated construct pG1Prom5U(mut) was validated by sequencing. As can be seen in Figure 3.39a, the double mutation of the AARE resulted in β-galactosidase activity that was about 30% less than that with pG1Prom5U in transfected cells stressed by 1 μM tunicamycin for 16 hours; this relative to the background activity from untransfected and transfected cells. Although the activity from pG1Prom5U in these experiments was less than that seen previously, correcting for background activity shows that the mutations result in a ~75% reduction in promoter activity (Figure 3.39). Comparing the β-galactosidase activity in cells transfected with the promoter-less control plasmid pNoPromLacZ to those transfected with pG1Prom5U(mut) shows a small but significant increase. In all, this suggests that although the AARE may be necessary for GlyT-1a transcription additional elements elsewhere in the GlyT-1a sequence are indeed required.
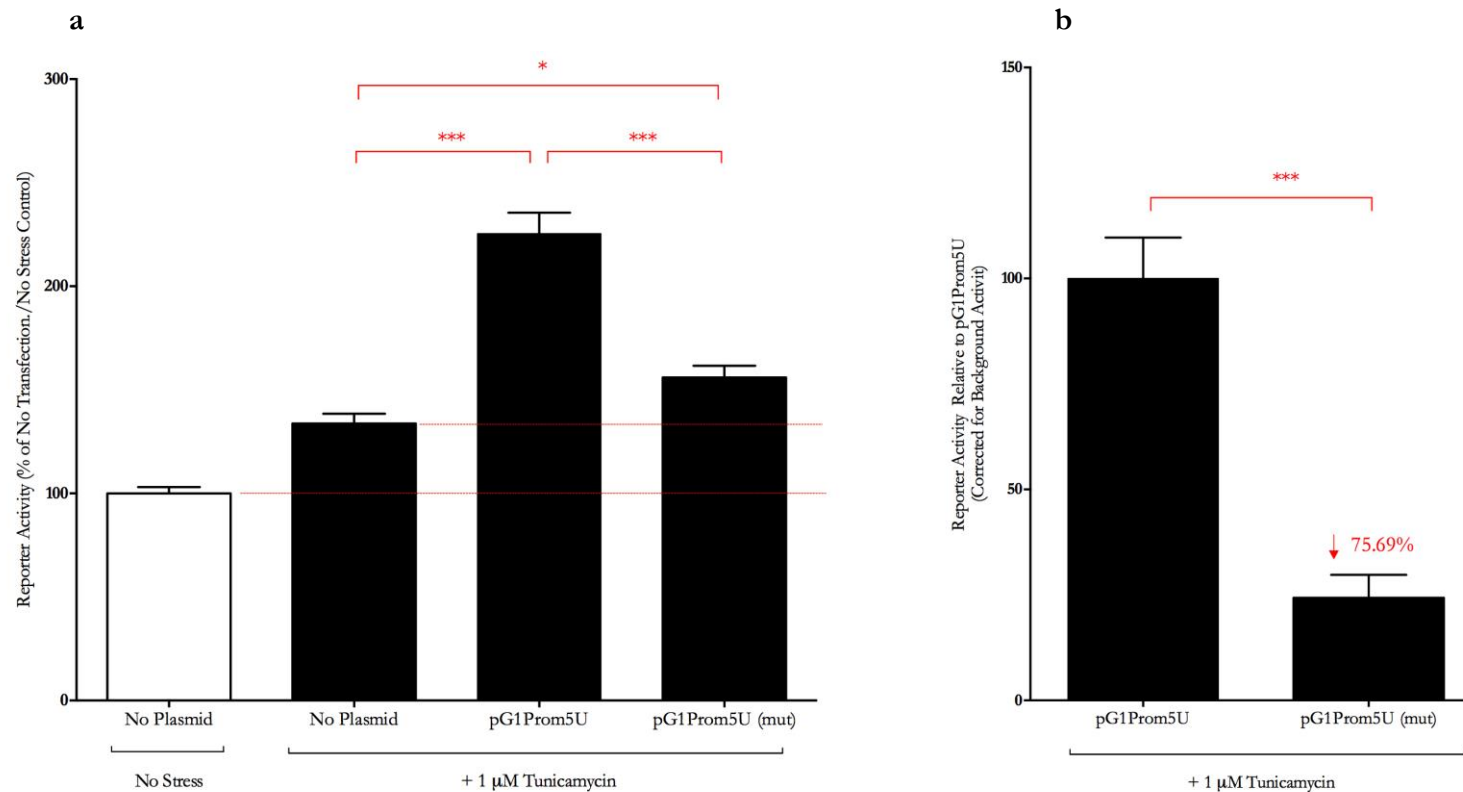
Figure 3.39: Mutation of the putative GlyT1a AARE reduces promoter activity. **a)** Caco-2 cells transfected with pG1Prom5U and treated with tunicamycin showed ~1.5 fold induction of β-galactosidase activity; in cells transfected with pG1Prom5U (mut) a small but significant increase in activity was observed when compared to the No Plasmid control. **b)** Correction for background activity from the No Plasmid control showed that in tunicamycin treated cells transfected with pG1Prom5U(mut) results in β-galactosidase activity that is 75% less when compared to the un-mutated probe. Note that in these experiments tunicamycin induced reporter activity of the pG1Prom5U construct was less than that observed previously. Data are pooled from three experiments each with three experimental replicates (N=3, n=9) and shown here as a percentage mean ± SEM relative to the average activity of the control. One-way ANOVA was used for comparisons. *** indicates a *P*-value of < 0.001.

3.4.2 Reporter constructs of the second exon UTR lack activity.

The challenge therefore was to identify which additional *cis*-regulatory motifs other than those shown to bind Atf-4 above may be necessary for GlyT-1a regulation. To begin, *in silico* analysis of the 86 untranslated bases of the second exon of GlyT-1a was performed to detect any additional binding sites. Using the TRANSFAC Composite Module Analyst Tool, a 21 base pair region of sequence TG<u>TGAGT</u>GTggtCCCCG<u>TCAC</u>CA contained a possible sense site for Atf-4 binding (Sequence: ccCGTCAcc with 85.4% similarity to the V\$ATF4_Q6 matrix). Additionally, C/EBP-β and C/EBP-α were also predicted to bind this sequence, however their matching matrix similarity scores were much lower than the 85% threshold score imposed by the analysis to minimise false negatives.

To determine if the UTR of the second exon of GlyT-1a (G1E2) contained all the binding sites required for transcriptional regulation of GlyT-1a, PCR with forward primer 5 and reverse primer 4 (Primer sequences are shown in Table 2.4, page59) was used to amplify the first 86 bases of exon 2 from cDNA. The resulting G1E2 amplicon was cloned upstream of the β-galactosidase coding sequence. No additional β-galactosidase activity was detected in total protein extracts from Caco-2 cells transfected with the pG1E2 reporter construct and stressed with tunicamycin (Figure 3.40); this when compared to activity from cells transfected with pNoPromLacZ or untransfected cells. Based on the fact that significant activity was observed in simultaneous control experiments using the pG1Prom5U construct, this result suggests that the sequence of the untranslated region of exon 2 does not contain motifs capable of independently regulating GlyT-1a transcription. Why then is this sequence required for activity of the pG1Prom5U construct? A possible explanation is that even if this sequence contained motifs necessary for transcription, additional distant TFBS are required in a billboard-like CRM for the optimal transcription of GlyT-1a following tunicamycin stress.
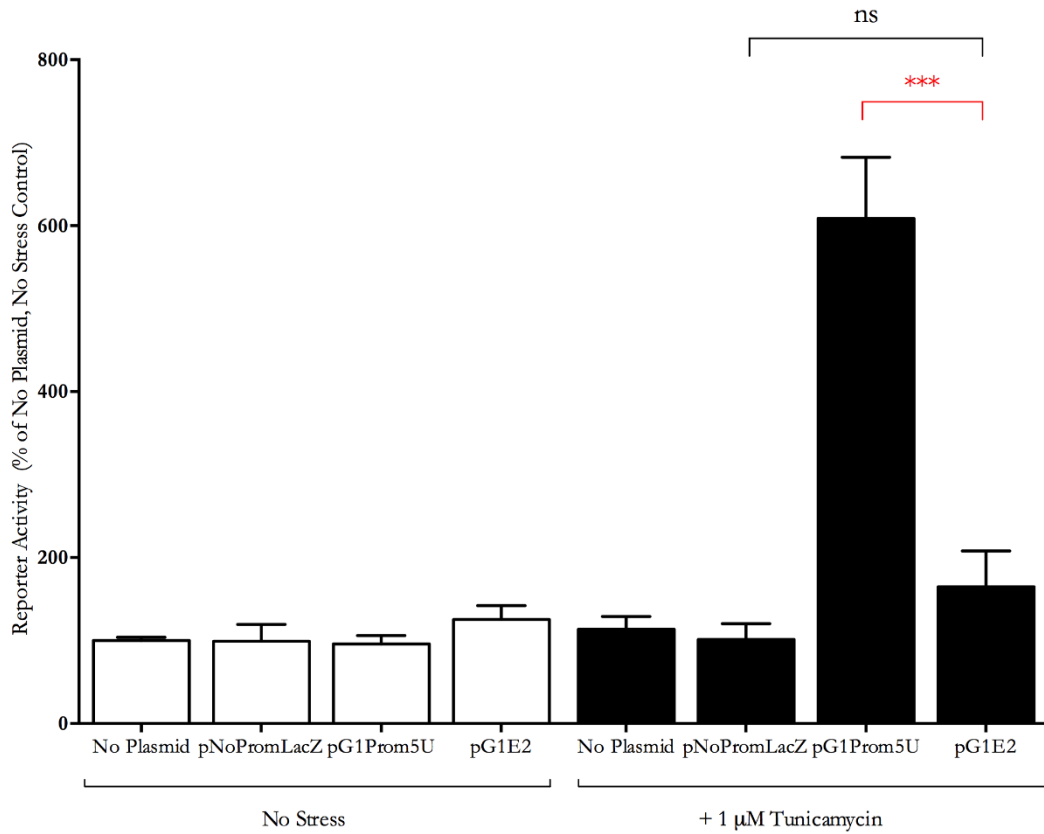
Figure 3.40: GlyT-1a exon 2 UTR alone does generate promoter activity No additional β-galactosidase activity was seen in Caco-2 cells transfected with pG1E2, a reporter construct of the first 86 bases of exon 2; this when compared to cells transfected with the pNoPromLacZ. Bars show the relative unstressed (open bars) and stressed (closed bars) activity in Caco-2 cell relative to the untransfected and unstressed (No Plasmid) control. Data are pooled from two experiments each with three experimental replicates (N=2, n=6) and shown here as percentage mean ± SEM relative to the average activity from the untransfected and unstressed control. One-way ANOVA was used for sample group comparisons. *** indicates a *P*-value of < 0.001; ns indicates a *P*-value > 0.05.

3.4.3 Activity from constructs of just the GlyT-1 5'UTR

A further reporter construct (pG15U) was developed to include the UTRs of both the first and second exons of GlyT-1a without the interrupting large intron sequence (Figure 3.41). pG15U was designed to assess if the GlyT-1a UTR alone was sufficient for transcription or if TFBS in the promoter region are also required. When pG15U was transfected into Caco-2 cells which were subsequently stressed with 1μM tunicamycin for 16 hours, the β-galactosidase activity detected in total extracts from these cells was three times higher than that from cells transfected with the pNoPromLacZ (Figure 3.42). However, when compared to cells transfected with pG1Prom5U and stressed in the same way the activity from pG15U was half that seen with pG1Prom5U. This suggests that whilst elements in the GlyT-1a 5'-UTR found on both the first and second exon are capable of driving transcription, additional motifs in the 1159 bases of sequence upstream of the TSS enhance it. Three additional CRMs with binding sites for Atf-4 and several of its dimerisation partners were predicted in this region by bioinformatic analysis (See CRM6, CRM7 and CRM8 discussed in Section 3.2).
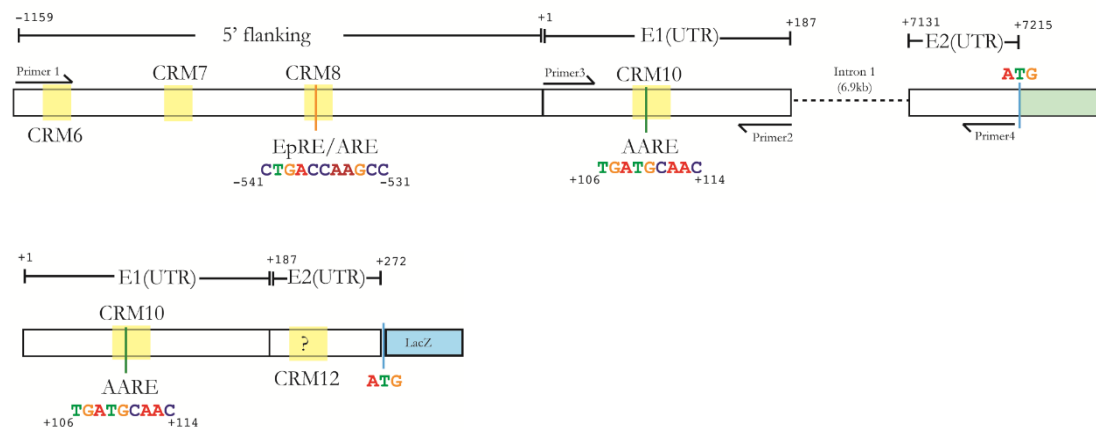
Figure 3.41: Illustration showing the features of the pG15U reporter. Primers 3 and 4 were used to amplify the UTR of exons 1 and 2 from cDNA. The resulting PCR product was cloned upstream of the β-galactosidase coding sequence (lacZ). Features are numbered relative to the TSS and the position of a putative amino acid response element (AARE) and electrophile or antioxidant response elements (EpRE/ARE) are given.
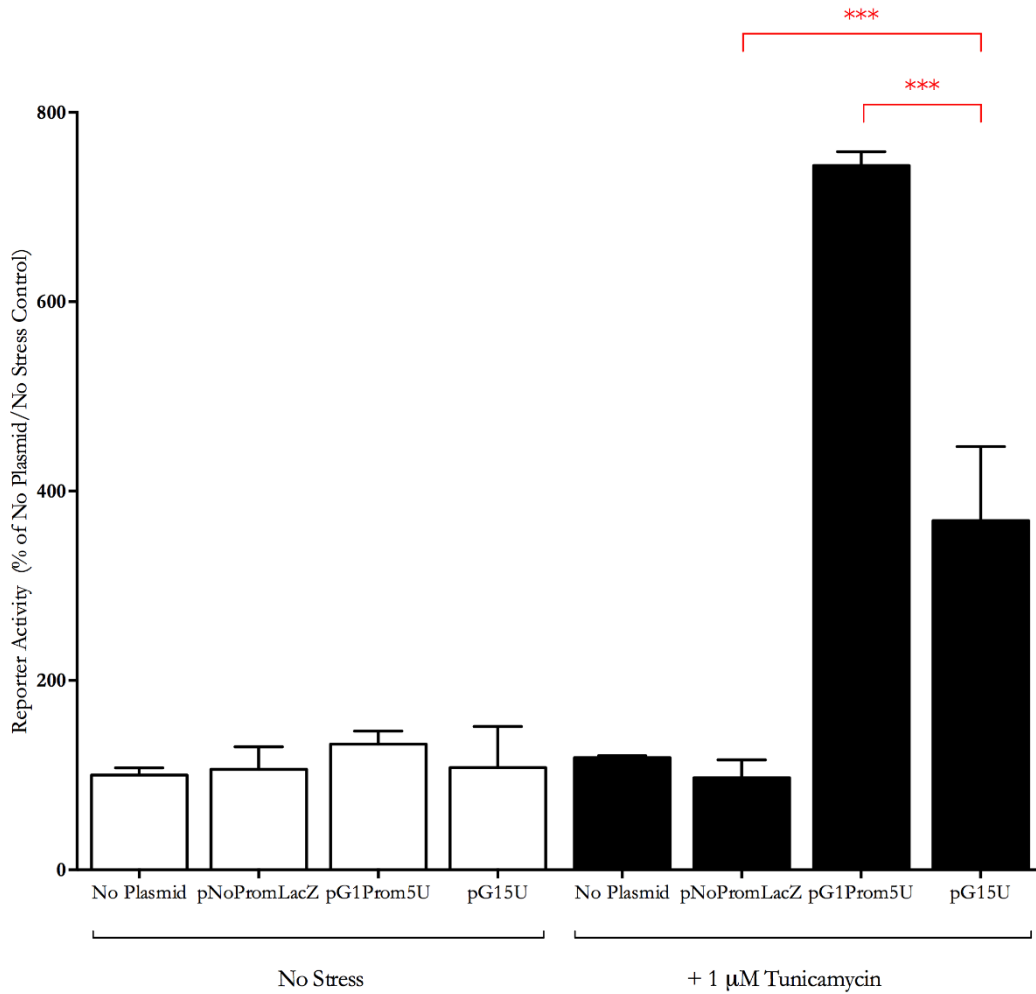
Figure 3.42: Reporter activity from the GlyT-1a UTR sequence. Caco-2 cells transfected with pG15U, a reporter construct of the 5'UTR of exon 1 and 2 showed β-galactosidase activity three times that from cells transfected with pNoPromLacZ. However, the activity from pG15U is approximately half that from pG1Prom5U. Bars show the relative unstressed (open bars) and stressed (closed bars) induced activity in Caco-2 cell relative to the untransfected and unstressed (No Plasmid) control. Data are pooled from two experiments each with three experimental replicates (N=2, n=6) and shown here as percentage mean ± SEM relative to the average activity from the untransfected and unstressed control. One-way ANOVA was used for sample group comparisons. *** indicates a P-value of < 0.001.

3.4.4 Activity from constructs including additional intronic elements

The role of intronic elements (INE) in the direct regulation of gene transcription is increasingly being elucidated. Functional AARE-like motifs have been described in the intronic sequence of the Atf-4 regulated Cat-1 gene (Huang *et al.*, 2009). Analysis of the large 6.9kb first intron of GlyT-1a revealed two additional AARE-like motifs. The two intronic AAREs are labelled INE-1 and INE-2 for convenience. INE-1 is located on the antisense strand between nucleotides 2551 and 2559 downstream from the GlyT-1a TSS. The INE-1 (reverse complement sequence TGATGAAT) differs from the AARE of the first exon (TGATGCAAC) in the occurrence of a thymine in place of the most 3' cytosine and lacks the cytosine at position 6. INE-2 of sequence TGATGAAC is located in the reverse orientation on the sense strand. INE-2 is identical to that described in the first intron of the Cat-1 gene shown to bind the Purine-rich Element Binding Protein A (Purα) (Huang *et al.*, 2009).

To determine if these sequences had any additional effect on the activity already observed with the pG1Prom5U, a new reporter construct (pG1PromInt15U) which is similar to pG1Prom5U but contained an additional 964bp intronic sequence inserted between exons 1 and 2 was made (Figure 3.43). pG1PromInt15U was generated in four steps. In the first step, the GlyT-1a promoter sequence starting 1159 upstream from the TSS, exon 1 and 209 bases of intronic donor sequence were amplified using forward primer 1 and reverse primer 6 (G1PromE1_). In the second step, the 964bp intronic sequence (_G1Int1_; between bases 2472 and 3436 downstream from the GlyT-1a TSS) was amplified using sense primer 7 and reverse primer 8. Primer 7 was designed to include a 5'-overhang sequence complementary to the 3' end of Primer 6. This allowed for ligation of these two fragments (G1PromE1_ and _G1Int1_) by LM-PCR in the next step to generate G1PromInt1_. A variant of primer 5 (primer 5b) was designed, identical to primer 5 in the 3' region, but including a 5'-overhang sequence complementary to the 3' end of primer 8. Primer 5b and primer 4 were then used to amplify the first 86 bases of GlyT-1 exon 2 (_G1E2). _G1E2 was subsequently joined to the G1PromIn1_ by LMPCR to generate a 1445bp product (G1PromInt5U). G1PromInt5U was cloned upstream of the β-galactosidase coding sequence as described earlier to generate the construct pG1PromInt5U.

When Caco-2 cells were transfected with this construct (pG1PromInt5U) and stressed with 1µM tunicamycin for 16 hours, the resulting reporter activity (from a single experiment) was no different than that observed with pG1Prom5U (Figure 3.44); suggesting that INE-1 and INE-2 did not have any additive effect on GlyT-1a transcription.
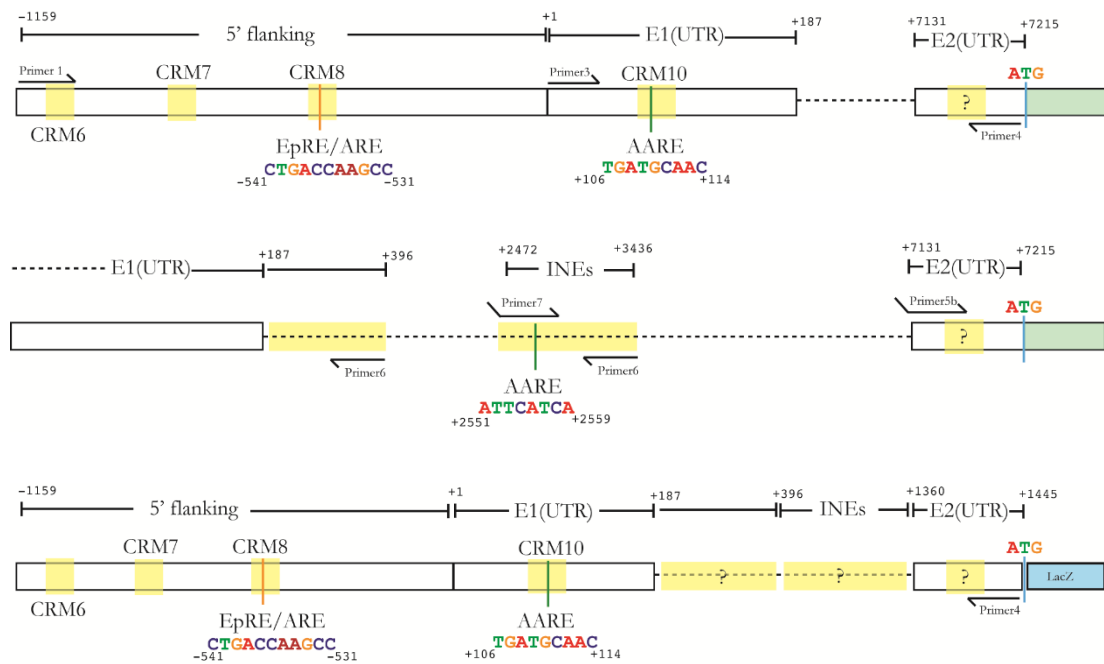
Figure 3.43: Illustrations showing features of the pG1PromInt15U reporter (bottom); The pG1PromInt15U construct which contains additional AARE-like sequences was generate by ligating three PCR products as described in the text. The resulting 1445 bp sequence was cloned upstream of the β-galactosidase coding sequence and the resulting construct validated by sequencing.
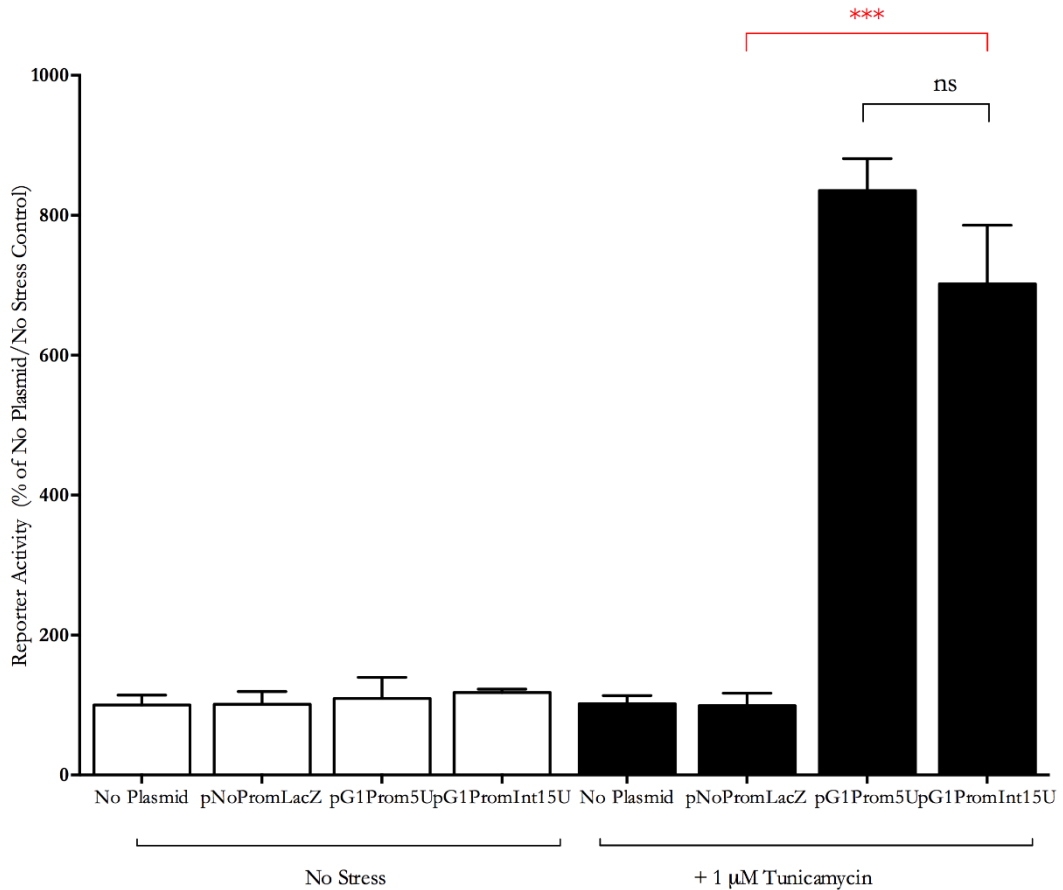
Figure 3.44: Activity of GlyT-1a constructs incorporating AAREs of intron 1. Bars represent the relative unstressed (open bars) and stressed (closed bars) induced β-galactosidase activity of constructs transfected into Caco-2 cells. The pG1PromInt15U construct (which is designed to incorporate two additional AAREs identified in intron 1) resulted in β-galactosidase activity that is comparable to that observed with pG1Prom5U and significantly higher than the unstressed control in tunicamycin treated cells. Data are from a single experiment with three experimental replicates (N=1, n=3) and shown here as percentage mean $\pm$ SEM relative to the average activity from the untransfected and unstressed control (No Plasmid). *** indicates a P-value of < 0.001 while ns represents a P-value > 0.05.

3.5   Regulation of GlyT-1a transcription by Nrf-2

Nrf-2, a cap-n-collar (CNC) BZIP factor, is an important transcriptional regulator of cyto-protective genes in response to several cellular insults. It was first identified by Moi *et al.* (1994) when screening for proteins which bound to a tandem Ap-1-like repeat in the β-globin gene. The Nrf-2 protein has since been found in several cell types and tissues (Lee *et al.*, 2005). Nrf-2 double knock-out mice show increased tumour growth in the intestinal tract when treated with the tar derived carcinogen benzo(a)pyrene (Ramos-Gomez *et al.*, 2001, Ramos-Gomez *et al.*, 2003). In the studies by Ramos-Gomez and colleagues, treating Nrf-2$^{-/-}$ mice with the tumour preventing drug oltipraz had little effect on benzo(a)pyrene induced tumour growth when compared to the wild type controls. Nrf-2 repression and the subsequent down-regulation of antioxidant genes correlates to significant accumulation of reactive oxygen species (ROS) in human mesenchymal stem cells (MSC) and breast cancer cell lines (Tsai *et al.*, 2013, Funes *et al.*, 2014). The constitutive activation of the Nrf-2 pathway is an effective chemo resistance and evasion strategy for several cancer types (Kwak *et al.*, 2002, Kwak *et al.*, 2003, Sporn and Liby, 2012). There are indications that Nrf-2 may be a direct PERK substrate (Cullinan *et al.*, 2003, Cullinan and Diehl, 2004, Cullinan and Diehl, 2006). The phosphorylation of eIF-2α by PERK and other ER bound kinases followed by the activation of Atf-4 marks a convergence point in the intergrated stress response (ISR) pathway. Thus, it is possible that the ISR pathway might result in Nrf-2 activation irrespective of the stress type. Together Nrf-2 and Atf-4 may mediate the differential regulation of target genes such as CHOP (Zong *et al.*, 2012) and xCT (Ye *et al.*, 2014).

Bioinformatics analysis of the GlyT-1a flanking sequence revealed a putative composite regulatory module (CRM8) consisting of TFBS sites for Nrf-2 (Illustrated in Figure 3.45). CRM8, located between 558 and 531 bases upstream from the GlyT-1a TSS, also contains binding sites for Atf-4 and small Maf proteins both of which are known Nrf-2 interacting partners (He *et al.*, 2001, Motohashi *et al.*, 2004).

As reviewed in Section 1.7.3, Nrf-2 activation is generally considered to contribute to regulatory networks induced by oxidative, chemical and xenobiotic insults. Our hypothesis posited that the expression of GlyT-1a may be regulated in a similar manner to xCT. Both GlyT-1 and xCT are equally important requirements for the supply of glycine and cysteine respectively, for the synthesis of the antioxidant glutathione (GSH).

Studies have demonstrated the regulation of xCT by Nrf-2 via antioxidant response elements (ARE) in its promoter (Sasaki *et al.*, 2002, Lewerenz *et al.*, 2009, Bell *et al.*, 2011). The same is true for transcriptional regulation of EAAT-3 (the principal glutamate transporter in the CNS which provides cysteine to neuronal cells that lack system X_c^- (Escartin *et al.*, 2011)).

In the last 20 years significant advancements have been made in understanding how Nrf-2 is activated. The general consensus is that Nrf-2 mRNA is constitutively transcribed in the cell, and the resulting protein, if unneeded, is degraded by a Keap-1 dependent pathway (Nguyen *et al.*, 2005, Misra *et al.*, 2013). Exactly how Nrf-2 escapes sequestration and degradation downstream of stress is still being investigated. That the translation of the always expressed Nrf-2 mRNA is itself redox sensitive has been suggested (Li *et al.*, 2010). Li *et al,* in another study, suggest that upon dissociation of the expressed protein from Keap-1, its dimerisation with other BZIP factors such as the small Maf proteins, mask the nuclear export signal (NES) motif within its leucine zipper domain (Li *et al.*, 2008). Such interactions effectively renders Nrf-2 nuclear bound, away from the cytoplasmic cytoskeleton bound Keap-1.
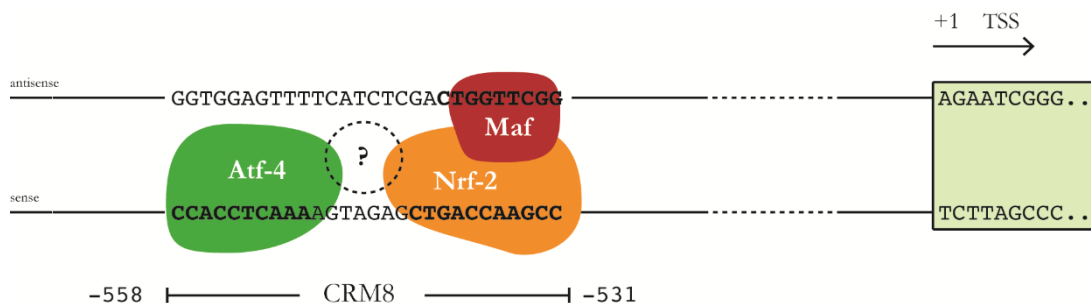
Figure 3.45: Hypothetical interaction of Nrf-2 and other factors at CRM 8.

Little is known about the transcriptional regulation (if any) of Nrf-2 following stress. Although the primary mode of Nrf-2 'activation' downstream of stress is in its post-translational dissociation from Keap-1 followed by nuclear stabilisation, some researchers report increases in Nrf-2 mRNA expression following stress (Miao *et al.*, 2005). There is a suggestion that Nrf-2 may in fact regulate its own expression via the presence of ARE-like sequences in its promoter (Kwak *et al.*, 2002, Chorley *et al.*, 2012). Thus to completely determine if Nrf-2 is involved in the regulation of GlyT-1a, its expression pattern

downstream of stress treatment was studied as well as the possible interaction of the translated Nrf-2 protein at the putative CRM8 found in the GlyT-1a promoter.

3.5.1 Basal Nrf2 mRNA expression is unchanged by stress

To determine if increases in Nrf-2 transcription coincided with its activation for the regulation of GlyT-1a downstream of stress treatment, Nrf-2 mRNA was quantified in total RNA of Caco-2 cells subject to varied stressors. QPCR was performed to quantify a representative 120bp region spanning the third and fourth exons of the Nrf-2 coding sequence. QPCR primers used were capable of detecting all three Nrf-2 transcript variants in total RNA extracted from stressed treated Caco-2 cells as described in Chapter 2. PCR conditions were optimised by endpoint PCR; the size of the resulting amplicon was checked on a 2% agarose gel (Figure 3.46). The integrity of the Nrf-2 amplicon (cloned in a pGEM-T-Easy plasmid for use as a QPCR standard) was validated by sequencing (Figure 3.47)
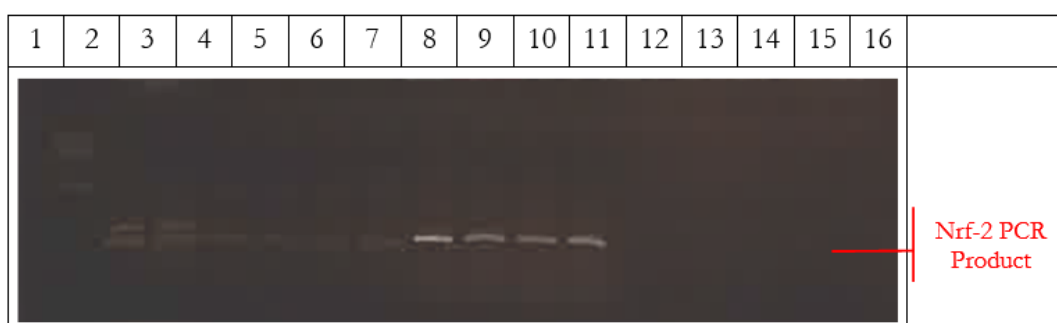
Figure 3.46: Gel image of cloned of pNRF2 standard. The 120bp PCR product seen in lanes 8-11 was generated by amplification of Caco-2 cDNA with Nrf-2 specific primers shown in Table 2.3. The eluted product was ligated into pGEM-T-easy, cloned in JM109 cells and characterised by sequencing (see Figure 3.47) and used to generate standard curves for use in QPCR.
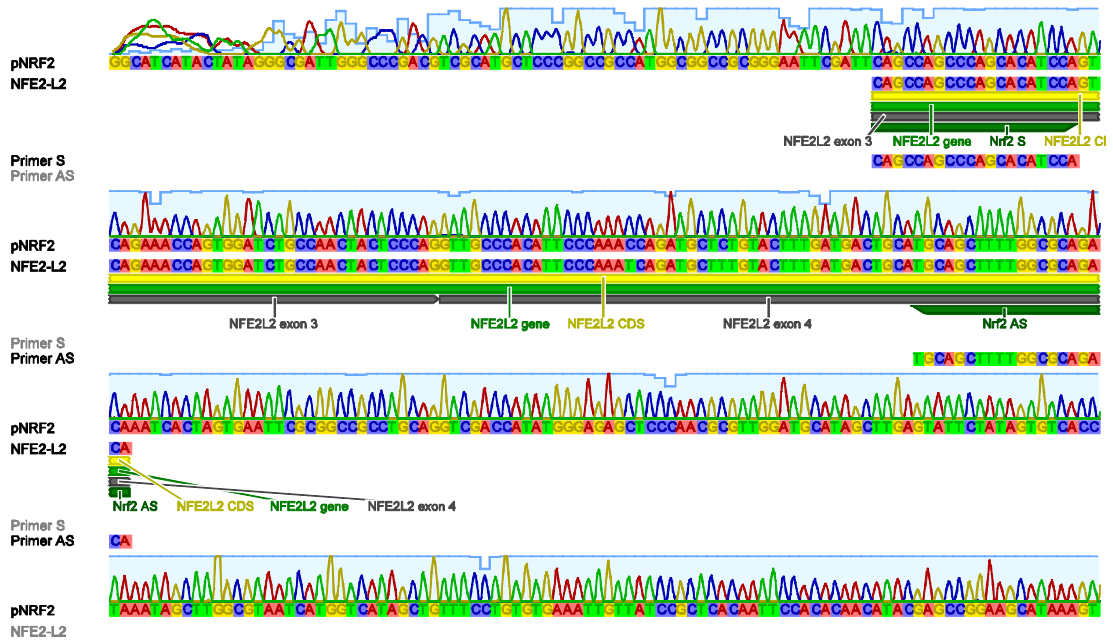
Figure 3.47: Validation of cloned pNrf-2 for use as standards in RT-QPCR. A 120bp region of Nrf-2 coding sequence was cloned into a pGEM-T-easy plasmid for use as QPCR standard. Sequencing results were aligned to the Nrf-2 (NFE2L2) sequence obtained from NCBI. The two PCR errors identified at positions 73 and 82 of the product were non critical for QPCR assays.

Given that in the UPR Nrf-2 may also be activated by PERK along with the observation that it is active downstream of prolonged ER stress suggests a pro-survival role for Nrf-2 (Cullinan and Diehl, 2004). However our data do not suggest Nrf-2 activation by the UPR correlates to increases in its transcription. As shown in Figure 3.48, the expression patterns observed for Nrf-2 show no significant changes in Nrf-2 mRNA expression in Caco-2 cells following either 1 μM tunicamycin, 50 μM thapsigargin for 16 hours, with 0.2mM DEM for four hours or eight hours of amino acid starvation.

ER stress induced by thapsigargin (a Ca^{2+}/ATPase inhibitor) has been associated with expression of downstream Nrf-2 target genes (Cullinan and Diehl, 2006). The rapid increase in intracellular Ca^{2+} concentration and subsequent accumulation in the mitochondria results in the production of ROS. Blocking the transcription of Nrf-2 in normal human bronchial epithelial (NHBE) cells with specific siRNA had consequential down-regulatory effects on the expression of its target genes following thapsigargin treatment (Yolanda et al., 2013). Similar decreases in the expression of target genes was

not observed when these NHBE cells were treated with siRNA against PERK. This suggests that whilst post-translational events are important to the activation of Nrf-2 downstream of oxidative stress, its mRNA expression may be equally important. Our data however does not suggest that a heightened oxidative stress burden induced by DEM results in any changes to Nrf-2 basal mRNA expression
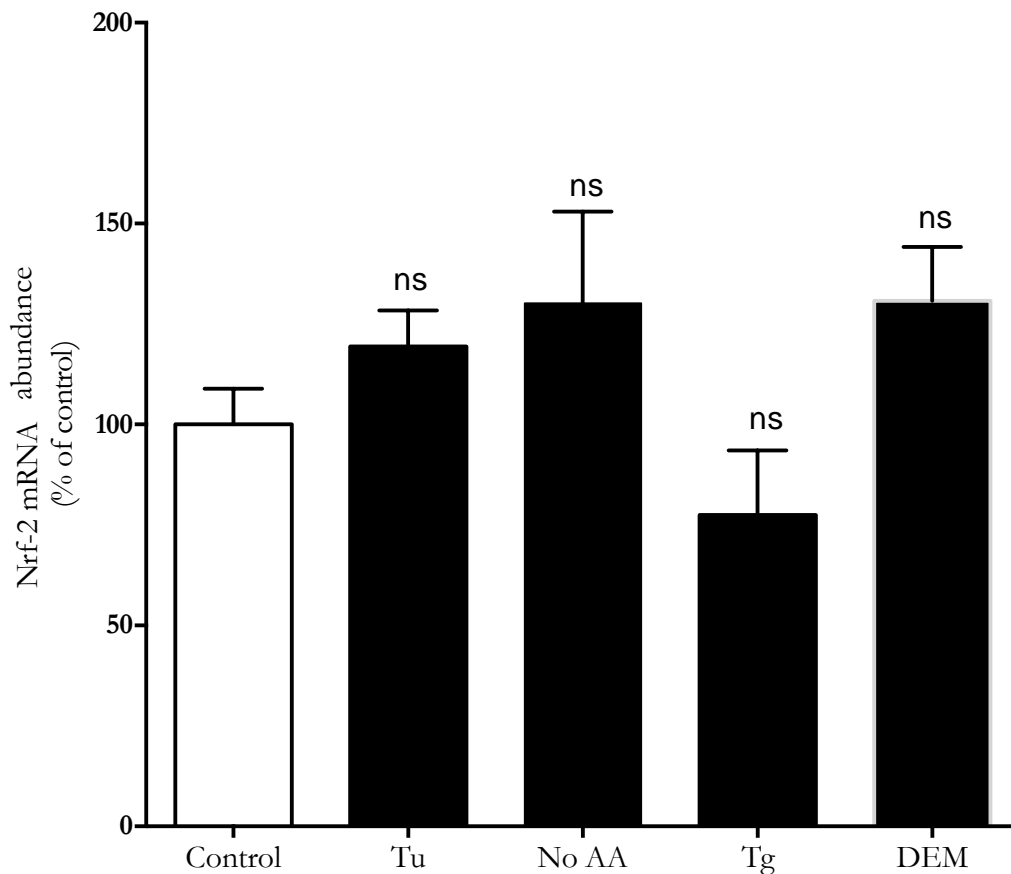
Figure 3.48: Nrf2 mRNA abundance is unchanged following stress in Caco-2 cells with either 16 hours of tunicamycin (Tu) or thapsigargin (Tg), 8 hours of amino acid starvation (No AA) or 4 hours of DEM. Values are pooled from multiple experiments and represent a percentage mean of the untreated control sample displayed as mean ± SEM, n=3 (N=2-3 for each treatment). One-way (non-parametric) analysis of variance (ANOVA) was performed. Dunnett's multiple comparison tests with pooled variance for each treatment was used for comparison to the control

3.5.2 Atf-4 knockdown increases Nrf-2 mRNA expression only with ER stress by thapsigargin and Nutrient stress by amino acid starvation.

The spatio-temporal and tissue-specific precision with which TFs link external stimuli to changes in expression of target genes, often requires an interplay of several other factors. Filtz *et* al suggest the interconnectedness of TFs is governed by a host of several post-translational modifications (PTMs) (Filtz *et al.*, 2014). It is possible that the PTMs regulating the activation of one transcription factor signals directly to the compensatory expression of other TFs. The redundancy of DNA binding site specificity for one TF points to the probability of other factors binding to the same TFBS in promoter regions of target genes to produce the same or similar regulatory effect (Hummler *et al.*, 1994, Mechta-Grigoriou *et al.*, 2001). Takahasi *et al* show that the binding affinities of TFs to their near-perfect required binding sites can be improved by increasing the concentrations of the TFs (Takahasi *et al.*, 2011). To establish whether increased Nrf-2 transcription may compensate for the down regulation of other BZIP TFs known to be active in the stress response, Nrf-2 mRNA level was quantified in total RNA extracted from Caco-2 cells pre-treated with specific siRNA to either Atf-4, Atf-6 or Xbp-1.

Following the transfection of Caco-2 cells with specific siRNA targeting Atf-4, a differential pattern of Nrf-2 mRNA expression was observed based on the type of stressor. Despite a significant reduction in Atf-4 mRNA expression in the cells (Figure 3.49), no changes in Nrf-2 mRNA expression were observed following either 16 hours of 1 μM tunicamycin treatment or 4 hours of 0.2 mM DEM (Figure 3.50). Subjecting the transfected Caco-2 cells to 16 hours of 200 μM thapsigargin, or 8 hours of amino acid starvation resulted in approximate doubling of Nrf-2 mRNA expression when compared to the untransfected controls. That Nrf-2 and Atf-4 can both bind to characteristic antioxidant response elements (AREs) has been established (He *et al.*, 2001). Beyond the observed protein-protein interaction between Nrf-2 and Atf-4, the transcriptional relationship of these proteins remains unclear. There is no evidence for the transcriptional regulation of Nrf-2 by Atf-4. Others have suggested that Nrf-2 mediates the transcriptional regulation of Atf-4 (Afonyushkin *et al.*, 2010, Miyamoto *et al.*, 2011). Nrf-2 was reported to bind to characteristic ARE-like sequence 300 bases upstream of the Atf-4 TSS (Miyamoto *et al.*, 2011). In that study, Miyamoto and colleagues also demonstrated that down regulation of Nrf-2 with specific siRNA completely abolished Atf-4 induction by anoxia.
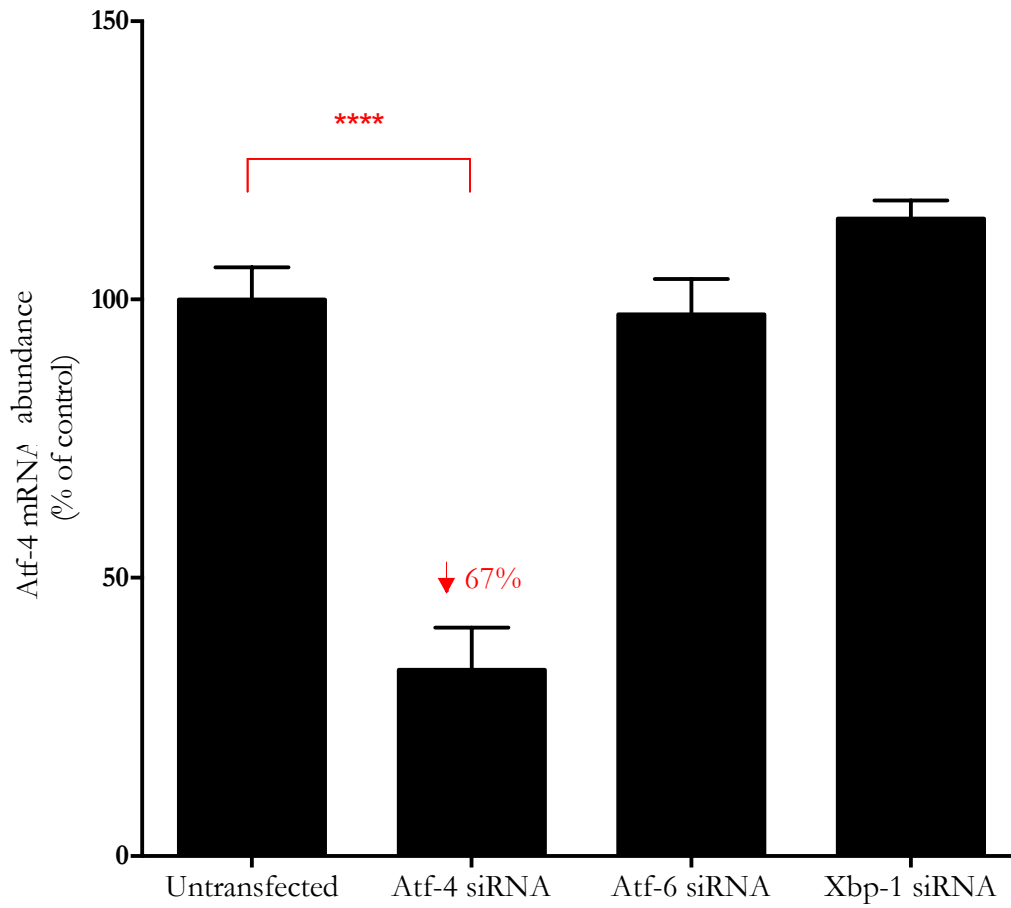
Figure 3.49: Specific siRNA knockdown of Atf-4 in Caco-2 cells. Transfection of Caco-2 cells with Atf-4 specific siRNAs reduced the expression of Atf-4 by about 67% ($P < 0.001$ are represented with a ****). Control transfection with siRNA against Atf-6 and Xbp-1 showed no effect on Atf-4 mRNA abundance. Data are pooled from multiple transfection experiments and are represented as a percentage mean $\pm$ SEM of the untransfected control, n=9 (N = 3 for each treatment). A one-way analysis of variance (ANOVA) was performed to compare changes in mRNA abundance for each transfected sample group.
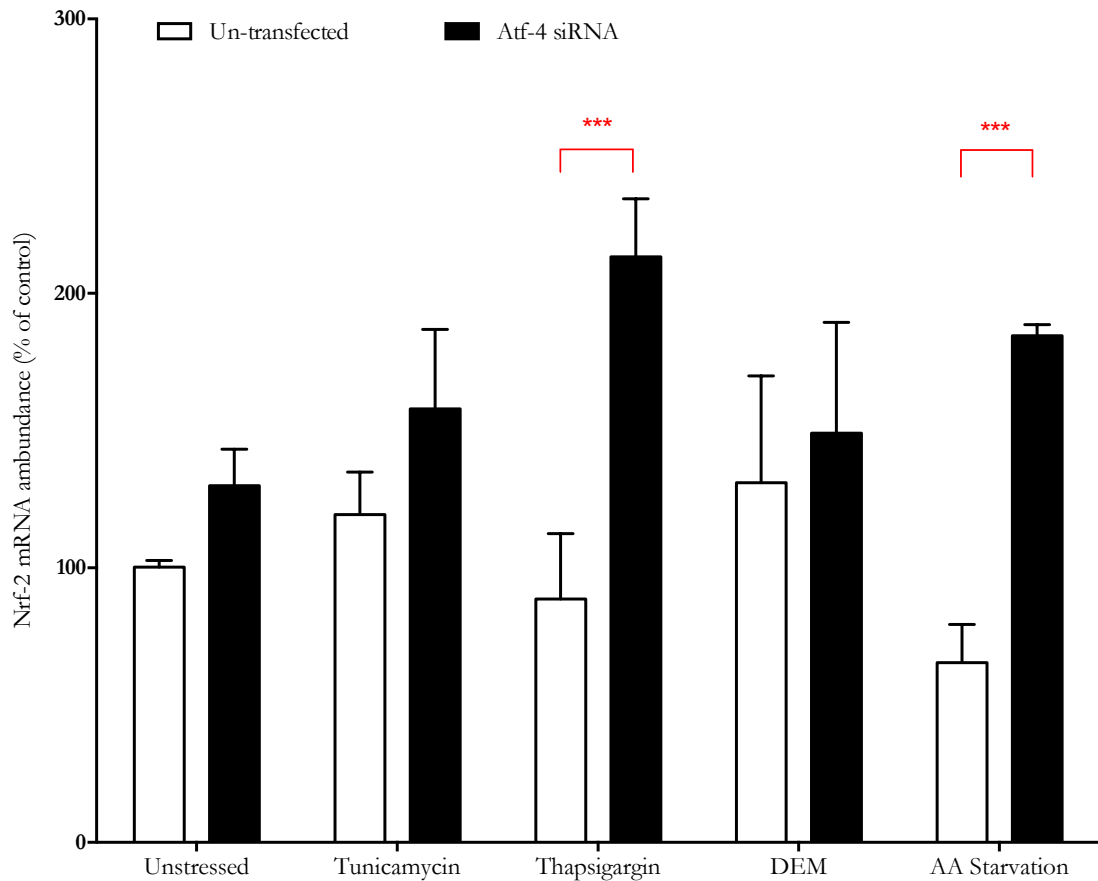
Figure 3.50: Effects of Atf-4 knockdown on Nrf-2 mRNA abundance in stressed Caco-2 cells. Twenty-four hours after plating, cells were either transfected with a specific siRNA to Atf-4 (black bars) or mock transfected (open bars). A further 24 hours after transfection cells were stressed with either tunicamycin, thapsigargin, DEM or starved of amino acids. Values are pooled from multiple transfection experiments and represent a percentage mean $\pm$ SEM of the un-transfected control. n=3 (N = 2-3 for each treatment). A two-way analysis of variance (ANOVA) was performed to compare changes in Nrf-2 abundance for each stress treatment group. Tukey's multiple comparison test, revealed significant differences only between the un-transfected and transfected expression levels of Nrf-2 following 16 hours of 50 μM thapsigargin treatment or four hours of amino acid starvation (*** indicates a p-value of < 0.001).

3.5.3 Knockdown of Atf-6 or Xbp-1 does not affect Nrf-2 expression.

Following the transfection of Caco-2 cells with specific siRNA targeting Atf-6 (Figure 3.51) no changes in Nrf-2 mRNA were observed following 1 μM tunicamycin treatment for 16 hours or 0.2 mM DEM for four hours (Figure 3.52). The same was true for Nrf-2 mRNA levels in these transfected cells following 16 hours of 50 μM of thapsigargin or 8 hours of amino acid starvation. Unlike with Atf-4 there is no published evidence for Nrf-2/Atf-6 interactions. Likewise there are no reports of the regulation of Nrf-2 by Atf-6. Data shown here indicates that indeed Atf-6 is not necessary for Nrf-2 induction.

Atf-6, a type II transmembrane ER resident BZIP factor is an important transducer of ER stress response signals. All cells possess two Atf-6 isoforms (Atf-6α and Atf-6β) with contradicting activator or repressor function. Unlike Atf-4, Atf-6 activation in the UPR is independent of PERK. Atf-6 is confined to the ER by its association with the ER chaperone BiP/GRP78 (Shen *et al.*, 2002, Sommer and Jarosch, 2002). During ER stress, the dissociation from BiP exposes the Atf-6 Golgi localisation signal (GLS) leading to its relocation to the Golgi. In the Golgi, Atf-6 is cleaved at two sites by site-1 and site-2 proteases (S1P, S2P) (Lee *et al.*, 2002). The N-terminal of the cleaved Atf-6 (Atf-6N) translocates to the nucleus where it dimerises with other BZIP factors to mediate the transcription of ER stress target genes.

Given that amongst the cascade of Atf-6 target genes are factors which may alter Nrf-2 activation (Yoshida *et al.*, 2000), the consequences of Atf-6 activation should be investigated as a possible link between ER stress, oxidative stress and apoptosis. For instance, Xbp-1 is an Atf-6 target (Yoshida *et al.*, 2000, Lee *et al.*, 2002). When activated in tissues from liver cirrhosis patients, Xbp-1 causes the transcriptional up-regulation of the synovial apoptosis inhibitor (SYVN-1 or Hrd-1) to protect against ER stress induced apoptosis (Wu *et al.*, 2014). Hrd-1 mediates enhanced post-translational degradation of Nrf-2 and consequent reduction in the expression of Nrf-2 target genes. Given that Atf-6 up regulates Xbp-1 which in turn down regulates Nrf-2 response pathways, it is possible that knocking down Atf-6 stabilises the Nrf-2 protein; hence, no requirement for additional Nrf-2 mRNA expression. Whether Xbp-1 activation and the consequent degradation in Nrf-2 protein results in increased Nrf-2 mRNA expression is not known.
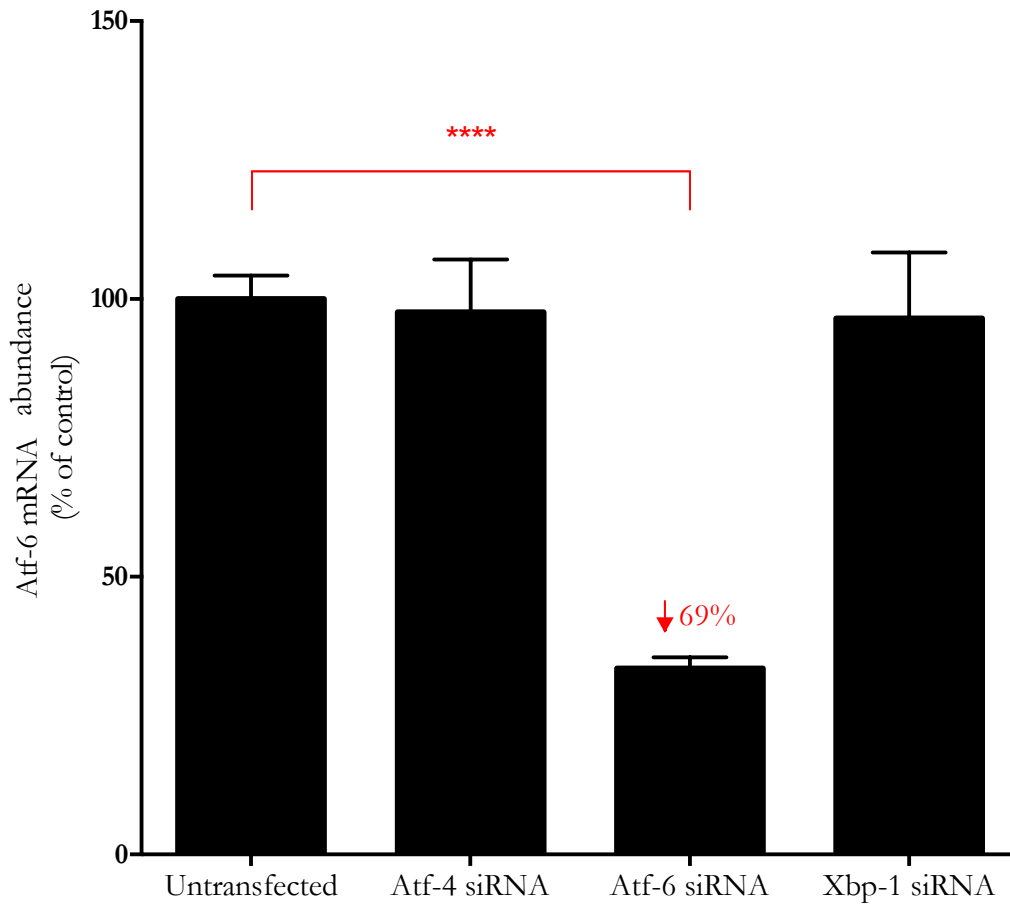
Figure 3.51: Specific siRNA knockdown of Atf-6 in Caco-2 cells. Transfection of these cells with specific siRNA to Atf-6 decreased Atf-6 mRNA abundance significantly by 69% ($P < 0.001$). Atf-6 expression was unchanged following transfection with siRNA specific for Atf-4 or Xbp-1. Data are pooled from three transfection experiments and represent a percentage mean $\pm$ SEM, n=3 (N = 3 for each treatment). A one-way analysis of variance (ANOVA) was performed to compare changes in mRNA abundance for each transfected sample group.
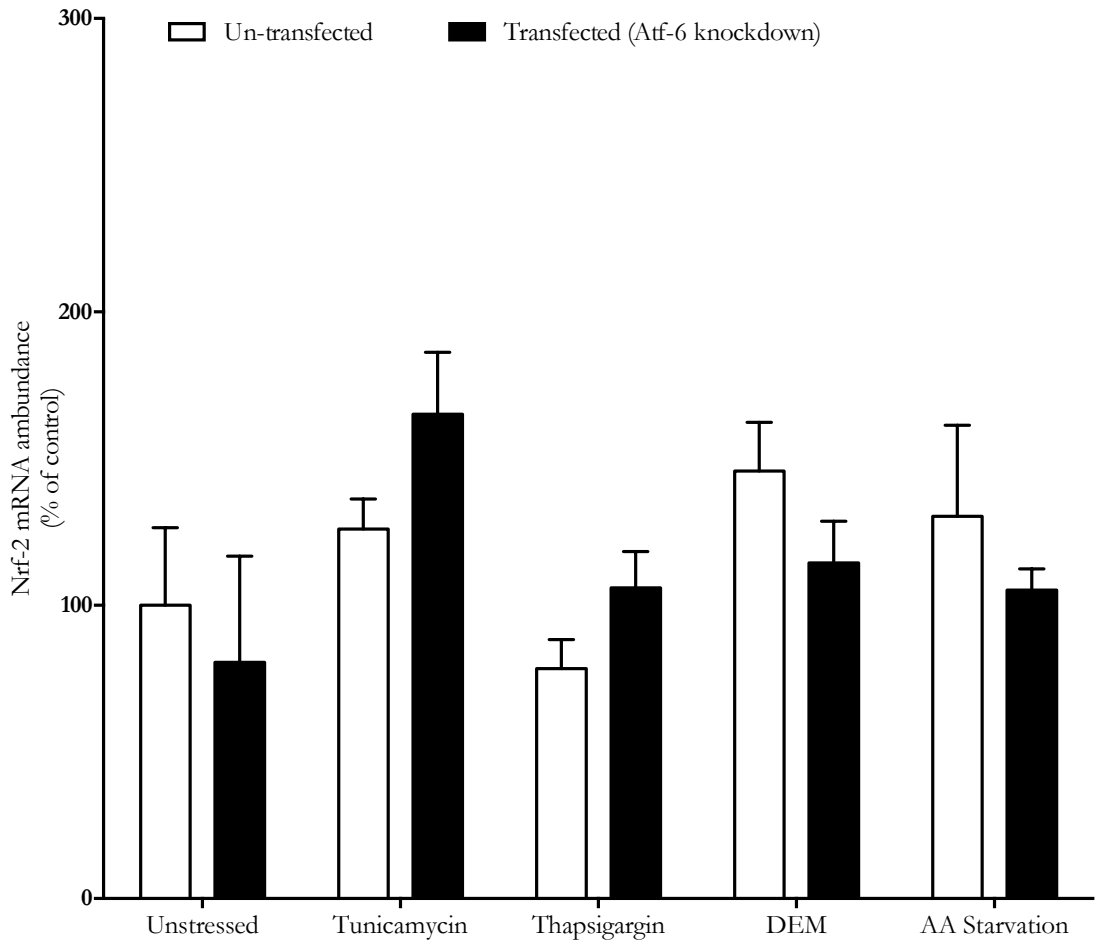
Figure 3.52: Effect of Atf-6 knockdown on Nrf-2 mRNA abundance in stressed Caco-2. Twenty-four hours after plating, cells were either transfected with a specific siRNA to Atf-6 (black bars) or mock transfected (open bars). A further 24 hours after transfection cells were stressed with either tunicamycin, thapsigargin, DEM or starved of amino acids. Relative abundance data was measured by QPCR and are pooled from multiple transfection experiments and represent a percentage mean ± SEM of the un-transfected control samples. n=3 (N = 3 for each treatment) are plotted. A two-way analysis of variance (ANOVA) was performed to compare changes in Nrf-2 abundance for each stress treatment group. Tukey's multiple comparison test, revealed no significant difference between stressed and unstressed samples of either the un-transfected and transfected groups.

Work by Liu *et al* showing increased oxidative burden and cell death in Xbp-1 deficient mouse embryonic fibroblast (MEF) cells somewhat contradicts the conclusions by Wu *et al.* (2014) that Xbp-1 induces degradation of Nrf-2 via Hrd-1 (Liu *et al.*, 2009). Following oxidative stress Xbp-1 has been shown to be regulated in an Nrf-2 dependent fashion (Kwak *et al.*, 2003). To determine if Xbp-1 knockdown had any effects on Nrf-2 mRNA expression Caco-2 cells were transfected with specific siRNA to Xbp-1 and Nrf-2 mRNA quantified under both stressed and unstressed conditions.

A significant knockdown of Xbp-1 mRNA expression was observed following treatment with specific siRNA targeting Xbp-1 (Figure 3.53). Xbp-1 mRNA expression was unchanged by treatment with other non-specific siRNA. The knockdown of Xbp-1 had a doubling effect on Nrf-2 mRNA only in unstressed Caco-2 cells (Figure 3.54). This effect was not observed following either 16 hours treatment with 1 μM tunicamycin or 50 μM of thapsigargin, treatment for 4 hours with 0.2 mM DEM treatment, or 8 hours of amino acid starvation.

Xbp-1, a BZIP factor, was originally detected as a factor bound to the *cis*-acting *x*-box sequence in the promoter regions of the human major histocompatibility complex (MHC) Class II genes (Liou *et al.*, 1990). Signalling via Xbp-1 provides an important link between ER function and the regulation of apoptosis. Downstream of ER stress, Xbp-1 mRNA induced by Atf-6 is spliced by the inositol-requiring enzyme 1 (IRE-1), an ER trans-membrane endo-ribonuclease (Yoshida *et al.*, 2001). There is extensive documented evidence on the role of the transcriptional activity of the spliced product Xbp-1s downstream of stress (Liu *et al.*, 2009, Williams *et al.*, 2014, Xu *et al.*, 2014). However, a post-translational role of un-spliced Xbp-1 (Xbp-1u) mRNA is emerging (Zhao *et al.*, 2013). In the latter study, Zhao and colleagues demonstrated the physical interaction between Xbp-1u and other proteins like members of the FOX protein family, with subsequent phosphorylation of Xbp-1u by ERK leading to degradation by the 20S proteasome. Xbp-1u also sequesters Xbp-1s and Atf-6 (but not Atf-4) for degradation in a similar manner (Yoshida *et al.*, 2006, Navon *et al.*, 2010). Does Xbp-1u and Nrf-2 interaction under unstressed conditions lead to its degradation in a similar manner? If true, the divergent roles of Xbp-1u in unstressed conditions and Xbp-1s in stressed conditions may explain the differences in Nrf-2 mRNA expression observed following Xbp-1 knockdown; however as discussed earlier, Nrf-2 degradation may not necessarily result in changes to its basal mRNA expression.
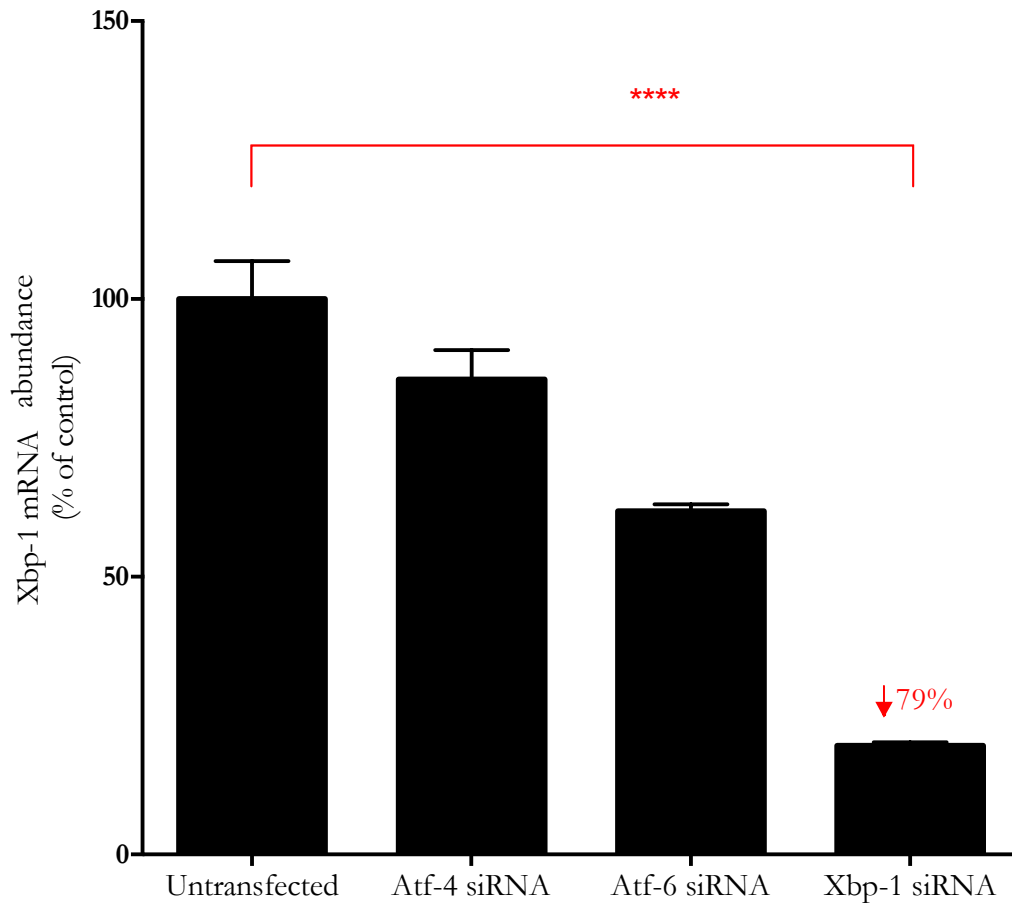
Figure 3.53: Specific siRNA knockdown of Xbp-1 In Caco-2 cells. XBP-1 reduction of 79% (*P* < 0.001) is achieved following transfection with Xbp-1 siRNA. Xpb-1 mRNA expression was unchanged following treatment with Atf-4 siRNA however a significant reduction in its mRNA abundance was observed following Atf-6 transfection in Caco-2 cells. Data are pooled from multiple transfection experiments and represent a percentage mean ± SEM, n=3 (N = 3 for each treatment). A one-way analysis of variance (ANOVA) was performed to compare changes in mRNA abundance for each transfected sample group.
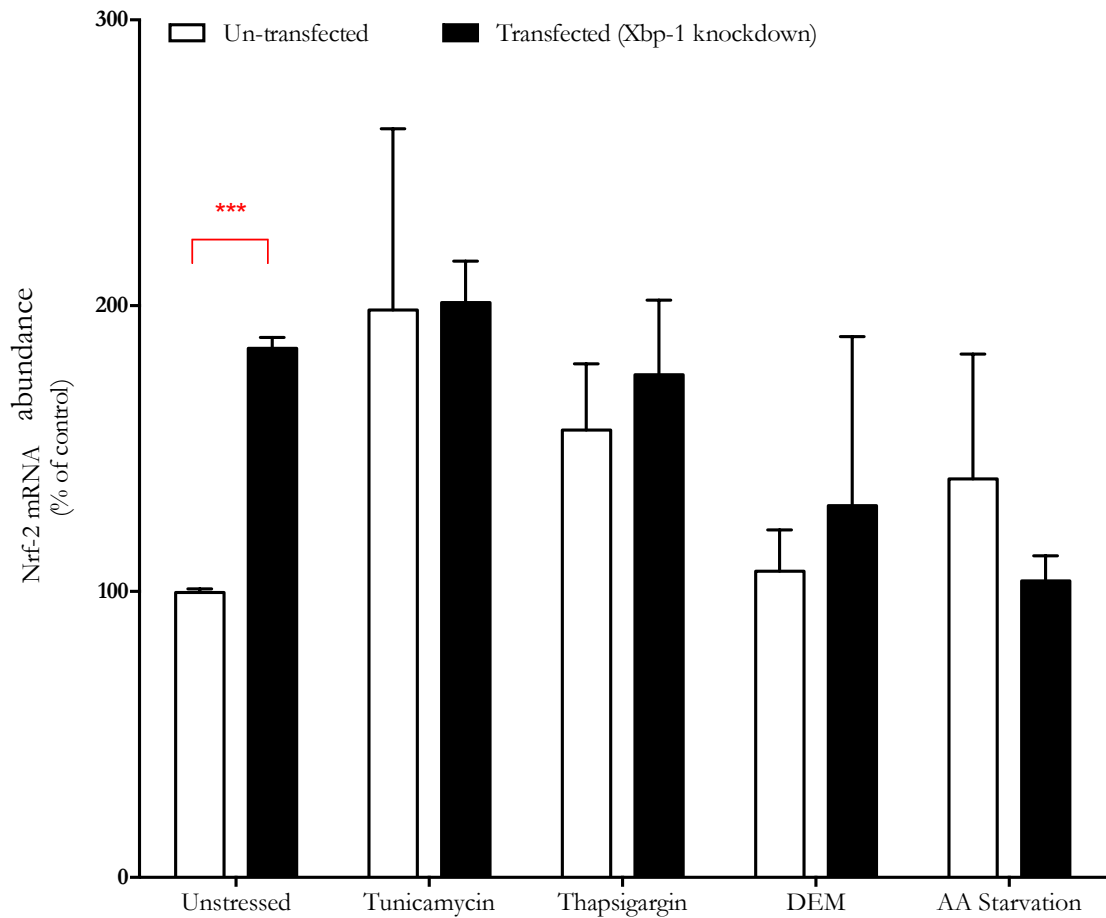
Figure 3.54: Effects of Xbp-1 knockdown on Nrf-2 abundance in stressed Caco-2 cells. Data are pooled from multiple transfection experiments and represent a percentage mean ± SEM of the un-transfected control samples displayed as open bars, or closed bars for the transfected samples (.n=3, N = 2-3 for each treatment). A two-way analysis of variance (ANOVA) was performed to compare changes in Nrf-2 abundance for each stress treatment group. Tukey's multiple comparison test, reveals no significant difference between stressed and unstressed samples of either the un-transfected and transfected groups. However the transfection with specific Xbp-1 siRNA without any stress showed increased expression of Nrf-2 over the un-transfected control. *** indicates a P-value of < 0.001

3.5.4 Nrf-2 protein abundance is increased by stress

The absence of any changes in Nrf-2 expression (as discussed above) is possibly due the complex mode of activation of Nrf-2 protein following stress (i.e. its escape from cytosolic sequestration by KEAP-1 and nuclear translocation; See section 1.7.3). Given that Nrf-2 activity is dependent on stabilisation of the protein, Nrf-2 protein abundance was quantified and its DNA interactions following stress treatment investigated. Despite no changes to Nrf-2 gene transcription (Figure 3.48), Western blot analysis showed an approximately 3-fold increase in the Nrf-2 protein abundance in total protein extracted from Caco-2 cells after 4 hours of 0.2mM DEM treatment (Figure 3.55a, b).
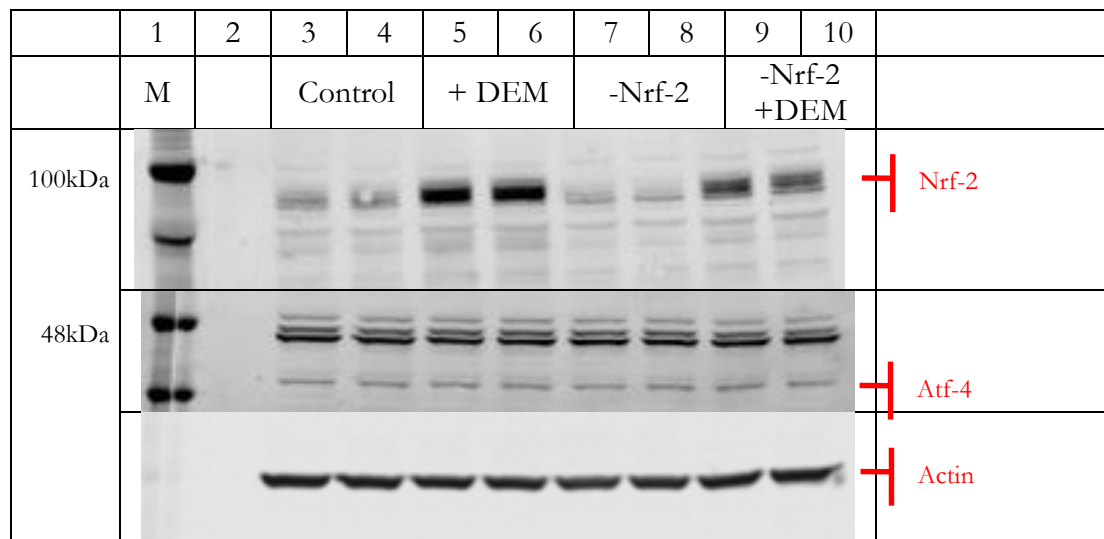
Figure 3.55a: Effects of DEM and Nrf-2 knockdown on Nrf-2 protein abundance. Caco-2 cells treated with 0.2mM DEM showed a large increase in Nrf-2 protein abundance (lanes 5 and 6) when compared to untransfected and unstressed control cells (lanes 3 and 4). Nrf-2 siRNA attenuated protein abundance in stress cells (lanes 9 and 10) when compared to untransfected yet stressed cells (lanes 5 and 6), however compared to the untransfected and unstressed controls (lanes 3 and 4), Nrf-2 protein is increased despite siRNA treatment in the transfected and stressed group (lanes 9 and 10). Abundance of Atf-4 protein remained at very low/undetectable level in the cells regardless of treatment. Actin loading control was used to ensure the same amount of protein was added to each well. Independent samples (n=2) were used for each treatment group. Blot is representative of four experiments conducted by Dr Alison Howard.
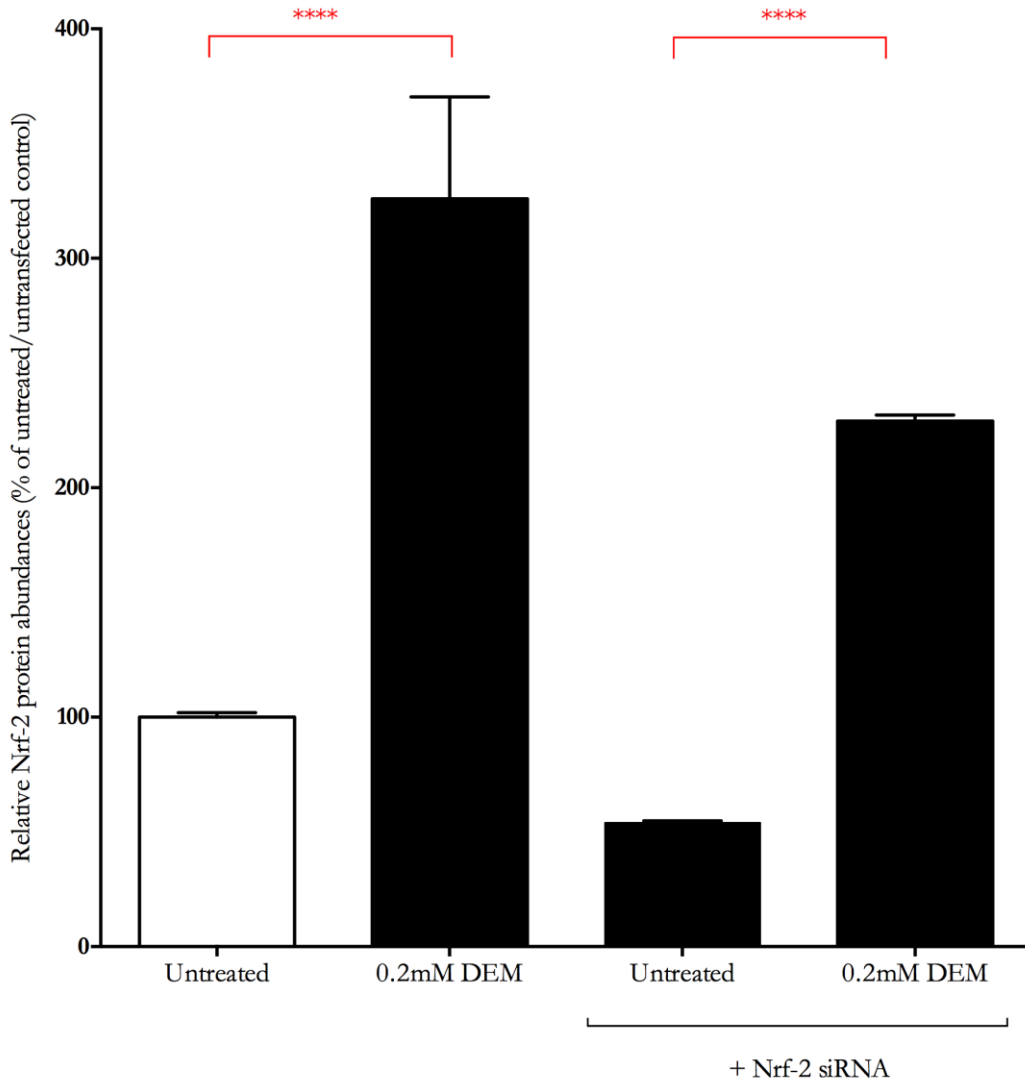
Figure 3.55b: Densitometry of total Nrf-2 protein following DEM treatment. Caco-2 cells treated with 0.2mM DEM showed significantly increased Nrf-2 protein abundance of about 3 fold. Pre-treating the cells with Nrf-2 siRNA attenuates the protein in both unstressed (untreated) and stressed. Despite siRNA transfection in these cells, DEM treatment increased Nrf-2 protein in a manner consistent with the untransfected controls. Data is from three independent experiments, n=2-3 per experiment. **** indicates value significantly different from control ($p < 0.001$).

3.5.5 Nrf-2 does not bind to an ARE identified in the GlyT-1a promoter

To determine if an outcome of the increases in Nrf-2 protein following DEM treatment was increased binding to the putative ARE-like (TGACCAAGCC) sequence identified in CRM8 of the human GlyT-1a gene, electrophoretic mobility shift assays (EMSA) using a 50bp DNA probe sequence from the GlyT-1a promoter were performed. As shown in Figure 3.56, the /5IRD800/G1Prom probe was designed to span between 559 and 509 bases upstream of the GlyT1a TSS. Sense and anti-sense strands were synthetically generated and annealed to form a duplex tagged at its 5'-end with an IRD-800 dye (/5IRD800/G1Prom). This probe was incubated with nuclear extracts from either unstressed or DEM treated Caco-2 cells and the resulting protein DNA compounds analysed by PAGE.
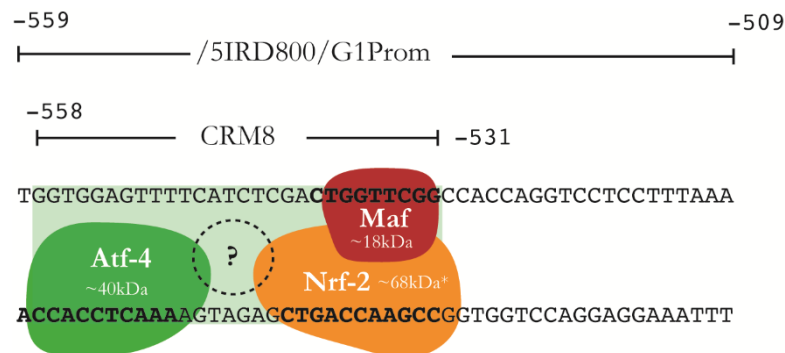
Figure 3.56: Illustration showing the /5IRD800/GlProm EMSA probe sequence for Nrf-2 interaction analysis. Sense and anti-sense oligonucleotides were synthesised to include sequence between 509 and 559 bases upstream from the GlyT-1a transcription start site (TSS). The arrangement of a predicted protein complex at the contained CRM is shown.

Although no unique bands of low mobility protein-DNA complexes were seen on representative gel images capturing the mobility of the /5IRD800/G1Prom probe, inspection of the 'flow through' of unbound probe in these images highlights some important differences (Figure 3.57). In unstressed cells, there is a clear difference in the amount of unbound probe (highest mobility) passing through the gel when the incubation included nuclear protein (NP; lane 8) than when it included cytosolic protein (CP; lane 9); the level of flow through obtained with NP in unstressed cells is similar to that seen in a reaction were excess competitor probe was included (lane 10).

The amount of unbound probe in unstressed NP is almost identical to competitive binding reactions containing excess unlabelled probe sequence (lane 10). Conversely, when cells are stressed, there is less flow-through with NP than CP (lane 2 – NP from DEM treated cells; lane 5 - NP from amino acid starved cells). One possible explanation is that stress results in nuclear translocation of cytosolic proteins which are then able to bind to the /5IRD800/G1Prom probe.

To improve the resolution of detection of protein-DNA interactions by EMSA, the incubation period for binding reactions was increased from 1 hour at room temperature to overnight (16 hours) at 4°C. As can be seen in Figure 3.58, whilst no retarded bands were detected in reactions containing only the /5IRD800/G1Prom probe (lane 1) multiple faint bands, possibly indicative of protein-DNA interactions are evident when the probe is incubated with crude NP extracts from unstressed Caco-2 cells (lane 2), those treated with 0.2 mM DEM for four hours (lane 3) or amino acid starvation for eight hours (lane 4).

To determine if any of these bands represented complexes of Nrf-2 bound to the /5IRD800/G1Prom probe, super-shift assays were performed using an anti-Nrf-2 antibody to further retard the migration of any Nrf-2 bound complex. As shown in Figure 3.59, no significant band shifts were detected in super-shift experiments using the /5IRD800/G1Prom probe with NP extracts from unstressed control cells, cells starved of amino acids for 8 hours, or cells treated with 0.2mM DEM for four hours. This suggests that Nrf-2 does not bind to the ARE-like sequence in the postulated CRM8 of GlyT-1a.

1	2	3	4	5	6	7	8	9	10	
	0.2mM DEM			AA starvation.			Unstressed			
+	+	+	+	+	+	+	+	+	+	/5IRD800/G1Prom
-	+	-	+	+	-	+	+	-	+	Nuclear proteins (NP)
-	-	+	-	-	+	-	-	+		Cytosolic proteins (CP)
-	-	-	+	-	-	+	-	-	+	Unlabelled competitor

Possible protein/DNA complexes but also seen in Lane 1

Note the differences in amount of unbound probes between sample groups
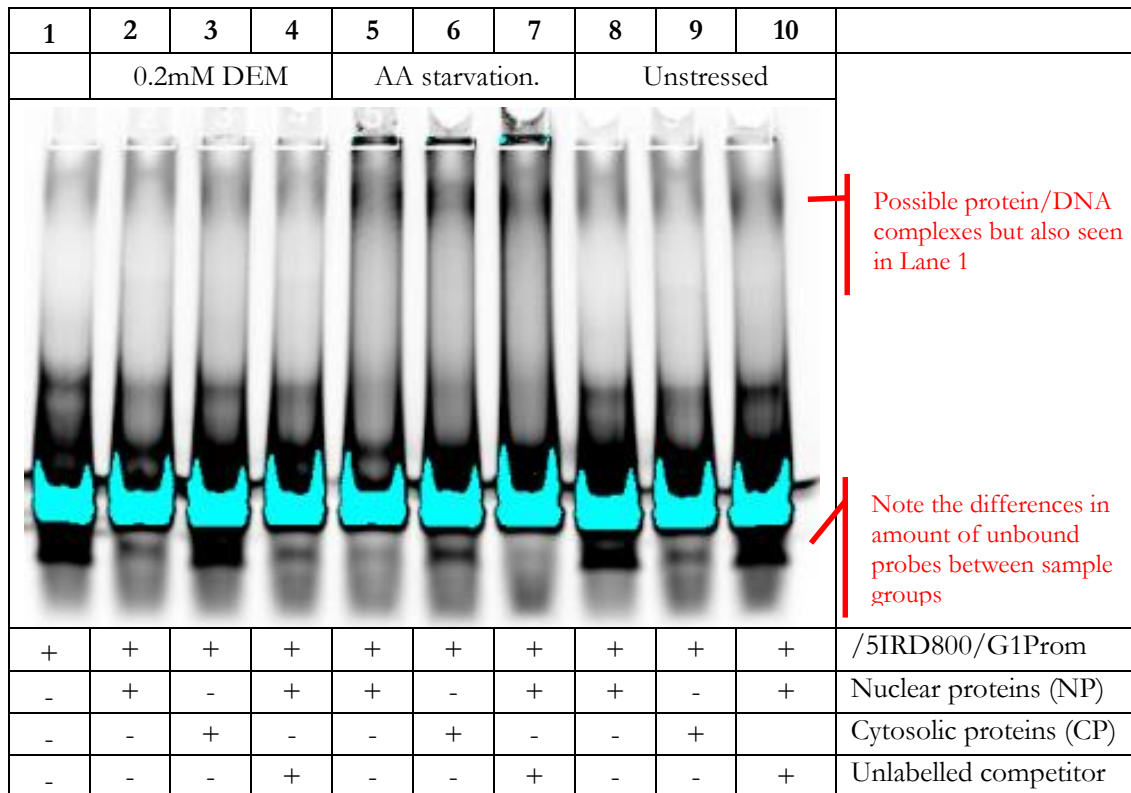
Figure 3.57: Gel image showing the binding to the /5IRD800/G1Prom in crude cytosolic (CP) and nuclear (NP) extracts from stressed and unstressed Caco-2 cells. The /5IRD800/G1Prom probe is representative of bases -559 to -509 relative to the TSS of the human Glyt-1a gene and labelled with IRD800. A low mobility band, suggestive of probe-DNA binding is observed in all lanes, but its appearance in lane 1 (no protein control) suggests this is an artefact of the probe. The amount of unbound probe in each reaction varies dependent on the nature of the protein with which the probe was incubated; in stressed cells, nuclear protein binds a greater proportion of the probe whereas in unstressed cells, cytosolic protein is more effective at binding the probe. Competition assays which included an excess of unlabelled probe (lanes 4, 7 and 10) showed no effect of the competitor on the binding profile, do not change the binding profile observed. Samples were collected in a single experiment but the depicted gel is representative of a total of three independent experiments.

	Unstressed	DEM	AA Starvation	
1	2	3	4	
				Possible protein/DNA complexes (seen in lanes 2, 3 and 4 but not in the probe only lane 1).
				Possible protein/DNA complexes seen in lane 4 only
				Protein/DNA complexes seen in lane 4 and lane 3 but not in lane 2.
+	+	+	+	/5IRD800/G1Prom
-	+	+	+	Nuclear proteins (NP)

Figure 3.58: Gel image showing the binding to the /5IRD800/G1Prom probe. Binding reactions were incubated overnight at 4°C. Lane 1 does not include any protein whilst lane 2, 3 and 4 include NP respectively from unstressed cells, cells treated with 0.2mM DEM for 4 hours and cells starved of amino acids for 8 hours.

Possible protein/DNA complexes but seen throughout lanes 17, 18, 19, 20 21. No band shifts are observed

Lane	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	
	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	/5IRD800/G1Prom
	-	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	Nuclear proteins (NP)
	-	-	-	+	-	+	-	-	+	-	+	-	-	+	-	+	-	-	+	-	+	-	-	+	-	+	Anti-Nrf-2 antibody
	-	-	-	-	+	+	-	-	-	+	+	-	-	-	+	+	-	-	-	+	+	-	-	-	+	+	Unlabelled competitor
		Unstressed					0.2mM DEM					1µM Tu					AA starvation					200mM Tg					
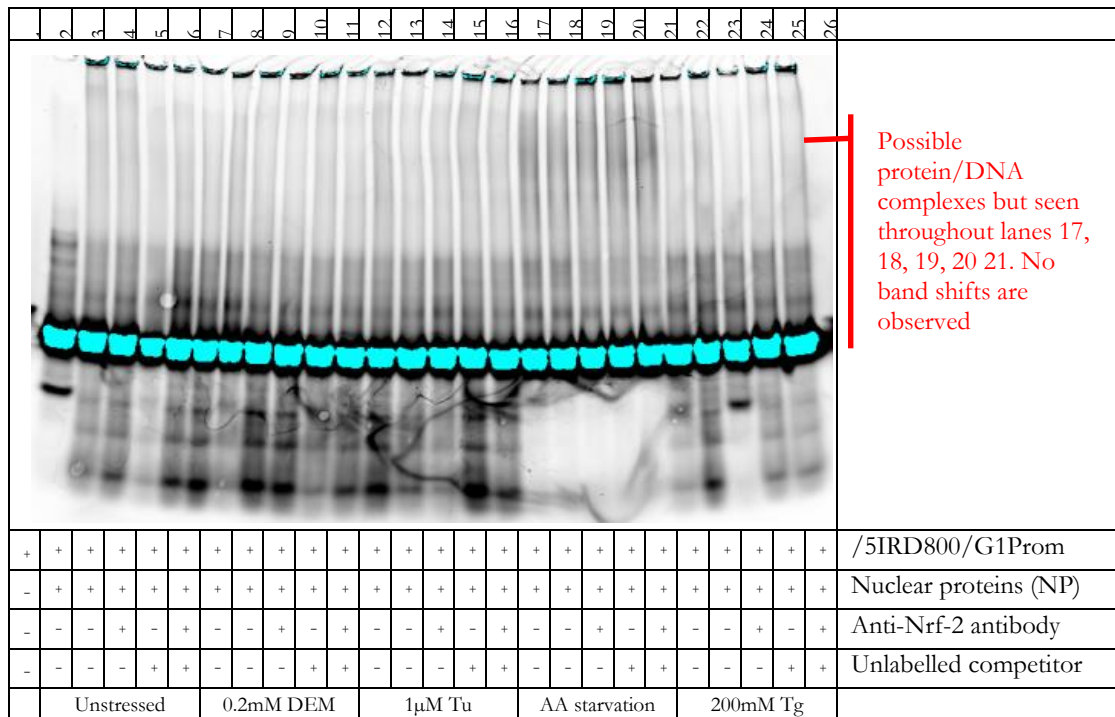
Figure 3.59: Gel image of super-shift assays with the /5IRD800/G1Prom probe and nuclear (NP) extracts from stressed and unstressed Caco-2 cells. Lane 1 does not include any NP or antibody. No band shifts are seen in super-shift experiments with NP from unstressed cells (lanes 2-6), DEM treated cells (lanes 7-11), cells treated with tunicamycin (lanes 12-16), cells starved of amino acids for 8 hours (lanes 17-21), and thapsigargin treated cells (lanes 21-26)

Taken together, that Nrf-2 mRNA levels are unchanged by stress and that despite increases in Nrf-2 protein following DEM stress (even in cells treated with siRNA to Nrf-2) it does not bind to a predicted GlyT-1a ARE, suggests that Nrf-2 is not necessary for transcriptional regulation of GlyT-1a. In a study to measure expression of Nrf-2 target genes in human lymphoid cells downstream of oxidative stress, Chorley *et al* sequenced Nrf-2 bound genomic regions captured by ChIP (Chorley *et al.* (2012) data in supplementary file). 96% of the 845 sequences to which Nrf-2 bound contained an ARE-like sequence. Whilst sequences upstream of GlyT-1 were not enriched by ChIP, Nrf-2 was detected bound to an ARE-like sequence 56 bases upstream from the transcriptional start site of xCT/SLC7A11. This reinforces the many reports on Nrf-2 regulation of xCT.

3.5.6 Nrf-2 knockdown does not affect GlyT-1a expression

Despite a significant reduction in Nrf-2 total cell protein following the treatment of Caco-2 cells with Nrf-2 siRNA, there was no effect on the expression of GlyT-1a in these cells (Figure 3.60).
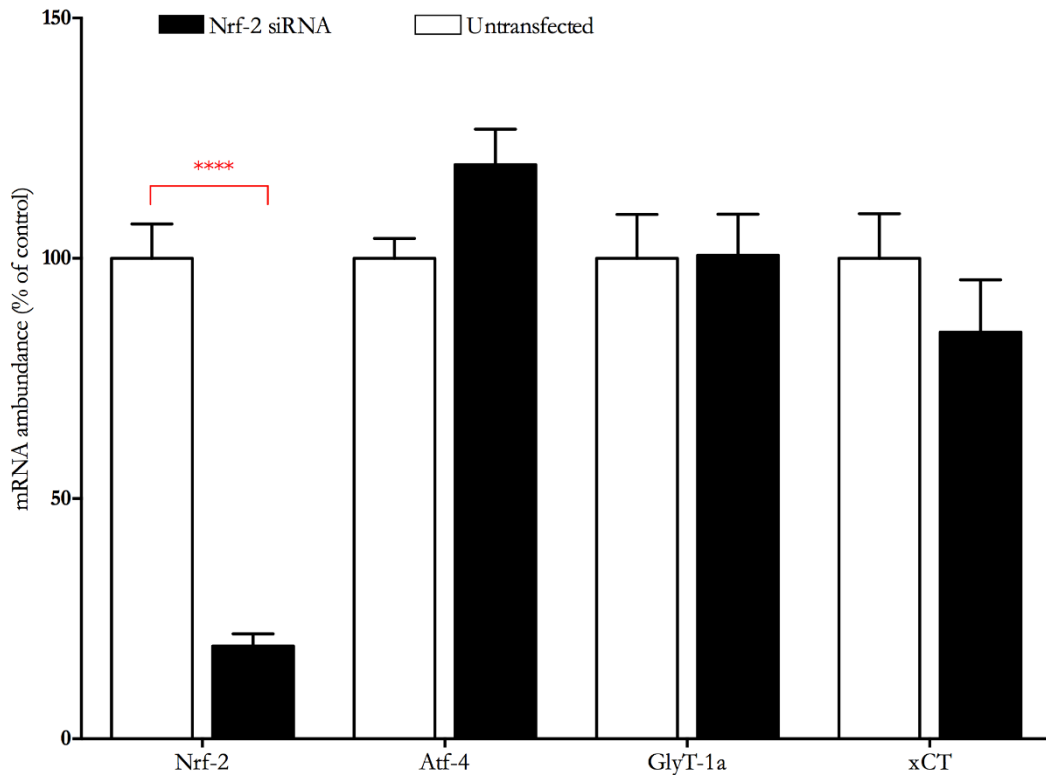
Figure 3.60: Nrf-2 knockdown has no effect on GlyT-1a mRNA. Although Nrf-2 expression is significantly decreased in Caco-2 cells transfected with specific Nrf-2 siRNA, this reduction has no effect on Atf-4, GlyT-1a or xCT expression. Bars represent mean percentage of the control $\pm$ SEM, n=2-3. One-way analysis of variance (ANOVA) was performed to compare changes in mRNA expression between transfected (solid bars) and untransfected (open bars) samples. **** indicates a *P*-value of < 0.001

3.6   Regulation of GlyT-1a transcription by Atf-4

Atf-4, a member of the large CREB/ATF subfamily of BZIP factors is an established transcriptional activator of the integrated stress response (ISR). As discussed earlier, in response to stress, the ISR connects a series of adaptive response pathways to the transcriptional regulation of proteins enabling eradication of the stressor or cell death in cases of prolonged stress. The convergence of these adaptive processes at phosphorylation of eIF-2α at the ser51 residue by several ER resident kinases results in the blocking of protein translation and the preferential up regulation of ATF4 mRNA translation via upstream open reading frames (uORF). Such translation regulation was first described in the Atf-4 yeast homolog GCN4 (Hinnebusch, 1984) and later shown to exist in mammalian cells (Harding *et al.*, 2000).

Atf-4 was first identified as a protein bound to the tax responsive elements (TRE) of the human T-cell leukaemia virus type 1 (HTLV-1) and was historically referred to as the tax responsive enhancer element B67 or TaxREB67 (Tsujimoto *et al.*, 1991). In the year after its identification, Karpinski *et al.* (1992) cloned and characterised a protein as a cAMP response elements (CRE) binding protein CREB-2 (note that a different protein from that described by Karpinski *et al.*, and now known to be Atf-2 was also historically referred to as CREB-2 or CRE-BP1 Ozawa *et al.* (1991)). Owing to the fact that CREB-2 (Atf-4) repressed the expression of the human pro-encephalin (PENK) gene when bound to CRE elements in its promoter it was originally described as a negative regulator of transcription (Karpinski *et al.*, 1992, Bartsch *et al.*, 1995). Whilst it is likely that by virtue of its dimerisation partners Atf-4 may repress gene expression (Soda *et al.*, 2013), Atf-4 is widely regarded as a transcriptional activator (Rutkowski and Kaufman, 2003). Upon activation, Atf-4 co-ordinates the transcriptional regulation of genes in response to several stress types, including nutrient deprivation, ER stress, hypoxia, viral infection etc. Although the primary mode of Atf-4 activation is by post-translational modification, its transcript is also up regulated by the ISR (Harding *et al.*, 2000, Harding *et al.*, 2003, Dey *et al.*, 2010). We therefore investigated the effect of stress on both ATF4 mRNA and protein abundance in Caco-2 cells to determine if it may be involved in the stress induced increase in Glyt1a mRNA abundance demonstrated in Section 3.1. As shown in Figure 3.61, treating Caco-2 cells with either 1μM tunicamycin or 50nM thapsigargin for 16 hours resulted in significant doubling of Atf-4 mRNA in these cells ($P < 0.001$).
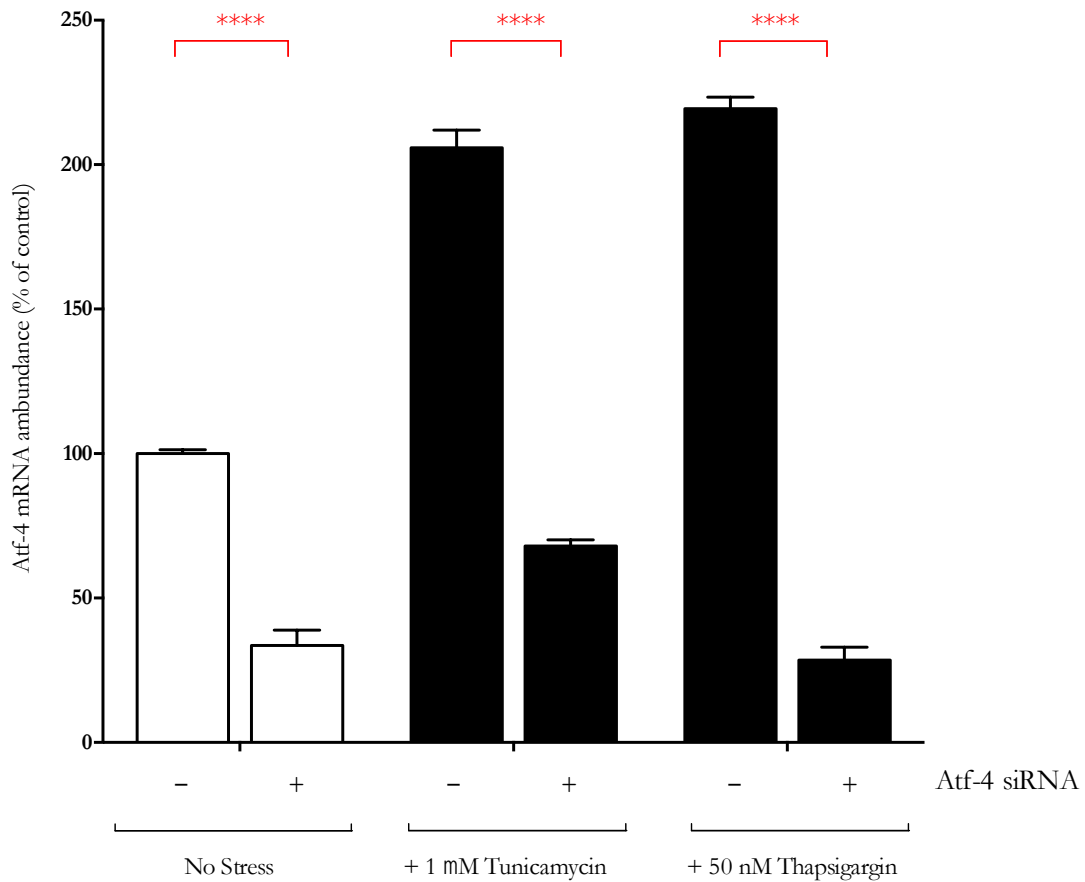
Figure 3.61: Atf-4 mRNA abundance is increased following stress. Treatment of Caco-2 cells with tunicamycin or thapsigargin results in a similar increase in Atf-4 mRNA abundance for both conditions as measured by QPCR. Pre-treating cells with specific siRNA to ATF-4 whilst effectively knocking down Atf-4 expression in unstressed cells (open bars) also significantly blunts the observed increases in Atf-4 expression following stress treatment (closed bars) by either tunicamycin or thapsigargin. Bars represent data expressed as a percent of the no stress, no siRNA control $\pm$ SEM (n=4-8. N=2-3). One-way analysis of variance (ANOVA) was performed to compare changes in Atf-4 expression between transfected and un-transfected samples within each treatment group. **** represents P-value of less than 0.001

As expected, the magnitude of this response and of basal level of Atf-4 mRNA were attenuated in cells transfected with siRNA targeted to Atf-4 prior to the initiation of stress. Experiments investigating the effect of Atf-4 knockdown used one of either two specific siRNA targeting different regions of the Atf-4 mRNA or a mixture containing three different Atf-4-specific siRNAs (See Section 2.4.1, page54 for further details). All

transfections, regardless of the siRNA used, gave the same result as that shown in Figure 3.61. These data suggests that there is either increased transcription of the Atf-4 gene during stress or that stress enhances the stability of pre-existing mRNA to prolong its lifespan. However the data described in Section 3.1 which shows that GlyT-1 mRNA is increased under the same conditions as that used here would support the former i.e. that increased transcription of Atf-4 may be required for additional activation of GlyT-1 mRNA during stress.

To determine if transcriptional changes in Atf-4 mRNA were consistent with changes at the protein level, Atf-4 protein abundance in Caco-2 cells treated with tunicamycin and transfected with ATF4 specific siRNAs was examined by Western blotting (Figure 3.62). In Caco-2 cells a band of 48kDa and at low intensity in control untransfected, untreated cells, was increased in density in cells treated with either tunicamycin or thapsigargin. This was diminished to control level in cells transfected with either ATF4 specific siRNA prior to treatment.

Amongst the many Atf-4 target genes are molecular chaperones such as GRP78, DNA damage proteins including the CCAAT/enhancer protein homology protein (CHOP or C/EBP-ζ) and amino acid transporters such as SNAT-2 (Palii *et al.*, 2006), xCT (Sato *et al.*, 2004) and Cat-1 (Fernandez *et al.*, 2003). Previous work showed increases in GlyT-1a mRNA in Caco-2 cells following stress treatment (Figure 3.1 page72). The knockdown of Atf-4 in Caco-2 cells by specific siRNA against Atf-4 significantly reduced Atf-4 mRNA expression (Figure 3.61) and total protein abundance (Figure 3.62). The effect of the reduced Atf-4 knockdown on GlyT-1a mRNA was investigated. Measurement of GlyT-1a mRNA abundance in the same samples indicated that the knockdown of Atf-4 also reduced the expression of GlyT-1a in both stressed and unstressed cells (Figure 3.63). This suggests that Atf-4 is required for both the basal and stress-induced GlyT-1a expression. As seen in lane 2 of the blot image in Figure 3.62, Atf-4 itself is protein abundance is increased when Caco-2 cells are stress when compared to unstressed control cells. Although the low protein abundance in unstressed cells (evident from the high differential between stress and unstressed cells, and the fact that the control band intensity is much lower than that in the knockdowns) may in part reflect the quality of the antibody used in this work, it may be a true indication of the abundance of the protein in the absence of stress. If so, this suggests that only a low Atf-4 titre is required to drive the basal transcription of the GlyT-1 gene.

	1	2	3	4	5	6	
	+	-	-	-	-	-	Molecular Ladder
	-	+	-	-	-	-	Control
	-	-	+	-	-	-	Tunicamycin (16 hours)
	-	-	-	+	-	-	Atf-4 siRNA (24 hours) + Tunicamycin
	-	-	-	-	+	-	Thapsigargin (16 hours)
	-	-	-	-	-	+	Atf-4 siRNA (24 hours) + Thapsigargin
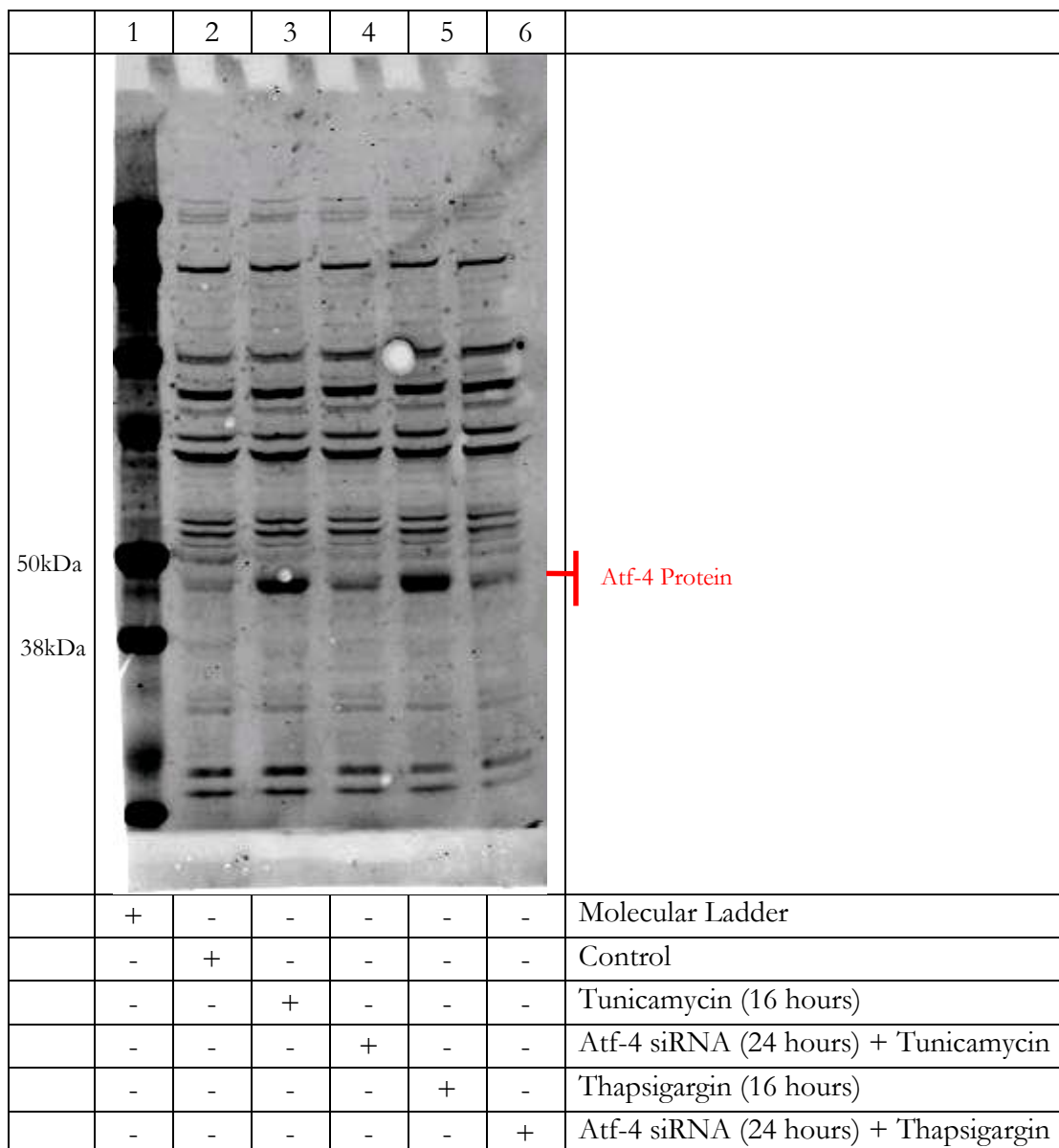
Figure 3.62: Treating cells with Atf-4 siRNA results in decreased Atf-4 protein. Caco-2 cells treated with tunicamycin (lane 3) or thapsigargin (lane 5), increased the intensity of the 48kDa band corresponding to Atf-4 when compared to the unstressed control (lane 2). Pre-transfecting the cells prior to stress with Atf-4 siRNA (lanes 4 and 6) diminished protein titres as shown by the reduced band intensity. Note that the commercially available Atf-4 antibodies detect multiple non-specific bands as well as the specific Atf-4 band which runs at the apparent molecular weight of 48kDA. All other additional bands shown on this blot are of unknown identify. The blot image shown here is representative of data from at least four similar experiment. Western blot experiment was conducted by Dr Alison Howard.

It is worth noting at this point that several attempts at using Western blots to follow changes in GlyT-1a abundance in stressed Caco-2 cells have been carried out. However, neither commercially available nor self-generated antibodies have been successful at detecting GlyT-1 in these cells. Control experiments using these antibodies on tissues with expected high GlyT-1 expression (such as brain tissue, spinal cord and neuronal derived cells) have given ambiguous results, suggesting that this is an issue of reagent quality rather than the absence of the GlyT-1 protein. Nonetheless, GlyT-1 expression at an mRNA and protein level has been confirmed in Caco-2, HCT-8 cells and human intestinal samples by immunohistochemistry using antibodies that are no longer commercially available (Christie *et al.*, 2001, Howard *et al.*, 2010).

That GlyT-1a expression is reduced following Atf-4 knockdown is consistent with microarray data from work carried out by Lange *et al* that shows a decrease in GlyT-1 differential expression in cortical neurones of Atf-4 knockout mouse when compared to wild-type controls (See Table S3 in Lange *et al.*, 2008). The reduction in GlyT-1a mRNA following the knockdown of Atf-4 was consistent in both unstressed and stressed (tunicamycin or thapsigargin treated) Caco-2 cells. A similar reduction in GlyT-1a mRNA was not seen following either Xbp-1, Atf-6 knockdown (in unpublished data from this lab) or Nrf-2 (see Section 3.5.6); suggesting that Atf-4 and none of these other transcription factors are crucial to transcriptional regulation of GlyT-1.

To determine if this regulation was part of a non-specific up-regulation of transporters driven simply by an increased requirement for amino acids, the effect of Atf-4 knock down on the mRNAs for transporters, xCT – which as described previously is thought like GlyT-1 to provide amino acids for glutathione synthesis; and PepT1 – which has no specific role in glutathione synthesis; was investigated. As shown in Figure 3.64, Atf-4 knockdown had the same effect on xCT mRNA as it did on GlyT-1a mRNA i.e. it is reduced in both stressed and unstressed Caco-2 cells. In contrast, expression of the peptide transporter was not influenced by Atf-4 knockdown in either basal or stressed conditions (Figure 3.65). That the knockdown of Atf-4 in these cells, had no real effect on the PepT1 mRNA following stress, but did on GlyT-1 and xCT supports the hypothesis that a specific coordinated transcription regulatory network controls the import of amino acids such as glycine and cysteine for function in roles beyond general protein synthesis.
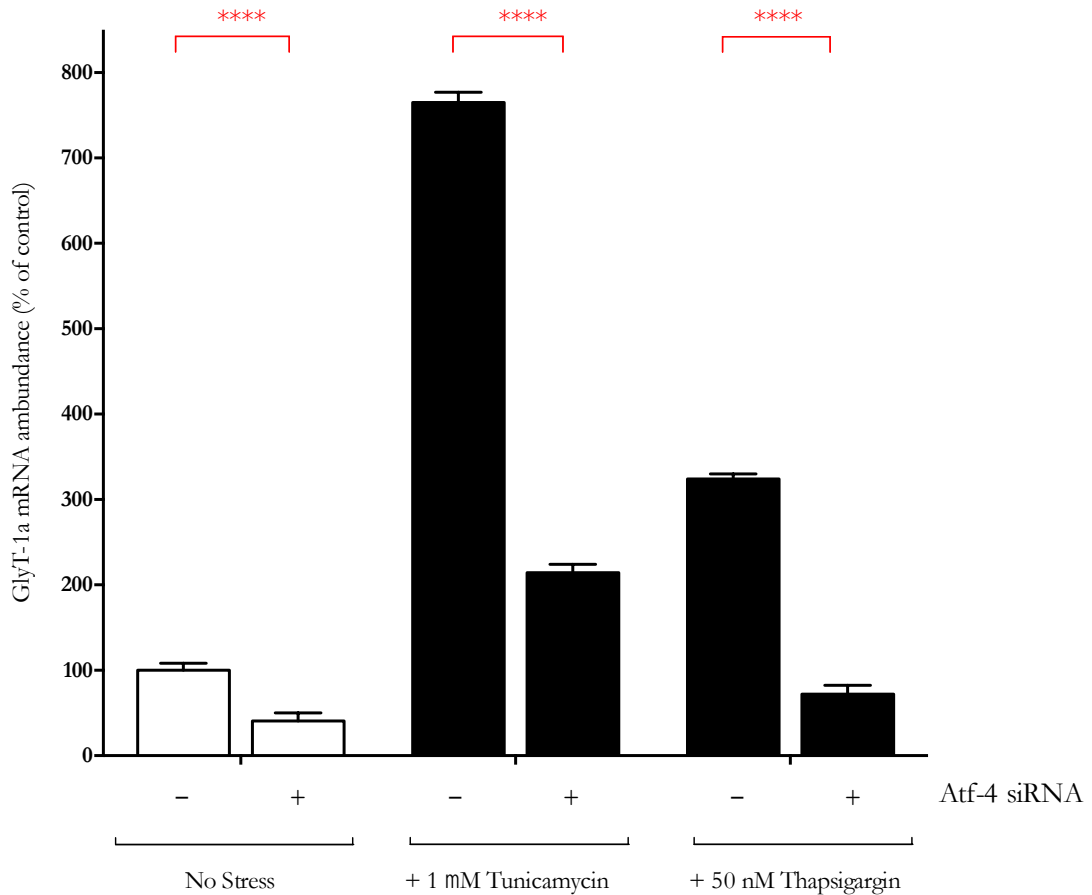
Figure 3.63: Atf-4 knockdown in Caco-2 cells decreases GlyT-1a expression. Whilst GlyT-1a expression is significantly increased in total RNA extracted from tunicamycin or thapsigargin treated Caco-2 cells (as measured by QPCR), transfecting these cells with specific siRNA to ATF-4 significantly decreases GlyT-1a mRNA abundance in unstressed cells (open bars) but also significantly attenuates the increased expression in stressed cells (closed bars). Data is presented as percent of no stress, no transfection control $\pm$ SEM (n=3-6, N=2). One-way analysis of variance (ANOVA) was performed to compare changes in GlyT-1a expression between transfected and untransfected samples within each treatment group (i.e. unstressed, tunicamycin and thapsigargin). **** indicates a P-value of < 0.001.
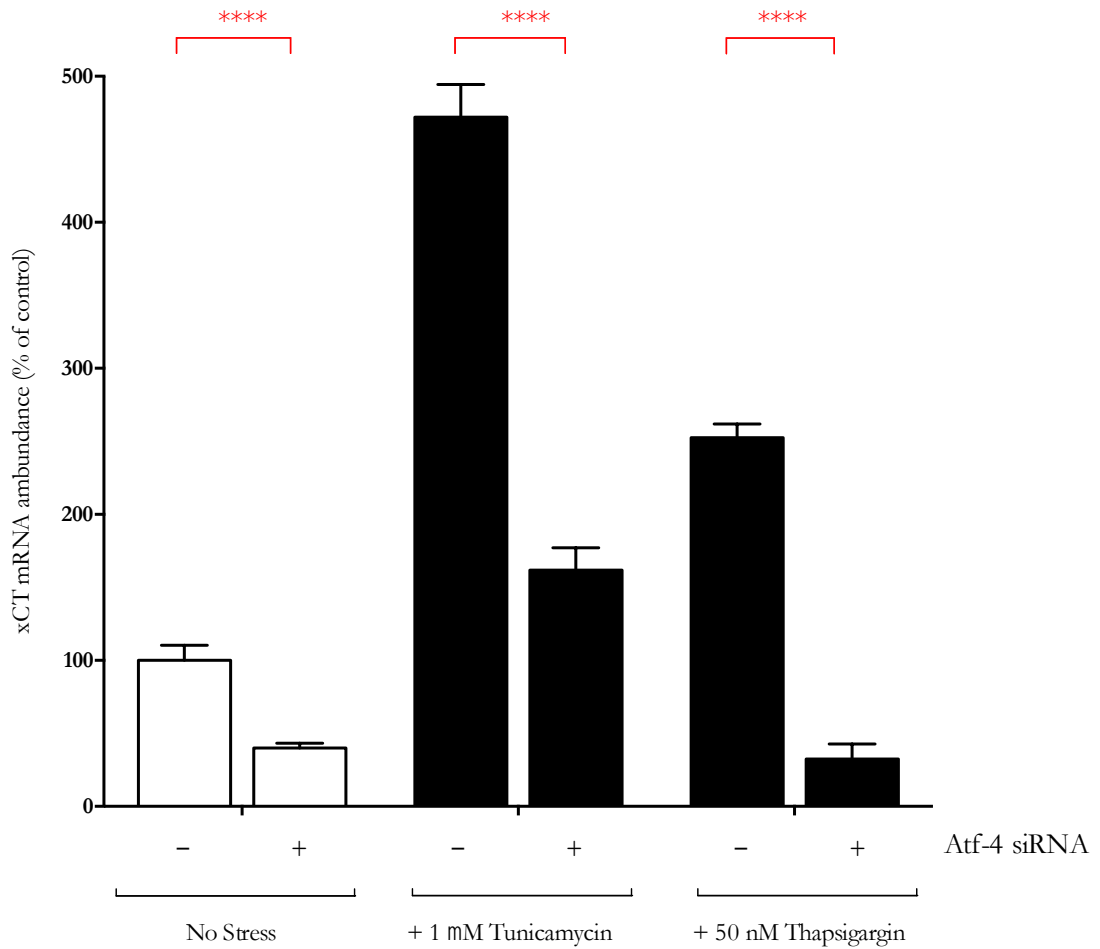
Figure 3.64: Atf-4 knockdown in Caco-2 cells decreases xCT expression. In untransfected cells xCT expression is increased in total RNA extracted from tunicamycin or thapsigargin treated Caco-2 cells (as measured by QPCR). Transfecting these cells with specific siRNA to ATF-4 significantly decreases xCT mRNA abundance in unstressed cells (open bars) but also significantly attenuates the increased expression in stressed cells (closed bars) in a pattern similar to those observed with GlyT-1a (Figure 3.63). Data is presented as a percent of the untransfected and unstressed control ± SEM, (n=3-6, N=2). One-way analysis of variance (ANOVA) was performed to compare changes in xCT expression between transfected and untransfected samples within each treatment group (i.e. unstressed, tunicamycin and thapsigargin). **** indicates a *P*-value of < 0.001.
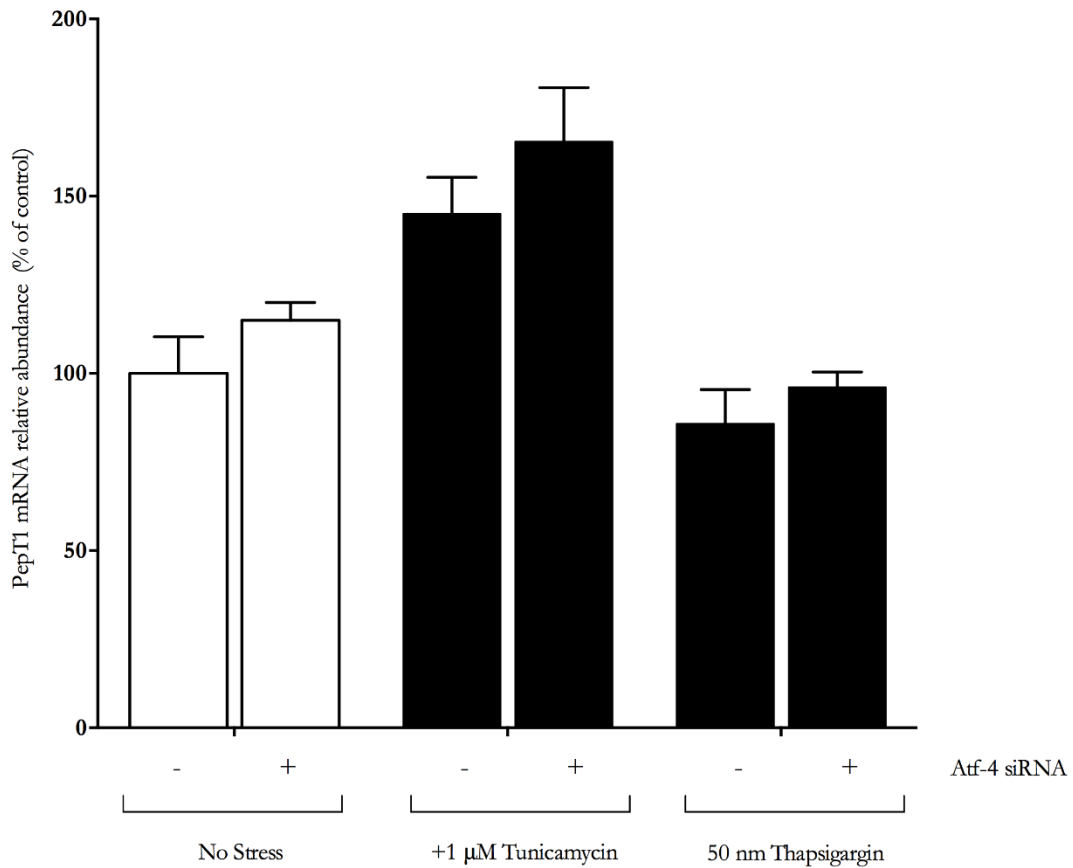
Figure 3.65: Atf-4 knockdown has no effect on PepT1 expression in Caco-2. Transfecting these cells with specific siRNA to ATF-4 had no effect on PepT1 mRNA abundance in unstressed cells (open bars) or stressed cells (closed bars) contrary to the effects of Atf-4 knockdown on GlyT-1a mRNA (Figure 3.63). Data is shown as percent of the untransfected and unstressed control $\pm$ SEM (n=3-5, N=3). One-way analysis of variance (ANOVA) was performed to compare changes in PepT1 mRNA abundance between transfected and untransfected samples within each treatment group (i.e. unstressed, tunicamycin and thapsigargin). No statically significant differences were noted.

Similar effects of Atf-4 knockdown have been described on other genes important to the response to several type of stress. In HeLa cells transfected with Atf-4 specific siRNA, CHOP mRNA was significantly reduced (See figure 4 in Averous *et al.*, 2004). In that study, the reporter activity of constructs containing the CHOP AARE required for Atf-4 mediated transcription was reduced in HeLa cells pre-transfected with Atf-4 siRNA and subject to amino acid starvation. Likewise, and consistent with the findings of this study, a reduction in activity of reporter constructs representing the xCT promoter was evident when these constructs where co-transfected with Atf-4 siRNA into mouse embryonic fibroblast cells (See figure 2G in Lewerenz and Maher, 2009). Subsequent work by Lewerenz *et al.*, also agrees with the results reported here, in that they found that Atf-4 siRNA treatment of Caco-2 cells reduced the mRNA expression of xCT in both stressed and unstressed cells (Figure 3.64). Atf-4 knockdown in HT-22 cells by two independent siRNA effectively reduced xCT mRNA expression (See supplementary figure 3 in Lewerenz *et al.*, 2012). Given that downstream of stress treatment xCT mRNA was up regulated co-ordinately with that of GlyT-1a (Section 3.1), and that as shown here Atf-4 knockdown resulted in the down regulation of both GlyT-1a and xCT, it is probable that both GlyT-1 and xCT are co-regulated by Atf-4. This is supported by similar data from Ait-Ghezala *et al.* who showed that increased Atf-4 activation associated with depletion of the eukaryotic release factor 3a (eRF-3α) in HCT-116 cells resulted in a 2.35 fold increase in GlyT-1 expression and a 3.08 fold increase in xCT expression (See table 2 in Ait Ghezala *et al.*, 2012).

Treating BV-2 cells with the cannabinoids tetrahydrocannabinol (THC) and cannabidiol (CBD) both known to deplete cellular glutathione (McKallip *et al.*, 2006) leads to the up regulation of both xCT and GlyT1 (See table 2 in Juknat *et al.*, 2012). In addition to depleting glutathione levels and induction of oxidative stress, cannabinoids are known to induce ER stress and subsequent apoptosis in an Atf-4 dependent way (Carracedo *et al.*, 2006, Salazar *et al.*, 2009). These observations point to the possibility of a common Atf-4 dependent regulatory programme for both xCT and GlyT-1 downstream of both oxidative, nutrient and ER stress.

The transcriptional control of xCT by Atf-4 has been studied. Sato *et al.* documented the tandem repeat of two characteristic AAREs in forward and reverse orientations upstream from the TSS of the mouse xCT gene (Sato *et al.*, 2004). EMSA assays showed that Atf-4 bound to probes containing either the most 5' AARE or both AAREs whilst the mutation

of one or both AAREs (which are separated by just nine bp) significantly diminished the activity observed in promoter constructs. Similar forward and reverse AAREs have been shown in the human xCT gene, between 64 and 90 bases upstream from the TSS (Ye *et al.*, 2014). 113 bases upstream from the human xCT TSS is a reverse anti-oxidant response element (ARE) to which Ye *et al* propose (without providing any experiment evidence) that Atf-4 as well as Nrf-2 may bind.

The GlyT-1a flanking sequence was analysed to computationally detect the occurrence of composite regulatory modules (CRMs), consisting of TFBS sites for Atf-4 like those found upstream of the xCT TSS. As discussed in Section 3.2, ten CRMs were identified within 2000 bases upstream from the GlyT1a sequence. Whilst each of these CRMs may warrant individual experimental validation, the characteristics and arrangement of potential binding sites in CRM10 (Figure 3.11c, page101) make it interesting for several reasons. Firstly, the Atf-4 binding site of sequence (ACTACGTTG) within this module is identical to the reverse complement of AARE sequences described in several other stress target genes (Table 3.2), including those in the promoter sequences of the asparagine synthetase (ASNS) and CHOP genes, and that in the first intron of the SNAT-2 gene. Secondly, this highly conserved putative regulatory module (CRM10), located in the first exon of GlyT-1a (between 104 and 134 bases downstream from the GlyT-1a TSS) also contains a TFBS validated by the ENCODE project to bind C/EBPβ (Consortium, 2012). Similarly positioned CRMs downstream of the TSS have been described in other Atf-4 regulated genes. Fernandez *et al.* (2003) reported on the transcriptional activity of Atf-4 bound at an AARE of sequence TGATGAAAC located in the first exon of the arginine/lysine transporter – Cat-1 gene. Dimerisation between Atf-4, C/EBP-β and Atf-3 at the AARE of Cat-1 gene are known to regulate its spatio and temporal transcription following physiological stress (Fernandez *et al.*, 2003, Lopez *et al.*, 2007, Huang *et al.*, 2009). A further example is the human *tribbles* homolog TRB3, a suspected pro-apoptotic protein kinase up-regulated by Atf-4 and known to mediate the degradation of C/EBP-β proteins (Hattori *et al.*, 2003). Three tandem repeats of C/EBP and Atf-4 composite sites of identical sequences (TGATGCAAA) were identified in the first exon of TRB3 gene (Ohoka *et al.*, 2005). Finally, sites for Atf-4, Atf-6 and C/EBPβ binding have also been identified in the first exon of the microtubule-associated protein 1 light chain 3β (MAP1LC3B) (Rzymski *et al.*, 2010).

Table 3.2: Amino acid response elements in stress target genes. Start stop locations are with respect to the transcriptional start site. Binding factors shown are those confirmed to interact with the respective sites. Table adapted from (Brasse-Lagnel *et al.*, 2009).

AARE Sequence	Regulated Gene	Featured Location	Start	Stop	Orientation	Binding factors	Cell Model
TGATGAAAC	ASNS	5'-flanking	-68	-60	Forward	C/EBPβ, Atf-4	HeLa
CATGATG	ASNS	5'-flanking	-70	-64	Reverse	?	HepG2
TGATGCAAC	Atf-3	5'-flanking	-23	-15	Forward	C/EBPβ, Atf-4, Atf-3	HepG2
ATTGCATCA	CHOP	5'-flanking	-310	-302	Reverse	C/EBPβ, Atf-4, Atf-2	HeLa
TGATGCAAA	xCT	5'-flanking	-94	-86	Forward	Atf-4	NIH/3T3
TTTGCATCA	xCT	5'-flanking	-78	-68	Reverse	Atf-4	NIH/3T3
TGATGAAAC	Cat-1	Exon 1	+45	+53	Forward	C/EBPβ, Atf-4, Atf-3	Rat C6
TGATGCAAA	TRB3	Exon 1	+228	+236	Forward	CHOP, Atf-4	
TGATGCAAA	TRB3	Exon 1	+262	+269	Forward	CHOP, Atf-4	
TGATGCAAA	TRB3	Exon 1	+294	+302	Forward	CHOP, Atf-4	
ATTGCATCA	SNAT-2	Intron 1	+712	+724	Reverse	CHOP, C/EBPβ, Atf-4	HepG2
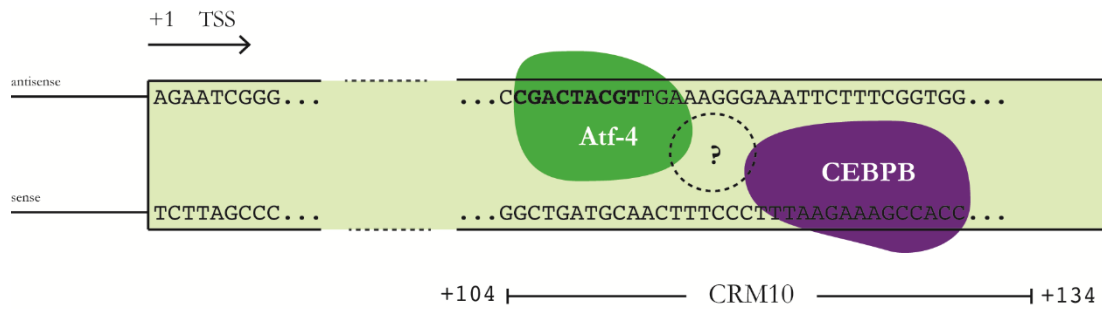
Figure 3.66: Hypothetical interaction of Atf4 and other factors at CRM10. The Atf-4 TFBS site of sequence ACTACGTTG is 85.9% similar to the consensus Atf-4 binding site described in the TRANSFAC database. Similar sequences have been described in other Atf-4 target genes. Downstream from the Atf-4 binding site in this module is a TFBS able to bind C/EBP. C/EBP-β protein was positively identified bound to the GlyT-1a sequence (AAGAAAGCCACC) in ChIP-seq experiments conducted by the ENCODE project. Dimers of Atf-4 and various C/EBP proteins have been shown to regulate expression of several genes downstream of amino acid stress, ER stress, oxidative stress and nutrient stress (Kilberg *et al.*, 2012, Kilberg *et al.*, 2009). As to whether the arrangement of the individual TFBS in this composite regulatory module (CRM) allows for such dimerisation between Atf-4 and C/EBP proteins needs to be experimentally tested.

3.6.1 Stress induced protein factors complex to a GlyT-1a probe

To determine if downstream of stress, the Atf-4 protein effectively bound to the putative AARE-like (TGACCAAGCC) sequence identified in CRM10 of the first exon of GlyT-1a (Section 3.2.4.3, page108), electrophoretic mobility shift assays (EMSA) were performed using an 88bp DNA probe sequence from the GlyT-1a promoter. As shown in Figure 3.67, the probe, name /5IRD700/G1Exon1, was designed to span between 97 and 187 bases downstream from the GlyT1a TSS. Sense and anti-sense strands tagged at the 5' end with IRD-700 dye where synthetically generated and annealed to form a duplex. This was incubated with extracts of nuclear protein from Caco-2 cells and the formation of protein-DNA complexes confirmed in gel shift assays by virtue of their relative migration on a polyacrylamide gel. From PAGE results of independent samples shown in Figure 3.68, incubating the /5IRD700/G1Exon1 probe with nuclear protein extracts (NP) from Caco-2 cells stressed for 16 hours with 1μM tunicamycin resulted in at least two bands. This is in line with expectations; if only the two predicted factors (Atf-4 and C/EBP-β) bind to the CRM in the probe, a maximum of four bands including that

representative of the unbound probe are possible based on the differing molecular sizes of the binding protein (Figure 3.67).
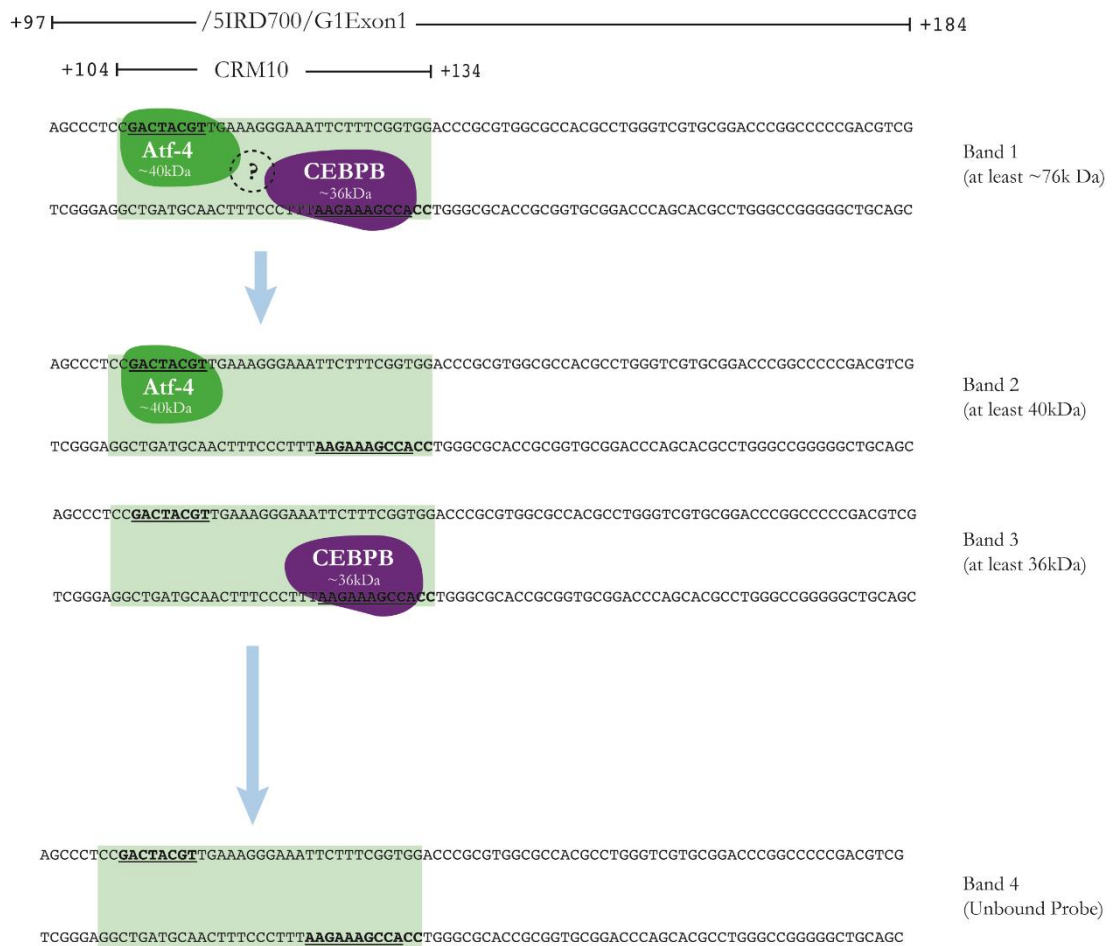
Figure 3.67: Illustration of the /5IRD700/GlExon1 probe sequence showing possible binding patterns for Atf-4 and C/EBP-β. Based on the hypothetical model of interaction, should just Atf-4 and C/EBP bind to the CRM contained in the probe four complexes of different mobility are predicted. Band 1, with maximal occupancy of the TFBS in the CRM should have lowest mobility, by virtue of the summarily high molecular weight. Monomeric interactions between both Atf-4 and C/EBP should result in two additional bands (Band 2 and 3), whilst Band 4 would represent the high mobility migration or 'flow through' of the unbound probe.

A band with the lowest relative mobility (RF) is expected, if both Atf-4 and C/EBP-β bind to their respective halves in the regulatory composite module. Should either Atf-4 or C/EBP-β monomers bind separately to the probe, two further bands of relatively higher mobility are expected as the only variability is in the molecular size of Atf-4 (~40kDa) and C/EBP-β (~36kDa) given that the molecular weight of the probe is constant and small in comparison to that of either protein. However it is possible that the 7.5% PAGE gel used here is unable to resolve the 4kDa difference in molecular weight between Atf-4 and C/EBP; so the bands may appear as one. Finally, a high mobility band of unbound probe should be visible at the bottom of the gel image.

As can be seen in Figure 3.68, a band of high mobility (identified as the unbound probe by comparison to the signal obtained in lane 1 from a reaction where no NP was added) is seen in all lanes but with varying band intensity. In all other lanes a further 2 bands of retarded mobility are observed. Lanes 3, 5 and 7 in Figure 3.68, represent competition reactions that include an unlabelled probe representative of the validated AARE contained in the CHOP gene. A 10-fold molar excess of the competitor was added to binding reactions containing the labelled /5IRD700/G1Exon1 and nuclear extracts from 1μM tunicamycin treated Caco-2 cells. This resulted in a visible difference in intensity of the higher mobility unbound probe at the bottom of the gel, but did not appear to affect the intensity of either of the two lower mobility bands. Looking only at lanes 2 and 3, the data would suggest that the competitor is able to displace the labelled probe from the bound protein resulting in a diminished intensity for band 1 and enhanced flow through of the free probe (band 3). However when lanes 4-7 (which contained the products of identical reactions to lanes 2 and 3 but with different NP samples), are considered this does not appear to be the case. In these samples, the competitor caused an increase in flow through of the unlabelled probe, suggesting it has been displaced from a binding protein, but there is no apparent reduction in the intensity of either of the shifted bands. Thus, although the gel image does show that factors in NP from tunicamycin stressed cells bind to the probe, interpretation from these data on the effect of the competitor is ambiguous.

1	2	3	4	5	6	7	8	9	10	
	Reactions with NP from cells treated with 1μM Tunicamycin									
+	+	+	+	+	+	+	+	-	+	/5IRD700/G1Exon1
-	+	+	+	+	+	+	+	-	+	Nuclear proteins (NP)
-	-	+	-	+	-	+	-	-	-	Unlabelled competitor

Band 1 of protein/DNA complexes. Seen in lanes 2-8 and lane 10.

Band 2 of protein/DNA complexes. Seen in lanes 2-8 and lane 10.

Band 3, unbound probes
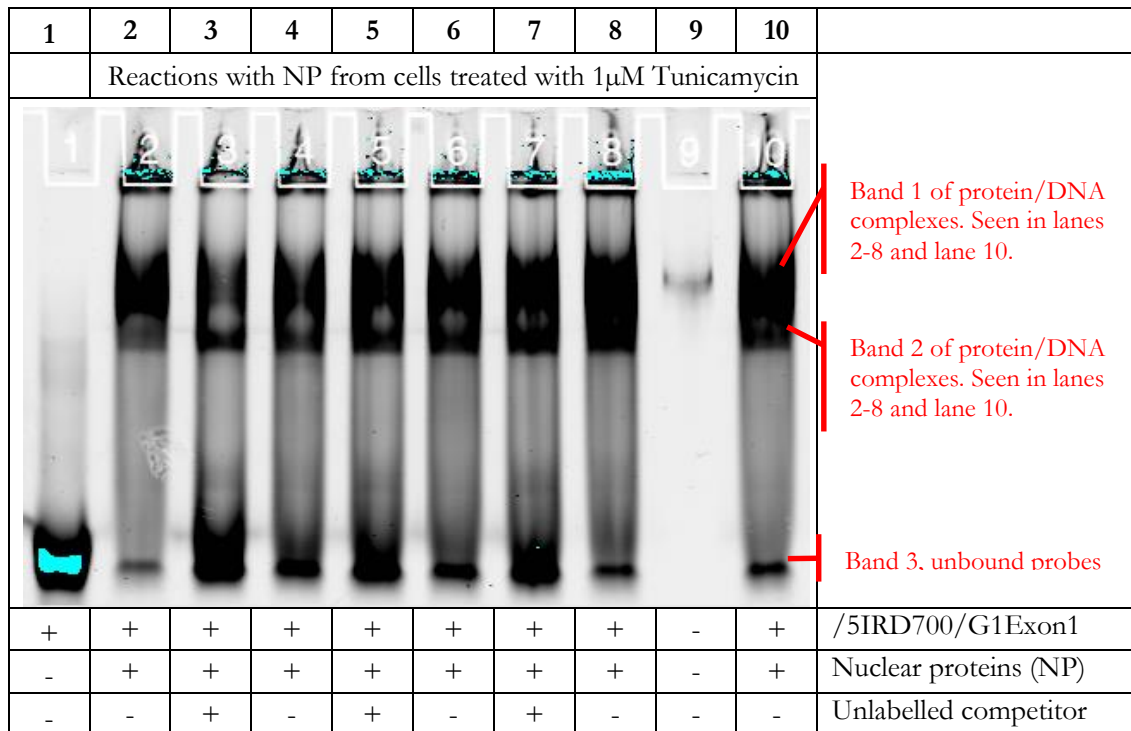
Figure 3.68: Transcription factors in nuclear extract from tunicamycin treated Caco-2 cells bind to the GlyT-1a CRM10 probe (/5IRD700/G1Exon1). The figure shows the result of PAGE imaged with a LICOR Odyssey infrared scanner. Lane 1 shows free probe, lanes 2, 4, 6, 8 and 10 the result of incubating nuclear extracts with the labelled probe and lanes 3, 5 and 7 the effect of including excess unlabelled competitor probe in the reaction. Note the nuclear extracts used for the reactions in lanes 2, 4, 6, 8 and 10 are from different cell samples; reactions for lanes 2 and 3, 4 and 5, and 6 and 7, were paired in that for example lanes 2 and 3 contained the same nuclear extract. For lanes 2-10 5μg of protein was loaded; for lane 1 the same amount of free probe as used in all other lanes was loaded. Well 9 of the gel was damaged and left empty. The image shows the migration of three bands corresponding to two DNA/protein complexes formed with each binding condition, and the excess unbound probe (indicated as bands 1-3 respectively in the right hand side panel). Note the differences in intensities of the unbound probe (band 3) in lanes to which the unlabelled competing CHOP AARE sequence was added. This gel is representative of two separate experiments (N = 2).

Comparing the mean of band 1 intensities for lanes 2, 4 and 6 with that of lanes 3, 5 and 7 showed no difference (Figure 3.69), suggesting that the competitor had no effect on probe binding. Experiments subsequent to this have given results similar to those in lanes 4-7 rather than in lanes 2 and 3. Similary attempts at optimisation, including altering the ratio of probe:NP or NP:competitor have not provided an explanation for this result. It is clear that the inclusion of the competitor in the binding reaction has an effect on probe binding but the nature of that effects remains to be determined.

.
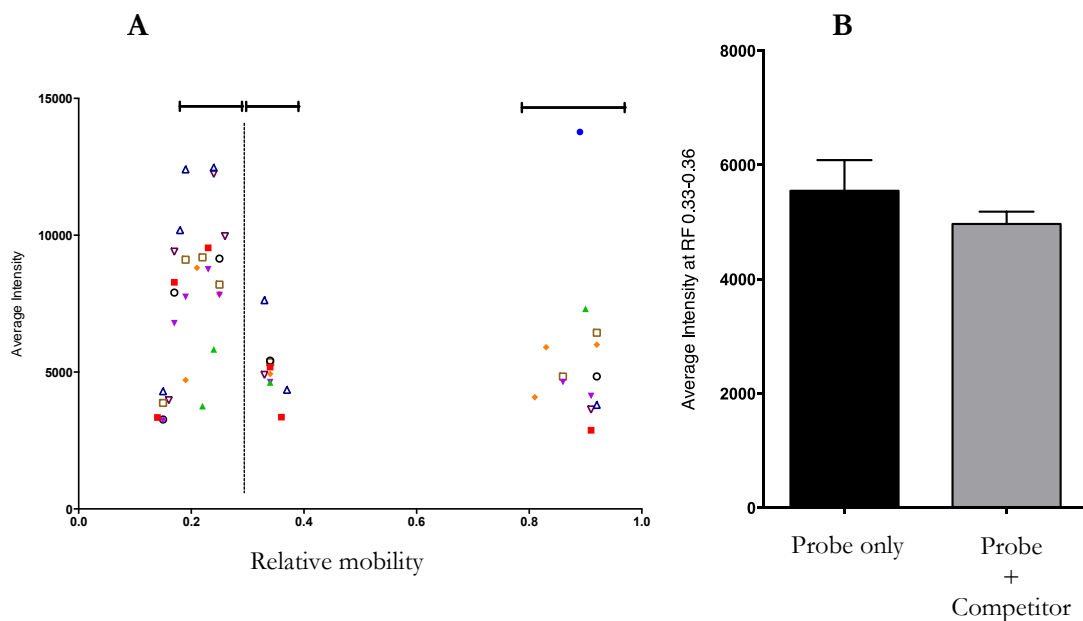
Figure 3.69: Plots of average intensities of EMSA gel bands from Figure 3.68. The average band intensity provides an approximation of how much of the probe is retained in the protein-DNA complex. The relative mobility (RF) score gives a percentage migration through the gel, where 0 is for bands with no migration at the top of the gel and 1 represents complete migration to the bottom of the gel. **(A)** The average intensity of each detected band peak (represented by each point on the graph) from each lane (represented using symbols of the same colour on the graph) of the gel image plotted against the RF representing the migration from the top of the gel [RF=0] as a function of molecular weight. Clustering reveals three distinct peak groups (indicative of bands) corresponding to the two DNA/Protein complexes with average RF of 0.2 and 0.35, and the excess DNA probe with average RF of 0.9. **(B)** Comparison of the average intensity of the bands with average RF of 0.35 in binding reactions to which the probe (/5IRD700/G1Exon1) and competitor were added (lanes 3, 5, 7 of Figure 3.68) or those to which only the probe was added (lanes 2, 4, 6, 8). The Student's t-test comparing the two groups revealed no statistical significance between the average intensities, suggesting similar binding affinities (n=3-4, $P > 0.05$)

In summary, these data suggests that a protein or other factors within the nuclear extracts of tunicamycin treated Caco-2 cells bind to sites contained in the /5IRD700/G1Exon1 probe, possibly at the AARE in the predicted composite regulatory module.

To determine if the bands observed are not just specific to tunicamycin treated Caco-2 cells, gel shift assays were then carried out using nuclear protein extracts from these cells following treatment with thapsigargin, DEM or amino acid starvation. Nuclear extracts from cells treated with 50nM thapsigargin for 16 hours were incubated with the probe as described above. As shown in Figure 3.70, unlike binding reactions with nuclear extracts from tunicamycin treated cells, where two distinct low mobility bands were present (Figure 3.68), nuclear extracts from thapsigargin treated cells only formed a single low mobility protein-DNA complex. It is therefore likely that the interaction of transcription factors at their response binding sites is dependent on the nature of the stress. In binding reactions of the /5IRD700/GlExon1 probe and nuclear extracts from Caco-2 cells treated for four hours with 0.2mM DEM, the intensity of the low mobility protein-DNA complex band is very low (Figure 3.71). This suggests that although transcription factors in the nuclear extracts from DEM treated cells may bind to TFBS in the probe, the affinity of interaction is not as high as with tunicamycin or thapsigargin. From these data it is impossible to determine if this is due to low titres of the bound TFs in the nuclear extract. The inclusion of excess unlabelled competitor probe into these binding reactions had no added effect on the intensities of either the retarded bands or the free probe, suggesting that protein binding to the /5IRD700/GlExon1 probe is at a site distinct from that contained in the competitor probe.

1	2	3	4	5	6	7	8	9	10	
	Reactions with NP from 50nM Thapsigargin treated cells									
+	+	+	+	+	+	+	+	-	+	/5IRD700/G1Exon1
-	+	+	+	+	+	+	+	-	+	Nuclear proteins (NP)
-	-	+	-	+	-	+	-	-	-	Unlabelled competitor

Band 1 of protein/DNA complexes in lanes 2-8 and lane 10.
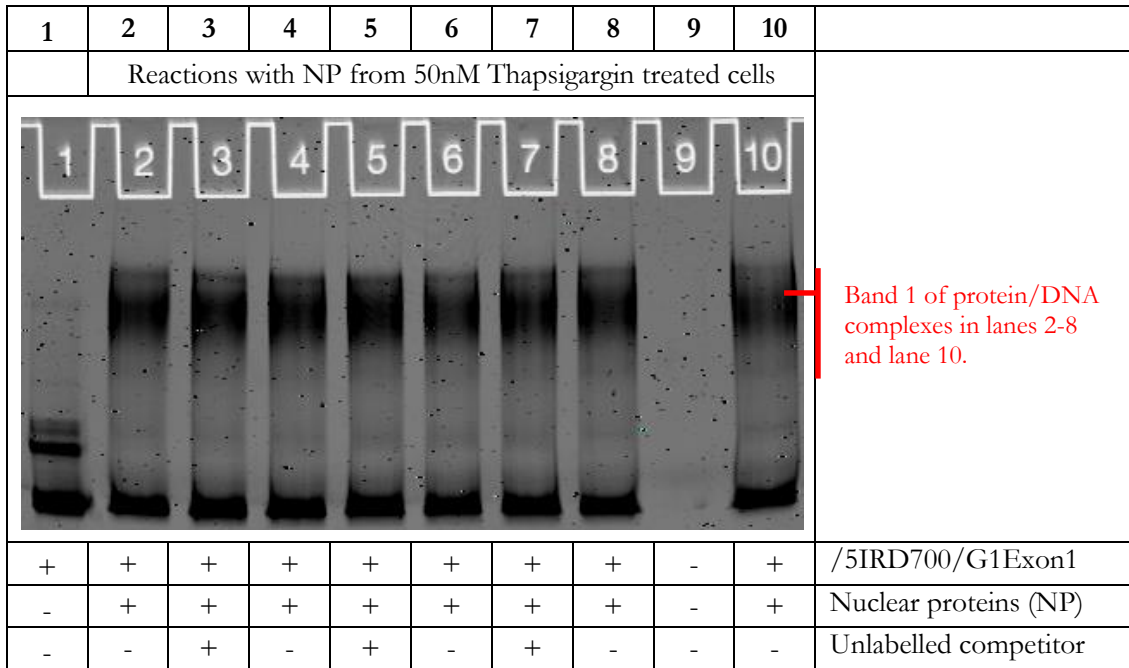
Figure 3.70: Transcription factors in nuclear extracts from thapsigargin treated Caco-2 cells bind to CRM10 of GlyT-1a. (/5IRD700/G1Exon1). PAGE gel imaged with a LICOR Odyssey infrared scanner showing the migration of two bands corresponding to a protein-DNA complex with low relative mobility and the excess unbound probe. Note that the addition of the unlabelled competitor had no effect on the binding intensities of the low mobility protein DNA/Complex. Different samples were used for each binding reaction and competition assay pair. Similar lane profiles were noted in two separate experiments (N = 2).

1	2	3	4	5	6	7	8	9	10	
	Reactions with NP from 0.2mM DEM treated cells									
										Note the low intensity diffuse band/s present in lanes 2-8 not visible in the probe only lane 1
+	+	+	+	+	+	+	+	-	-	/5IRD700/G1Exon1
-	+	+	+	+	+	+	+	-	-	Nuclear proteins (NP)
-	-	+	-	+	-	+	-	-	-	Unlabelled competitor
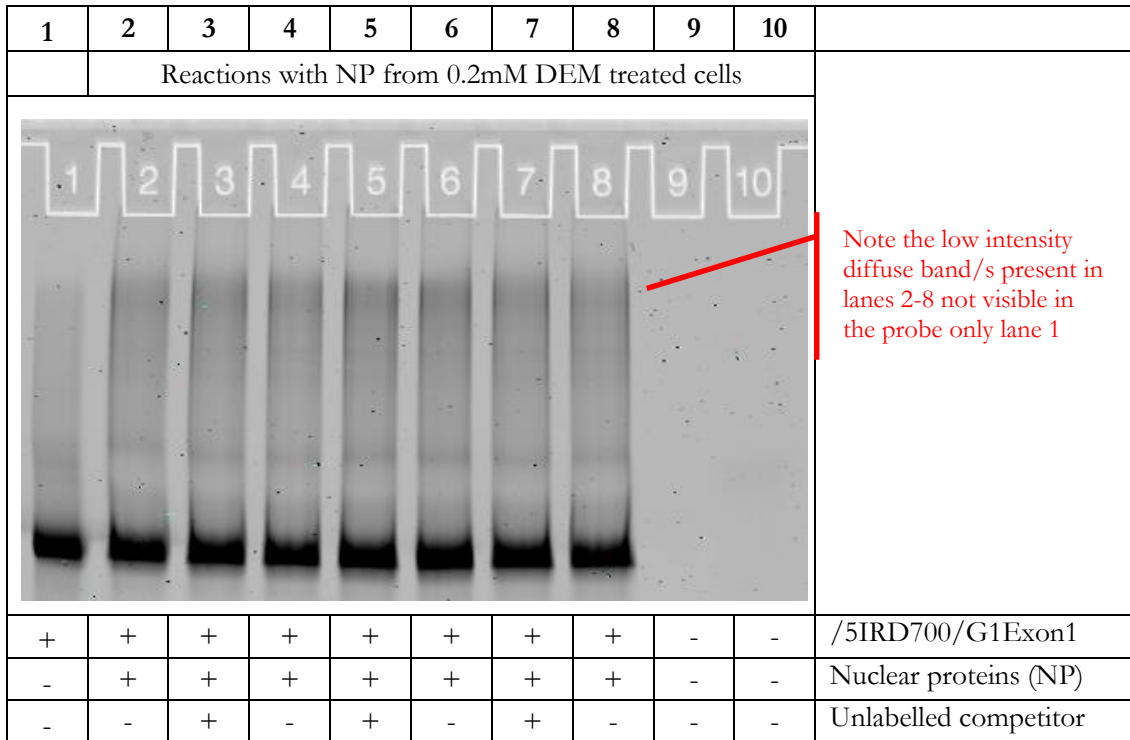
Figure 3.71: Transcription factors in DEM treated Caco-2 cells bind to GlyT-1a CMR 10 with very low intensity. PAGE gel imaged with a LICOR Odyssey infrared scanner showing the migration of two bands corresponding to a protein-DNA complexes with low relative mobility and the excess unbound probe. Note that the addition of the unlabelled competitor had no effect on the binding intensities of the low mobility protein DNA/Complex. Gel is representative of two experiments (N = 2)

1	2	3	4	5	6	7	8	9	10	
	Reactions with NP from cells incubated in Amino acid free medium for 4 hours.									
+	+	+	+	+	+	+	++	++	-	/5IRD700/G1Exon1
-	+	+	+	+	+	+	++	++	-	Nuclear proteins (NP)
-	-	+	-	+	-	+	-	+	-	Unlabelled competitor
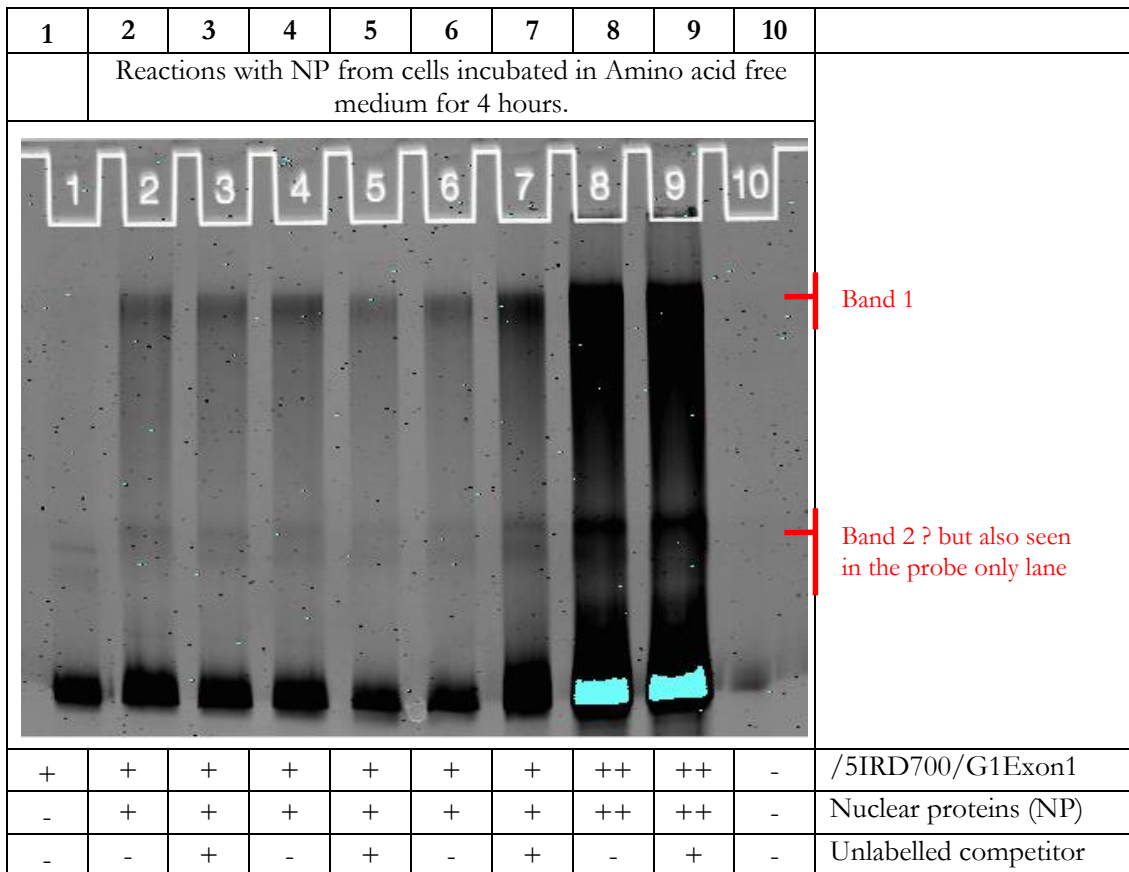
Figure 3.72: Transcription factors in nuclear extract from Caco-2 cells (incubated in amino acid free medium for 4 hours) bind to a GlyT1a probe (/5IRD700/G1Exon1). PAGE gel imaged with a LICOR Odyssey infrared scanner showing the migration of two bands corresponding to protein-DNA complex with low relative mobility and the excess unbound probe. Note that the addition of the unlabelled competitor had limited effect on the binding intensities of the low mobility protein DNA/Complex. Similar lane profiles were noted in three separate experiments (N = 3). NB: twice the amount of nuclear protein NP (++), or probe (++) was added to lanes 8 and 9.

Despite the fact that protein-DNA complex bands were seen in all EMSA assays of binding reactions using nuclear extracts from stressed Caco-2 cells, the data presented in Figure 3.68 - Figure 3.72 does not determine if any or all of the complexes identified result from interaction at the postulated AARE contained in the probe. To determine its role, if any, an 88 base pair probe (/5IRD700/G1Exon1/Mut) was generated. This probe spans the identical region of GlyT-1a as the /5IRD700/G1Exon1. The only difference between the two probes is in the double mutation of the predicted AARE from TG**A**TGCA**A**C; or reverse complement AC**T**ACGTT**G** to TG**T**TGCA**G**C; or reverse

complement AC**A**ACGT**C**G. The decision of which bases to change was guided by base-wise conservation scores in the Atf-4 binding consensus from the TRANSFAC database which showed these bases to be highly conserved.

As shown in Figure 3.73, binding reactions using either /5IRD700/G1Exon1 or /5IRD700/G1exon1/Mut resulted in varying numbers of shifted bands with differing intensities (Quantification traces are shown in Figure 3.74). Lanes 1 and 2 in Figure 3.73, represent incubations of just the wild type and the mutated probes respectively, and demonstrate that only a high mobility band representative of the unbound probe is seen for each. Using nuclear extracts from unstressed cells (lanes 3 and 4), generates two shifted bands (band 1 and 2 discussed above) which are evident for both the mutated and the wild type probe. This suggests that even in the basal state, nuclear proteins bind to this region of the GlyT-1a gene and stress may not be a pre-requisite for interaction at this site. The intensities of the shifted bands generated by both stressed and unstressed conditions is much reduced when the mutated probe was used, suggesting that the mutation reduces the ability of the probe to bind to, whatever DNA binding proteins.

When nuclear extracts from tunicamycin cells were including in binding assays with the wild-type probe (lane 5) four low mobility bands were evident. This differs from the previous result (Figure 3.68) were tunicamycin treatment resulted in two shifted bands, but was consistent between the present data and one further experiment. Of the four shifted bands, three (those with the lowest mobility) were reduced to very low intensities when the mutated probe was used (lane 6). This suggests that the mutation blocks the ability of the probe to bind the TFs responsible for the shift and that the AARE is crucial for binding of factors induced by tunicamycin treatment. The fourth highest mobility shift band (band 2, lane 6) had an apparent increase in intensity when compared to lane 5. A possible explanation for this observation is that the mutated AARE prevents certain TFs from binding and frees up more of the probe for binding to factors which interact at another site, thus increasing the intensity of this band. It is also possible that the mutation introduces a biologically functional site for other TFs which may be responsible for the band. Interestingly, as noted previously, this band is evident throughout all reactions using nuclear extracts from both stressed and unstressed samples (lanes 3-10). A similar increase in intensity of band 2 when the mutated probe was used, was seen only with nuclear extracts from DEM treated cells (lane 8). In untreated cells and those starved of amino

acids (lanes 3, 4 and 9, 10 respectively) the intensity of band 2 was reduced when the mutated probe was used in binding reactions.

The appearance of shifted bands generated with nuclear extracts from DEM treated cells and the wild-type probe (/5IRD700/G1Exon1) is similar to that obtained previously (Figure 3.71) in that there is a faint, diffuse band of very low mobility (band 1 in this figure). However, as previously noted, there is an additional band (band 2) that is increased in intensity when the wild-type probe is substituted for the mutated probe (lane 8). Furthermore, another band of intermediate mobility is evident with the mutated probe that is not present when using the wild-type probe. This suggests that the mutated AARE introduces an additional binding site for a protein factor active downstream of DEM treatment. The intensity of the diffuse, lower mobility band (band 1 in the figure) is not visibly altered between the two probes, suggesting that the mutation, unlike in nuclear extracts from tunicamycin treated cells, has no real effect on the binding of the factors involved here.

When cells were subjected to amino acid starvation, binding of nuclear extracts to the wild-type probe (lane 9) resulted in two shifted bands (bands 1 and 2, discussed above); with band 1 most likely equivalent to the lowest mobility, diffuse band evident in lanes 2 to 9 in Figure 3.72. The intensity of either of these two bands was not altered by mutating the probe (lane 10). As with DEM treatment, the mutated probe allows the formation of new DNA-protein complexes not evident with the wild-type probe. These differ in size from the additional bands seen with nuclear extracts of DEM treated cells, indicating that the array of TFs activated following amino acid starvation differs from that generated by oxidative stress.

Overall, the data described in this section demonstrates that the AARE is indeed required for both the basal and tunicamycin stimulated GlyT-1a expression, but appears to be of little significance in DNA-protein interaction following either DEM induced oxidative stress or the stress induced by amino acid starvation. In this way it highlights that the cells respond in different ways to the different stressors used in these experiments, however it does not show definitively whether binding of Atf-4 by the AARE is involved in any or all of them. The experiments described in the next section set out to determine this.

1	2	3	4	5	6	7	8	9	10	
		Control		Tunica.		DEM		No-AA.		
+	-	+	-	+	-	+	-	+	-	/5IRD700/G1Exon1
-	+	-	+	-	+	-	+	-	+	/5IRD700/G1Exon1/Mut
-	-	+	+	+	+	+	+	+	+	Nuclear proteins (NP)
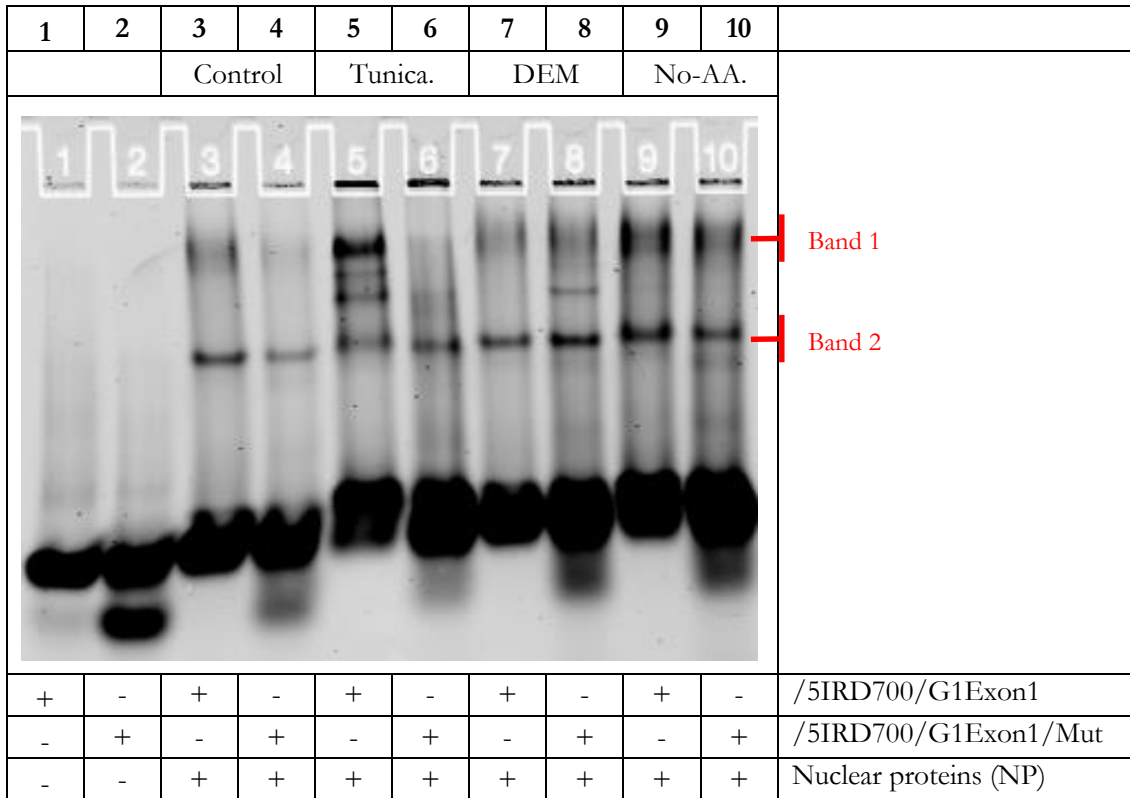
Figure 3.73: Gel image showing binding of nuclear proteins to the 5IRD700/G1Exon1/Mut probe. Nuclear protein (NP) extracts from control cells (untreated; lanes 3 and 4), tunicamycin treated cells (lanes 5 and 6), DEM treated cells (lanes 7 and 8) and amino acid starved cells (lanes 9 and 10) were incubated with either the mutated or the un-mutated probe (/5IRD700/G1Exon1). Gel image is representative of two separate experiments.

Control lanes	Tunicamycin	DEM	Amino acid starvation
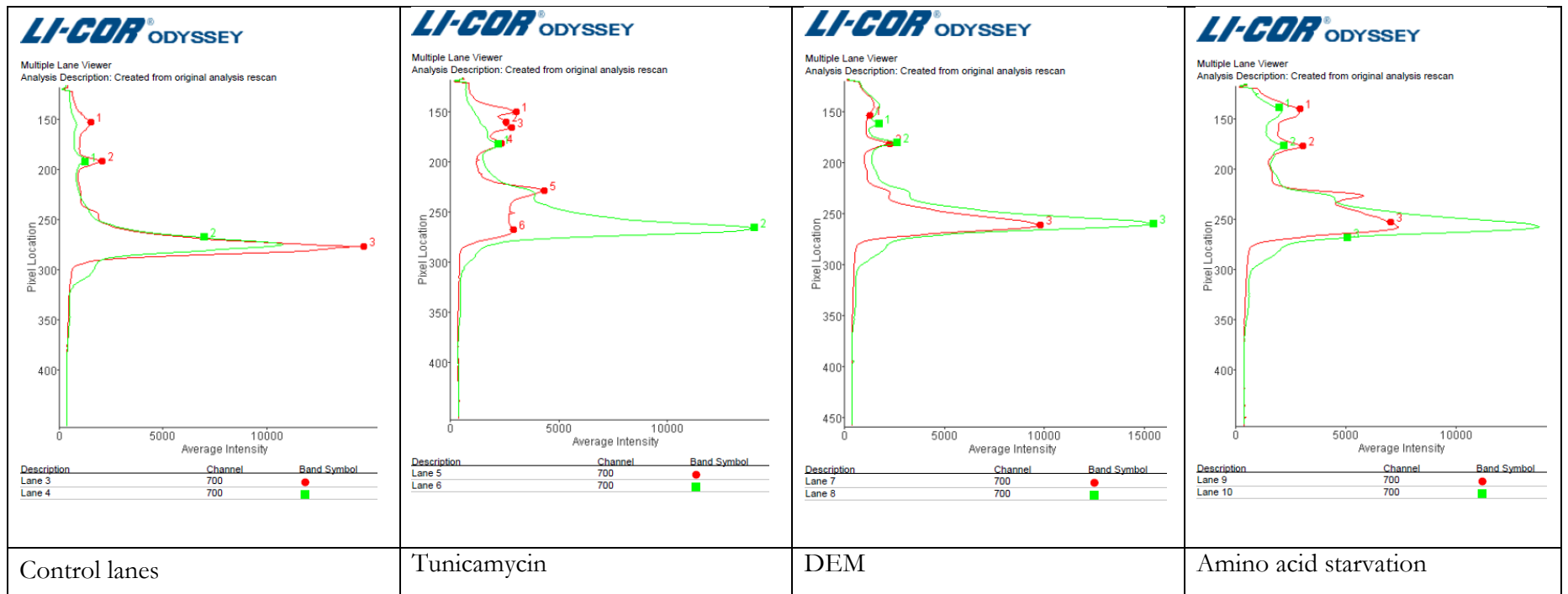
Figure 3.74: Lane profile analysis of gel shift assays in Figure 3.73 using either a probe with mutated AARE sequence (/5IRD700/G1Exon1/Mut; green trace) or wild-type probe (/5IRD700/G1Exon1; red trace). The polyacrylamide gel shown in Figure 3.73 was analysed using the LICOR Odyssey analysis software to determine the average intensity per pixel location of the gel.

3.6.2 Atf-4 binds to GlyT-1a AARE after tunicamycin and thapsigargin stress

To provide confirmatory evidence that Atf-4 interacts at the GlyT-1a AARE, gel super-shift assays and ChIP-PCR analysis were performed and are described in this section. Gel super-shift assays work on the principle that the binding of an antibody to a TF bound to its recognition site on the probe increases the molecular weight of the complex, thus decreasing its relative mobility through a PAGE gel. As can be seen in Figure 3.75, when nuclear protein extracts from tunicamycin (panel A), thapsigargin (panel B) and DEM (panel C) treated cells are incubated with the probe, protein-DNA complex bands 2 and 3 are evident. However when the anti-Atf-4 antibody is added to the binding reactions, an extra band (band 1) of even lower mobility is seen in lanes 3 and 4 (panel A), which is absent from lane 1 (panel A). This observation confirms that Atf-4 binds to the /5IRD700/G1Exon1 probe following tunicamycin treatment. Similarly, with nuclear extracts from thapsigargin treated cells an extra band (band 1) is seen in lane 5 (panel B) further retarded on the PAGE gel, than the low mobility bands (band 2 and 3) seen in lane 2 (panel B). Significantly, no further band shifts were seen following the incubation of nuclear extracts from DEM treated Caco-2 cells, the probe and the anti-Atf-4 antibody (panel C) suggesting that the transcription factors binding to the probe and producing bands 1 and 2 following DEM stress are not Atf-4.

Having confirmed the AARE in the GlyT1a first exon was a functional binding site for BZIP dimers of at least one Atf-44 subunit, the next experimental aim was to determine if this AARE was transcriptionally active following stress. To investigate this, ChIP analysis was performed in Caco-2 cells stressed by tunicamycin, amino acid limitation, or DEM as described in the Chapter 2, Section 2.7. DNA isolated from the ChIP assays using an antibody for Atf-4 were analysed by QPCR to determine the relative enrichment of a 100bp sequence spanning the GlyT-1a AARE.

QPCR assays show that following tunicamycin stress, the GlyT1a AARE sequence in ChIP DNA isolated using an Atf-4 antibody is enriched nine times compared to that in unstressed cells (Figure 3.76). With amino acid starvation and DEM treatment, this enrichment is respectively five and two times that in controls. This is consistent with the lower intensity of the protein-DNA complex bands seen in gel shift assays of binding reactions with protein extract from DEM treated or amino acid starved Caco-2 cells (Figure 3.71 and Figure 3.72).

A					B					C					
Tunicamycin					Thapsigargin					DEM					
1	2	3	4	5	1	2	3	4	5	1	2	3	4	5	
+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	/5IRD700/G1Exon1
+	-	+	+	-	-	+	+	-	+	-	+	+	-	+	Nuclear proteins (NP)
-	+	+	+	-	-	+	-	+	+	-	-	+	+	+	Anti-Atf-4 antibody
-	+	-	-	-	-	-		+	-	-	-	-	+	-	BSA control
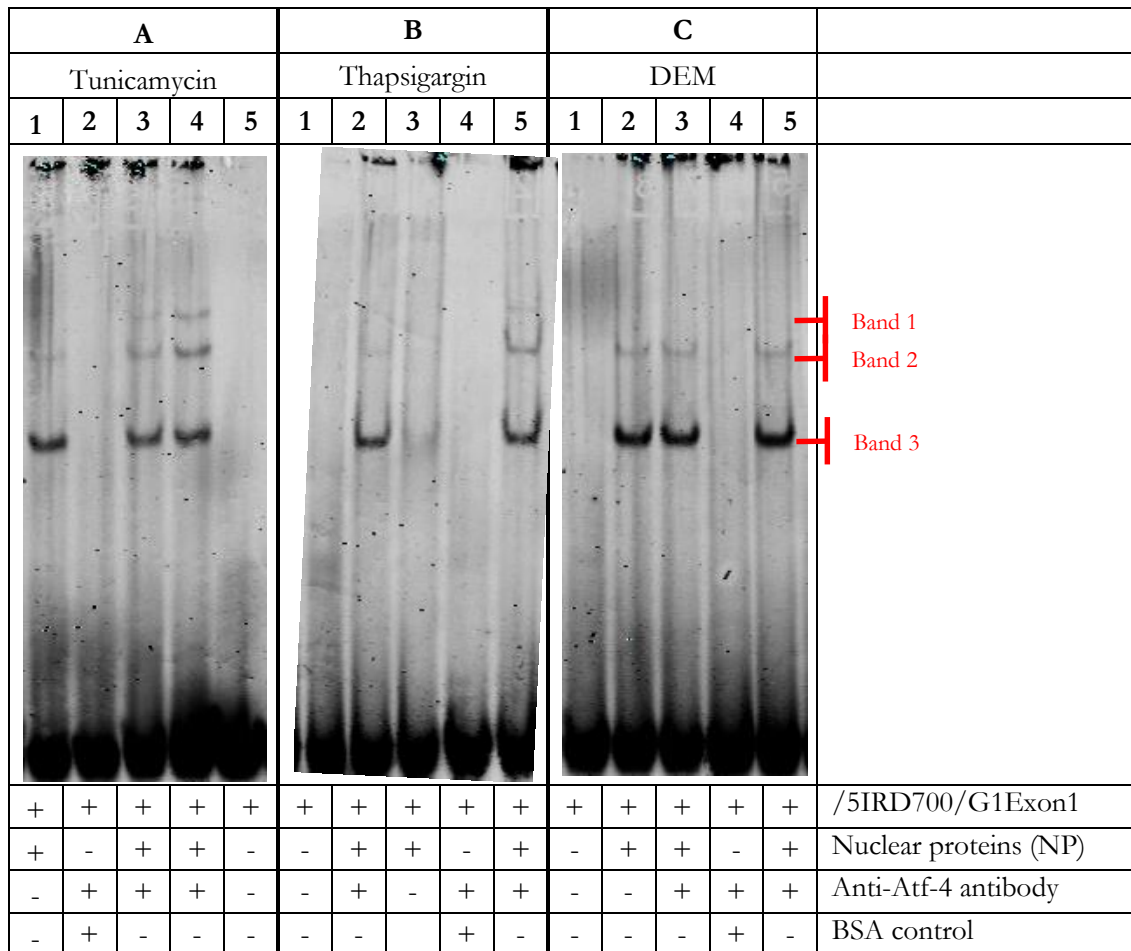
Figure 3.75: Super shift assay showing Atf-4 binds the /5IRD700/G1Exon1 probe sequence between positions +97 to +185 of human GlyT-1a exon 1 sequence and containing the potential AARE. When incubated with Caco-2 nuclear protein extract and in the absence of the antibody, two shifted bands are visible (bands 2 and 3); a third super-shifted band (band 1) with low mobility is visible above band 2 when an anti-Atf-4 antibody was included in binding reactions between the probe and extracts from tunicamycin (lanes 3 and 4 or panel A) and thapsigargin (lanes 2 and 5 or panel B) treated cells. No super-shifted bands are visible in panel C corresponding to binding reactions of nuclear extracts from DEM treated cells. Images are representative of observations seen in three separate experiments.
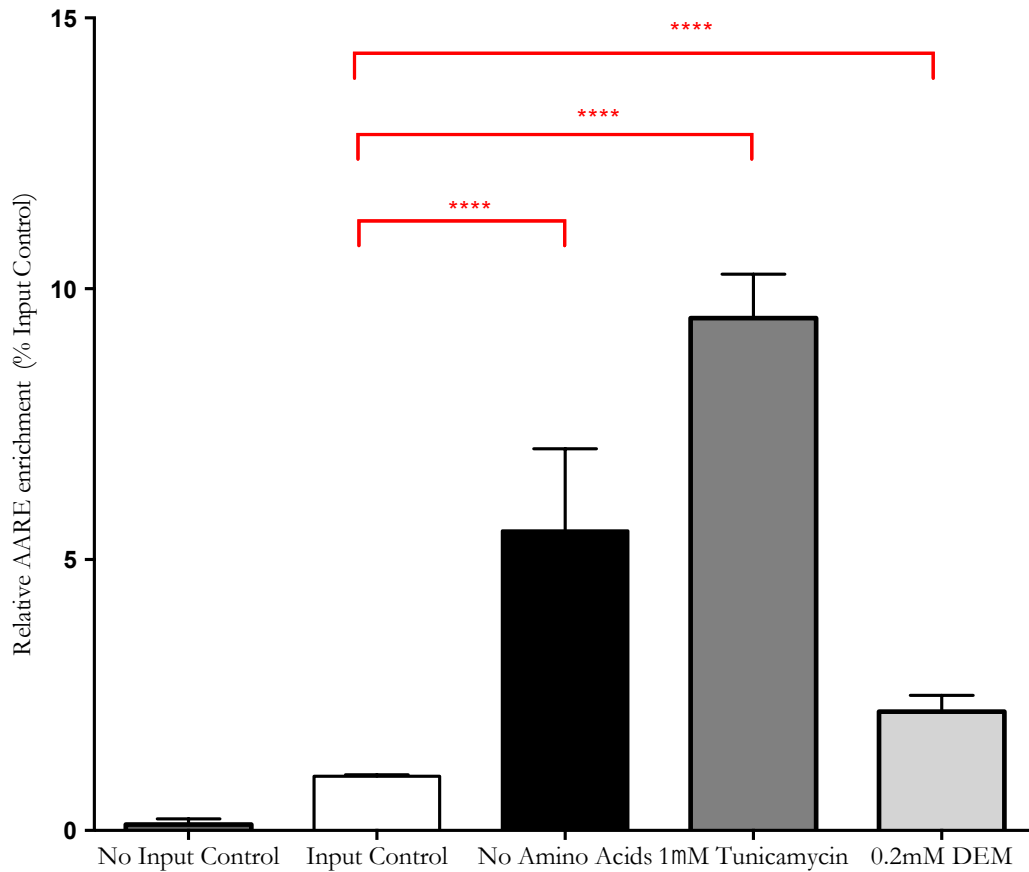
Figure 3.76: ChIP-QPCR analysis showing enrichment of GlyT-1a AARE in immune-precipitated DNA/protein complexes captured with a specific Atf-4 antibody. Bars represent the relative mean amounts ± SEM of a representative 100bp sequence spanning the GlyT-1a AARE in "input" vs "no-input" samples. "Input" samples represent aliquots of ChIP protein-DNA extracts from ~1x10^6 Caco-2 cells incubated overnight using 0.5μg/mL of Atf-4 antibody linked to protein-A magnetic Dynabeads. The "Input controls" data is from chromatin samples of unstressed cells. "No input" control samples represent DNA isolated from ChIP samples not treated with the Atf-4 antibody. Stress protocols were maintained as used throughout the text. One way ANOVA was used for statistical comparisons between the sample groups. *P*-values of less than 0.001 are indicated with ****.

CHAPTER 4.

CONCLUDING DISCUSSIONS

Outline: In this chapter, I discuss the main highlights from the work described in this thesis. This general discussion focuses on the merits and short comings of the observations described in Chapter 3. It provides a brief commentary on the theoretical implications of these results with respect to existing knowledge of GlyT-1a regulation. It goes on to describe avenues for further investigation into the spatio-temporal regulation of GlyT-1a; reflecting on a possible role for other transcription factors known to dimerise with Atf-4.

Based on the understanding that the intracellular accumulation of glycine (as a function of the GlyT-1 transporter) is important to resist stress in epithelial cells of the human intestine, the primary aim of the work described in this thesis was to investigate how amino acid starvation, oxidative stress, the unfolded protein response and ER stress resulted in increases in GlyT-1a mRNA expression. To achieve this aim, Caco-2 cells treated with tunicamycin, thapsigargin, DEM and tBHQ or subjected to altered amino acid availability were used as models of stress in the intestinal epithelium.

Gene activation in response to extracellular signals requires a highly complex integrated network of molecular interaction which directs the transcriptional machinery for the appropriate gene. In the work described here, particular attention was paid to the pathways of the integrated stress response, and transcription factors known to be active in the stress response, notably the BZIP factors Atf-4 and Nrf-2.

4.1 Atf-4 is involved in the regulation of GlyT-1a mRNA expression

Results from this work (discussed in Chapter 3) shows that the knockdown of Atf-4 in Caco-2 cells by specific siRNA decreased its mRNA by ~70%. This decrease in mRNA was consistent with a reduction in the protein titres detected by Western blot. A consequence of Atf-4 knockdown in these cells was an overall reduction in GlyT-1a mRNA expression relative to control transfected cells. A putative GlyT-1a AARE was identified in the GlyT-1a first exon by computational analysis (CRM 10). Reporter constructs containing this AARE showed significant reporter activity upon stress treatment. An 88bp probe containing the same AARE showed DNA/protein complex bands in gel shift assays while super shift assays using an anti-Atf-4 antibody confirmed the presence of Atf-4 in the protein-DNA complexes formed. Double mutation of the AARE sequence resulted in a ~75% reduction in transcriptional induction. Consistent with these findings, a high signal enrichment of a 100bp sequence spanning the predicted GlyT-1a AARE was obtained in QPCR analysis of DNA isolated from chromatin immune-precipitates of stress treated Caco-2 cells using an anti-Atf-4 antibody.

As the activation of Atf-4 represents a central convergence point in the integrated stress response (Harding *et al.*, 2000, Harding *et al.*, 2003), studies into the regulation of several other amino acid transporters by stress have largely focused on a possible role for Atf-4. Having shown that the knockdown of Atf-4 in Caco-2 and HCT-8 cells repressed GlyT-1a mRNA expression, and that Atf-4 mRNA was up-regulated by stress in these cells, the challenge was to map the direct molecular interactions between Atf-4 and enhancer elements in the GlyT-1a sequence. Importantly, the effects seen with Atf-4 knockdown on GlyT-1a mRNA were not evident with either Xbp-1 or Atf-6 knockdown.

For the identification of transcriptional regulatory elements associated with a gene, it is typical to perform *in silico* analysis of its core, proximal and distal promoter regions (Wasserman and Sandelin, 2004). There are an ever increasing number of computational strategies for such analysis. However, there exists significant variation in sensitivity in detection of biologically relevant transcription factor binding sites (TFBS) amongst these methods. The strategy employed here was designed to minimise false negatives as well as false positives from computational analysis. In this work, multiple sources of evidences and classification techniques were used to improve the significance of predicted TFBS. The promoter regions of genes potentially co-regulated with GlyT-1a were analysed for

the over-representation of TFBS when compared to a control subset of housekeeping genes. Caselle *et al* argue that the statistical relationship between motif over-representation in a subset of co-regulated genes to the biological activity of that motif is significant (Caselle *et al.*, 2002). Haverty *et al* have used such statistical relationships in developing an algorithm for computational inference of gene regulatory networks from expression profiles from micro-array studies (Haverty *et al.*, 2004).

In the absence of micro-array data from stressed Caco-2 cells it is impossible to accurately catalogue genes that have similar spatio-temporal patterns of expression to GlyT-1a when treated with the stress agents used in this work. Thus deciphering a transcriptional regulatory network for GlyT-1 in these cells is particularly challenging. It is also not directly possible to determine such information from publicly available micro-array data sets. To overcome this challenge, a subset of genes was identified in public micro-array datasets which had condition dependent correlation in their expression profiles to that of GlyT-1a. Based on the assumption that similar regulatory factors resulted in these patterns of expression, it was concluded that over-represented TFBS in the promoter regions of these genes would be important to condition specific regulation. Over a hundred enriched motifs were returned from the promoter analysis of co-regulated genes. It would be laborious to study the individual biological validity of every one of the returned motifs. Given that BZIP factors such as Atf-4 are known to function as dimers, assumptions were made on the nearness of individual TFBS within composite regulatory modules (CRM) for each monomer in a BZIP dimer. Such an approach has also been suggested as a way of improving the significance of predicted transcription factor binding sites (Syed *et al.*, 2009).

In the work presented here, ten CRMs were identified in the sequence region spanning 1159 bases upstream from the GlyT-1a start site, the first exon and the untranslated region (UTR) of the second exon. These modules each contained an arrangement of TFBS which may allow for the binding of BZIP dimers which include at least one Atf-4 subunit. The binding of Atf-4 to the amino acid response element (AARE; sequence TGATGCAAC) in the first exon of GlyT-1a has been confirmed by gel shift assays. It is highly likely that this AARE regulates the transcription of GlyT-1a given that nuclear extracts from tunicamycin, thapsigargin, DEM, and amino acid starved Caco-2 cells bound to probes containing it. The mutation of this AARE sequence both in the EMSA probe and β-galactosidase reporter constructs showed diminished binding and reporter

activity respectively, suggesting that it is an important motif for the regulation of the transcript following stress.

As discussed earlier (Section 3.2.3.1 page 95), owing to the fact that the first half of this AARE sequence resembles the original CRE binding motif, and that it was subsequently described as a binding site for several CREB/ATF proteins, it is commonly referred to as a CREB/ATF or simply CRE/ATF composite site. With the understanding that during the amino acid stress response protein members of the C/EBP subfamily may also bind to one half of the characteristic CREB/ATF motifs like the one shown here, they are often referred to as CARE motifs (short for C/EBP/ATF regulatory motifs) (Wolfgang *et al.*, 1997, Fawcett *et al.*, 1999). Kilberg *et al.* (2009) aligned the CARE motifs from 14 Atf-4 target genes and concluded that the high conservation of the Atf-4 binding half (TGATG) and a fairly divergent C/EBP half (*HAAH*; where *H* is either a C, A or T according to the IUPAC nucleotide code) suggests that different Atf-4 dimerisation partners provide transcriptional specificity to the stress signal. It is also possible that interplay of factors at this site may control the spatio-temporal regulation of GlyT-1a by specific stress signals. It has been shown that the binding of Atf-3 to a similarly placed binding motif in the first exon of the Cat-1 gene negatively regulates its transcription in response to amino acid starvation, with Atf-3 replacing Atf-4 at the Atf-4 half-site when the duration of stress was extensive (Lopez *et al.*, 2007). The negative isoform of Atf-3, a stress induced gene (Hai *et al.*, 1999, Hai and Hartman, 2001, Okamoto *et al.*, 2001), also repressed Atf-4 mediated induction of arsenate-induced activation of the Gadd153 promoter (Fawcett *et al.*, 1999). Could Atf-3 negatively regulate GlyT-1a via this site in a similar manner?

To fully understand the topology and dynamics of a spatial and temporal GlyT-1a regulatory network involving Atf-4, there is need to understand how Atf-4 expression and interactions with other factors such as C/EBPβ or Atf-3 might change with time. Such temporal changes in the expression of Atf-4, Atf-3 and C/EBPβ have been documented for the regulation of the SNAT-2 gene by histidine limitation (Palii *et al.*, 2004, Palii *et al.*, 2006, Thiaville *et al.*, 2008b). In these studies, ChIP analysis show the binding of Atf-4 to the SNAT-2 intronic AARE rises sharply after histidine starvation in HepG2 cells. The peak in Atf-4 interaction at between 2 and 12 hours after stress overlaps with a peak in C/EBPβ interaction at the same site. Interestingly, a drop in both Atf-4 and C/EBPβ interaction at the SNAT-2 AARE after 12 hours of stress overlaps a peak in Atf-3 protein

enrichment at the same AARE. Whilst it is impossible to directly infer such changes in TF interactions at the GlyT-1a AARE, it is highly likely given the sequence similarity of the GlyT-1a AARE to that of SNAT-2 and the fact that similar transcription factors are at play following stress in Caco-2 cells. A project to investigate this was undertaken in this laboratory earlier this year by Manasvi Bhasvar (MRes student, titled: Co-ordination of cell stress responses: Identification of Atf-4 binding partners in response to different stresses). From this work it appears Atf-3 may be an important player in regulating GlyT-1, however early data from this project are inconclusive and will require further investigations, by QPCR, ChIP or EMSA to confirm its involvement.

In contrast to the intronic AARE of SNAT-2 and the AARE located in the CHOP promoter (Bruhat *et al.*, 2002, Palii *et al.*, 2004), mutation of the GlyT-1a AARE did not completely abolish basal transcription despite a reduction in protein complexes detected by gel shift assays. However, the rate of basal GlyT-1a promoter induction was reduced by about ~75%. This observation suggested that although the GlyT-1a AARE is required for transcriptional induction following stress, other as of yet unidentified motifs may be needed for optimal transcriptional induction of GlyT-1. This is characteristic of flexible 'billboard' enhancer motifs (Arnosti and Kulkarni, 2005), where combinatorial effects of transcription factors binding to near and distant motifs control the activation or repression of transcription.

With genome-wide DNase I hypersensitivity data and ChIP-Seq data from projects such as ENCODE, there is further evidence of the importance of long-range regulatory elements such as enhancers, insulators and silencers in transcriptional regulation (Gerstein *et al.*, 2012). It has been suggested that the transcriptional pre-initiation complexes (PIC) formed at proximal, but also distal enhancers may be important in regulating the timing and onset of transcription (Szutorisz *et al.*, 2005). It has long been known that the binding of some BZIP and other transcription factors to DNA, results in the bending and wrapping of the DNA strand around the polymerase, increasing the proximity of previously distant motifs (Gustafson *et al.*, 1989, Horikoshi *et al.*, 1992, Coulombe and Burton, 1999). The DNA-looping model has been suggested as a transcriptional mechanism allowing for the regulation of a genes by very distant regulatory motifs (Vilar and Saiz, 2005). The high mobility group of nuclear proteins (HMGN) which are known to regulate unraveling of the chromatin for transcription of GlyT-1 (West *et al.*, 2004), have also been shown to result in DNA bending in favor of specific transcription factor

assembly at functional enhancers (Lorenz *et al.*, 1999). Could the binding of factors at the postulated GlyT-1a CRMs or other motifs, result in 'DNA-loops', bringing distant motifs in closer proximity to the core-promoter in an arrangement conducive to either repression or activation of the gene? From the work discussed here, it remains unclear exactly which other TFBS in the many predicted CRMs of the TATA-less GlyT-1a gene, together with the AARE in exon 1 cooperatively coordinate GlyT-1a transcriptional induction. However a systematic strategy of site directed mutagenesis of these CRMs, together with reporter experiments as used in these work, should confirm the relevance of the other predicted motifs. High throughput chromatin confirmation capture (3C) methods (Simonis *et al.*, 2007), could also be used to confirm the exact arrangement of these transcriptional regulatory regions downstream of stress.

4.2   Nrf-2 does not directly regulate GlyT-1a

One of the founding hypotheses for the work described in this thesis, was the suggestion that similar regulatory modules may regulate both GlyT-1a and xCT following stress. The occurrence of AARE in both xCT and GlyT-1a genes, and the demonstration that Atf-4 mediates the regulation of both xCT and Glyt-1a following oxidative and ER stress as well as amino acid starvation, largely confirms this hypothesis. That this regulation is not common to PepT1 also confirms that this is not just a general consequence of the requirement for increased amino acids for protein synthesis. Owing to the fact that Nrf-2 is known to regulate other amino acid transporters including xCT downstream of redox perturbations, it was expected that Nrf-2 may be directly involved in mediating the increase in GlyT-1a mRNA following oxidative stress. The data presented here however do not suggest this is the case. Although Nrf-2 mRNA was not increased in Caco-2 cells following stress, there was a large increase in total Nrf-2 protein in these cells after DEM treatment. However, Nrf-2 protein was not detected bound to a predicted antioxidant element (ARE) upstream of the GlyT-1a TSS, and the knockdown of Nrf-2 did not have any repressive effects on GlyT-1a mRNA.

Nrf-2 protein activation is controlled by a very complex epigenetic, transcriptional and post-transcriptional network. Whilst the data shown here does not implicate Nrf-2 in the direct transcriptional regulation of GlyT-1a, there may still be as yet unknown effects of Nrf-2 protein activation on other factors involved in the post-transcriptional regulation of GlyT-1a. There are indications that Nrf-2 may directly regulate, and interact with Atf-

4 (He *et al.*, 2001). The co-regulation of both Nrf-2 and Atf-4 by the flavonoid fisetien is known to mediate enhancement of glutathione levels following stress (Ehren and Maher, 2013). Ye and colleagues using T24 Bladder carcinoma cells showed an increase in Atf-4 and Nrf-2 mRNA and protein levels following treatment of these cells with sulfasalazine (Ye *et al.*, 2014). In the study by Ye *et al*, Atf-4 and Nrf-2 both mediated up-regulation of xCT via separate AARE and ARE binding sites in its promoter. Interestingly, in luciferase reporter experiments with xCT promoter constructs containing the mutated ARE sequence, Nrf-2 further increased xCT transcription mediated by Atf4; however when the two xCT AAREs were mutated reporter activity was completely abolished. In another study, Zong and colleagues showed that Nrf-2, by virtue of its interactions with Atf-4 may regulate its binding to the two AAREs in the CHOP promoter (Zong *et al.*, 2012). Like xCT, the CHOP promoter also contains two AAREs and one ARE in the reverse orientation. In contrast to the study by Ye *et al*, Zong and colleagues observed a significant reduction in reporter activity using constructs of the CHOP promoter containing a mutated ARE element. They conclude that in the absence of an effective Nrf-2 binding site, Nrf-2 interaction with Atf-4 sequesters it from binding to the AAREs, hence the diminished CHOP induction.

The fact that Nrf-2 may block Atf-4 mediated transcription of genes lacking a functional TFBS for Nrf-2 has led to exploration of a possible dual role for Nrf-2 in controlling a switch between protective/survival and apoptotic signalling (Cullinan and Diehl, 2006). Mozzini and colleagues suggest that as to whether Nrf-2 is repressive or activating may be related to a dose-dependent increase in ROS accumulation (Mozzini *et al.*, 2014). Although the data collected here appears to be indicative that Nrf-2 is not involved in GlyT-1a transcription, it remains unclear if there is any role for an Atf-4 and Nrf-2 interaction either as a transcriptional activating or repressive module for GlyT-1a or indeed xCT regulation.

4.3 Significance and future perspectives

Despite detailed understanding of the kinetics of GlyT-1 transport, very little is known about its transcriptional regulation. The work described in this thesis shows for the first time that Atf-4, a factor important to the transcriptional induction of several other stress response genes, binds to a novel amino acid response element (AARE) in the first exon of GlyT-1a. These findings lay the foundation for further investigation into whether

interplays of other Atf-4 interacting BZIP factors such as Atf-3, C/EBPβ and Nrf-2 oversee the spatial and temporal regulation of GlyT-1 from its AARE sequence.

That the transcription of GlyT-1 is up-regulated in an Atf-4 dependent manner despite an overall repression of protein synthesis following stress, is testament that the transporter is important for resistance to stress. Beyond its role in the synthesis of the antioxidant glutathione, the supply and metabolism of glycine is also pivotal to cancer progression (Jain *et al.*, 2012, di Salvo *et al.*, 2013, Locasale, 2013). It is possible that several cancers may hijack protective pathways such as the Atf-4 dependent up-regulation of glycine transport by GlyT-1 (or other glycine transporters) to evade the damaging effects of prolonged stress (Amelio *et al.*, 2014). As such understanding precisely the transcriptional modules important to GlyT-1a induction or repression could provide a new avenue for precise control of pro-survival or pro-apoptotic outcomes in these cells. It is evident from this work that knocking-down Atf-4 in Caco-2 cells diminishes expression of GlyT-1a. On the flip side, Atf-4 activation by ER resident kinases such as PERK up-regulate GlyT-1a mRNA. The activation of Atf-4 by such ER kinases has been shown to protect colonic epithelium in mice against UPR-induced colitis (Cao *et al.*, 2012). Could the up-regulation of the PERK-Atf-4 pathway and consequently increase in GlyT-1 have protective effects against colitis and other forms of stress induced bowel disease?

The role of stress signalling in the pathogenesis of several degenerative diseases and cancer is an active area of research. Last year, it was reported that the knockdown of PERK by a proprietary compound prevented the clinical presentation of Prion disease in mice, although there was no prevention or decrease in further accumulation of unfolded proteins in the ER (Moreno *et al.*, 2013). The work by Moreno and colleagues set out on the foundation that the UPR induced phosphorylation of PERK and resulted in disrupted proteostasis and neuronal death. The authors hypothesised that PERK knockdown represented an important therapeutic target to halt the progression of prion disease owing to PERK mediated repression of proteins important for effective synaptic function. However, these mice developed significant hypoglycaemia and weight loss 'side effects'. These may in fact be a major consequence of blocking key protective outcomes downstream of PERK phosphorylation such as the activation of Atf-4 and transcriptional induction of GlyT-1.

A detailed understanding of the regulation of pro-survival and pro-apoptotic mechanisms downstream of stress induction may provide better therapeutic targets, whilst minimising debilitating side effects. It will be of interest to further investigate the involvement of other transcription factors in regulating the temporal aspects of GlyT-1 expression as well as the switch between pro-survival and pro-apoptosis. If resources become permitting ChIP-seq experiments could effectively confirm factors bound to the other predicted CRMs in the GlyT-1a sequence under investigation following stress treatment. Coupling such data to pathway analysis of expression or protein profiles of factors following stress treatment, would provide important insights into the possible use of therapeutic modulation of glycine supply.

In conclusion, the work described in this thesis – whilst complete in itself and providing many novel insights into regulation of the human GlyT-1a gene, has by virtue of confirming its role as a significant target of the ISR validated the need for further investigation of this transporter and its substrate. This with the aim of better understand their role in the processes of significant clinical relevance such as in stress induced degenerative diseases.

APPENDICES

Appendix A

BIOINFORMATICS SUPPLEMENTS

A.1 Classification of GlyT-1a co-regulated genes returned by CORD

To identify genes potentially co-regulated with GlyT-1 the gene CO-regulation database (CORD) was searched. As described in the text (page85, Section 3.2.2), 185 genes with correlation co-efficient of greater than 84% were returned. To determine if these genes had similar function to that of GlyT-1 they were classified using gene ontology (GO) terms representing their function. As can be seen in the table below most of these terms overlap GlyT-1a function.

GO Identifier	Gene symbol	GO Term
GO:0071705	ADORA3, DRD2, SLC14A1, SLC1A2, SLC1A3, SLC22A4, SLC35E4, SLC38A3, SLC6A12, SLC6A8, SLC6A9, SLC7A10, SNAP25	nitrogen compound transport
GO:0015837	ADORA3, DRD2, SLC1A2, SLC1A3, SLC22A4, SLC38A3, SLC6A12, SLC6A8, SLC6A9, SLC7A10, SNAP25	amine transport
GO:0006836	DRD2, SLC1A2, SLC1A3, SLC6A12, SLC6A8, SLC6A9, SNAP25, SV2B	neurotransmitter transport
GO:0046942	BSG, GOT2, PLA2G4C, SLC1A2, SLC1A3, SLC22A4, SLC38A3, SLC6A12, SLC6A8, SLC6A9, SLC7A10	carboxylic acid transport
GO:0071702	ABCD1, ABCG4, ADORA3, BSG, DRD2, GOT2, PLA2G4C, SLC1A2, SLC1A3, SLC22A4, SLC2A1, SLC35E4, SLC38A3, SLC6A12, SLC6A8, SLC6A9, SLC7A10, SNAP25	organic substance transport
GO:0006811	AQP1, ATP1A2, ATP1B2, BSG, C3AR1, CHP, CLCN4, CYBRD1, DRD2, EHD3, ELN, FXYD1, HTR3B, KCNC3, KCNK12, KCNN2, SCN1B, SLC1A2, SLC1A3, SLC22A4, SLC2A1, SLC35E4, SLC6A12, SLC6A8, SLC6A9, TRPM3	ion transport
GO:0006865	**SLC1A2, SLC1A3, SLC22A4, SLC38A3, SLC6A12, SLC6A8, SLC6A9, SLC7A10**	**amino acid transport**
GO:0015849	BSG, GOT2, PLA2G4C, SLC1A2, SLC1A3, SLC22A4, SLC38A3, SLC6A12, SLC6A8, SLC6A9, SLC7A10	organic acid transport
GO:0006778	ALAD, BLVRB, CPOX, HMBS, UROD	porphyrin metabolic process

GO Identifier	Gene symbol	GO Term
GO:0033013	ALAD, BLVRB, CPOX, HMBS, UROD	tetrapyrrole metabolic process
GO:0015672	ATP1A2, ATP1B2, CHP, DRD2, KCNC3, KCNK12, KCNN2, SCN1B, SLC22A4, SLC6A12, SLC6A8, SLC6A9, TRPM3	monovalent inorganic cation transport
GO:0006783	ALAD, CPOX, HMBS, UROD	heme biosynthetic process
GO:0006438	VARS, VARS2	valyl-tRNA aminoacylation
GO:0006779	ALAD, CPOX, HMBS, UROD	porphyrin biosynthetic process
GO:0033014	ALAD, CPOX, HMBS, UROD	tetrapyrrole biosynthetic process
GO:0003008	ADORA3, AOC2, AQP1, ATP1A2, C1QTNF5, C3AR1, CA14, CRYAB, CRYBB2, DIP2A, DMPK, DRD2, ELN, FXYD1, HTR3B, KCNC3, NPBWR2, NPPC, OR10H5, OR1C1, OR6B3, PPP1R12B, PYY, RD3, RDH5, ROM1, SCG2, SCN1B, SIX6, SLC1A2, SLC1A3, SLC22A4, SLC6A8, SLC6A9, SNAP25, SNCB, SV2B, TRH, TRPM3, ZNF385A	system process
GO:0042168	ALAD, CPOX, HMBS, UROD	heme metabolic process
GO:0007601	AOC2, C1QTNF5, CRYAB, CRYBB2, RD3, RDH5, ROM1, SCG2, SIX6, ZNF385A	visual perception
GO:0050953	AOC2, C1QTNF5, CRYAB, CRYBB2, RD3, RDH5, ROM1, SCG2, SIX6, ZNF385A	sensory perception of light stimulus
GO:0001504	SLC1A2, SLC1A3, SLC6A8	neurotransmitter uptake
GO:0045967	GABARAPL1, TENC1	negative regulation of growth rate
GO:0007268	CA14, DRD2, HTR3B, KCNC3, NPBWR2, PYY, SCN1B, SLC1A2, SLC1A3, SLC6A8, SLC6A9, SNAP25, SNCB, SV2B	synaptic transmission
GO:0045768	DRD2, GSN, NPBWR2, SNCB	positive regulation of anti-apoptosis
GO:0006814	ATP1A2, CHP, SCN1B, SLC6A12, SLC6A8, SLC6A9, TRPM3	sodium ion transport
GO:0019226	CA14, DRD2, HTR3B, KCNC3, NPBWR2, PYY, SCN1B, SLC1A2, SLC1A3, SLC6A8, SLC6A9, SNAP25, SNCB, SV2B	transmission of nerve impulse
GO:0035637	CA14, DRD2, HTR3B, KCNC3, NPBWR2, PYY, SCN1B, SLC1A2, SLC1A3, SLC6A8, SLC6A9, SNAP25, SNCB, SV2B	multicellular organismal signalling
GO:0001505	DRD2, SLC1A2, SLC1A3, SLC6A8, SNAP25, SV2B	regulation of neurotransmitter levels

GO Identifier	Gene symbol	GO Term
GO:0050877	ADORA3, AOC2, C1QTNF5, CA14, CRYAB, CRYBB2, DRD2, HTR3B, KCNC3, NPBWR2, OR10H5, OR1C1, OR6B3, PYY, RD3, RDH5, ROM1, SCG2, SCN1B, SIX6, SLC1A2, SLC1A3, SLC6A8, SLC6A9, SNAP25, SNCB, SV2B, TRH, ZNF385A	neurological system process
GO:0006812	AQP1, ATP1A2, ATP1B2, C3AR1, CHP, CYBRD1, DRD2, EHD3, ELN, KCNC3, KCNK12, KCNN2, SCN1B, SLC22A4, SLC6A12, SLC6A8, SLC6A9, TRPM3	cation transport
GO:0051234	ABCD1, ABCG4, ADORA3, AGFG2, AQP1, ATP1A2, ATP1B2, ATP5D, BSG, C3AR1, CADM2, CHP, CLCN4, CTSD, CYBRD1, DGKG, DRD2, EHD3, ELN, FGF1, FXYD1, GABARAPL1, GOT2, GSN, HBD, HTR3B, KCNC3, KCNK12, KCNN2, KLF4, NPPC, PACSIN3, PLA2G4C, PPARA, PPP1R3C, S100B, SCG2, SCN1B, SLC14A1, SLC1A2, SLC1A3, SLC22A4, SLC2A1, SLC35E4, SLC38A3, SLC6A12, SLC6A8, SLC6A9, SLC7A10, SNAP25, SNCB, SNCG, SV2B, TENC1, TRAK2, TRH, TRPM3, VPS18	establishment of localization
GO:0046329	DUSP3, EPB49, SNCG	negative regulation of JNK cascade
GO:0015838	SLC22A4, SLC6A12	betaine transport
GO:0042462	NXNL1, ROM1	eye photoreceptor cell development
GO:0006386	NFIC, NFIX	termination of RNA polymerase III transcription
GO:0040009	GABARAPL1, TENC1	regulation of growth rate
GO:0006810	ABCD1, ABCG4, ADORA3, AGFG2, AQP1, ATP1A2, ATP1B2, ATP5D, BSG, C3AR1, CADM2, CHP, CLCN4, CTSD, CYBRD1, DGKG, DRD2, EHD3, ELN, FGF1, FXYD1, GABARAPL1, GOT2, GSN, HBD, HTR3B, KCNC3, KCNK12, KCNN2, NPPC, PACSIN3, PLA2G4C, PPARA, PPP1R3C, SCG2, SCN1B, SLC14A1, SLC1A2, SLC1A3, SLC22A4, SLC2A1, SLC35E4, SLC38A3, SLC6A12, SLC6A8, SLC6A9, SLC7A10, SNAP25, SNCB, SNCG, SV2B, TRAK2, TRH, TRPM3, VPS18	transport
GO:0045767	DRD2, GSN, NPBWR2, SNCB	regulation of anti-apoptosis
GO:0030001	ATP1A2, C3AR1, CHP, CYBRD1, DRD2, EHD3, ELN, KCNC3, KCNK12, KCNN2, SCN1B, SLC6A12, SLC6A8, SLC6A9, TRPM3	metal ion transport

GO Identifier	Gene symbol	GO Term
GO:0070303	DUSP3, EPB49, SNCG	negative regulation of stress-activated protein kinase
GO:0001754	NXNL1, ROM1	eye photoreceptor cell differentiation
GO:0046148	ALAD, CPOX, HMBS, UROD	pigment biosynthetic process
GO:0006956	C1QC, C3AR1, C4B, C4B	complement activation
GO:0015804	SLC38A3, SLC6A9, SLC7A10	neutral amino acid transport
GO:0060284	BSG, FGF1, KLF4, MGAT5B, ROM1, S100B, SCG2	regulation of cell development
GO:0048592	NXNL1, ROM1	eye morphogenesis
GO:0042440	ALAD, CPOX, HMBS, UROD	pigment metabolic process
GO:0072376	C1QC, C3AR1, C4B, C4B	protein activation cascade
GO:0042461	NXNL1, ROM1	photoreceptor cell development
GO:0006813	ATP1A2, CHP, DRD2, KCNC3, KCNK12, KCNN2	potassium ion transport
GO:0044092	ADORA3, BSG, CRYAB, DRD2, DUSP3, EPB49, GAB1, GNG7, NPBWR2, PRELP, PYY, S100B, SEC14L2, SNCB, SNCG, TENC1	negative regulation of molecular function
GO:0015697	SLC22A4, SLC6A12	quaternary ammonium group transport
GO:0009268	AMELY, CLCN4, FMO2	response to pH

A.2 Enriched TFBS in the promoter of GlyT-1 co-regulated genes

Stress associated factors used in composite module analysis are shown in bold. Using the TRANSFAC F-MATCH algorithm, analysis of the promoter regions (spanning 2000 bases upstream to 1000 bases downstream of the TSS) of co-regulated genes identified by CORD was performed. F-MATCH compares the occurrence of these over-represented TFBS in the 'test' or 'YES' list of differentially expressed gene promoters to a 'control' or 'NO' list of housekeeping genes to correct for multiplicity (errors in inference, *i.e.* where the frequency of a given TFBS is higher in the NO set than the YES set it is omitted from the result). FDR: false discovery rate.

Positional Weight Matrices	Yes (sites per 1kbp)	Yes/ No	P-value	Matched promoters p-value	From	To	FDR	Matched promoters FDR
V$AHR_Q5	0.1896	1.3837	0.0097	0.0902	-1700	-900	0.0156	0.1530
V$AHRHIF_Q6	0.0462	2.1646	0.0073	0.0691	-1800	-1000	0.0135	0.1202
V$AP1_Q2_01	**0.1064**	**1.6042**	**0.0070**	**0.0069**	**-1900**	**-1600**	**0.0131**	**0.0194**
V$AP1_Q4_01	**0.0486**	**2.1262**	**0.0070**	**0.0042**	**-600**	**100**	**0.0132**	**0.0156**
V$AP2_Q6	**0.0717**	**2.5605**	**0.0001**	**0.0096**	**-1600**	**-1300**	**0.0011**	**0.0241**
V$AP2_Q6_01	**0.0578**	**2.6156**	**0.0005**	**0.0143**	**-1500**	**-1200**	**0.0022**	**0.0321**
V$AP2ALPHA_01	**0.0902**	**3.0603**	**0.0000**	**0.0000**	**-300**	**0**	**0.0001**	**0.0018**
V$AP4_01	0.0624	2.2301	0.0015	0.0262	-2000	-1700	0.0047	0.0489
V$AR_Q2	0.0994	1.9009	0.0009	0.0023	-1200	-900	0.0033	0.0101
V$AREB6_03	0.0601	1.8978	0.0085	0.0166	-900	-200	0.0142	0.0352
V$BCL6_Q3	0.0439	2.1299	0.0100	0.0079	-1600	-700	0.0159	0.0214
V$BRCA_01	0.0624	2.0178	0.0041	0.0082	0	300	0.0091	0.0220
V$CACD_01	0.1734	1.4899	0.0034	0.0113	-1300	-1000	0.0081	0.0267
V$CDP_02	0.0532	2.1233	0.0050	0.0691	100	500	0.0106	0.1202

Positional Weight Matrices	Yes (sites per 1kbp)	Yes/ No	P-value	Matched promoters p-value	From	To	FDR	Matched promoters FDR
V$CDPCR1_01	0.0416	2.5681	0.0031	0.3995	-400	-100	0.0080	0.6361
V$CDPCR3_01	0.1017	1.8172	0.0015	0.0013	200	500	0.0047	0.0081
V$CDXA_02	0.0439	6.6262	0.0000	0.0006	100	400	0.0001	0.0055
V$C/EBP_Q2_01	**0.1873**	**1.589**	**0.0006**	**0.0126**	**-500**	**-200**	**0.0026**	**0.0287**
V$C/EBPA_01	**0.148**	**1.5334**	**0.0039**	**0.1174**	**100**	**400**	**0.0088**	**0.1944**
V$C/EBPG_Q6	**0.074**	**2.2827**	**0.0004**	**0.0183**	**-300**	**0**	**0.0024**	**0.0374**
V$CIZ_01	0.0509	2.0925	0.0067	0.0030	-1300	500	0.0129	0.0116
V$CKROX_Q2	0.0462	2.511	0.0023	0.0125	-1000	-700	0.0063	0.0287
V$CMAF_01	0.0439	2.2937	0.0058	0.0044	-1600	-1200	0.0119	0.0158
V$COREBINDINGFACTOR_Q6	0.0786	1.7212	0.0088	0.0253	-800	-500	0.0145	0.0484
V$COUP_DR1_Q6	0.0671	1.8963	0.0057	0.0028	-900	-600	0.0119	0.0113
V$CP2_02	0.0786	1.7495	0.0074	0.0214	-300	0	0.0131	0.0418
V$CREB_02	**0.1064**	**1.6596**	**0.0044**	**0.0045**	**-900**	**-600**	**0.0096**	**0.0158**
V$CREBATF_Q6	**0.0971**	**1.7346**	**0.0035**	**0.0028**	**-900**	**-600**	**0.0083**	**0.0115**
V$DBP_Q6	0.0578	2.5312	0.0006	0.0021	-700	-400	0.0027	0.0101
V$DR1_Q3	0.104	1.6424	0.0056	0.0065	-1900	-1600	0.0118	0.0191
V$DR3_Q4	0.0462	2.1646	0.0073	0.0143	-1400	-1000	0.0135	0.0321
V$DR4_Q2	0.0462	2.9893	0.0005	0.0069	-600	0	0.0025	0.0194
V$E12_Q6	0.0416	2.5681	0.0031	0.0017	-800	-100	0.0080	0.0091
V$E2A_Q2	0.0439	3.3131	0.0003	0.0003	-700	-400	0.0020	0.0043
V$E2F_03	0.0786	1.8722	0.0034	0.0564	-1800	-1500	0.0081	0.0989
V$E2F_Q6_01	0.0671	2.4601	0.0003	0.0194	-1800	-1500	0.0020	0.0389
V$EBF_Q6	0.0416	3.5311	0.0003	0.0034	-2000	-1700	0.0020	0.0129

Positional Weight Matrices	Yes (sites per 1kbp)	Yes/ No	P-value	Matched promoters p-value	From	To	FDR	Matched promoters FDR
V$EBOX_Q6_01	0.1064	1.6986	0.0032	0.0017	-700	-400	0.0078	0.0089
V$EFC_Q6	0.074	1.7621	0.0085	0.0077	-1400	-1100	0.0142	0.0213
V$EGR1_01	0.0416	3.3234	0.0004	0.0050	-1600	-1000	0.0024	0.0157
V$ER_Q6	0.0532	2.4893	0.0012	0.0022	-1900	-1200	0.0039	0.0103
V$ETF_Q6	0.0532	2.1876	0.0038	0.0342	-1500	-800	0.0088	0.0609
V$ETS_Q6	0.178	1.5011	0.0026	0.0150	-1400	-1100	0.0070	0.0323
V$FAC1_01	0.0948	1.9207	0.0010	0.0260	-100	200	0.0036	0.0495
V$FOXJ2_02	0.0416	2.5681	0.0031	0.0020	200	500	0.0080	0.0103
V$FXR_Q3	0.0416	3.7665	0.0002	0.0001	-100	200	0.0013	0.0035
V$GATA_C	0.0532	2.6737	0.0006	0.0005	-500	0	0.0026	0.0057
V$GATA_Q6	0.0486	2.6365	0.0012	0.0213	-300	100	0.0039	0.0419
V$GEN_INI3_B	0.3052	1.307	0.0064	0.0066	-1900	-1600	0.0126	0.0190
V$GFI1_Q6	0.0486	2.7464	0.0008	0.0002	-200	200	0.0032	0.0042
V$GRE_C	0.0416	4.3459	0.0001	0.0001	-600	500	0.0007	0.0031
V$HIC1_03	0.0439	2.2937	0.0058	0.0066	-1800	-1400	0.0119	0.0192
V$HIF1_Q3	0.0509	2.5575	0.0012	0.0048	-2000	-1700	0.0039	0.0166
V$HMGIY_Q6	0.296	1.3044	0.0075	0.1106	-1900	-1600	0.0132	0.1846
V$HNF3_Q6_01	0.0439	2.2937	0.0058	0.0474	200	500	0.0119	0.0838
V$HNF6_Q6	0.0416	3.5311	0.0003	0.0007	-1700	-1200	0.0020	0.0063
V$HSF2_01	0.0763	1.9181	0.0029	0.0182	-1200	-900	0.0077	0.0377
V$IK_Q5	0.2035	1.4537	0.0028	0.0172	-700	-400	0.0073	0.0360
V$IPF1_Q4_01	0.0555	2.354	0.0016	0.0053	-1000	500	0.0046	0.0163
V$KAISO_01	0.0439	3.1387	0.0005	0.0017	-2000	-900	0.0023	0.0095

Positional Weight Matrices	Yes (sites per 1kbp)	Yes/ No	P-value	Matched promoters p-value	From	To	FDR	Matched promoters FDR
V$KID3_01	0.8023	1.189	0.0036	0.5211	-1000	-700	0.0085	0.8234
V$KROX_Q6	0.0416	3.5311	0.0003	0.0049	-1700	-1100	0.0020	0.0161
V$LEF1_Q2_01	0.1064	1.5866	0.0082	0.0049	-700	100	0.0138	0.0163
V$LRF_Q2	0.1156	2.4144	0.0000	0.0056	-1600	-1300	0.0001	0.0170
V$LRH1_Q5	0.0416	2.354	0.0059	0.0089	-700	300	0.0118	0.0230
V$LXR_Q3	0.0486	2.8658	0.0006	0.0011	-600	-300	0.0026	0.0074
V$MAF_Q6_01	**0.0416**	**2.6903**	**0.0022**	**0.0008**	**-700**	**0**	**0.0062**	**0.0063**
V$MAZ_Q6	**0.0671**	**2.3953**	**0.0005**	**0.0082**	**-1000**	**-700**	**0.0022**	**0.0220**
V$MEF2_Q6_01	0.0416	2.2599	0.0079	0.0024	-1800	-800	0.0138	0.0101
V$MEIS1BHOXA9_02	0.074	2.5754	0.0001	0.0023	100	400	0.0009	0.0101
V$MINI19_B	0.0809	2.0344	0.0011	0.0103	-500	-200	0.0038	0.0255
V$MRF2_01	0.0486	2.354	0.0031	0.0042	-1200	500	0.0079	0.0156
V$MYOGNF1_01	0.0671	1.8205	0.0086	0.0059	-1800	-1000	0.0142	0.0176
V$NFAT_Q4_01	**0.0532**	**2.5782**	**0.0009**	**0.0022**	**-100**	**300**	**0.0033**	**0.0103**
V$NFY_01	0.0439	2.8398	0.0011	0.0008	-1500	-1100	0.0038	0.0063
V$NKX22_01	0.0439	2.2937	0.0058	0.0153	0	400	0.0119	0.0327
V$OCT_Q6	0.0439	3.9757	0.0001	0.0007	-300	400	0.0008	0.0063
V$OCT1_02	0.0439	9.9393	0.0000	0.0001	0	300	0.0000	0.0030
V$OCT1_03	0.1457	1.6208	0.0016	0.0091	100	400	0.0047	0.0233
V$OCT1_07	0.0994	2.6464	0.0000	0.0004	200	500	0.0001	0.0052
V$OCT1_Q5_01	0.0439	3.7272	0.0001	0.0010	-300	300	0.0011	0.0078
V$OCT4_01	0.0416	4.0355	0.0001	0.0003	-300	500	0.0009	0.0045
V$OCT4_02	0.0624	2.1186	0.0025	0.0113	200	500	0.0069	0.0269

Positional Weight Matrices	Yes (sites per 1kbp)	Yes/ No	P-value	Matched promoters p-value	From	To	FDR	Matched promoters FDR
V$OG2_01	0.0416	2.2599	0.0079	0.0048	-100	500	0.0138	0.0166
V$OSF2_Q6	0.0694	1.9217	0.0043	0.0113	-800	-500	0.0095	0.0271
V$P300_01	**0.0462**	**5.2312**	**0.0000**	**0.0000**	**-1400**	**-700**	**0.0001**	**0.0020**
V$P53_02	0.1595	1.4535	0.0073	0.0309	-1100	-800	0.0135	0.0561
V$PAX2_01	0.1133	1.5855	0.0066	0.0301	-900	400	0.0127	0.0551
V$PAX3_01	0.0509	3.6343	0.0000	0.0001	-1400	-1100	0.0006	0.0031
V$PAX4_03	0.0486	3.8773	0.0000	0.0711	0	300	0.0006	0.1217
V$PAX5_01	0.0439	2.7107	0.0016	0.0034	-700	-400	0.0047	0.0129
V$PAX6_Q2	0.3098	1.4968	0.0001	0.0012	-300	0	0.0010	0.0073
V$PAX8_01	0.0486	2.2729	0.0041	0.0022	-2000	-1500	0.0091	0.0103
V$PBX_Q3	0.0416	7.0621	0.0000	0.1836	-400	500	0.0001	0.2969
V$PBX1_02	0.0416	4.7081	0.0000	0.0020	-200	500	0.0005	0.0103
V$PLZF_02	0.0555	3.9647	0.0000	0.0016	-300	0	0.0002	0.0095
V$POU1F1_Q6	0.074	3.24	0.0000	0.0011	100	400	0.0001	0.0076
V$POU3F2_01	0.0902	4.221	0.0000	0.0000	200	500	0.0000	0.0018
V$POU3F2_02	0.0948	3.4781	0.0000	0.0001	100	400	0.0000	0.0031
V$POU6F1_01	0.0416	2.9735	0.0010	0.0196	-200	500	0.0037	0.0391
V$PPAR_DR1_Q2	0.0717	2.5605	0.0001	0.0001	-1900	-1600	0.0011	0.0028
V$PPARA_02	0.0624	1.8423	0.0098	0.1423	-300	200	0.0157	0.2337
V$PPARG_02	0.0509	2.2275	0.0040	0.0262	-300	0	0.0090	0.0492
V$R_01	0.0416	2.5681	0.0031	0.0049	-900	400	0.0080	0.0157
V$RUSH1A_02	0.0416	2.8249	0.0015	0.0011	-1800	-1100	0.0048	0.0076
V$S8_01	0.0416	2.8249	0.0015	0.0186	-800	400	0.0048	0.0378

Positional Weight Matrices	Yes (sites per 1kbp)	Yes/ No	P-value	Matched promoters p-value	From	To	FDR	Matched promoters FDR
V$SMAD3_Q6	0.0462	2.7293	0.0011	0.0017	-800	-300	0.0039	0.0093
V$SOX9_B1	0.2543	1.5143	0.0003	0.0904	100	500	0.0020	0.1522
V$SP1_Q6	0.0486	2.5351	0.0016	0.2564	-2000	-1700	0.0047	0.4114
V$SP3_Q3	0.0486	3.1387	0.0003	0.0002	-1100	-400	0.0019	0.0044
V$SPZ1_01	0.0439	4.9697	0.0000	0.0001	-1900	-1600	0.0002	0.0031
V$SREBP_Q3	0.1017	1.5874	0.0094	0.0220	-800	-300	0.0154	0.0425
V$SREBP1_01	0.1249	1.5408	0.0070	0.0311	-700	-400	0.0132	0.0560
V$STAF_02	0.0439	2.9818	0.0007	0.0005	-1100	-300	0.0030	0.0060
V$STAT_Q6	0.1711	1.4426	0.0065	0.0120	-700	-300	0.0126	0.0280
V$STAT1_01	0.0416	3.3234	0.0004	0.0049	-1600	-1300	0.0024	0.0161
V$SZF11_01	0.0532	2.7766	0.0004	0.0008	-1900	-1400	0.0025	0.0061
V$TAL1BETAE47_01	0.0578	3.1387	0.0001	0.0003	200	500	0.0007	0.0041
V$TAXCREB_02	0.0416	2.2599	0.0079	0.0294	-1600	-1200	0.0138	0.0543
V$TBP_Q6	0.0717	2.7028	0.0001	0.0005	200	500	0.0008	0.0058
V$TBX5_01	0.0439	2.1299	0.0100	0.0110	-600	500	0.0159	0.0268
V$TCF11_01	0.0578	3.2695	0.0000	0.0003	-1100	-700	0.0006	0.0041
V$TEL2_Q6	0.1225	1.7511	0.0010	0.0019	-1100	-800	0.0036	0.0101
V$TST1_01	0.0439	2.7107	0.0016	0.0098	0	300	0.0047	0.0243
V$VDR_Q3	0.0416	3.3234	0.0004	0.0007	-1000	-700	0.0024	0.0062
V$WT1_Q6	0.0439	3.3131	0.0003	0.1668	-1900	-1500	0.0020	0.2719
V$XFD2_01	0.0416	3.3234	0.0004	0.0010	-600	500	0.0024	0.0078
V$XVENT1_01	0.0462	2.1646	0.0073	0.0143	-2000	-1300	0.0135	0.0321
V$YY1_Q6	0.0763	1.7556	0.0079	0.0085	-1700	-1400	0.0138	0.0223

Positional Weight Matrices	Yes (sites per 1kbp)	Yes/ No	P-value	Matched promoters p-value	From	To	FDR	Matched promoters FDR
V$ZF5_B	0.0416	3.1387	0.0007	0.0144	-2000	-1300	0.0029	0.0314
V$ZIC3_01	0.0416	3.1387	0.0007	0.0024	-500	-200	0.0029	0.0099

A.3 GlyT-1a/SLC6A9 orthologous transcripts.

GlyT-1a/SLC6A9 orthologous sequences were obtained from **OrthoDB** 7 – A database cataloguing orthologous groups of 64 vertebrates, 57 arthropods , 175 fungi, 14 basal metazoans and 1115 bacteria species (as of the 22nd of July 2013). For complete documentation of this database and current release. see (Waterhouse *et al.*, 2013). OrthoDB does not differentiate between the two Cow GlyT-1a/SLC6A9 paralogous genes, both of which were used in the phylogenetic analysis. Transcripts of the Zebra Finch and Platypus have no GlyT-1a/SLC6A9 5'UTR defined in their respective genome assembly whilst care was taken to minimize the impact of the 'frameshifts' introduced by Ensembl's genome build program to the upstream sequences of transcripts from Dolphin, Zebra Finch, Turkey, Chicken, Kangaroo Rat, Lesser Hedgehog Tenrec, Platypus and Tarsier.

Common Name	Scientific Name	SLC6A9 Ensembl Gene ID	Orthologous Ensembl Transcript ID	Protein Length	Ensembl Database
Human	Homo sapiens	ENSG00000196517	ENST00000372310	633	hsapiens_gene_ensembl
Panda	Ailuropoda melanoleuca	ENSAMEG00000001028	ENSAMET00000001148	696	amelanoleuca_gene_ensembl
Anole lizard	Anolis carolinensis	ENSACAG00000014488	ENSACAT00000014677	636	acarolinensis_gene_ensembl
Cave fish	Astyanax mexicanus	ENSAMXG00000012229	ENSAMXT00000012594	649	amexicanus_gene_ensembl
Cow**	Bos taurus	ENSBTAG00000006043	ENSBTAT00000011456	757	btaurus_gene_ensembl
Cow	Bos taurus	ENSBTAG00000008690	ENSBTAT00000011454	633	btaurus_gene_ensembl
Marmoset	Callithrix jacchus	ENSCJAG00000003879	ENSCJAT00000007493	573	cjacchus_gene_ensembl
Dog	Canis lupus familiaris	ENSCAFG00000004826	ENSCAFT00000007773	632	cfamiliaris_gene_ensembl
Guinea Pig	Cavia porcellus	ENSCPOG00000013598	ENSCPOT00000013734	647	cporcellus_gene_ensembl
Zebrafish	Danio rerio	ENSDARG00000018534	ENSDART00000077462	653	drerio_gene_ensembl
Armadillo	Dasypus novemcinctus	ENSDNOG00000037887	ENSDNOT00000043792	411	dnovemcinctus_gene_ensembl
Kangaroo rat	Dipodomys ordii	ENSDORG00000005475	ENSDORT00000005474	687	dordii_gene_ensembl
Lesser hedgehog tenrec	Echinops telfairi	ENSETEG00000018374	ENSETET00000018377	705	etelfairi_gene_ensembl

Common Name	Scientific Name	SLC6A9 Ensembl Gene ID	Orthologous Ensembl Transcript ID	Protein Length	Ensembl Database
Horse	Equus caballus	ENSECAG00000002225	ENSECAT00000005457	623	ecaballus_gene_ensembl
Hedgehog	Erinaceus europaeus	ENSEEUG00000002019	ENSEEUT00000002027	693	eeuropaeus_gene_ensembl
Cat	Felis catus	ENSFCAG00000006493	ENSFCAT00000006495	761	fcatus_gene_ensembl
Flycatcher	Ficedula albicollis	ENSFALG00000005298	ENSFALT00000005539	633	falbicollis_gene_ensembl
Chicken	Gallus gallus	ENSGALG00000010098	ENSGALT00000038518	633	ggallus_gene_ensembl
Stickleback	Gasterosteus aculeatus	ENSGACG00000014462	ENSGACT00000019111	628	gaculeatus_gene_ensembl
Gorilla	Gorilla gorilla gorilla	ENSGGOG00000003651	ENSGGOT00000003674	706	ggorilla_gene_ensembl
Squirrel	Ictidomys tridecemlineatus	ENSSTOG00000004456	ENSSTOT00000004456	689	itridecemlineatus_gene_ensembl
Coelacanth	Latimeria chalumnae	ENSLACG00000004060	ENSLACT00000004601	565	lchalumnae_gene_ensembl
Spotted gar	Lepisosteus oculatus	ENSLOCG00000002674	ENSLOCT00000003165	639	loculatus_gene_ensembl
Elephant	Loxodonta africana	ENSLAFG00000000144	ENSLAFT00000000144	633	lafricana_gene_ensembl
Macaque	Macaca mulatta	ENSMMUG00000015158	ENSMMUT00000021254	709	mmulatta_gene_ensembl
Wallaby	Macropus eugenii	ENSMEUG00000015649	ENSMEUT00000015696	659	meugenii_gene_ensembl
Turkey	Meleagris gallopavo	ENSMGAG00000010003	ENSMGAT00000011249	635	mgallopavo_gene_ensembl
Mouse Lemur	Microcebus murinus	ENSMICG00000011037	ENSMICT00000011036	707	mmurinus_gene_ensembl
Opossum	Monodelphis domestica	ENSMODG00000002931	ENSMODT00000039267	573	mdomestica_gene_ensembl
Mouse	Mus musculus	ENSMUSG00000028542	ENSMUST00000030269	633	mmusculus_gene_ensembl
Ferret	Mustela putorius furo	ENSMPUG00000013573	ENSMPUT00000013686	633	mfuro_gene_ensembl
Gibbon	Nomascus leucogenys	ENSNLEG00000013923	ENSNLET00000017767	712	nleucogenys_gene_ensembl
Pika	Ochotona princeps	ENSOPRG00000013803	ENSOPRT00000013813	698	oprinceps_gene_ensembl
Tilapia	Oreochromis niloticus	ENSONIG00000009062	ENSONIT00000011406	659	oniloticus_gene_ensembl
Platypus	Ornithorhynchus anatinus	ENSOANG00000013788	ENSOANT00000021747	432	oanatinus_gene_ensembl
Rabbit	Oryctolagus cuniculus	ENSOCUG00000023733	ENSOCUT00000031972	706	ocuniculus_gene_ensembl
Medaka	Oryzias latipes	ENSORLG00000012525	ENSORLT00000015683	638	olatipes_gene_ensembl
Bushbaby	Otolemur garnettii	ENSOGAG00000001182	ENSOGAT00000001183	706	ogarnettii_gene_ensembl
Sheep	Ovis aries	ENSOARG00000000739	ENSOART00000000788	441	oaries_gene_ensembl

Common Name	Scientific Name	SLC6A9 Ensembl Gene ID	Orthologous Ensembl Transcript ID	Protein Length	Ensembl Database
Chimpanzee	Pan troglodytes	ENSPTRG00000022644	ENSPTRT00000047245	706	ptroglodytes_gene_ensembl
Orangutan	Pongo abelii	ENSPPYG00000001435	ENSPPYT00000001719	652	pabelii_gene_ensembl
Hyrax	Procavia capensis	ENSPCAG00000012744	ENSPCAT00000012781	699	pcapensis_gene_ensembl
Megabat	Pteropus vampyrus	ENSPVAG00000009510	ENSPVAT00000009510	706	pvampyrus_gene_ensembl
Rat	Rattus norvegicus	ENSRNOG00000019484	ENSRNOT00000048770	633	rnorvegicus_gene_ensembl
Tasmanian devil	Sarcophilus harrisii	ENSSHAG00000006607	ENSSHAT00000007664	443	sharrisii_gene_ensembl
Zebra Finch	Taeniopygia guttata	ENSTGUG00000007574	ENSTGUT00000007899	492	tguttata_gene_ensembl
Fugu	Takifugu rubripes	ENSTRUG00000000570	ENSTRUT00000001360	649	trubripes_gene_ensembl
Tarsier	Tarsius syrichta	ENSTSYG00000012739	ENSTSYT00000012738	624	tsyrichta_gene_ensembl
Tetraodon	Tetraodon nigroviridis	ENSTNIG00000007959	ENSTNIT00000002211	692	tnigroviridis_gene_ensembl
Tree Shrew	Tupaia belangeri	ENSTBEG00000004257	ENSTBET00000004291	674	tbelangeri_gene_ensembl
Dolphin	Tursiops truncatus	ENSTTRG00000001295	ENSTTRT00000001295	705	ttruncatus_gene_ensembl
Xenopus	Xenopus tropicalis	ENSXETG00000003268	ENSXETT00000007095	633	xtropicalis_gene_ensembl
Platyfish	Xiphophorus maculatus	ENSXMAG00000019099	ENSXMAT00000019205	660	xmaculatus_gene_ensembl

A.4   Plasmids used for reporter construction and PCR standards
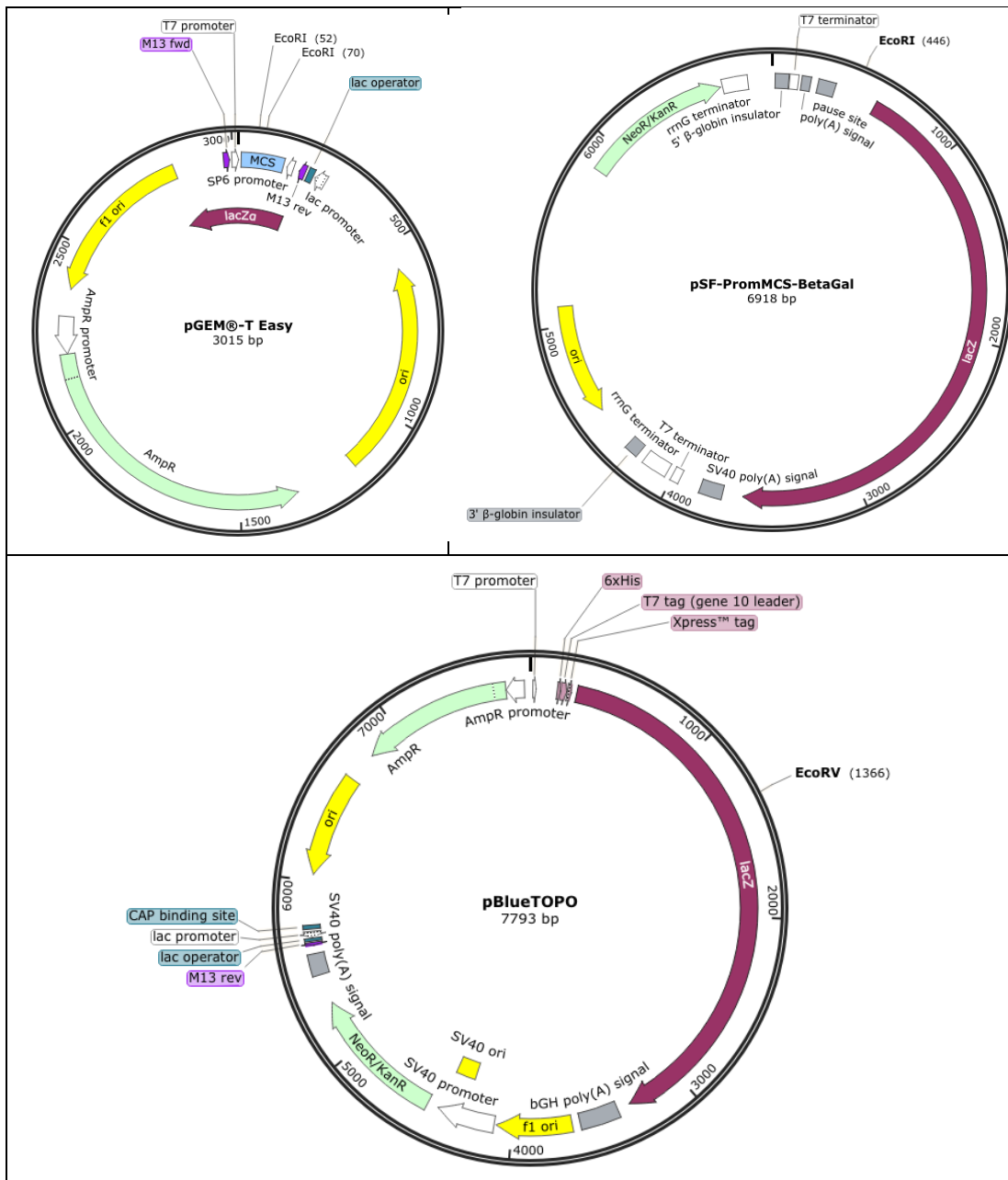
BIBLIOGRAPHY

Adamopoulos, C., Farmaki, E., Spilioti, E., Kiaris, H., Piperi, C. & Papavassiliou, A. G. 2014. **Advanced glycation end-products induce endoplasmic reticulum stress in human aortic endothelial cells.** *Clin Chem Lab Med,* 52**,** 151-60.

Adams, R. H., Sato, K., Shimada, S., Tohyama, M., Puschel, A. W. & Betz, H. 1995. **Gene structure and glial expression of the glycine transporter GlyT1 in embryonic and adult rodents.** *J Neurosci,* 15**,** 2524-32.

Afonyushkin, T., Oskolkova, O. V., Philippova, M., Resink, T. J., Erne, P., Binder, B. R. & Bochkov, V. N. 2010. **Oxidized phospholipids regulate expression of ATF4 and VEGF in endothelial cells via NRF2-dependent mechanism: novel point of convergence between electrophilic and unfolded protein stress pathways.** *Arterioscler Thromb Vasc Biol,* 30**,** 1007-13.

Agerberth, B. & Gudmundsson, G. H. 2006. **Host antimicrobial defence peptides in human disease.** *Curr Top Microbiol Immunol,* 306**,** 67-90.

Ait Ghezala, H., Jolles, B., Salhi, S., Castrillo, K., Carpentier, W., Cagnard, N., Bruhat, A., Fafournoux, P. & Jean-Jean, O. 2012. **Translation termination efficiency modulates ATF4 response by regulating ATF4 mRNA translation at 5' short ORFs.** *Nucleic Acids Res,* 40**,** 9557-70.

Albers, A., Broer, A., Wagner, C. A., Setiawan, I., Lang, P. A., Kranz, E. U., Lang, F. & Broer, S. 2001. **Na+ transport by the neural glutamine transporter ATA1.** *Pflugers Arch,* 443**,** 92-101.

Alexander, S. P., Benson, H. E., Faccenda, E., Pawson, A. J., Sharman, J. L., Spedding, M., Peters, J. A., Harmar, A. J. & Collaborators, C. 2013. **The Concise Guide to PHARMACOLOGY 2013/14: transporters.** *Br J Pharmacol,* 170**,** 1706-96.

Alfadda, A. A. & Sallam, R. M. 2012. **Reactive oxygen species in health and disease.** *J Biomed Biotechnol,* 2012**,** 936486.

Alfieri, R. R., Bonelli, M. A., Petronini, P. G., Desenzani, S., Cavazzoni, A., Borghetti, A. F. & Wheeler, K. P. 2005. **Hypertonic stress and amino acid deprivation both increase expression of mRNA for amino Acid transport system A.** *J Gen Physiol,* 125**,** 37-9.

Alikhani, M., Maclellan, C. M., Raptis, M., Vora, S., Trackman, P. C. & Graves, D. T. 2007. **Advanced glycation end products induce apoptosis in fibroblasts through activation of ROS, MAP kinases, and the FOXO1 transcription factor.** *Am J Physiol Cell Physiol,* 292**,** C850-6.

Allocco, D. J., Kohane, I. S. & Butte, A. J. 2004. **Quantifying the relationship between co-expression, co-regulation and gene function.** *Bmc Bioinformatics,* 5**,** 18.

Altamura, C., Maes, M., Dai, J. & Meltzer, H. Y. 1995. **Plasma concentrations of excitatory amino acids, serine, glycine, taurine and histidine in major depression.** *Eur Neuropsychopharmacol,* 5 Suppl**,** 71-5.

Alteheld, B., Evans, M. E., Gu, L. H., Ganapathy, V., Leibach, F. H., Jones, D. P. & Ziegler, T. R. 2005. **Alanylglutamine dipeptide and growth hormone maintain PepT1-mediated transport in oxidatively stressed Caco-2 cells.** *J Nutr,* 135**,** 19-26.

Amelio, I., Cutruzzola, F., Antonov, A., Agostini, M. & Melino, G. 2014. **Serine and glycine metabolism in cancer.** *Trends Biochem Sci,* 39**,** 191-8.

Amin, J., Ananthan, J. & Voellmy, R. 1988. **Key features of heat shock regulatory elements.** *Mol Cell Biol,* 8**,** 3761-9.

Amoutzias, G. D., Veron, A. S., Weiner, J., 3rd, Robinson-Rechavi, M., Bornberg-Bauer, E., Oliver, S. G. & Robertson, D. L. 2007. **One billion years of bZIP transcription factor evolution: conservation and change in dimerization and DNA-binding site specificity.** *Mol Biol Evol,* 24**,** 827-35.

Arnosti, D. N. & Kulkarni, M. M. 2005. **Transcriptional enhancers: Intelligent enhanceosomes or flexible billboards?** *J Cell Biochem,* 94**,** 890-8.

Arts, I. C. & Hollman, P. C. 2005. **Polyphenols and disease risk in epidemiologic studies.** *Am J Clin Nutr,* 81**,** 317S-325S.

Austin, S. A., Pollard, J. W., Jagus, R. & Clemens, M. J. 1986. **Regulation of polypeptide chain initiation and activity of initiation factor eIF-2 in Chinese-hamster-ovary cell mutants containing temperature-sensitive aminoacyl-tRNA synthetases.** *Eur J Biochem,* 157**,** 39-47.

Averous, J., Bruhat, A., Jousse, C., Carraro, V., Thiel, G. & Fafournoux, P. 2004. **Induction of CHOP expression by amino acid limitation requires both ATF4 expression and ATF2 phosphorylation.** *J Biol Chem,* 279**,** 5288-97.

Avivar-Valderas, A., Salas, E., Bobrovnikova-Marjon, E., Diehl, J. A., Nagi, C., Debnath, J. & Aguirre-Ghiso, J. A. 2011. **PERK integrates autophagy and oxidative stress responses to promote survival during extracellular matrix detachment.** *Mol Cell Biol,* 31**,** 3616-29.

Ayoubi, T. A. & Van De Ven, W. J. 1996. **Regulation of gene expression by alternative promoters.** *FASEB J,* 10**,** 453-60.

Bailey, D. & O'hare, P. 2007. **Transmembrane bZIP transcription factors in ER stress signaling and the unfolded protein response.** *Antioxid Redox Signal,* 9**,** 2305-21.

Baliova, M. & Jursky, F. 2004a. **Phosphatase inhibitors influence proteolytic cleavage pattern of glycine transporter GlyT2 N-terminus.** *Biologia,* 59**,** 843-845.

Baliova, M. & Jursky, F. 2004b. **Use of calpain for native GlyT2 N-terminal region separation and its potential use in transporter N-terminus interaction studies.** *Biologia,* 59**,** 839-842.

Baliova, M. & Jursky, F. 2005. **Calpain sensitive regions in the N-terminal cytoplasmic domains of glycine transporters GlyT1A and GlyT1B.** *Neurochem Res,* 30**,** 1093-100.

Baliova, M. & Jursky, F. 2010. **Calcium dependent modification of distal C-terminal sequences of glycine transporter GlyT1.** *Neurochem Int,* 57**,** 254-61.

Baltz, J. M. & Zhou, C. 2012. **Cell volume regulation in mammalian oocytes and preimplantation embryos.** *Mol Reprod Dev,* 79**,** 821-31.

Banjac, A., Perisic, T., Sato, H., Seiler, A., Bannai, S., Weiss, N., Kolle, P., Tschoep, K., Issels, R. D., Daniel, P. T., Conrad, M. & Bornkamm, G. W. 2008. **The cystine/cysteine cycle: a redox cycle regulating susceptibility versus resistance to cell death.** *Oncogene,* 27**,** 1618-28.

Bar-Joseph, Z., Gerber, G. K., Lee, T. I., Rinaldi, N. J., Yoo, J. Y., Robert, F., Gordon, D. B., Fraenkel, E., Jaakkola, T. S., Young, R. A. & Gifford, D. K. 2003. **Computational discovery of gene modules and regulatory networks.** *Nat Biotechnol,* 21**,** 1337-42.

Bar-Peled, L., Chantranupong, L., Cherniack, A. D., Chen, W. W., Ottina, K. A., Grabiner, B. C., Spear, E. D., Carter, S. L., Meyerson, M. & Sabatini, D. M. 2013. **A Tumor suppressor complex with GAP activity for the Rag GTPases that signal amino acid sufficiency to mTORC1.** *Science,* 340**,** 1100-6.

Barkess, G., Postnikov, Y., Campos, C. D., Mishra, S., Mohan, G., Verma, S., Bustin, M. & West, K. L. 2012. **The chromatin-binding protein HMGN3 stimulates histone acetylation and transcription across the Glyt1 gene.** *Biochem J,* 442**,** 495-505.

Barrett, J. C., Hansoul, S., Nicolae, D. L., Cho, J. H., Duerr, R. H., Rioux, J. D., Brant, S. R., Silverberg, M. S., Taylor, K. D., Barmada, M. M., Bitton, A., Dassopoulos, T., Datta, L. W., Green, T., Griffiths, A. M., Kistner, E. O., Murtha, M. T., Regueiro, M. D., Rotter, J. I., Schumm, L. P., Steinhart, A. H., Targan, S. R., Xavier, R. J., Consortium, N. I. G., Libioulle, C., Sandor, C., Lathrop, M., Belaiche, J., Dewit, O., Gut, I., Heath, S., Laukens, D., Mni, M., Rutgeerts, P., Van Gossum, A., Zelenika, D., Franchimont, D., Hugot, J. P., De Vos, M., Vermeire, S., Louis, E., Belgian-French, I. B. D. C., Wellcome Trust Case Control, C., Cardon, L. R., Anderson, C. A., Drummond, H., Nimmo, E., Ahmad, T., Prescott, N. J., Onnie, C. M., Fisher, S. A., Marchini, J., Ghori, J., Bumpstead, S., Gwilliam, R., Tremelling, M., Deloukas, P., Mansfield, J., Jewell, D., Satsangi, J., Mathew, C. G., Parkes, M., Georges, M. & Daly, M. J. 2008. **Genome-wide association defines more than 30 distinct susceptibility loci for Crohn's disease.** *Nat Genet,* 40**,** 955-62.

Barrett, T., Wilhite, S. E., Ledoux, P., Evangelista, C., Kim, I. F., Tomashevsky, M., Marshall, K. A., Phillippy, K. H., Sherman, P. M., Holko, M., Yefanov, A., Lee, H., Zhang, N., Robertson, C. L., Serova, N., Davis, S. & Soboleva, A. 2013. **NCBI GEO: archive for functional genomics data sets--update.** *Nucleic Acids Res,* 41**,** D991-5.

Bartsch, D., Ghirardi, M., Skehel, P. A., Karl, K. A., Herder, S. P., Chen, M., Bailey, C. H. & Kandel, E. R. 1995. **Aplysia Creb2 Represses Long-Term Facilitation - Relief of Repression Converts Transient Facilitation into Long-Term Functional and Structural-Change.** *Cell,* 83**,** 979-992.

Bas, H., Kalender, S. & Pandir, D. 2014. **In vitro effects of quercetin on oxidative stress mediated in human erythrocytes by benzoic acid and citric acid.** *Folia Biol (Krakow),* 62**,** 59-66.

Beckman, M. L., Bernstein, E. M. & Quick, M. W. 1998. **Protein kinase C regulates the interaction between a GABA transporter and syntaxin 1A.** *Journal of Neuroscience,* 18**,** 6103-6112.

Bell, K. F. S., Fowler, J. H., Al-Mubarak, B., Horsburgh, K. & Hardingham, G. E. 2011. **Activation of Nrf2-Regulated Glutathione Pathway Genes by Ischemic Preconditioning.** *Oxidative Medicine and Cellular Longevity,* 2011**,** 1-7.

Berlanga, J. J., Santoyo, J. & De Haro, C. 1999. **Characterization of a mammalian homolog of the GCN2 eukaryotic initiation factor 2alpha kinase.** *Eur J Biochem,* 265**,** 754-62.

Bermingham, J. R., Jr. & Pennington, J. 2004. **Organization and expression of the SLC36 cluster of amino acid transporter genes.** *Mamm Genome,* 15**,** 114-25.

Berthold, H. K., Jahoor, F., Klein, P. D. & Reeds, P. J. 1995. **Estimates of the effect of feeding on whole-body protein degradation in women vary with the amino acid used as tracer.** *The Journal of nutrition.*

Betz, H. & Eulenburg, V. 2012. **Transport activities and expression patterns of glycine transporters 1 and 2 in the developing murine brain stem and spinal cord.** *Biochemical and biophysical research communications,* 423**,** 661-666.

Betz, H., Gomeza, J., Armsen, W., Scholze, P. & Eulenburg, V. 2006. **Glycine transporters: essential regulators of synaptic transmission.** *Biochemical Society Transactions,* 34**,** 55-58.

Bitensky, L. 1990. **Glutathione: Chemical, biochemical, and medical aspects.** *Cell Biochemistry and Function,* 8**,** 645-727.

Bohmer, C., Broer, A., Munzinger, M., Kowalczuk, S., Rasko, J. E., Lang, F. & Broer, S. 2005. **Characterization of mouse amino acid transporter B0AT1 (slc6a19).** *Biochem J,* 389**,** 745-51.

Boll, M., Daniel, H. & Gasnier, B. 2004. **The SLC36 family: proton-coupled transporters for the absorption of selected amino acids from extracellular and intracellular proteolysis.** *Pflugers Arch,* 447**,** 776-9.

Boll, M., Foltz, M., Anderson, C. M., Oechsler, C., Kottra, G., Thwaites, D. T. & Daniel, H. 2003a. **Substrate recognition by the mammalian proton-dependent amino acid transporter PAT1.** *Mol Membr Biol,* 20**,** 261-9.

Boll, M., Foltz, M., Rubio-Aliaga, I. & Daniel, H. 2003b. **A cluster of proton/amino acid transporter genes in the human and mouse genomes.** *Genomics,* 82**,** 47-56.

Bolte, G., Beuermann, K. & Stern, M. 1997. **THE CELL LINES CACO-2, T84, AND HT-29: MODELS OF ENTEROCYTIC DIFFERENTIATION AND FUNCTION.** *Journal of Pediatric Gastroenterology and Nutrition,* 24**,** 473.

Borowsky, B. & Hoffman, B. J. 1998. **Analysis of a gene encoding two glycine transporter variants reveals alternative promoter usage and a novel gene structure.** *J Biol Chem,* 273**,** 29077-85.

Borowsky, B., Mezey, E. & Hoffman, B. J. 1993. **Two glycine transporter variants with distinct localization in the CNS and peripheral tissues are encoded by a common gene.** *Neuron,* 10**,** 851-63.

Bower, R. H., Cerra, F. B., Bershadsky, B., Licari, J. J., Hoyt, D. B., Jensen, G. L., Van Buren, C. T., Rothkopf, M. M., Daly, J. M. & Adelsberg, B. R. 1995. **Early enteral administration of a formula (Impact Registered Trademark) supplemented with arginine, nucleotides, and fish oil in intensive care unit patients: Results of a multicenter, prospective, randomized, clinical trial.** *Critical Care Medicine,* 23**,** 436-449.

Boza, J. J., Jahoor, F. & Reeds, P. J. 1996. **Ribonucleic acid nucleotides in maternal and fetal tissues derive almost exclusively from synthesis de novo in pregnant mice.** *Journal of Nutrition,* 126**,** 1749-1758.

Brasse-Lagnel, C., Lavoinne, A. & Husson, A. 2009. **Control of mammalian gene expression by amino acids, especially glutamine.** *FEBS J,* 276**,** 1826-44.

Broberg, M., Holm, R., Tonsberg, H., Frolund, S., Ewon, K. B., Nielsen, A., Brodin, B., Jensen, A., Kall, M. A., Christensen, K. V. & Nielsen, C. U. 2012. **Function and expression of the proton-coupled amino acid transporter PAT1 along the rat gastrointestinal tract: implications for intestinal absorption of gaboxadol.** *Br J Pharmacol,* 167**,** 654-65.

Brocker, C., Thompson, D. C. & Vasiliou, V. 2012. **The role of hyperosmotic stress in inflammation and disease.** *Biomol Concepts,* 3**,** 345-364.

Broer, A., Albers, A., Setiawan, I., Edwards, R. H., Chaudhry, F. A., Lang, F., Wagner, C. A. & Broer, S. 2002. **Regulation of the glutamine transporter SN1 by extracellular pH and intracellular sodium ions.** *J Physiol,* 539**,** 3-14.

Broer, A., Cavanaugh, J. A., Rasko, J. E. & Broer, S. 2006. **The molecular basis of neutral aminoacidurias.** *Pflugers Arch,* 451**,** 511-7.

Broer, S. 2008a. **Amino acid transport across mammalian intestinal and renal epithelia.** *Physiol Rev,* 88**,** 249-86.

Broer, S. 2008b. **Apical transporters for neutral amino acids: physiology and pathophysiology.** *Physiology (Bethesda),* 23**,** 95-103.

Broer, S., Bailey, C. G., Kowalczuk, S., Ng, C., Vanslambrouck, J. M., Rodgers, H., Auray-Blais, C., Cavanaugh, J. A., Broer, A. & Rasko, J. E. 2008. **Iminoglycinuria and hyperglycinuria are discrete human phenotypes resulting from complex mutations in proline and glycine transporters.** *J Clin Invest,* 118**,** 3881-92.

Broer, S. & Gether, U. 2012. **The solute carrier 6 family of transporters.** *Br J Pharmacol,* 167**,** 256-78.

Bruhat, A., Averous, J., Carraro, V., Zhong, C., Reimold, A. M., Kilberg, M. S. & Fafournoux, P. 2002. **Differences in the molecular mechanisms involved in the transcriptional activation of the CHOP and asparagine synthetase genes in response to amino acid deprivation or activation of the unfolded protein response.** *J Biol Chem,* 277**,** 48107-14.

Bruhat, A., Cherasse, Y., Maurin, A. C., Breitwieser, W., Parry, L., Deval, C., Jones, N., Jousse, C. & Fafournoux, P. 2007. **ATF2 is required for amino acid-regulated transcription by orchestrating specific histone acetylation.** *Nucleic Acids Res,* 35**,** 1312-21.

Bulyk, M. L. 2007. **Protein Binding Microarrays for the Characterization of Protein-DNA Interactions.** *Advances in biochemical engineering/biotechnology,* 104**,** 65.

Burg, M. B., Ferraris, J. D. & Dmitrieva, N. I. 2007. **Cellular Response to Hyperosmotic Stresses**.

Burg, M. B., Kwon, E. D. & Kultz, D. 1997. **Regulation of gene expression by hypertonicity.** *Annu Rev Physiol,* 59**,** 437-55.

Burke, T. W. & Kadonaga, J. T. 1996. **Drosophila TFIID binds to a conserved downstream basal promoter element that is present in many TATA-box-deficient promoters.** *Genes & Development,* 10**,** 711-724.

Busch, S. J. & Sassonecorsi, P. 1990. **Fos, Jun and Creb Basic-Domain Peptides Have Intrinsic DNA-Binding Activity Enhanced by a Novel Stabilizing Factor.** *Oncogene,* 5**,** 1549-1556.

Butler, J. E. & Kadonaga, J. T. 2002. **The RNA polymerase II core promoter: a key component in the regulation of gene expression.** *Genes Dev,* 16**,** 2583-92.

Cao, S. S., Song, B. & Kaufman, R. J. 2012. **PKR protects colonic epithelium against colitis through the unfolded protein response and prosurvival signaling.** *Inflamm Bowel Dis,* 18**,** 1735-42.

Carey, M. F., Peterson, C. L. & Smale, S. T. 2009. **Chromatin immunoprecipitation (ChIP).** *Cold Spring Harb Protoc,* 2009**,** pdb prot5279.

Carracedo, A., Gironella, M., Lorente, M., Garcia, S., Guzman, M., Velasco, G. & Iovanna, J. L. 2006. **Cannabinoids induce apoptosis of pancreatic tumor cells via endoplasmic reticulum stress-related genes.** *Cancer Res,* 66**,** 6748-55.

Carracedo, A., Ma, L., Teruya-Feldstein, J., Rojo, F., Salmena, L., Alimonti, A., Egia, A., Sasaki, A. T., Thomas, G., Kozma, S. C., Papa, A., Nardella, C., Cantley, L. C., Baselga, J. & Pandolfi, P. P. 2008. **Inhibition of mTORC1 leads to MAPK pathway activation through a PI3K-dependent feedback loop in human cancer.** *J Clin Invest,* 118**,** 3065-74.

Caselle, M., Di Cunto, F. & Provero, P. 2002. **Correlating overrepresented upstream motifs to gene expression: a computational approach to regulatory element discovery in eukaryotes.** *BMC Bioinformatics,* 3**,** 7.

Catalioto, R. M., Maggi, C. A. & Giuliani, S. 2011. **Intestinal epithelial barrier dysfunction in disease and possible therapeutical interventions.** *Curr Med Chem,* 18**,** 398-426.

Chantret, I., Barbat, A., Dussaulx, E., Brattain, M. G. & Zweibaum, A. 1988. **Epithelial polarity, villin expression, and enterocytic differentiation of cultured human colon carcinoma cells: a survey of twenty cell lines.** *Cancer Res,* 48**,** 1936-42.

Chen, C., Dudenhausen, E., Chen, H., Pan, Y. X., Gjymishka, A. & Kilberg, M. S. 2005. **Amino-acid limitation induces transcription from the human C/EBPbeta gene via an enhancer activity located downstream of the protein coding sequence.** *Biochem J,* 391**,** 649-58.

Chen, H., Pan, Y. X., Dudenhausen, E. E. & Kilberg, M. S. 2004. **Amino acid deprivation induces the transcription rate of the human asparagine synthetase gene through a timed program of expression and promoter binding of nutrient-responsive basic region/leucine zipper transcription factors as well as localized histone acetylation.** *J Biol Chem,* 279**,** 50829-39.

Chen, K. Y. & Canellakis, E. S. 1977. **Enzyme regulation in neuroblastoma cells in a salts/glucose medium: induction of ornithine decarboxylase by asparagine and glutamine.** *Proc Natl Acad Sci U S A,* 74**,** 3791-5.

Chen, Z., Fei, Y. J., Anderson, C. M. H., Wake, K. A., Miyauchi, S., Huang, W., Thwaites, D. T. & Ganapathy, V. 2002. **Structure, function and immunolocalization of a proton-coupled amino acid transporter (hPAT1) in the human intestinal cell line Caco-2.** *The Journal of Physiology,* 546**,** 349-361.

Cheng, A. S., Cheng, Y. H., Chiou, C. H. & Chang, T. L. 2012. **Resveratrol upregulates Nrf2 expression to attenuate methylglyoxal-induced insulin resistance in Hep G2 cells.** *J Agric Food Chem,* 60**,** 9180-7.

Cheng, X., Liu, H., Jiang, C. C., Fang, L., Chen, C., Zhang, X. D. & Jiang, Z. W. 2014. **Connecting endoplasmic reticulum stress to autophagy through IRE1/JNK/beclin-1 in breast cancer cells.** *Int J Mol Med,* 34**,** 772-81.

Cherasse, Y., Maurin, A. C., Chaveroux, C., Jousse, C., Carraro, V., Parry, L., Deval, C., Chambon, C., Fafournoux, P. & Bruhat, A. 2007. **The p300/CBP-associated**

factor (PCAF) is a cofactor of ATF4 for amino acid-regulated transcription of CHOP.** *Nucleic Acids Res,* 35**,** 5954-65.

Cheung, K. L., Yu, S., Pan, Z., Ma, J., Wu, T. Y. & Kong, A. N. 2011. **tBHQ-induced HO-1 expression is mediated by calcium through regulation of Nrf2 binding to enhancer and polymerase II to promoter region of HO-1.** *Chem Res Toxicol,* 24**,** 670-6.

Choi, D. Y., Lee, Y. J., Hong, J. T. & Lee, H. J. 2012. **Antioxidant properties of natural polyphenols and their therapeutic potentials for Alzheimer's disease.** *Brain Res Bull,* 87**,** 144-53.

Chojnacka, K., Saeid, A., Witkowska, Z. & Tuhy, L. 2012. **Biologically active compounds in seaweed extracts—the prospects for the application.** *The Open Conference Proceedings Journal,* 3**,** 20-28.

Chorley, B. N., Campbell, M. R., Wang, X., Karaca, M., Sambandan, D., Bangura, F., Xue, P., Pi, J., Kleeberger, S. R. & Bell, D. A. 2012. **Identification of novel NRF2-regulated genes by ChIP-Seq: influence on retinoid X receptor alpha.** *Nucleic Acids Res,* 40**,** 7416-29.

Chow, C. W., Grinstein, S. & Rotstein, O. D. 1995. **Signaling events in monocytes and macrophages.** *New Horiz,* 3**,** 342-51.

Christie, G. R., Ford, D., Howard, A., Clark, M. A. & Hirst, B. H. 2001. **Glycine supply to human enterocytes mediated by high-affinity basolateral GLYT1.** *Gastroenterology,* 120**,** 439-48.

Chue, P. 2013. **Glycine Reuptake Inhibition as a New Therapeutic Approach in Schizophrenia: Focus on the Glycine Transporter 1 (GlyT1).** *Current Pharmaceutical Design,* 19**,** 1311-1320.

Ciccocioppo, R., Vanoli, A., Klersy, C., Imbesi, V., Boccaccio, V., Manca, R., Betti, E., Cangemi, G. C., Strada, E., Besio, R., Rossi, A., Falcone, C., Ardizzone, S., Fociani, P., Danelli, P. & Corazza, G. R. 2013. **Role of the advanced glycation end products receptor in Crohn's disease inflammation.** *World J Gastroenterol,* 19**,** 8269-81.

Clements, C. M., Mcnally, R. S., Conti, B. J., Mak, T. W. & Ting, J. P. 2006. **DJ-1, a cancer- and Parkinson's disease-associated protein, stabilizes the antioxidant transcriptional master regulator Nrf2.** *Proc Natl Acad Sci U S A,* 103**,** 15091-6.

Consortium, T. E. P. 2012. **An integrated encyclopedia of DNA elements in the human genome.** *The EMBO journal,* 489**,** 57-74.

Cooper, G. M. & Brown, C. D. 2008. **Qualifying the relationship between sequence conservation and molecular function.** *Genome Res,* 18**,** 201-5.

Cos, P., Ying, L., Calomme, M., Hu, J. P., Cimanga, K., Van Poel, B., Pieters, L., Vlietinck, A. J. & Vanden Berghe, D. 1998. **Structure-activity relationship

and classification of flavonoids as inhibitors of xanthine oxidase and superoxide scavengers. *J Nat Prod,* 61**,** 71-6.

Coulombe, B. & Burton, Z. F. 1999. **DNA bending and wrapping around RNA polymerase: a "revolutionary" model describing transcriptional mechanisms.** *Microbiol Mol Biol Rev,* 63**,** 457-78.

Creyghton, M. P., Cheng, A. W., Welstead, G. G., Kooistra, T., Carey, B. W., Steine, E. J., Hanna, J., Lodato, M. A., Frampton, G. M., Sharp, P. A., Boyer, L. A., Young, R. A. & Jaenisch, R. 2010. **Histone H3K27ac separates active from poised enhancers and predicts developmental state.** *Proc Natl Acad Sci U S A,* 107**,** 21931-6.

Cubelos, B., Gimenez, C. & Zafra, F. 2005. **Localization of the GLYT1 glycine transporter at glutamatergic synapses in the rat brain.** *Cereb Cortex,* 15**,** 448-59.

Cullen, M. E. & Barton, P. J. 2007. **Mapping transcriptional start sites and in silico DNA footprinting.** *Methods Mol Biol,* 366**,** 203-16.

Cullinan, S. B. & Diehl, J. A. 2004. **PERK-dependent activation of Nrf2 contributes to redox homeostasis and cell survival following endoplasmic reticulum stress.** *J Biol Chem,* 279**,** 20108-17.

Cullinan, S. B. & Diehl, J. A. 2006. **Coordination of ER and oxidative stress signaling: the PERK/Nrf2 signaling pathway.** *Int J Biochem Cell Biol,* 38**,** 317-32.

Cullinan, S. B., Zhang, D., Hannink, M., Arvisais, E., Kaufman, R. J. & Diehl, J. A. 2003. **Nrf2 is a direct PERK substrate and effector of PERK-dependent cell survival.** *Molecular and Cellular Biology,* 23**,** 7198-7209.

Daniel, H. & Kottra, G. 2004. **The proton oligopeptide cotransporter family SLC15 in physiology and pharmacology.** *Pflugers Arch,* 447**,** 610-8.

Dawson, K. M. & Baltz, J. M. 1997. **Organic osmolytes and embryos: substrates of the Gly and beta transport systems protect mouse zygotes against the effects of raised osmolarity.** *Biol Reprod,* 56**,** 1550-8.

Dawson, K. M., Collins, J. L. & Baltz, J. M. 1998. **Osmolarity-dependent glycine accumulation indicates a role for glycine as an organic osmolyte in early preimplantation mouse embryos.** *Biology of Reproduction,* 59**,** 225-232.

Day, C. P. & James, O. F. 1998. **Hepatic steatosis: innocent bystander or guilty party?** *Hepatology,* 27**,** 1463-6.

Day, C. P. & Saksena, S. 2002. **Non-alcoholic steatohepatitis: definitions and pathogenesis.** *J Gastroenterol Hepatol,* 17 Suppl 3**,** S377-84.

Deaton, A. M. & Bird, A. 2011. **CpG islands and the regulation of transcription.** *Genes Dev,* 25**,** 1010-22.

Delie, F. & Rubas, W. 1997. **A human colonic cell line sharing similarities with enterocytes as a model to examine oral absorption: Advantages and limitations of the Caco-2 model.** *Critical Reviews in Therapeutic Drug Carrier Systems,* 14**,** 221-286.

Den Butter, G., Lindell, S. L., Sumimoto, R., Schilling, M. K., Southard, J. H. & Belzer, F. O. 1993. **Effect of glycine in dog and rat liver transplantation.** *Transplantation,* 56**,** 817-22.

Den Butter, G., Marsh, D. C., Lindell, S. L., Belzer, F. O. & Southard, J. H. 1994. **Effect of glycine on isolated, perfused rabbit livers following 48-hour preservation in University of Wisconsin solution without glutathione.** *Transplant international : official journal of the European Society for Organ Transplantation,* 7**,** 195-200.

Deppmann, C. D., Alvania, R. S. & Taparowsky, E. J. 2006. **Cross-species annotation of basic leucine zipper factor interactions: Insight into the evolution of closed interaction networks.** *Mol Biol Evol,* 23**,** 1480-92.

Dey, S., Baird, T. D., Zhou, D., Palam, L. R., Spandau, D. F. & Wek, R. C. 2010. **Both transcriptional regulation and translational control of ATF4 are central to the integrated stress response.** *J Biol Chem,* 285**,** 33165-74.

Di Salvo, M. L., Contestabile, R., Paiardini, A. & Maras, B. 2013. **Glycine consumption and mitochondrial serine hydroxymethyltransferase in cancer cells: the heme connection.** *Med Hypotheses,* 80**,** 633-6.

Dickhout, J. G., Carlisle, R. E., Jerome, D. E., Mohammed-Ali, Z., Jiang, H., Yang, G., Mani, S., Garg, S. K., Banerjee, R., Kaufman, R. J., Maclean, K. N., Wang, R. & Austin, R. C. 2012. **Integrated stress response modulates cellular redox state via induction of cystathionine gamma-lyase: cross-talk between integrated stress response and thiol metabolism.** *J Biol Chem,* 287**,** 7603-14.

Doring, F., Walter, J., Will, J., Focking, M., Boll, M., Amasheh, S., Clauss, W. & Daniel, H. 1998. **Delta-aminolevulinic acid transport by intestinal and renal peptide transporters and its physiological and clinical implications.** *J Clin Invest,* 101**,** 2761-7.

Duret, L., Dorkeld, F. & Gautier, C. 1993. **Strong conservation of non-coding sequences during vertebrates evolution: potential involvement in post-transcriptional regulation of gene expression.** *Nucleic Acids Res,* 21**,** 2315-22.

Edgar, R., Domrachev, M. & Lash, A. E. 2002. **Gene Expression Omnibus: NCBI gene expression and hybridization array data repository.** *Nucleic Acids Res,* 30**,** 207-10.

Eftekharzadeh, B., Maghsoudi, N. & Khodagholi, F. 2010. **Stabilization of transcription factor Nrf2 by tBHQ prevents oxidative stress-induced amyloid beta formation in NT2N neurons.** *Biochimie,* 92**,** 245-53.

Ehren, J. L. & Maher, P. 2013. **Concurrent regulation of the transcription factors Nrf2 and ATF4 mediates the enhancement of glutathione levels by the flavonoid fisetin.** *Biochem Pharmacol,* 85**,** 1816-26.

El Hafidi, M., Perez, I., Zamora, J., Soto, V., Carvajal-Sandoval, G. & Banos, G. 2004. **Glycine intake decreases plasma free fatty acids, adipose cell size, and blood pressure in sucrose-fed rats.** *Am J Physiol Regul Integr Comp Physiol,* 287**,** R1387-93.

Emami, K. H., Burke, T. W. & Smale, S. T. 1998. **Sp1 activation of a TATA-less promoter requires a species-specific interaction involving transcription factor IID.** *Nucleic Acids Res,* 26**,** 839-46.

Engle, M. J., Goetz, G. S. & Alpers, D. H. 1998. **Caco-2 cells express a combination of colonocyte and enterocyte phenotypes.** *J Cell Physiol,* 174**,** 362-9.

Erickson, A. M., Nevarea, Z., Gipp, J. J. & Mulcahy, R. T. 2002. **Identification of a variant antioxidant response element in the promoter of the human glutamate-cysteine ligase modifier subunit gene. Revision of the ARE consensus sequence.** *J Biol Chem,* 277**,** 30730-7.

Escartin, C., Won, S. J., Malgorn, C., Auregan, G., Berman, A. E., Chen, P. C., Deglon, N., Johnson, J. A., Suh, S. W. & Swanson, R. A. 2011. **Nuclear factor erythroid 2-related factor 2 facilitates neuronal glutathione synthesis by upregulating neuronal excitatory amino acid transporter 3 expression.** *J Neurosci,* 31**,** 7392-401.

Fafournoux, P., Bruhat, A. & Jousse, C. 2000. **Amino acid regulation of gene expression.** *Biochem J,* 351**,** 1-12.

Fagerberg, L., Hallstrom, B. M., Oksvold, P., Kampf, C., Djureinovic, D., Odeberg, J., Habuka, M., Tahmasebpoor, S., Danielsson, A., Edlund, K., Asplund, A., Sjostedt, E., Lundberg, E., Szigyarto, C. A., Skogs, M., Takanen, J. O., Berling, H., Tegel, H., Mulder, J., Nilsson, P., Schwenk, J. M., Lindskog, C., Danielsson, F., Mardinoglu, A., Sivertsson, A., Von Feilitzen, K., Forsberg, M., Zwahlen, M., Olsson, I., Navani, S., Huss, M., Nielsen, J., Ponten, F. & Uhlen, M. 2014. **Analysis of the human tissue-specific expression by genome-wide integration of transcriptomics and antibody-based proteomics.** *Mol Cell Proteomics,* 13**,** 397-406.

Fawcett, T. W., Martindale, J. L., Guyton, K. Z., Hai, T. & Holbrook, N. J. 1999. **Complexes containing activating transcription factor (ATF)/cAMP-responsive-element-binding protein (CREB) interact with the CCAAT enhancer-binding protein (C/EBP)-ATF composite site to regulate Gadd153 expression during the stress response.** *Biochemical Journal,* 339**,** 135-141.

Fei, Y. J., Sugawara, M., Nakanishi, T., Huang, W., Wang, H., Prasad, P. D., Leibach, F. H. & Ganapathy, V. 2000. **Primary structure, genomic organization, and functional and electrogenic characteristics of human system N 1, a Na+- and H+-coupled glutamine transporter.** *J Biol Chem,* 275**,** 23707-17.

Feil, R. & Fraga, M. F. 2011. **Epigenetics and the environment: emerging patterns and implications.** *Nat Rev Genet,* 13**,** 97-109.

Fernandez, J., Lopez, A. B., Wang, C., Mishra, R., Zhou, L., Yaman, I., Snider, M. D. & Hatzoglou, M. 2003. **Transcriptional control of the arginine/lysine transporter, cat-1, by physiological stress.** *J Biol Chem,* 278**,** 50000-9.

Fernandez-Sanchez, E., Martinez-Villarreal, J., Gimenez, C. & Zafra, F. 2009. **Constitutive and regulated endocytosis of the glycine transporter GLYT1b is controlled by ubiquitination.** *J Biol Chem,* 284**,** 19482-92.

Ferraris, J. D. & Garcia-Perez, A. 2001. **Osmotically responsive genes: The mammalian osmotic response element (ORE).** *American Zoologist,* 41**,** 734-742.

Filtz, T. M., Vogel, W. K. & Leid, M. 2014. **Regulation of transcription factor activity by interconnected post-translational modifications.** *Trends Pharmacol Sci,* 35**,** 76-85.

Fogh, J., Fogh, J. M. & Orfeo, T. 1977. **One hundred and twenty-seven cultured human tumor cell lines producing tumors in nude mice.** *J Natl Cancer Inst,* 59**,** 221-6.

Foltz, M., Boll, M., Raschka, L., Kottra, G. & Daniel, H. 2004. **A novel bifunctionality: PAT1 and PAT2 mediate electrogenic proton/amino acid and electroneutral proton/fatty acid symport.** *FASEB J,* 18**,** 1758-60.

Fornes, A., Nunez, E., Alonso-Torres, P., Aragon, C. & Lopez-Corcuera, B. 2008. **Trafficking properties and activity regulation of the neuronal glycine transporter GLYT2 by protein kinase C.** *Biochem J,* 412**,** 495-506.

Fornes, A., Nunez, E., Aragon, C. & Lopez-Corcuera, B. 2004. **The second intracellular loop of the glycine transporter 2 contains crucial residues for glycine transport and phorbol ester-induced regulation.** *J Biol Chem,* 279**,** 22934-43.

Franceschini, A., Szklarczyk, D., Frankild, S., Kuhn, M., Simonovic, M., Roth, A., Lin, J., Minguez, P., Bork, P., Von Mering, C. & Jensen, L. J. 2013. **STRING v9.1: protein-protein interaction networks, with increased coverage and integration.** *Nucleic Acids Res,* 41**,** D808-15.

Franchi-Gazzola, R., Gaccioli, F., Bevilacqua, E., Visigalli, R., Dall'asta, V., Sala, R., Varoqui, H., Erickson, J. D., Gazzola, G. C. & Bussolati, O. 2004. **The synthesis of SNAT2 transporters is required for the hypertonic stimulation of system A transport activity.** *Biochim Biophys Acta,* 1667**,** 157-66.

Franchi-Gazzola, R., Visigalli, R., Dall'asta, V., Sala, R., Woo, S. K., Kwon, H. M., Gazzola, G. C. & Bussolati, O. 2001. **Amino acid depletion activates TonEBP and sodium-coupled inositol transport.** *Am J Physiol Cell Physiol,* 280**,** C1465-74.

Freeman, H. J. & Kim, Y. S. 1978. **Digestion and absorption of protein.** *Annu Rev Med,* 29**,** 99-116.

Friling, R. S., Bergelson, S. & Daniel, V. 1992. **Two adjacent AP-1-like binding sites form the electrophile-responsive element of the murine glutathione S-transferase Ya subunit gene.** *Proc Natl Acad Sci U S A,* 89**,** 668-72.

Funes, J. M., Henderson, S., Kaufman, R., Flanagan, J. M., Robson, M., Pedley, B., Moncada, S. & Boshoff, C. 2014. **Oncogenic transformation of mesenchymal stem cells decreases Nrf2 expression favoring in vivo tumor growth and poorer survival.** *Molecular Cancer,* 13**,** 20.

Gaccioli, F., Huang, C. C., Wang, C., Bevilacqua, E., Franchi-Gazzola, R., Gazzola, G. C., Bussolati, O., Snider, M. D. & Hatzoglou, M. 2006. **Amino acid starvation induces the SNAT2 neutral amino acid transporter by a mechanism that involves eukaryotic initiation factor 2alpha phosphorylation and cap-independent translation.** *J Biol Chem,* 281**,** 17929-40.

Gaemers, I. C., Stallen, J. M., Kunne, C., Wallner, C., Van Werven, J., Nederveen, A. & Lamers, W. H. 2011. **Lipotoxicity and steatohepatitis in an overfed mouse model for non-alcoholic fatty liver disease.** *Biochim Biophys Acta,* 1812**,** 447-58.

Gan, L., Johnson, D. A. & Johnson, J. A. 2010. **Keap1-Nrf2 activation in the presence and absence of DJ-1.** *Eur J Neurosci,* 31**,** 967-77.

Gatley, S. J. & Sherratt, H. S. 1977. **The synthesis of hippurate from benzoate and glycine by rat liver mitochondria. Submitochondrial localization and kinetics.** *Biochem J,* 166**,** 39-47.

Geboes, K., Leo, M., Fanni, D. & Faa, G. 2014. **Inflammatory Bowel Diseases**. *In:* GEBOES, K., NEMOLATO, S., LEO, M. & FAA, G. (eds.) *Colitis.* Springer International Publishing.

Geerlings, A., Lopez-Corcuera, B. & Aragon, C. 2000. **Characterization of the interactions between the glycine transporters GLYT1 and GLYT2 and the SNARE protein syntaxin 1A.** *Febs Letters,* 470**,** 51-54.

Geerlings, A., Nunez, E., Lopez-Corcuera, B. & Aragon, C. 2001. **Calcium- and syntaxin 1-mediated trafficking of the neuronal glycine transporter GLYT2.** *J Biol Chem,* 276**,** 17584-90.

Geillinger, K. E., Kipp, A. P., Schink, K., Roder, P. V., Spanier, B. & Daniel, H. 2014. **Nrf2 regulates the expression of the peptide transporter PEPT1 in the human colon carcinoma cell line Caco-2.** *Biochim Biophys Acta,* 1840**,** 1747-54.

Gerstein, M. B., Kundaje, A., Hariharan, M., Landt, S. G., Yan, K. K., Cheng, C., Mu, X. J., Khurana, E., Rozowsky, J., Alexander, R., Min, R., Alves, P., Abyzov, A., Addleman, N., Bhardwaj, N., Boyle, A. P., Cayting, P., Charos, A., Chen, D. Z., Cheng, Y., Clarke, D., Eastman, C., Euskirchen, G., Frietze, S., Fu, Y., Gertz, J., Grubert, F., Harmanci, A., Jain, P., Kasowski, M., Lacroute, P., Leng, J., Lian, J., Monahan, H., O'geen, H., Ouyang, Z., Partridge, E. C., Patacsil, D., Pauli, F.,

Raha, D., Ramirez, L., Reddy, T. E., Reed, B., Shi, M., Slifer, T., Wang, J., Wu, L., Yang, X., Yip, K. Y., Zilberman-Schapira, G., Batzoglou, S., Sidow, A., Farnham, P. J., Myers, R. M., Weissman, S. M. & Snyder, M. 2012. **Architecture of the human regulatory network derived from ENCODE data.** *Nature,* 489**,** 91-100.

Gibney, E. R. & Nolan, C. M. 2010. **Epigenetics and gene expression.** *Heredity (Edinb),* 105**,** 4-13.

Gietzen, D. W., Ross, C. M., Hao, S. & Sharp, J. W. 2004. **Phosphorylation of eIF2alpha is involved in the signaling of indispensable amino acid deficiency in the anterior piriform cortex of the brain in rats.** *The Journal of nutrition,* 134**,** 717-723.

Glover, J. N. & Harrison, S. C. 1995. **Crystal structure of the heterodimeric bZIP transcription factor c-Fos-c-Jun bound to DNA.** *Nature,* 373**,** 257-61.

Glover, J. N. M., Chen, L. & Verdine, G. L. 1995. **A Structural Analysis of the bZIP Family of Transcription Factors: Dimerization, DNA Binding and Interactions with Other Transcription Factors.** *J Biomol Struct Dyn***,** 23-33.

Gohrbandt, B., Fischer, S., Warnecke, G., Avsar, M., Sommer, S. P., Haverich, A. & Strueber, M. 2006. **Glycine intravenous donor preconditioning is superior to glycine supplementation to low-potassium dextran flush preservation and improves graft function in a large animal lung transplantation model after 24 hours of cold ischemia.** *J Thorac Cardiovasc Surg,* 131**,** 724-9.

Gomeza, J., Hulsmann, S., Ohno, K., Eulenburg, V., Szoke, K., Richter, D. & Betz, H. 2003a. **Inactivation of the glycine transporter 1 gene discloses vital role of glial glycine uptake in glycinergic inhibition.** *Neuron,* 40**,** 785-96.

Gomeza, J., Ohno, K., Hulsmann, S., Armsen, W., Eulenburg, V., Richter, D. W., Laube, B. & Betz, H. 2003b. **Deletion of the mouse glycine transporter 2 results in a hyperekplexia phenotype and postnatal lethality.** *Neuron,* 40**,** 797-806.

Gomeza, J., Zafra, F., Olivares, L., Gimenez, C. & Aragon, C. 1995. **Regulation by phorbol esters of the glycine transporter (GLYT1) in glioblastoma cells.** *Biochim Biophys Acta,* 1233**,** 41-6.

Graber, T., Kluge, H., Hirche, F., Broz, J. & Stangl, G. I. 2012. **Effects of dietary benzoic acid and sodium-benzoate on performance, nitrogen and mineral balance and hippuric acid excretion of piglets.** *Arch Anim Nutr,* 66**,** 227-36.

Gregus, Z., Fekete, T., Halaszi, E. & Klaassen, C. D. 1996. **Lipoic acid impairs glycine conjugation of benzoic acid and renal excretion of benzoylglycine.** *Drug Metab Dispos,* 24**,** 682-8.

Gregus, Z., Fekete, T., Varga, F. & Klaassen, C. D. 1992. **Availability of glycine and coenzyme A limits glycine conjugation in vivo.** *Drug Metab Dispos,* 20**,** 234-40.

Gregus, Z., Fekete, T., Varga, F. & Klaassen, C. D. 1993. **Dependence of glycine conjugation on availability of glycine: role of the glycine cleavage system.** *Xenobiotica,* 23**,** 141-53.

Guay, F., Matte, J. J., Girard, C. L., Palin, M. F., Giguere, A. & Laforest, J. P. 2002. **Effect of folic acid and glycine supplementation on embryo development and folate metabolism during early pregnancy in pigs.** *J Anim Sci,* 80**,** 2134-43.

Guenin, S., Mauriat, M., Pelloux, J., Van Wuytswinkel, O., Bellini, C. & Gutierrez, L. 2009. **Normalization of qRT-PCR data: the necessity of adopting a systematic, experimental conditions-specific, validation of references.** *J Exp Bot,* 60**,** 487-93.

Gustafson, T. A., Taylor, A. & Kedes, L. 1989. **DNA Bending Is Induced by a Transcription Factor That Interacts with the Human C-Fos and Alpha-Actin Promoters.** *Proceedings of the National Academy of Sciences of the United States of America,* 86**,** 2162-2166.

Guturu, H., Doxey, A. C., Wenger, A. M. & Bejerano, G. 2013. **Structure-aided prediction of mammalian transcription factor complexes in conserved non-coding elements.** *Philos Trans R Soc Lond B Biol Sci,* 368**,** 20130029.

Guzman, J. N., Sanchez-Padilla, J., Wokosin, D., Kondapalli, J., Ilijic, E., Schumacker, P. T. & Surmeier, D. J. 2010. **Oxidant stress evoked by pacemaking in dopaminergic neurons is attenuated by DJ-1.** *Nature,* 468**,** 696-U119.

Hai, T. & Hartman, M. G. 2001. **The molecular biology and nomenclature of the activating transcription factor/cAMP responsive element binding family of transcription factors: activating transcription factor proteins and homeostasis.** *Gene,* 273**,** 1-11.

Hai, T., Wolfgang, C. D., Marsee, D. K., Allen, A. E. & Sivaprasad, U. 1999. **ATF3 and stress responses.** *Gene Expr,* 7**,** 321-35.

Hanley, J. G., Jones, E. M. & Moss, S. J. 2000. **GABA receptor rho1 subunit interacts with a novel splice variant of the glycine transporter, GLYT-1.** *J Biol Chem,* 275**,** 840-6.

Harding, H. P., Novoa, I., Zhang, Y., Zeng, H., Wek, R., Schapira, M. & Ron, D. 2000. **Regulated translation initiation controls stress-induced gene expression in mammalian cells.** *Mol Cell,* 6**,** 1099-108.

Harding, H. P., Zhang, Y., Zeng, H., Novoa, I., Lu, P. D., Calfon, M., Sadri, N., Yun, C., Popko, B., Paules, R., Stojdl, D. F., Bell, J. C., Hettmann, T., Leiden, J. M. & Ron, D. 2003. **An integrated stress response regulates amino acid metabolism and resistance to oxidative stress.** *Mol Cell,* 11**,** 619-33.

Hatanaka, T., Huang, W., Ling, R., Prasad, P. D., Sugawara, M., Leibach, F. H. & Ganapathy, V. 2001. **Evidence for the transport of neutral as well as cationic amino acids by ATA3, a novel and liver-specific subtype of**

amino acid transport system A. *Biochimica Et Biophysica Acta-Biomembranes,* 1510**,** 10-17.

Hatanaka, T., Huang, W., Wang, H., Sugawara, M., Prasad, P. D., Leibach, F. H. & Ganapathy, V. 2000. **Primary structure, functional characteristics and tissue expression pattern of human ATA2, a subtype of amino acid transport system A.** *Biochim Biophys Acta,* 1467**,** 1-6.

Hattori, T., Ohoka, N., Inoue, Y., Hayashi, H. & Onozaki, K. 2003. **C/EBP family transcription factors are degraded by the proteasome but stabilized by forming dimer.** *Oncogene,* 22**,** 1273-80.

Haverty, P. M., Hansen, U. & Weng, Z. 2004. **Computational inference of transcriptional regulatory networks from expression profiling and transcription factor binding site identification.** *Nucleic Acids Res,* 32**,** 179-88.

He, C. H., Gong, P., Hu, B., Stewart, D., Choi, M. E., Choi, A. M. & Alam, J. 2001. **Identification of activating transcription factor 4 (ATF4) as an Nrf2-interacting protein. Implication for heme oxygenase-1 gene regulation.** *J Biol Chem,* 276**,** 20858-65.

He, X., Mishchuk, D. O., Shah, J., Weimer, B. C. & Slupsky, C. M. 2013. **Cross-talk between E. coli strains and a human colorectal adenocarcinoma-derived cell line.** *Sci Rep,* 3**,** 3416.

Heim, K. E., Tagliaferro, A. R. & Bobilya, D. J. 2002. **Flavonoid antioxidants: chemistry, metabolism and structure-activity relationships.** *J Nutr Biochem,* 13**,** 572-584.

Herdon, H. J., Godfrey, F. M., Brown, A. M., Coulton, S., Evans, J. R. & Cairns, W. J. 2001. **Pharmacological assessment of the role of the glycine transporter GlyT-1 in mediating high-affinity glycine uptake by rat cerebral cortex and cerebellum synaptosomes.** *Neuropharmacology,* 41**,** 88-96.

Hestand, M. S., Van Galen, M., Villerius, M. P., Van Ommen, G. J., Den Dunnen, J. T. & T Hoen, P. A. 2008. **CORE_TF: a user-friendly interface to identify evolutionary conserved transcription factor binding sites in sets of co-regulated genes.** *BMC Bioinformatics,* 9**,** 495.

Hilgers, A. R., Conradi, R. A. & Burton, P. S. 1990. **Caco-2 Cell Monolayers as a Model for Drug Transport across the Intestinal-Mucosa.** *Pharmaceutical Research,* 7**,** 902-910.

Hillman, R. E., Albrecht, I. & Rosenberg, L. E. 1968. **Identification and analysis of multiple glycine transport systems in isolated mammalian renal tubules.** *J Biol Chem,* 243**,** 5566-71.

Hinnebusch, A. G. 1984. **Evidence for translational regulation of the activator of general amino acid control in yeast.** *Proc Natl Acad Sci U S A,* 81**,** 6442-6.

Horikoshi, M., Bertuccioli, C., Takada, R., Wang, J., Yamamoto, T. & Roeder, R. G. 1992. **Transcription Factor Tfiid Induces DNA Bending Upon Binding to the Tata Element.** *P Natl Acad Sci USA,* 89**,** 1060-1064.

Horton, N. & Quick, M. W. 2001. **Syntaxin 1A up-regulates GABA transporter expression by subcellular redistribution.** *Mol Membr Biol,* 18**,** 39-44.

Hosoya, K. I., Tomi, M., Ohtsuki, S., Takanaga, H., Saeki, S., Kanai, Y., Endou, H., Naito, M., Tsuruo, T. & Terasaki, T. 2002. **Enhancement of L-cystine transport activity and its relation to xCT gene induction at the blood-brain barrier by diethyl maleate treatment.** *Journal of Pharmacology and Experimental Therapeutics,* 302**,** 225-231.

Howard, A. & Hirst, B. H. 2011. **Regulation of GLYT1 and xCT in human intestinal cells by ATF4 but not XBP1 or ATF6.** *Faseb Journal,* 25.

Howard, A., Tahir, I., Javed, S., Waring, S. M., Ford, D. & Hirst, B. H. 2010. **Glycine transporter GLYT1 is essential for glycine-mediated protection of human intestinal epithelial cells against oxidative damage.** *J Physiol,* 588**,** 995-1009.

Hsin, Y. H., Tang, C. H., Lai, H. T. & Lee, T. H. 2011. **The role of TonEBP in regulation of AAD expression and dopamine production in renal proximal tubule cells upon hypertonic challenge.** *Biochem Biophys Res Commun,* 414**,** 598-603.

Huang, C. C., Chiribau, C. B., Majumder, M., Chiang, C. M., Wek, R. C., Kelm, R. J., Jr., Khalili, K., Snider, M. D. & Hatzoglou, M. 2009. **A bifunctional intronic element regulates the expression of the arginine/lysine transporter Cat-1 via mechanisms involving the purine-rich element binding protein A (Pur alpha).** *J Biol Chem,* 284**,** 32312-20.

Hummler, E., Cole, T. J., Blendy, J. A., Ganss, R., Aguzzi, A., Schmid, W., Beermann, F. & Schutz, G. 1994. **Targeted Mutation of the Creb Gene - Compensation within the Creb/Atf Family of Transcription Factors.** *Proceedings of the National Academy of Sciences of the United States of America,* 91**,** 5647-5651.

Ikejima, K., Iimuro, Y., Forman, D. T. & Thurman, R. G. 1996. **A diet containing glycine improves survival in endotoxin shock in the rat.** *Am J Physiol,* 271**,** G97-103.

Ikejima, K., Qu, W., Stachlewitz, R. F. & Thurman, R. G. 1997. **Kupffer cells contain a glycine-gated chloride channel.** *Am J Physiol,* 272**,** G1581-6.

Imhoff, B. R. & Hansen, J. M. 2010. **Tert-butylhydroquinone induces mitochondrial oxidative stress causing Nrf2 activation.** *Cell Biol Toxicol,* 26**,** 541-51.

Inigo, C., Barber, A. & Lostao, M. P. 2006. **Na+ and pH dependence of proline and beta-alanine absorption in rat small intestine.** *Acta Physiol (Oxf),* 186**,** 271-8.

Inoue, R., Jian, Z. & Kawarabayashi, Y. 2009. **Mechanosensitive TRP channels in cardiovascular pathophysiology.** *Pharmacol Ther,* 123**,** 371-85.

Ito, T., Fujio, Y., Matsuda, T., Takahashi, K. & Azuma, J. 2005. **Transcription factor NFAT5 is necessary for regulation of taurine transporter and cell survival in cardiomyocytes.** *Journal of Cardiac Failure,* 11**,** S279-S279.

Ito, T., Fujio, Y., Schaffer, S. W. & Azuma, J. 2009. **Involvement of transcriptional factor TonEBP in the regulation of the taurine transporter in the cardiomyocyte.** *Adv Exp Med Biol,* 643**,** 523-32.

Iyer, S. S., Ramirez, A. M., Ritzenthaler, J. D., Torres-Gonzalez, E., Roser-Page, S., Mora, A. L., Brigham, K. L., Jones, D. P., Roman, J. & Rojas, M. 2009. **Oxidation of extracellular cysteine/cystine redox state in bleomycin-induced lung fibrosis.** *Am J Physiol Lung Cell Mol Physiol,* 296**,** L37-45.

Jackson, A. A. 1986. **Blood glutathione in severe malnutrition in childhood.** *Trans R Soc Trop Med Hyg,* 80**,** 911-3.

Jackson, A. A., Gibson, N. R., Lu, Y. & Jahoor, F. 2004. **Synthesis of erythrocyte glutathione in healthy adults consuming the safe amount of dietary protein.** *Am J Clin Nutr,* 80**,** 101-7.

Jackson, A. A., Persaud, C., Hall, M., Smith, S., Evans, N. & Rutter, N. 1997. **Urinary excretion of 5-L-oxoproline (pyroglutamic acid) during early life in term and preterm infants.** *Arch Dis Child Fetal Neonatal Ed,* 76**,** F152-7.

Jaenisch, R. & Bird, A. 2003. **Epigenetic regulation of gene expression: how the genome integrates intrinsic and environmental signals.** *Nature Genetics,* 33**,** 245-254.

Jain, M., Nilsson, R., Sharma, S., Madhusudhan, N., Kitami, T., Souza, A. L., Kafri, R., Kirschner, M. W., Clish, C. B. & Mootha, V. K. 2012. **Metabolite profiling identifies a key role for glycine in rapid cancer cell proliferation.** *Science,* 336**,** 1040-4.

Javahery, R., Khachi, A., Lo, K., Zenzie-Gregory, B. & Smale, S. T. 1994. **DNA sequence requirements for transcriptional initiator activity in mammalian cells.** *Mol Cell Biol,* 14**,** 116-27.

Jewell, J. L., Russell, R. C. & Guan, K. L. 2013. **Amino acid signalling upstream of mTOR.** *Nat Rev Mol Cell Biol,* 14**,** 133-9.

Jeziorska, D. M., Jordan, K. W. & Vance, K. W. 2009. **A systems biology approach to understanding cis-regulatory module function.** *Semin Cell Dev Biol,* 20**,** 856-62.

Jiang, H. Y., Wek, S. A., Mcgrath, B. C., Scheuner, D., Kaufman, R. J., Cavener, D. R. & Wek, R. C. 2003. **Phosphorylation of the a subunit of eukaryotic initiation factor 2 is required for activation of NF-kappa B in response to diverse cellular stresses.** *Molecular and Cellular Biology,* 23**,** 5651-5663.

Johnson, D. S., Mortazavi, A., Myers, R. M. & Wold, B. 2007. **Genome-wide mapping of in vivo protein-DNA interactions.** *Science,* 316**,** 1497-502.

Jones, R. H. & Jones, N. C. 1989. **Mammalian cAMP-responsive element can activate transcription in yeast and binds a yeast factor(s) that resembles the mammalian transcription factor ANF.** *Proceedings of the National Academy of Sciences of the United States of America,* 86**,** 2176-2180.

Ju, P., Aubrey, K. R. & Vandenberg, R. J. 2004. **Zn2+ inhibits glycine transport by glycine transporter subtype 1b.** *J Biol Chem,* 279**,** 22983-91.

Juknat, A., Pietr, M., Kozela, E., Rimmerman, N., Levy, R., Coppola, G., Geschwind, D. & Vogel, Z. 2012. **Differential transcriptional profiles mediated by exposure to the cannabinoids cannabidiol and Delta9-tetrahydrocannabinol in BV-2 microglial cells.** *Br J Pharmacol,* 165**,** 2512-28.

Jursky, F. & Nelson, N. 2002. **Developmental Expression of the Glycine Transporters GLYT1 and GLYT2 in Mouse Brain.** *Journal of neurochemistry,* 67**,** 336-344.

Juven-Gershon, T. & Kadonaga, J. T. 2010. **Regulation of gene expression via the core promoter and the basal transcriptional machinery.** *Dev Biol,* 339**,** 225-9.

Kadonaga, J. T. 2002. **The DPE, a core promoter element for transcription by RNA polymerase II.** *Exp Mol Med,* 34**,** 259-64.

Kaminsky, L. S. & Zhang, Q. Y. 2003. **The small intestine as a xenobiotic-metabolizing organ.** *Drug Metab Dispos,* 31**,** 1520-5.

Kang, Y. J., Lu, M. K. & Guan, K. L. 2011. **The TSC1 and TSC2 tumor suppressors are required for proper ER stress response and protect cells from ER stress-induced apoptosis.** *Cell Death Differ,* 18**,** 133-44.

Karpinski, B. A., Morle, G. D., Huggenvik, J., Uhler, M. D. & Leiden, J. M. 1992. **Molecular-Cloning of Human Creb-2 - an Atf/Creb Transcription Factor That Can Negatively Regulate Transcription from the Camp Response Element.** *Proceedings of the National Academy of Sciences of the United States of America,* 89**,** 4820-4824.

Kato, H., Nakajima, S., Saito, Y., Takahashi, S., Katoh, R. & Kitamura, M. 2012. **mTORC1 serves ER stress-triggered apoptosis via selective activation of the IRE1-JNK pathway.** *Cell Death Differ,* 19**,** 310-20.

Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., Buxton, S., Cooper, A., Markowitz, S., Duran, C., Thierer, T., Ashton, B., Meintjes, P. & Drummond, A. 2012. **Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data.** *Bioinformatics,* 28**,** 1647-9.

Kennedy, D. J., Gatfield, K. M., Winpenny, J. P., Ganapathy, V. & Thwaites, D. T. 2005. **Substrate specificity and functional characterisation of the H+/amino acid transporter rat PAT2 (Slc36a2).** *Br J Pharmacol,* 144**,** 28-41.

Kilberg, M. S., Balasubramanian, M., Fu, L. & Shan, J. 2012. **The transcription factor network associated with the amino acid response in mammalian cells.** *Adv Nutr,* 3**,** 295-306.

Kilberg, M. S., Shan, J. & Su, N. 2009. **ATF4-dependent transcription mediates signaling of amino acid limitation.** *Trends Endocrinol Metab,* 20**,** 436-43.

Kim, J., Zwieb, C., Wu, C. & Adhya, S. 1989. **Bending of DNA by Gene-Regulatory Proteins - Construction and Use of a DNA Bending Vector.** *Gene,* 85**,** 15-23.

Kim, K. C., Kang, K. A., Zhang, R., Piao, M. J., Kim, G. Y., Kang, M. Y., Lee, S. J., Lee, N. H., Surh, Y. J. & Hyun, J. W. 2010. **Up-regulation of Nrf2-mediated heme oxygenase-1 expression by eckol, a phlorotannin compound, through activation of Erk and PI3K/Akt.** *Int J Biochem Cell Biol,* 42**,** 297-305.

Kim, K. M., Kingsmore, S. F., Han, H., Yang-Feng, T. L., Godinot, N., Seldin, M. F., Caron, M. G. & Giros, B. 1994. **Cloning of the human glycine transporter type 1: molecular and pharmacological characterization of novel isoform variants and chromosomal localization of the gene in the human and mouse genomes.** *Mol Pharmacol,* 45**,** 608-17.

Knights, K. M., Sykes, M. J. & Miners, J. O. 2007. **Amino acid conjugation: contribution to the metabolism and toxicity of xenobiotic carboxylic acids.** *Expert Opin Drug Metab Toxicol,* 3**,** 159-68.

Kouzarides, T. 2007. **Chromatin modifications and their function.** *Cell,* 128**,** 693-705.

Kristensen, A. S., Andersen, J., Jorgensen, T. N., Sorensen, L., Eriksen, J., Loland, C. J., Stromgaard, K. & Gether, U. 2011. **SLC6 neurotransmitter transporters: structure, function, and regulation.** *Pharmacol Rev,* 63**,** 585-640.

Krotova, K. Y., Zharikov, S. I. & Block, E. R. 2003. **Classical isoforms of PKC as regulators of CAT-1 transporter activity in pulmonary artery endothelial cells.** *Am J Physiol Lung Cell Mol Physiol,* 284**,** L1037-44.

Kucharzik, T., Lugering, N., Rautenberg, K., Lugering, A., Schmidt, M. A., Stoll, R. & Domschke, W. 2000. **Role of M cells in intestinal barrier function.** *Ann N Y Acad Sci,* 915**,** 171-83.

Kuntz, S. G., Williams, B. A., Sternberg, P. W. & Wold, B. J. 2012. **Transcription factor redundancy and tissue-specific regulation: evidence from functional and physical network connectivity.** *Genome Res,* 22**,** 1907-19.

Kurek-Gorecka, A., Rzepecka-Stojko, A., Gorecki, M., Stojko, J., Sosada, M. & Swierczek-Zieba, G. 2013. **Structure and antioxidant activity of polyphenols derived from propolis.** *Molecules,* 19**,** 78-101.

Kurokawa, H., Motohashi, H., Sueno, S., Kimura, M., Takagawa, H., Kanno, Y., Yamamoto, M. & Tanaka, T. 2009. **Structural basis of alternative DNA recognition by Maf transcription factors.** *Mol Cell Biol,* 29**,** 6232-44.

Kwak, M. K., Itoh, K., Yamamoto, M. & Kensler, T. W. 2002. **Enhanced expression of the transcription factor Nrf2 by cancer chemopreventive agents: role of antioxidant response element-like sequences in the nrf2 promoter.** *Mol Cell Biol,* 22**,** 2883-92.

Kwak, M. K., Wakabayashi, N., Itoh, K., Motohashi, H., Yamamoto, M. & Kensler, T. W. 2003. **Modulation of gene expression by cancer chemopreventive dithiolethiones through the Keap1-Nrf2 pathway. Identification of novel gene clusters for cell survival.** *J Biol Chem,* 278**,** 8135-45.

Lange, P. S., Chavez, J. C., Pinto, J. T., Coppola, G., Sun, C. W., Townes, T. M., Geschwind, D. H. & Ratan, R. R. 2008. **ATF4 is an oxidative stress-inducible, prodeath transcription factor in neurons in vitro and in vivo.** *J Exp Med,* 205**,** 1227-42.

Larionov, A., Krause, A. & Miller, W. 2005. **A standard curve based method for relative real time PCR data processing.** *BMC Bioinformatics,* 6**,** 62.

Lecca, M. R., Wagner, U., Patrignani, A., Berger, E. G. & Hennet, T. 2005. **Genome-wide analysis of the unfolded protein response in fibroblasts from congenital disorders of glycosylation type-I patients.** *FASEB J,* 19**,** 240-2.

Lee, J. M., Li, J., Johnson, D. A., Stein, T. D., Kraft, A. D., Calkins, M. J., Jakel, R. J. & Johnson, J. A. 2005. **Nrf2, a multi-organ protector?** *FASEB J,* 19**,** 1061-6.

Lee, K., Tirasophon, W., Shen, X., Michalak, M., Prywes, R., Okada, T., Yoshida, H., Mori, K. & Kaufman, R. J. 2002. **IRE1-mediated unconventional mRNA splicing and S2P-mediated ATF6 cleavage merge to regulate XBP1 in signaling the unfolded protein response.** *Genes Dev,* 16**,** 452-66.

Lee, O.-H., Yoon, K.-Y., Kim, K.-J., You, S. & Lee, B.-Y. 2011. **SEAWEED EXTRACTS AS A POTENTIAL TOOL FOR THE ATTENUATION OF OXIDATIVE DAMAGE IN OBESITY-RELATED PATHOLOGIES1.** *Journal of Phycology,* 47**,** 548-556.

Leung, K. S., Wong, K. C., Chan, T. M., Wong, M. H., Lee, K. H., Lau, C. K. & Tsui, S. K. 2010. **Discovering protein-DNA binding sequence patterns using association rule mining.** *Nucleic Acids Res,* 38**,** 6324-37.

Lewerenz, J., Albrecht, P., Tien, M. L., Henke, N., Karumbayaram, S., Kornblum, H. I., Wiedau-Pazos, M., Schubert, D., Maher, P. & Methner, A. 2009. **Induction of Nrf2 and xCT are involved in the action of the neuroprotective antibiotic ceftriaxone in vitro.** *J Neurochem,* 111**,** 332-43.

Lewerenz, J. & Maher, P. 2009. **Basal levels of eIF2alpha phosphorylation determine cellular antioxidant status by regulating ATF4 and xCT expression.** *J Biol Chem,* 284**,** 1106-15.

Lewerenz, J., Sato, H., Albrecht, P., Henke, N., Noack, R., Methner, A. & Maher, P. 2012. **Mutation of ATF4 mediates resistance of neuronal cell lines against oxidative stress by inducing xCT expression.** *Cell Death Differ,* 19**,** 847-58.

Lewis, R. M., Godfrey, K. M., Jackson, A. A., Cameron, I. T. & Hanson, M. A. 2005. **Low serine hydroxymethyltransferase activity in the human placenta has important implications for fetal glycine supply.** *J Clin Endocrinol Metab,* 90**,** 1594-8.

Li, H., Chen, D. & Zhang, J. 2012. **Analysis of intron sequence features associated with transcriptional regulation in human genes.** *PLoS One,* 7**,** e46784.

Li, J., Johnson, D., Calkins, M., Wright, L., Svendsen, C. & Johnson, J. 2005. **Stabilization of Nrf2 by tBHQ confers protection against oxidative stress-induced cell death in human neural stem cells.** *Toxicol Sci,* 83**,** 313-28.

Li, J. & Johnson, J. A. 2002. **Time-dependent changes in ARE-driven gene expression by use of a noise-filtering process for microarray data.** *Physiol Genomics,* 9**,** 137-44.

Li, W., Thakor, N., Xu, E. Y., Huang, Y., Chen, C., Yu, R., Holcik, M. & Kong, A. N. 2010. **An internal ribosomal entry site mediates redox-sensitive translation of Nrf2.** *Nucleic Acids Res,* 38**,** 778-88.

Li, W., Yu, S., Liu, T., Kim, J. H., Blank, V., Li, H. & Kong, A. N. 2008. **Heterodimerization with small Maf proteins enhances nuclear retention of Nrf2 via masking the NESzip motif.** *Biochim Biophys Acta,* 1783**,** 1847-56.

Linares, A. F., Loikkanen, J., Jorge, M. F., Soria, R. B. & Novoa, A. V. 2004. **Antioxidant and neuroprotective activity of the extract from the seaweed, Halimeda incrassata (Ellis) Lamouroux, against in vitro and in vivo toxicity induced by methyl-mercury.** *Vet Hum Toxicol,* 46**,** 1-5.

Lindemann, G., Grohs, M., Stange, E. F. & Fellermann, K. 2001. **Limited heat-shock protein 72 induction in Caco-2 cells by L-glutamine.** *Digestion,* 64**,** 81-6.

Liou, H. C., Boothby, M. R., Finn, P. W., Davidon, R., Nabavi, N., Zeleznikle, N. J., Ting, J. P. Y. & Glimcher, L. H. 1990. **A New Member of the Leucine Zipper Class of Proteins That Binds to the Hla Dr-Alpha Promoter.** *Science,* 247**,** 1581-1584.

Liu, H. & Gu, L. 2012. **Phlorotannins from brown algae (Fucus vesiculosus) inhibited the formation of advanced glycation endproducts by scavenging reactive carbonyls.** *J Agric Food Chem,* 60**,** 1326-34.

Liu, Q. R., Lopez-Corcuera, B., Mandiyan, S., Nelson, H. & Nelson, N. 1993. **Cloning and expression of a spinal cord- and brain-specific glycine transporter with novel structural features.** *J Biol Chem,* 268**,** 22802-8.

Liu, Y., Adachi, M., Zhao, S., Hareyama, M., Koong, A. C., Luo, D., Rando, T. A., Imai, K. & Shinomura, Y. 2009. **Preventing oxidative stress: a new role for XBP1.** *Cell Death Differ,* 16**,** 847-57.

Livingstone, C., Patel, G. & Jones, N. 1995. **ATF-2 contains a phosphorylation-dependent transcriptional activation domain.** *EMBO J,* 14**,** 1785-97.

Lo, K. & Smale, S. T. 1996. **Generality of a functional initiator consensus sequence.** *Gene,* 182**,** 13-22.

Locasale, J. W. 2013. **Serine, glycine and one-carbon units: cancer metabolism in full circle.** *Nat Rev Cancer,* 13**,** 572-83.

Lock, J. T., Sinkins, W. G. & Schilling, W. P. 2011. **Effect of protein S-glutathionylation on Ca2+ homeostasis in cultured aortic endothelial cells.** *Am J Physiol Heart Circ Physiol,* 300**,** H493-506.

Lodish, H., Berk, A. & Zipursky, S. L. 2000. ***Collagen: The Fibrous Proteins of the Matrix****,* W. H. Freeman.

Lopez, A. B., Wang, C., Huang, C. C., Yaman, I., Li, Y., Chakravarty, K., Johnson, P. F., Chiang, C. M., Snider, M. D., Wek, R. C. & Hatzoglou, M. 2007. **A feedback transcriptional mechanism controls the level of the arginine/lysine transporter cat-1 during amino acid starvation.** *Biochem J,* 402**,** 163-73.

Lopez-Corcuera, B., Aragon, C. & Geerlings, A. 2001a. **Regulation of glycine transporters.** *Biochem Soc Trans,* 29**,** 742-5.

Lopez-Corcuera, B., Geerlings, A. & Aragon, C. 2001b. **Glycine neurotransmitter transporters: an update.** *Mol Membr Biol,* 18**,** 13-20.

Lopez-Fontanals, M., Rodriguez-Mulero, S., Casado, F. J., Derijard, B. & Pastor-Anglada, M. 2003. **The osmoregulatory and the amino acid-regulated responses of system A are mediated by different signal transduction pathways.** *J Gen Physiol,* 122**,** 5-16.

Lopez-Rodriguez, C., Antos, C. L., Shelton, J. M., Richardson, J. A., Lin, F., Novobrantseva, T. I., Bronson, R. T., Igarashi, P., Rao, A. & Olson, E. N. 2004. **Loss of NFAT5 results in renal atrophy and lack of tonicity-responsive gene expression.** *Proc Natl Acad Sci U S A,* 101**,** 2392-7.

López-Rodríguez, C., Aramburu, J., Rakeman, A. S. & Rao, A. 1999. **NFAT5, a constitutively nuclear NFAT protein that does not cooperate with Fos and Jun.** *Proceedings of the National Academy of Sciences,* 96**,** 7214-7219.

Lorenz, M., Hillisch, A., Payet, D., Buttinelli, M., Travers, A. & Diekmann, S. 1999. **DNA bending induced by high mobility group proteins studied by fluorescence resonance energy transfer.** *Biochemistry,* 38**,** 12150-8.

Luo, J. Q., Chen, D. W. & Yu, B. 2013. **Upregulation of amino acid transporter expression induced by L-leucine availability in L6 myotubes is associated with ATF4 signaling through mTORC1-dependent mechanism.** *Nutrition,* 29**,** 284-90.

Lyons, J., Rauh-Pfeiffer, A., Yu, Y. M., Lu, X. M., Zurakowski, D., Tompkins, R. G., Ajami, A. M., Young, V. R. & Castillo, L. 2000. **Blood glutathione synthesis rates in healthy adults receiving a sulfur amino acid-free diet.** *Proc Natl Acad Sci U S A,* 97**,** 5071-6.

Ma, T. Y. 1997. **Intestinal epithelial barrier dysfunction in Crohn's disease.** *Proc Soc Exp Biol Med,* 214**,** 318-27.

Mackenzie, B. & Erickson, J. D. 2004. **Sodium-coupled neutral amino acid (System N/A) transporters of the SLC38 gene family.** *Pflugers Arch,* 447**,** 784-95.

Maes, M., Verkerk, R., Vandoolaeghe, E., Lin, A. & Scharpe, S. 1998. **Serum levels of excitatory amino acids, serine, glycine, histidine, threonine, taurine, alanine and arginine in treatment-resistant depression: modulation by treatment with antidepressants and prediction of clinical responsivity.** *Acta Psychiatr Scand,* 97**,** 302-8.

Maher, P. 2006. **A comparison of the neurotrophic activities of the flavonoid fisetin and some of its derivatives.** *Free Radic Res,* 40**,** 1105-11.

Maher, P. & Hanneken, A. 2005. **Flavonoids protect retinal ganglion cells from oxidative stress-induced death.** *Invest Ophthalmol Vis Sci,* 46**,** 4796-803.

Malik, P. & Mukherjee, T. K. 2014. **Structure-Function Elucidation of Antioxidative and Prooxidative Activities of the Polyphenolic Compound Curcumin.** *Chinese Journal of Biology,* 2014**,** 1-8.

Mangino, J. E., Kotadia, B. & Mangino, M. J. 1996. **Characterization of hypothermic intestinal ischemia-reperfusion injury in dogs. Effects of glycine.** *Transplantation,* 62**,** 173-8.

Mariadason, J. M., Rickard, K. L., Barkla, D. H., Augenlicht, L. H. & Gibson, P. R. 2000. **Divergent phenotypic patterns and commitment to apoptosis of Caco-2 cells during spontaneous and butyrate-induced differentiation.** *J Cell Physiol,* 183**,** 347-54.

Martinac, B. 2004. **Mechanosensitive ion channels: molecules of mechanotransduction.** *J Cell Sci,* 117**,** 2449-60.

Martinez-Maza, R., Poyatos, I., Lopez-Corcuera, B., E, N. U., Gimenez, C., Zafra, F. & Aragon, C. 2001. **The role of N-glycosylation in transport to the plasma membrane and sorting of the neuronal glycine transporter GLYT2.** *J Biol Chem,* 276**,** 2168-73.

Mason, M. J., Garciarodriguez, C. & Grinstein, S. 1991. **Coupling between Intracellular Ca2+ Stores and the Ca2+ Permeability of the Plasma-Membrane - Comparison of the Effects of Thapsigargin, 2,5-Di-(Tert-Butyl)-1,4-Hydroquinone, and Cyclopiazonic Acid in Rat Thymic Lymphocytes.** *Journal of Biological Chemistry,* 266**,** 20856-20862.

Matanjun, P., Mohamed, S., Mustapha, N. M., Muhammad, K. & Ming, C. H. 2008. **Antioxidant activities and phenolics content of eight species of seaweeds from north Borneo.** *Journal of Applied Phycology,* 20**,** 367-373.

Matys, V., Fricke, E., Geffers, R., Gossling, E., Haubrock, M., Hehl, R., Hornischer, K., Karas, D., Kel, A. E., Kel-Margoulis, O. V., Kloos, D. U., Land, S., Lewicki-Potapov, B., Michael, H., Munch, R., Reuter, I., Rotert, S., Saxel, H., Scheer, M.,

Thiele, S. & Wingender, E. 2003. **TRANSFAC: transcriptional regulation, from patterns to profiles.** *Nucleic Acids Res,* 31**,** 374-8.

Maurel, M., Chevet, E., Tavernier, J. & Gerlo, S. 2014. **Getting RIDD of RNA: IRE1 in cell fate regulation.** *Trends Biochem Sci,* 39**,** 245-54.

Mccarty, M. F., Barroso-Aranda, J. & Contreras, F. 2009. **The hyperpolarizing impact of glycine on endothelial cells may be anti-atherogenic.** *Med Hypotheses,* 73**,** 263-4.

Mcguckin, M. A., Eri, R., Simms, L. A., Florin, T. H. & Radford-Smith, G. 2009. **Intestinal barrier dysfunction in inflammatory bowel diseases.** *Inflamm Bowel Dis,* 15**,** 100-13.

Mckallip, R. J., Jia, W., Schlomer, J., Warren, J. W., Nagarkatti, P. S. & Nagarkatti, M. 2006. **Cannabidiol-induced apoptosis in human leukemia cells: A novel role of cannabidiol in the regulation of p22phox and Nox4 expression.** *Mol Pharmacol,* 70**,** 897-908.

Mechta-Grigoriou, F., Gerald, D. & Yaniv, M. 2001. **The mammalian Jun proteins: redundancy and specificity.** *Oncogene,* 20**,** 2378-89.

Meireles-Filho, A. C. & Stark, A. 2009. **Comparative genomics of gene regulation-conservation and divergence of cis-regulatory information.** *Curr Opin Genet Dev,* 19**,** 565-70.

Meng, F. & Lowell, C. A. 1997. **Lipopolysaccharide (LPS)-induced macrophage activation and signal transduction in the absence of Src-family kinases Hck, Fgr, and Lyn.** *J Exp Med,* 185**,** 1661-70.

Metzner, L., Neubert, K. & Brandsch, M. 2006. **Substrate specificity of the amino acid transporter PAT1.** *Amino Acids,* 31**,** 111-7.

Mezler, M., Hornberger, W., Mueller, R., Schmidt, M., Amberg, W., Braje, W., Ochse, M., Schoemaker, H. & Behl, B. 2008. **Inhibitors of GlyT1 affect glycine transport via discrete binding sites.** *Mol Pharmacol,* 74**,** 1705-15.

Miao, W., Hu, L., Scrivens, P. J. & Batist, G. 2005. **Transcriptional regulation of NF-E2 p45-related factor (NRF2) expression by the aryl hydrocarbon receptor-xenobiotic response element signaling pathway: direct cross-talk between phase I and II drug-metabolizing enzymes.** *J Biol Chem,* 280**,** 20340-8.

Minami, H., Miyamoto, K., Fujii, Y., Nakabou, Y. & Hagihira, H. 1985. **Induction of intestinal ornithine decarboxylase by single amino acid feeding.** *J Biochem,* 98**,** 133-9.

Misra, J. R., Lam, G. & Thummel, C. S. 2013. **Constitutive activation of the Nrf2/Keap1 pathway in insecticide-resistant strains of Drosophila.** *Insect Biochem Mol Biol,* 43**,** 1116-24.

Mitchell, J. B., Russo, A., Biaglow, J. E. & Mcpherson, S. 1983. **Cellular glutathione depletion by diethyl maleate or buthionine sulfoximine: no effect of glutathione depletion on the oxygen enhancement ratio.** *Radiat Res, 96,* 422-8.

Mitchell, P. J. & Tjian, R. 1989. **Transcriptional regulation in mammalian cells by sequence-specific DNA binding proteins.** *Science, 245,* 371-8.

Miyakawa, H., Woo, S. K., Dahl, S. C., Handler, J. S. & Kwon, H. M. 1999. **Tonicity-responsive enhancer binding protein, a Rel-like protein that stimulates transcription in response to hypertonicity.** *Proceedings of the National Academy of Sciences of the United States of America, 96,* 2538-2542.

Miyamoto, N., Izumi, H., Miyamoto, R., Bin, H., Kondo, H., Tawara, A., Sasaguri, Y. & Kohno, K. 2011. **Transcriptional regulation of activating transcription factor 4 under oxidative stress in retinal pigment epithelial ARPE-19/HPV-16 cells.** *Invest Ophthalmol Vis Sci, 52,* 1226-34.

Moi, P., Chan, K., Asunis, I., Cao, A. & Kan, Y. W. 1994. **Isolation of Nf-E2-Related Factor-2 (Nrf2), a Nf-E2-Like Basic Leucine-Zipper Transcriptional Activator That Binds to the Tandem Nf-E2/Ap1 Repeat of the Beta-Globin Locus-Control Region.** *Proceedings of the National Academy of Sciences of the United States of America, 91,* 9926-9930.

Montminy, M. R., Sevarino, K. A., Wagner, J. A., Mandel, G. & Goodman, R. H. 1986. **Identification of a Cyclic-Amp-Responsive Element within the Rat Somatostatin Gene.** *Proceedings of the National Academy of Sciences of the United States of America, 83,* 6682-6686.

Moreno, J. A., Halliday, M., Molloy, C., Radford, H., Verity, N., Axten, J. M., Ortori, C. A., Willis, A. E., Fischer, P. M., Barrett, D. A. & Mallucci, G. R. 2013. **Oral treatment targeting the unfolded protein response prevents neurodegeneration and clinical disease in prion-infected mice.** *Sci Transl Med, 5,* 206ra138.

Morimoto, R. I. 1998. **Regulation of the heat shock transcriptional response: cross talk between a family of heat shock factors, molecular chaperones, and negative regulators.** *Genes & Development, 12,* 3788-3796.

Morimoto, R. I., Sarge, K. D. & Abravaya, K. 1992. **Transcriptional regulation of heat shock genes. A paradigm for inducible genomic responses.** *J Biol Chem, 267,* 21987-90.

Morioka, N., Abdin, J. M., Morita, K., Kitayama, T., Nakata, Y. & Dohi, T. 2008. **The regulation of glycine transporter GLYT1 is mainly mediated by protein kinase Calpha in C6 glioma cells.** *Neurochem Int, 53,* 248-54.

Motohashi, H., Katsuoka, F., Engel, J. D. & Yamamoto, M. 2004. **Small Maf proteins serve as transcriptional cofactors for keratinocyte differentiation in the Keap1-Nrf2 regulatory pathway.** *Proc Natl Acad Sci U S A, 101,* 6379-84.

Motohashi, H., O'connor, T., Katsuoka, F., Engel, J. D. & Yamamoto, M. 2002. **Integration and diversity of the regulatory network composed of Maf and CNC families of transcription factors.** *Gene,* 294**,** 1-12.

Mozzini, C., Fratta Pasini, A., Garbin, U., Stranieri, C., Pasini, A., Vallerio, P. & Cominacini, L. 2014. **Increased endoplasmic reticulum stress and Nrf2 repression in peripheral blood mononuclear cells of patients with stable coronary artery disease.** *Free Radic Biol Med,* 68**,** 178-85.

Nair, S., Xu, C., Shen, G., Hebbar, V., Gopalakrishnan, A., Hu, R., Jain, M. R., Liew, C., Chan, J. Y. & Kong, A. N. 2007. **Toxicogenomics of endoplasmic reticulum stress inducer tunicamycin in the small intestine and liver of Nrf2 knockout and C57BL/6J mice.** *Toxicol Lett,* 168**,** 21-39.

Nakajima, S., Kato, H., Takahashi, S., Johno, H. & Kitamura, M. 2011. **Inhibition of NF-kappaB by MG132 through ER stress-mediated induction of LAP and LIP.** *FEBS Lett,* 585**,** 2249-54.

Nakanishi, T., Sugawara, M., Huang, W., Martindale, R. G., Leibach, F. H., Ganapathy, M. E., Prasad, P. D. & Ganapathy, V. 2001. **Structure, function, and tissue expression pattern of human SN2, a subtype of the amino acid transport system N.** *Biochem Biophys Res Commun,* 281**,** 1343-8.

Narkewicz, M. R., Jones, G. & Morales, D. 2000. **Serine and glycine transport in fetal ovine hepatocytes.** *Biochim Biophys Acta,* 1474**,** 41-6.

Navon, A., Gatushkin, A., Zelcbuch, L., Shteingart, S., Farago, M., Hadar, R. & Tirosh, B. 2010. **Direct proteasome binding and subsequent degradation of unspliced XBP-1 prevent its intracellular aggregation.** *FEBS Lett,* 584**,** 67-73.

Nelson, E. J., Li, C. C., Bangalore, R., Benson, T., Kass, R. S. & Hinkle, P. M. 1994. **Inhibition of L-type calcium-channel activity by thapsigargin and 2,5-t-butylhydroquinone, but not by cyclopiazonic acid.** *The Biochemical journal,* 302 (Pt 1)**,** 147-154.

Neph, S., Vierstra, J., Stergachis, A. B., Reynolds, A. P., Haugen, E., Vernot, B., Thurman, R. E., John, S., Sandstrom, R., Johnson, A. K., Maurano, M. T., Humbert, R., Rynes, E., Wang, H., Vong, S., Lee, K., Bates, D., Diegel, M., Roach, V., Dunn, D., Neri, J., Schafer, A., Hansen, R. S., Kutyavin, T., Giste, E., Weaver, M., Canfield, T., Sabo, P., Zhang, M., Balasundaram, G., Byron, R., Maccoss, M. J., Akey, J. M., Bender, M. A., Groudine, M., Kaul, R. & Stamatoyannopoulos, J. A. 2012. **An expansive human regulatory lexicon encoded in transcription factor footprints.** *Nature,* 489**,** 83-90.

Newburger, D. E. & Bulyk, M. L. 2009. **UniPROBE: an online database of protein binding microarray data on protein-DNA interactions.** *Nucleic Acids Res,* 37**,** D77-82.

Nguyen, T., Sherratt, P. J., Nioi, P., Yang, C. S. & Pickett, C. B. 2005. **Nrf2 controls constitutive and inducible expression of ARE-driven genes through a**

dynamic pathway involving nucleocytoplasmic shuttling by Keap1. *J Biol Chem,* 280**,** 32485-92.

Nijveldt, R. J., Van Nood, E., Van Hoorn, D. E., Boelens, P. G., Van Norren, K. & Van Leeuwen, P. A. 2001. **Flavonoids: a review of probable mechanisms of action and potential applications.** *Am J Clin Nutr,* 74**,** 418-25.

Nishimura, M. & Naito, S. 2005. **Tissue-specific mRNA Expression Profiles of Human ATP-binding Cassette and Solute Carrier Transporter Superfamiies.** *Drug Metabolism and Pharmacokinetics,* 20**,** 452-477.

Nissim, I., Hardy, M., Pleasure, J., Nissim, I. & States, B. 1992. **A mechanism of glycine and alanine cytoprotective action: stimulation of stress-induced HSP70 mRNA.** *Kidney Int,* 42**,** 775-82.

Nowak, T. S., Jr. 1990. **Protein synthesis and the heart shock/stress response after ischemia.** *Cerebrovasc Brain Metab Rev,* 2**,** 345-66.

Ochi, H., Tamai, T., Nagano, H., Kawaguchi, A., Sudou, N. & Ogino, H. 2012. **Evolution of a tissue-specific silencer underlies divergence in the expression of pax2 and pax8 paralogues.** *Nat Commun,* 3**,** 848.

Ogmundsdottir, M. H., Heublein, S., Kazi, S., Reynolds, B., Visvalingam, S. M., Shaw, M. K. & Goberdhan, D. C. 2012. **Proton-assisted amino acid transporter PAT1 complexes with Rag GTPases and activates TORC1 on late endosomal and lysosomal membranes.** *PLoS One,* 7**,** e36616.

Ohoka, N., Yoshii, S., Hattori, T., Onozaki, K. & Hayashi, H. 2005. **TRB3, a novel ER stress-inducible gene, is induced via ATF4-CHOP pathway and is involved in cell death.** *EMBO J,* 24**,** 1243-55.

Okamoto, Y., Chaves, A., Chen, J. C., Kelley, R., Jones, K., Weed, H. G., Gardner, K. L., Gangi, L., Yamaguchi, M., Klomkleaw, W., Nakayama, T., Hamlin, R. L., Carnes, C., Altschuld, R., Bauer, J. & Hai, T. 2001. **Transgenic mice with cardiac-specific expression of activating transcription factor 3, a stress-inducible gene, have conduction abnormalities and contractile dysfunction.** *American Journal of Pathology,* 159**,** 639-650.

Okinaga, S., Takahashi, K., Takeda, K., Yoshizawa, M., Fujita, H., Sasaki, H. & Shibahara, S. 1996. **Regulation of human heme oxygenase-1 gene expression under thermal stress.** *Blood,* 87**,** 5074-84.

Okubo, T., Yokoyama, Y., Kano, K. & Kano, I. 2003. **Cell death induced by the phenolic antioxidant tert-butylhydroquinone and its metabolite tert-butylquinone in human monocytic leukemia U937 cells.** *Food Chem Toxicol,* 41**,** 679-88.

Olivares, L., Aragon, C., Gimenez, C. & Zafra, F. 1994. **Carboxyl-Terminus of the Glycine Transporter Glyt1 Is Necessary for Correct Processing of the Protein.** *Journal of Biological Chemistry,* 269**,** 28400-28404.

Olivares, L., Aragon, C., Gimenez, C. & Zafra, F. 1995. **The role of N-glycosylation in the targeting and activity of the GLYT1 glycine transporter.** *J Biol Chem,* 270**,** 9437-42.

Omasa, M., Fukuse, T., Toyokuni, S., Mizutani, Y., Yoshida, H., Ikeyama, K., Hasegawa, S. & Wada, H. 2003. **Glycine ameliorates lung reperfusion injury after cold preservation in an ex vivo rat lung model.** *Transplantation,* 75**,** 591-8.

Ozawa, K., Sudo, T., Soeda, E., Yoshida, M. C. & Ishii, S. 1991. **Assignment of the human CREB2 (CRE-BP1) gene to 2q32.** *Genomics,* 10**,** 1103-4.

Palii, S. S., Chen, H. & Kilberg, M. S. 2004. **Transcriptional control of the human sodium-coupled neutral amino acid transporter system A gene by amino acid availability is mediated by an intronic element.** *J Biol Chem,* 279**,** 3463-71.

Palii, S. S., Thiaville, M. M., Pan, Y. X., Zhong, C. & Kilberg, M. S. 2006. **Characterization of the amino acid response element within the human sodium-coupled neutral amino acid transporter 2 (SNAT2) System A transporter gene.** *Biochem J,* 395**,** 517-27.

Paller, M. S. & Patten, M. 1992. **Protective effects of glutathione, glycine, or alanine in an in vitro model of renal anoxia.** *J Am Soc Nephrol,* 2**,** 1338-44.

Pandey, K. B. & Rizvi, S. I. 2009. **Plant polyphenols as dietary antioxidants in human health and disease.** *Oxid Med Cell Longev,* 2**,** 270-8.

Pastore, A., Federici, G., Bertini, E. & Piemonte, F. 2003. **Analysis of glutathione: implication in redox and detoxification.** *Clin Chim Acta,* 333**,** 19-39.

Patel, D. K., Ogunbona, A., Notarianni, L. J. & Bennett, P. N. 1990. **Depletion of plasma glycine and effect of glycine by mouth on salicylate metabolism during aspirin overdose.** *Hum Exp Toxicol,* 9**,** 389-95.

Pelechano, V., Wei, W., Jakob, P. & Steinmetz, L. M. 2014. **Genome-wide identification of transcript start and end sites by transcript isoform sequencing.** *Nat Protoc,* 9**,** 1740-59.

Pelham, H. R. 1986. **Speculations on the functions of the major heat shock and glucose-regulated proteins.** *Cell,* 46**,** 959-61.

Peng, T. I. & Jou, M. J. 2010. **Oxidative stress caused by mitochondrial calcium overload.** *Ann N Y Acad Sci,* 1201**,** 183-8.

Perez-Siles, G., Morreale, A., Leo-Macias, A., Pita, G., Ortiz, A. R., Aragon, C. & Lopez-Corcuera, B. 2011. **Molecular basis of the differential interaction with lithium of glycine transporters GLYT1 and GLYT2.** *J Neurochem,* 118**,** 195-204.

Peyret, N., Seneviratne, P. A., Allawi, H. T. & Santalucia, J., Jr. 1999. **Nearest-neighbor thermodynamics and NMR of DNA sequences with internal A.A, C.C, G.G, and T.T mismatches.** *Biochemistry,* 38**,** 3468-77.

Podust, L. M., Krezel, A. M. & Kim, Y. 2000. **Crystal structure of C/EBPbeta-ATF4 bZIP heterodimer in the absence of DNARunning title: the basic region of the ATF4 bZIP is an alpha-Helix.** *The Journal of biological chemistry.*

Ponce, J., Biton, B., Benavides, J., Avenet, P. & Aragon, C. 2000. **Transmembrane domain III plays an important role in ion binding and permeation in the glycine transporter GLYT2.** *J Biol Chem,* 275**,** 13856-62.

Ponce, J., Poyatos, I., Aragon, C., Gimenez, C. & Zafra, F. 1998. **Characterization of the 5' region of the rat brain glycine transporter GLYT2 gene: identification of a novel isoform.** *Neurosci Lett,* 242**,** 25-8.

Portela, A. & Esteller, M. 2010. **Epigenetic modifications and human disease.** *Nat Biotechnol,* 28**,** 1057-68.

Poyatos, I., Ruberti, F., Martinez-Maza, R., Gimenez, C., Dotti, C. G. & Zafra, F. 2000. **Polarized distribution of glycine transporter isoforms in epithelial and neuronal cells.** *Mol Cell Neurosci,* 15**,** 99-111.

Pramod, A. B., Foster, J., Carvelli, L. & Henry, L. K. 2013. **SLC6 transporters: structure, function, regulation, disease association and therapeutics.** *Mol Aspects Med,* 34**,** 197-219.

Qi, R. B., Zhang, J. Y., Lu, D. X., Wang, H. D., Wang, H. H. & Li, C. J. 2007. **Glycine receptors contribute to cytoprotection of glycine in myocardial cells.** *Chin Med J (Engl),* 120**,** 915-21.

Quick, A. J. & Cooper, W. T. T. a. O. M. A. 1931. **THE CONJUGATION OF BENZOIC ACID IN MAN.** *Journal of Biological Chemistry.*

Raghavendran, H. R. B., Sathivel, A. & Devaki, T. 2004. **Hepatoprotective nature of seaweed alcoholic extract on acetaminophen induced hepatic oxidative stress.** *Journal of Health Science,* 50**,** 42-46.

Ramos-Gomez, M., Dolan, P. M., Itoh, K., Yamamoto, M. & Kensler, T. W. 2003. **Interactive effects of nrf2 genotype and oltipraz on benzo[a]pyrene-DNA adducts and tumor yield in mice.** *Carcinogenesis,* 24**,** 461-7.

Ramos-Gomez, M., Kwak, M. K., Dolan, P. M., Itoh, K., Yamamoto, M., Talalay, P. & Kensler, T. W. 2001. **Sensitivity to carcinogenesis is increased and chemoprotective efficacy of enzyme inducers is lost in nrf2 transcription factor-deficient mice.** *Proc Natl Acad Sci U S A,* 98**,** 3410-5.

Ray, P. D., Huang, B. W. & Tsuji, Y. 2012. **Reactive oxygen species (ROS) homeostasis and redox regulation in cellular signaling.** *Cell Signal,* 24**,** 981-90.

Reddy, J. K. & Sambasiva Rao, M. 2006. *Lipid Metabolism and Liver Inflammation. II. Fatty liver disease and fatty acid oxidation.*

Redondo, P. C., Jardin, I., Lopez, J. J., Salido, G. M. & Rosado, J. A. 2008. **Intracellular Ca2+ store depletion induces the formation of macromolecular complexes involving hTRPC1, hTRPC6, the type II IP3 receptor and SERCA3 in human platelets.** *Biochim Biophys Acta,* 1783**,** 1163-76.

Reeds, P. J., Burrin, D. G., Stoll, B. & Van Goudoever, J. B. 2000. **Role of the gut in the amino acid economy of the host.** *Nestle Nutr Workshop Ser Clin Perform Programme,* 3**,** 25-40; discussion 40-6.

Reilly, C. A., Johansen, M. E., Lanza, D. L., Lee, J., Lim, J. O. & Yost, G. S. 2005. **Calcium-dependent and independent mechanisms of capsaicin receptor (TRPV1)-mediated cytokine production and cell death in human bronchial epithelial cells.** *J Biochem Mol Toxicol,* 19**,** 266-75.

Reinke, A. W., Baek, J., Ashenberg, O. & Keating, A. E. 2013. **Networks of bZIP protein-protein interactions diversified over a billion years of evolution.** *Science,* 340**,** 730-4.

Riento, K., Galli, T., Jansson, S., Ehnholm, C., Lehtonen, E. & Olkkonen, V. M. 1998. **Interaction of Munc-18-2 with syntaxin 3 controls the association of apical SNAREs in epithelial cells.** *J Cell Sci,* 111 (Pt 17)**,** 2681-8.

Rousseau, F., Aubrey, K. R. & Supplisson, S. 2008. **The glycine transporter GlyT2 controls the dynamics of synaptic vesicle refilling in inhibitory spinal cord neurons.** *J Neurosci,* 28**,** 9755-68.

Rousset, M. 1986. **The human colon carcinoma cell lines HT-29 and Caco-2: two in vitro models for the study of intestinal differentiation.** *Biochimie,* 68**,** 1035-40.

Roux, M. J. & Supplisson, S. 2000. **Neuronal and Glial Glycine Transporters Have Different Stoichiometries.** *Neuron,* 25**,** 373-383.

Roy, B. & Lee, A. S. 1999. **The mammalian endoplasmic reticulum stress response element consists of an evolutionarily conserved tripartite structure and interacts with a novel stress-inducible complex.** *Nucleic Acids Res,* 27**,** 1437-43.

Roy, S., Wapinski, I., Pfiffner, J., French, C., Socha, A., Konieczka, J., Habib, N., Kellis, M., Thompson, D. & Regev, A. 2013. **Arboretum: reconstruction and analysis of the evolutionary history of condition-specific transcriptional modules.** *Genome Res,* 23**,** 1039-50.

Rubinstein, M. & De Souza, F. S. 2013. **Evolution of transcriptional enhancers and animal diversity.** *Philos Trans R Soc Lond B Biol Sci,* 368**,** 20130017.

Rutkowski, D. T. & Kaufman, R. J. 2003. **All roads lead to ATF4.** *Developmental Cell,* 4**,** 442-444.

Rychlik, W., Spencer, W. J. & Rhoads, R. E. 1990. **Optimization of the annealing temperature for DNA amplification in vitro.** *Nucleic Acids Res,* 18**,** 6409-12.

Rzymski, T., Milani, M., Pike, L., Buffa, F., Mellor, H. R., Winchester, L., Pires, I., Hammond, E., Ragoussis, I. & Harris, A. L. 2010. **Regulation of autophagy by ATF4 in response to severe hypoxia.** *Oncogene,* 29**,** 4424-35.

Sagne, C., Agulhon, C., Ravassard, P., Darmon, M., Hamon, M., El Mestikawy, S., Gasnier, B. & Giros, B. 2001. **Identification and characterization of a lysosomal transporter for small neutral amino acids.** *Proc Natl Acad Sci U S A,* 98**,** 7206-11.

Salazar, M., Carracedo, A., Salanueva, I. J., Hernandez-Tiedra, S., Lorente, M., Egia, A., Vazquez, P., Blazquez, C., Torres, S., Garcia, S., Nowak, J., Fimia, G. M., Piacentini, M., Cecconi, F., Pandolfi, P. P., Gonzalez-Feria, L., Iovanna, J. L., Guzman, M., Boya, P. & Velasco, G. 2009. **Cannabinoid action induces autophagy-mediated cell death through stimulation of ER stress in human glioma cells.** *Journal of Clinical Investigation,* 119**,** 1359-1372.

Samali, A., Fitzgerald, U., Deegan, S. & Gupta, S. 2010. **Methods for monitoring endoplasmic reticulum stress and the unfolded protein response.** *Int J Cell Biol,* 2010**,** 830307.

Sambuy, Y., De Angelis, I., Ranaldi, G., Scarino, M. L., Stammati, A. & Zucco, F. 2005. **The Caco-2 cell line as a model of the intestinal barrier: influence of cell and culture-related factors on Caco-2 cell functional characteristics.** *Cell Biol Toxicol,* 21**,** 1-26.

Sandelin, A., Alkema, W., Engstrom, P., Wasserman, W. W. & Lenhard, B. 2004. **JASPAR: an open-access database for eukaryotic transcription factor binding profiles.** *Nucleic Acids Res,* 32**,** D91-4.

Sano, R. & Reed, J. C. 2013. **ER stress-induced cell death mechanisms.** *Biochim Biophys Acta,* 1833**,** 3460-70.

Santalucia, J., Jr. 1998. **A unified view of polymer, dumbbell, and oligonucleotide DNA nearest-neighbor thermodynamics.** *Proc Natl Acad Sci U S A,* 95**,** 1460-5.

Sasaki, H., Sato, H., Kuriyama-Matsumura, K., Sato, K., Maebara, K., Wang, H., Tamba, M., Itoh, K., Yamamoto, M. & Bannai, S. 2002. **Electrophile response element-mediated induction of the cystine/glutamate exchange transporter gene expression.** *J Biol Chem,* 277**,** 44765-71.

Sasaki, N., Fukatsu, R., Tsuzuki, K., Hayashi, Y., Yoshida, T., Fujii, N., Koike, T., Wakayama, I., Yanagihara, R., Garruto, R., Amano, N. & Makita, Z. 1998. **Advanced glycation end products in Alzheimer's disease and other neurodegenerative diseases.** *Am J Pathol,* 153**,** 1149-55.

Sato, H., Nomura, S., Maebara, K., Sato, K., Tamba, M. & Bannai, S. 2004. **Transcriptional control of cystine/glutamate transporter gene by amino acid deprivation.** *Biochem Biophys Res Commun,* 325**,** 109-16.

Sato, H., Shiiya, A., Kimata, M., Maebara, K., Tamba, M., Sakakura, Y., Makino, N., Sugiyama, F., Yagami, K., Moriguchi, T., Takahashi, S. & Bannai, S. 2005. **Redox imbalance in cystine/glutamate transporter-deficient mice.** *J Biol Chem,* 280**,** 37423-9.

Sato, K., Adams, R., Betz, H. & Schloss, P. 2002. **Modulation of a Recombinant Glycine Transporter (GLYT1b) by Activation of Protein Kinase C.** *Journal of neurochemistry,* 65**,** 1967-1973.

Scamps, F., Vigues, S., Restituito, S., Campo, B., Roig, A., Charnet, P. & Valmier, J. 2000. **Sarco-endoplasmic ATPase blocker 2,5-di(tert-butyl)-1,4-benzohydroquinone inhibits N-, P-, and Q- but not T-, L-, or R-Type calcium currents in central and peripheral neurons.** *Molecular Pharmacology,* 58**,** 18-26.

Schachter, D. & Taggart, J. V. 1953. **Benzoyl coenzyme A and hippurate synthesis.** *J Biol Chem,* 203**,** 925-34.

Schilling, T. & Eder, C. 2004. **A novel physiological mechanism of glycine-induced immunomodulation: Na+-coupled amino acid transporter currents in cultured brain macrophages.** *J Physiol,* 559**,** 35-40.

Schioth, H. B., Roshanbin, S., Hagglund, M. G. & Fredriksson, R. 2013. **Evolutionary origin of amino acid transporter families SLC32, SLC36 and SLC38 and physiological, pathological and therapeutic aspects.** *Mol Aspects Med,* 34**,** 571-85.

Schroder, M. & Kaufman, R. J. 2005. **ER stress and the unfolded protein response.** *Mutat Res,* 569**,** 29-63.

Seong, K. H., Li, D., Shimizu, H., Nakamura, R. & Ishii, S. 2011. **Inheritance of stress-induced, ATF-2-dependent epigenetic change.** *Cell,* 145**,** 1049-61.

Seow, H. F., Broer, S., Broer, A., Bailey, C. G., Potter, S. J., Cavanaugh, J. A. & Rasko, J. E. J. 2004. **Hartnup disorder is caused by mutations in the gene encoding the neutral amino acid transporter SLC6A19.** *Nature Genetics,* 36**,** 1003-1007.

Shan, J., Fu, L., Balasubramanian, M. N., Anthony, T. & Kilberg, M. S. 2012. **ATF4-dependent regulation of the JMJD3 gene during amino acid deprivation can be rescued in Atf4-deficient cells by inhibition of deacetylation.** *J Biol Chem,* 287**,** 36393-403.

Shan, J., Lopez, M. C., Baker, H. V. & Kilberg, M. S. 2010. **Expression profiling after activation of amino acid deprivation response in HepG2 human hepatoma cells.** *Physiol Genomics,* 41**,** 315-27.

Shen, J., Chen, X., Hendershot, L. & Prywes, R. 2002. **ER stress regulation of ATF6 localization by dissociation of BiP/GRP78 binding and unmasking of Golgi localization signals.** *Dev Cell,* 3**,** 99-111.

Siepel, A., Bejerano, G., Pedersen, J. S., Hinrichs, A. S., Hou, M., Rosenbloom, K., Clawson, H., Spieth, J., Hillier, L. W., Richards, S., Weinstock, G. M., Wilson, R.

K., Gibbs, R. A., Kent, W. J., Miller, W. & Haussler, D. 2005. **Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes.** *Genome Res,* 15**,** 1034-50.

Silk, D. B. A., Grimble, G. K. & Rees, R. G. **Protein digestion and amino acid and peptide absorption**. Proceedings of the Nutrition Society, 1985/00/01 1985. 63-72.

Simmen, T., Lynes, E. M., Gesson, K. & Thomas, G. 2010. **Oxidative protein folding in the endoplasmic reticulum: tight links to the mitochondria-associated membrane (MAM).** *Biochim Biophys Acta,* 1798**,** 1465-73.

Simonis, M., Kooren, J. & De Laat, W. 2007. **An evaluation of 3C-based methods to capture DNA interactions.** *Nat Methods,* 4**,** 895-901.

Singer, D., Camargo, S. M., Huggel, K., Romeo, E., Danilczyk, U., Kuba, K., Chesnov, S., Caron, M. G., Penninger, J. M. & Verrey, F. 2009. **Orphan transporter SLC6A18 is renal neutral amino acid transporter B0AT3.** *J Biol Chem,* 284**,** 19953-60.

Skipper, M., Dhand, R. & Campbell, P. 2012. **Presenting ENCODE.** *Nature,* 489**,** 45.

Sloan, J. L. & Mager, S. 1999. **Cloning and functional expression of a human Na+ and Cl--dependent neutral and cationic amino acid transporter B0+.** *Journal of Biological Chemistry,* 274**,** 23740-23745.

Smale, S. T. 2001. **Core promoters: active contributors to combinatorial gene regulation.** *Genes Dev,* 15**,** 2503-8.

Smale, S. T. & Baltimore, D. 1989. **The "initiator" as a transcription control element.** *Cell,* 57**,** 103-113.

Soda, T., Frank, C., Ishizuka, K., Baccarella, A., Park, Y. U., Flood, Z., Park, S. K., Sawa, A. & Tsai, L. H. 2013. **DISC1-ATF4 transcriptional repression complex: dual regulation of the cAMP-PDE4 cascade by DISC1.** *Mol Psychiatry,* 18**,** 898-908.

Soderholm, J. D. & Perdue, M. H. 2001. **Stress and gastrointestinal tract. II. Stress and intestinal barrier function.** *Am J Physiol Gastrointest Liver Physiol,* 280**,** G7-G13.

Sommer, T. & Jarosch, E. 2002. **BiP binding keeps ATF6 at bay.** *Developmental Cell,* 3**,** 1-2.

Spellman, P. T., Sherlock, G., Zhang, M. Q., Iyer, V. R., Anders, K., Eisen, M. B., Brown, P. O., Botstein, D. & Futcher, B. 1998. **Comprehensive identification of cell cycle-regulated genes of the yeast Saccharomyces cerevisiae by microarray hybridization.** *Mol Biol Cell,* 9**,** 3273-97.

Sporn, M. B. & Liby, K. T. 2012. **NRF2 and cancer: the good, the bad and the importance of context.** *Nat Rev Cancer,* 12**,** 564-71.

Stark, C., Breitkreutz, B. J., Reguly, T., Boucher, L., Breitkreutz, A. & Tyers, M. 2006. **BioGRID: a general repository for interaction datasets.** *Nucleic Acids Res,* 34**,** D535-9.

Steeves, C. L. & Baltz, J. M. 2005. **Regulation of intracellular glycine as an organic osmolyte in early preimplantation mouse embryos.** *J Cell Physiol,* 204**,** 273-9.

Steeves, C. L., Hammer, M. A., Walker, G. B., Rae, D., Stewart, N. A. & Baltz, J. M. 2003. **The glycine neurotransmitter transporter GLYT1 is an organic osmolyte transporter regulating cell volume in cleavage-stage embryos.** *Proc Natl Acad Sci U S A,* 100**,** 13982-7.

Stoffels, B., Turler, A., Schmidt, J., Nazir, A., Tsukamoto, T., Moore, B. A., Schnurr, C., Kalff, J. C. & Bauer, A. J. 2011. **Anti-inflammatory role of glycine in reducing rodent postoperative inflammatory ileus.** *Neurogastroenterol Motil,* 23**,** 76-87, e8.

Supplisson, S. & Roux, M. J. 2002. **Why glycine transporters have different stoichiometries.** *Febs Letters,* 529**,** 93-101.

Suzuki, Y., Tsunoda, T., Sese, J., Taira, H., Mizushima-Sugano, J., Hata, H., Ota, T., Isogai, T., Tanaka, T., Nakamura, Y., Suyama, A., Sakaki, Y., Morishita, S., Okubo, K. & Sugano, S. 2001. **Identification and characterization of the potential promoter regions of 1031 kinds of human genes.** *Genome Res,* 11**,** 677-84.

Syed, Z., Indyk, P. & Guttag, J. 2009. **Learning Approximate Sequential Patterns for Classification.** *Journal of Machine Learning Research,* 10**,** 1913-1936.

Szutorisz, H., Dillon, N. & Tora, L. 2005. **The role of enhancers as centres for general transcription factor recruitment.** *Trends Biochem Sci,* 30**,** 593-9.

Takahasi, K. R., Matsuo, T. & Takano-Shimizu-Kouno, T. 2011. **Two types of cis-trans compensation in the evolution of transcriptional regulation.** *Proc Natl Acad Sci U S A,* 108**,** 15276-81.

Takayanagi, S., Fukuda, R., Takeuchi, Y., Tsukada, S. & Yoshida, K. 2013. **Gene regulatory network of unfolded protein response genes in endoplasmic reticulum stress.** *Cell Stress Chaperones,* 18**,** 11-23.

Takeuchi, M., Kikuchi, S., Sasaki, N., Suzuki, T., Watai, T., Iwaki, M., Bucala, R. & Yamagishi, S. 2004. **Involvement of advanced glycation end-products (AGEs) in Alzheimer's disease.** *Curr Alzheimer Res,* 1**,** 39-46.

Taylor, S., Wakem, M., Dijkman, G., Alsarraj, M. & Nguyen, M. 2010. **A practical approach to RT-qPCR-Publishing data that conform to the MIQE guidelines.** *Methods,* 50**,** S1-5.

Thiaville, M. M., Dudenhausen, E. E., Awad, K. S., Gjymishka, A., Zhong, C. & Kilberg, M. S. 2008a. **Activated transcription via mammalian amino acid response elements does not require enhanced recruitment of the Mediator complex.** *Nucleic Acids Res,* 36**,** 5571-80.

Thiaville, M. M., Dudenhausen, E. E., Zhong, C., Pan, Y. X. & Kilberg, M. S. 2008b. **Deprivation of protein or amino acid induces C/EBPbeta synthesis and binding to amino acid response elements, but its action is not an absolute requirement for enhanced transcription.** *Biochem J,* 410**,** 473-84.

Thiaville, M. M., Pan, Y. X., Gjymishka, A., Zhong, C., Kaufman, R. J. & Kilberg, M. S. 2008c. **MEK signaling is required for phosphorylation of eIF2alpha following amino acid limitation of HepG2 human hepatoma cells.** *J Biol Chem,* 283**,** 10848-57.

Tiganis, T. 2011. **Reactive oxygen species and insulin resistance: the good, the bad and the ugly.** *Trends Pharmacol Sci,* 32**,** 82-9.

Tomasi, M. L., Ryoo, M., Yang, H., Iglesias Ara, A., Ko, K. S. & Lu, S. C. 2014. **Molecular mechanisms of lipopolysaccharide-mediated inhibition of glutathione synthesis in mice.** *Free Radic Biol Med,* 68**,** 148-58.

Tompkins, W. A., Watrach, A. M., Schmale, J. D., Schultz, R. M. & Harris, J. A. 1974. **Cultural and antigenic properties of newly established cell strains derived from adenocarcinomas of the human colon and rectum.** *J Natl Cancer Inst,* 52**,** 1101-10.

Tonkiss, J. & Calderwood, S. K. 2005. **Regulation of heat shock gene transcription in neuronal cells.** *Int J Hyperthermia,* 21**,** 433-44.

Trama, J., Lu, Q. J., Hawley, R. G. & Ho, S. N. 2000. **The NFAT-related protein NFATL1 (TonEBP/NFAT5) is induced upon T cell activation in a calcineurin-dependent manner.** *Journal of Immunology,* 165**,** 4884-4894.

Tsai, G., Ralph-Williams, R. J., Martina, M., Bergeron, R., Berger-Sweeney, J., Dunham, K. S., Jiang, Z., Caine, S. B. & Coyle, J. T. 2004. **Gene knockout of glycine transporter 1: characterization of the behavioral phenotype.** *Proc Natl Acad Sci U S A,* 101**,** 8485-90.

Tsai, J. J., Dudakov, J. A., Takahashi, K., Shieh, J. H., Velardi, E., Holland, A. M., Singer, N. V., West, M. L., Smith, O. M., Young, L. F., Shono, Y., Ghosh, A., Hanash, A. M., Tran, H. T., Moore, M. A. & Van Den Brink, M. R. 2013. **Nrf2 regulates haematopoietic stem cell function.** *Nat Cell Biol,* 15**,** 309-16.

Tsujimoto, A., Nyunoya, H., Morita, T., Sato, T. & Shimotohno, K. 1991. **Isolation of Cdnas for DNA-Binding Proteins Which Specifically Bind to a Tax-Responsive Enhancer Element in the Long Terminal Repeat of Human T-Cell Leukemia-Virus Type-I.** *Journal of Virology,* 65**,** 1420-1426.

Tsukada, J., Yoshida, Y., Kominato, Y. & Auron, P. E. 2011. **The CCAAT/enhancer (C/EBP) family of basic-leucine zipper (bZIP) transcription factors is a multifaceted highly-regulated system for gene regulation.** *Cytokine,* 54**,** 6-19.

Tsune, I., Ikejima, K., Hirose, M., Yoshikawa, M., Enomoto, N., Takei, Y. & Sato, N. 2003. **Dietary glycine prevents chemical-induced experimental colitis in the rat.** *Gastroenterology,* 125**,** 775-85.

Tu, B. P. & Weissman, J. S. 2004. **Oxidative protein folding in eukaryotes: mechanisms and consequences.** *J Cell Biol,* 164**,** 341-6.

Turjanski, A. G., Vaque, J. P. & Gutkind, J. S. 2007. **MAP kinases and the control of nuclear events.** *Oncogene,* 26**,** 3240-53.

Umbricht, D., Alberati, D., Martin-Facklam, M., Borroni, E., Youssef, E. A., Ostland, M., Wallace, T. L., Knoflach, F., Dorflinger, E., Wettstein, J. G., Bausch, A., Garibaldi, G. & Santarelli, L. 2014. **Effect of bitopertin, a glycine reuptake inhibitor, on negative symptoms of schizophrenia: a randomized, double-blind, proof-of-concept study.** *JAMA Psychiatry,* 71**,** 637-46.

Vagin, O., Kraut, J. A. & Sachs, G. 2009. **Role of N-glycosylation in trafficking of apical membrane proteins in epithelia.** *Am J Physiol Renal Physiol,* 296**,** F459-69.

Van Dam, H., Wilhelm, D., Herr, I., Steffen, A., Herrlich, P. & Angel, P. 1995. **ATF-2 is preferentially activated by stress-activated protein kinases to mediate c-jun induction in response to genotoxic agents.** *EMBO J,* 14**,** 1798-811.

Van Den Eynden, J., Ali, S. S., Horwood, N., Carmans, S., Brone, B., Hellings, N., Steels, P., Harvey, R. J. & Rigo, J. M. 2009. **Glycine and glycine receptor signalling in non-neuronal cells.** *Front Mol Neurosci,* 2**,** 9.

Van Duynhoven, J., Vaughan, E. E., Jacobs, D. M., Kemperman, R. A., Van Velzen, E. J., Gross, G., Roger, L. C., Possemiers, S., Smilde, A. K., Dore, J., Westerhuis, J. A. & Van De Wiele, T. 2011. **Metabolic fate of polyphenols in the human superorganism.** *Proc Natl Acad Sci U S A,* 108 Suppl 1**,** 4531-8.

Van Hoorde, L., Pocard, M., Maryns, I., Poupon, M. F. & Mareel, M. 2000. **Induction of invasion in vivo of alpha-catenin-positive HCT-8 human colon-cancer cells.** *Int J Cancer,* 88**,** 751-8.

Vargas-Medrano, J. 2010. ***Phosphorylation of the glycine transporter 1.*** PhD, The University of Texas at El Paso.

Vargas-Medrano, J., Castrejon-Tellez, V., Plenge, F., Ramirez, I. & Miranda, M. 2011. **PKCbeta-dependent phosphorylation of the glycine transporter 1.** *Neurochem Int,* 59**,** 1123-32.

Vaughn, L. S., Snee, B. & Patel, R. C. 2014. **Inhibition of PKR protects against tunicamycin-induced apoptosis in neuroblastoma cells.** *Gene,* 536**,** 90-6.

Vauzour, D., Rodriguez-Mateos, A., Corona, G., Oruna-Concha, M. J. & Spencer, J. P. 2010. **Polyphenols and human health: prevention of disease and mechanisms of action.** *Nutrients,* 2**,** 1106-31.

Vente, J. P., Von Meyenfeldt, M. F., Van Eijk, H. M., Van Berlo, C. L., Gouma, D. J., Van Der Linden, C. J. & Soeters, P. B. 1989. **Plasma-amino acid profiles in sepsis and stress.** *Ann Surg,* 209**,** 57-62.

Vermeulen, S. J., Chen, T. R., Speleman, F., Nollet, F., Van Roy, F. M. & Mareel, M. M. 1998. **Did the Four Human Cancer Cell Lines DLD-1, HCT-15, HCT-8, and HRT-18 Originate from One and the Same Patient?** *Cancer Genetics and Cytogenetics,* 107**,** 76-79.

Vermeulen, S. J., Nollet, F., Teugels, E., Philippe, J., Speleman, F., Van Roy, F. M., Bracke, M. E. & Mareel, M. M. 1997. **Mutation of alpha-catenin results in invasiveness of human HCT-8 colon cancer cells.** *Ann N Y Acad Sci,* 833**,** 186-9.

Vilar, J. M. & Saiz, L. 2005. **DNA looping in gene regulation: from the assembly of macromolecular complexes to the control of transcriptional noise.** *Curr Opin Genet Dev,* 15**,** 136-44.

Vita, J. A. 2005. **Polyphenols and cardiovascular disease: effects on endothelial and platelet function.** *Am J Clin Nutr,* 81**,** 292S-297S.

Viu, E., Zapata, A., Capdevila, J. L., Fossom, L. H., Skolnick, P. & Trullas, R. 1998. **Glycine site antagonists and partial agonists inhibit N-methyl-D-aspartate receptor-mediated [3H]arachidonic acid release in cerebellar granule cells.** *J Pharmacol Exp Ther,* 285**,** 527-32.

Waas, W. F., Lo, H. H. & Dalby, K. N. 2001. **The kinetic mechanism of the dual phosphorylation of the ATF2 transcription factor by p38 mitogen-activated protein (MAP) kinase alpha. Implications for signal/response profiles of MAP kinase pathways.** *J Biol Chem,* 276**,** 5676-84.

Walker, R. I. & Porvaznik, M. J. 1978. **Disruption of the permeability barrier (zonula occludens) between intestinal epithelial cells by lethal doses of endotoxin.** *Infection and immunity.*

Walters, M. C., Fiering, S., Eidemiller, J., Magis, W., Groudine, M. & Martin, D. I. 1995. **Enhancers increase the probability but not the level of gene expression.** *Proceedings of the National Academy of Sciences of the United States of America,* 92**,** 7125-7129.

Wang, C. C. & Tsou, C. L. 1993. **Protein Disulfide-Isomerase Is Both an Enzyme and a Chaperone.** *Faseb Journal,* 7**,** 1515-1517.

Wang, Q., Mora-Jensen, H., Weniger, M. A., Perez-Galan, P., Wolford, C., Hai, T., Ron, D., Chen, W., Trenkle, W., Wiestner, A. & Ye, Y. 2009. **ERAD inhibitors integrate ER stress with an epigenetic mechanism to activate BH3-only protein NOXA in cancer cells.** *Proc Natl Acad Sci U S A,* 106**,** 2200-5.

Wang, R., Paul, V. J. & Luesch, H. 2013a. **Seaweed extracts and unsaturated fatty acid constituents from the green alga Ulva lactuca as activators of the cytoprotective Nrf2-ARE pathway.** *Free Radic Biol Med,* 57**,** 141-53.

Wang, T., Jonsdottir, R., Liu, H., Gu, L., Kristinsson, H. G., Raghavan, S. & Olafsdottir, G. 2012. **Antioxidant capacities of phlorotannins extracted from the brown algae Fucus vesiculosus.** *J Agric Food Chem,* 60**,** 5874-83.

Wang, T. & Stormo, G. D. 2003. **Combining phylogenetic data with co-regulated genes to identify regulatory motifs.** *Bioinformatics,* 19**,** 2369-80.

Wang, W., Wu, Z., Dai, Z., Yang, Y., Wang, J. & Wu, G. 2013b. **Glycine metabolism in animals and humans: implications for nutrition and health.** *Amino Acids,* 45**,** 463-77.

Wasserman, W. W. & Fahl, W. E. 1997. **Functional antioxidant responsive elements.** *Proc Natl Acad Sci U S A,* 94**,** 5361-6.

Wasserman, W. W. & Sandelin, A. 2004. **Applied bioinformatics for the identification of regulatory elements.** *Nat Rev Genet,* 5**,** 276-87.

Waterhouse, R. M., Tegenfeldt, F., Li, J., Zdobnov, E. M. & Kriventseva, E. V. 2013. **OrthoDB: a hierarchical catalog of animal, fungal and bacterial orthologs.** *Nucleic Acids Res,* 41**,** D358-65.

Weinberg, J. M. 1990. **The effect of amino acids on ischemic and toxic injury to the kidney.** *Semin Nephrol,* 10**,** 491-500.

Weinberg, J. M., Davis, J. A., Abarzua, M., Kiani, T. & Kunkel, R. 1990. **Protection by Glycine of Proximal Tubules from Injury Due to Inhibitors of Mitochondrial Atp Production.** *American Journal of Physiology,* 258**,** C1127-C1140.

Weinberg, J. M., Davis, J. A., Abarzua, M. & Rajan, T. 1987. **Cytoprotective effects of glycine and glutathione against hypoxic injury to renal tubules.** *J Clin Invest,* 80**,** 1446-54.

West, K. L. 2004. **HMGN proteins play roles in DNA repair and gene expression in mammalian cells.** *Biochem Soc Trans,* 32**,** 918-9.

West, K. L., Castellini, M. A., Duncan, M. K. & Bustin, M. 2004. **Chromosomal proteins HMGN3a and HMGN3b regulate the expression of glycine transporter 1.** *Mol Cell Biol,* 24**,** 3747-56.

Wheeler, M., Stachlewitz, R. F., Yamashina, S., Ikejima, K., Morrow, A. L. & Thurman, R. G. 2000. **Glycine-gated chloride channels in neutrophils attenuate calcium influx and superoxide production.** *FASEB J,* 14**,** 476-84.

Wheeler, M. D., Ikejema, K., Enomoto, N., Stacklewitz, R. F., Seabra, V., Zhong, Z., Yin, M., Schemmer, P., Rose, M. L., Rusyn, I., Bradford, B. & Thurman, R. G. 1999. **Glycine: a new anti-inflammatory immunonutrient.** *Cell Mol Life Sci,* 56**,** 843-56.

Wijnhoven, B. P., Dinjens, W. N. & Pignatelli, M. 2000. **E-cadherin-catenin cell-cell adhesion complex and human cancer.** *Br J Surg,* 87**,** 992-1005.

Willem H Mager, A. H. D. B. M. H. S. H.-P. V. 2000. **Cellular responses to oxidative and osmotic stress.** *Cell Stress & Chaperones,* 5**,** 73.

Williams, H. R., Cox, I. J., Walker, D. G., Cobbold, J. F., Taylor-Robinson, S. D., Marshall, S. E. & Orchard, T. R. 2010. **Differences in gut microbial metabolism are responsible for reduced hippurate synthesis in Crohn's disease.** *BMC Gastroenterol,* 10**,** 108.

Williams, H. R., Cox, I. J., Walker, D. G., North, B. V., Patel, V. M., Marshall, S. E., Jewell, D. P., Ghosh, S., Thomas, H. J., Teare, J. P., Jakobovits, S., Zeki, S., Welsh, K. I., Taylor-Robinson, S. D. & Orchard, T. R. 2009. **Characterization of inflammatory bowel disease with urinary metabolic profiling.** *Am J Gastroenterol,* 104**,** 1435-44.

Williams, K. W., Liu, T., Kong, X., Fukuda, M., Deng, Y., Berglund, E. D., Deng, Z., Gao, Y., Liu, T., Sohn, J. W., Jia, L., Fujikawa, T., Kohno, D., Scott, M. M., Lee, S., Lee, C. E., Sun, K., Chang, Y., Scherer, P. E. & Elmquist, J. K. 2014. **Xbp1s in Pomc Neurons Connects ER Stress with Energy Balance and Glucose Homeostasis.** *Cell Metab,* 20**,** 471-82.

Wolfgang, C. D., Chen, B. P., Martindale, J. L., Holbrook, N. J. & Hai, T. 1997. **gadd153/Chop10, a potential target gene of the transcriptional repressor ATF3.** *Mol Cell Biol,* 17**,** 6700-7.

Wu, G. 2010. **Functional amino acids in growth, reproduction, and health.** *Adv Nutr,* 1**,** 31-7.

Wu, T., Zhao, F., Gao, B., Tan, C., Yagishita, N., Nakajima, T., Wong, P. K., Chapman, E., Fang, D. & Zhang, D. D. 2014. **Hrd1 suppresses Nrf2-mediated cellular protection during liver cirrhosis.** *Genes Dev,* 28**,** 708-22.

Wuensch, T., Ullrich, S., Schulz, S., Chamaillard, M., Schaltenberg, N., Rath, E., Goebel, U., Sartor, R. B., Prager, M., Buning, C., Bugert, P., Witt, H., Haller, D. & Daniel, H. 2014. **Colonic expression of the peptide transporter PEPT1 is downregulated during intestinal inflammation and is not required for NOD2-dependent immune activation.** *Inflamm Bowel Dis,* 20**,** 671-84.

Xiao, H. & Lis, J. T. 1988. **Germline Transformation Used to Define Key Features of Heat-Shock Response Elements.** *Science,* 239**,** 1139-1142.

Xu, T., Yang, L., Yan, C., Wang, X., Huang, P., Zhao, F., Zhao, L., Zhang, M., Jia, W., Wang, X. & Liu, Y. 2014. **The IRE1alpha-XBP1 pathway regulates metabolic stress-induced compensatory proliferation of pancreatic beta-cells.** *Cell Res,* 24**,** 1137-40.

Yamabe, S., Hirose, J., Uehara, Y., Okada, T., Okamoto, N., Oka, K., Taniwaki, T. & Mizuta, H. 2013. **Intracellular accumulation of advanced glycation end products induces apoptosis via endoplasmic reticulum stress in chondrocytes.** *FEBS J,* 280**,** 1617-29.

Yancey, P. H. 2005. **Organic osmolytes as compatible, metabolic and counteracting cytoprotectants in high osmolarity and other stresses.** *J Exp Biol,* 208**,** 2819-30.

Yang, C., Bolotin, E., Jiang, T., Sladek, F. M. & Martinez, E. 2007. **Prevalence of the initiator over the TATA box in human and yeast genes and identification of DNA motifs enriched in human TATA-less core promoters.** *Gene,* 389**,** 52-65.

Ye, P., Mimura, J., Okada, T., Sato, H., Liu, T., Maruyama, A., Ohyama, C. & Itoh, K. 2014. **Nrf2- and ATF4-Dependent Upregulation of xCT Modulates the Sensitivity of T24 Bladder Carcinoma Cells to Proteasome Inhibition.** *Mol Cell Biol,* 34**,** 3421-34.

Yolanda, S. L.-B., Marie, P., William, W. & Roberto, S. 2013. **NRF2 Modulates Anti-Oxidant And Cytoprotective Gene Expression In Normal Human Bronchial Epithelial Cells Exposed To Cigarette Smoke And Endoplasmic Reticulum Stress**. *A27. SMOKE SIGNALS ON PULMONARY CELLS.* American Thoracic Society.

Yordy, J. S. & Muise-Helmericks, R. C. 2000. **Signal transduction and the Ets family of transcription factors.** *Oncogene,* 19**,** 6503-13.

Yoshida, H., Matsui, T., Yamamoto, A., Okada, T. & Mori, K. 2001. **XBP1 mRNA is induced by ATF6 and spliced by IRE1 in response to ER stress to produce a highly active transcription factor.** *Cell,* 107**,** 881-891.

Yoshida, H., Okada, T., Haze, K., Yanagi, H., Yura, T., Negishi, M. & Mori, K. 2000. **ATF6 activated by proteolysis binds in the presence of NF-Y (CBF) directly to the cis-acting element responsible for the mammalian unfolded protein response.** *Mol Cell Biol,* 20**,** 6755-67.

Yoshida, H., Oku, M., Suzuki, M. & Mori, K. 2006. **pXBP1(U) encoded in XBP1 pre-mRNA negatively regulates unfolded protein response activator pXBP1(S) in mammalian ER stress response.** *J Cell Biol,* 172**,** 565-75.

Yu, Y. M., Burke, J. F., Vogt, J. A., Chambers, L. & Young, V. R. 1992. **Splanchnic and whole body L-[1-13C,15N]leucine kinetics in relation to enteral and parenteral amino acid supply.** *Am J Physiol,* 262**,** E687-94.

Yu, Y. M., Young, V. R., Tompkins, R. G. & Burke, J. F. 1995. **Comparative-Evaluation of the Quantitative Utilization of Parenterally and Enterally Administered Leucine and L-[1-13c,15n]Leucine within the Whole-Body and the Splanchnic Region.** *Journal of Parenteral and Enteral Nutrition,* 19**,** 209-215.

Zafra, F. & Gimenez, C. 2008. **Glycine transporters and synaptic function.** *IUBMB Life,* 60**,** 810-7.

Zaia, K. A. & Reimer, R. J. 2009. **Synaptic Vesicle Protein NTT4/XT1 (SLC6A17) Catalyzes Na+-coupled Neutral Amino Acid Transport.** *J Biol Chem,* 284**,** 8439-48.

Zhai, X., Lin, M., Zhang, F., Hu, Y., Xu, X., Li, Y., Liu, K., Ma, X., Tian, X. & Yao, J. 2013. **Dietary flavonoid genistein induces Nrf2 and phase II**

detoxification gene expression via ERKs and PKC pathways and protects against oxidative stress in Caco-2 cells. *Mol Nutr Food Res,* 57**,** 249-59.

Zhang, C., Xuan, Z., Otto, S., Hover, J. R., Mccorkle, S. R., Mandel, G. & Zhang, M. Q. 2006. **A clustering property of highly-degenerate transcription factor binding sites in the mammalian genome.** *Nucleic Acids Res,* 34**,** 2238-46.

Zhang, M. I. & O'neil, R. G. 1996. **An L-type calcium channel in renal epithelial cells.** *J Membr Biol,* 154**,** 259-66.

Zhao, Y., Li, X., Cai, M. Y., Ma, K., Yang, J., Zhou, J., Fu, W., Wei, F. Z., Wang, L., Xie, D. & Zhu, W. G. 2013. **XBP-1u suppresses autophagy by promoting the degradation of FoxO1 in cancer cells.** *Cell Res,* 23**,** 491-507.

Zhong, Z., Wheeler, M. D., Li, X., Froh, M., Schemmer, P., Yin, M., Bunzendaul, H., Bradford, B. & Lemasters, J. J. 2003. **L-Glycine: a novel antiinflammatory, immunomodulatory, and cytoprotective agent.** *Curr Opin Clin Nutr Metab Care,* 6**,** 229-40.

Zhou, H., Zarubin, T., Ji, Z., Min, Z., Zhu, W., Downey, J. S., Lin, S. & Han, J. 2005. **Frequency and distribution of AP-1 sites in the human genome.** *DNA Res,* 12**,** 139-50.

Zipper, L. M. & Mulcahy, R. T. 2003. **Erk activation is required for Nrf2 nuclear localization during pyrrolidine dithiocarbamate induction of glutamate cysteine ligase modulatory gene expression in HepG2 cells.** *Toxicol Sci,* 73**,** 124-34.

Zong, Z. H., Du, Z. X., Li, N., Li, C., Zhang, Q., Liu, B. Q., Guan, Y. & Wang, H. Q. 2012. **Implication of Nrf2 and ATF4 in differential induction of CHOP by proteasome inhibition in thyroid cancer cells.** *Biochim Biophys Acta,* 1823**,** 1395-404.