Identification of topological determinants in the N-terminal domain of transcription factor Nrf1 that control its orientation in the endoplasmic reticulum membrane

Yiguo Zhang, John D Hayes

Identification of topological determinants in the N-terminal domain of transcription factor Nrf1 that control its orientation in the endoplasmic reticulum membrane

Yiguo Zhang[†¶§], and John D. Hayes[†]

[†]Biomedical Research Institute, Ninewells Hospital and Medical School, University of Dundee, Dundee DD1 9SY, Scotland, United Kingdom, and [¶]Laboratory of Cell Biochemistry and Gene Regulation, College of Bioengineering and Life Sciences, University of Chongqing, 400044, China. [§]Correspondence should be addressed to Yiguo Zhang (email: y.z.zhang@dundee.ac.uk).

Running head: Topological determinants of Nrf1 N-terminal domain within the ER

Key words: Nrf1, topology, endoplasmic reticulum, N-glycosylation, detergent-resistant membrane and nuclear envelope

ABSTRACT

NF-E2-related factor 1 (Nrf1) is a cap'n'collar (CNC) basic-region leucine zipper (bZIP) transcription factor that is tethered to endoplasmic reticulum (ER) and nuclear envelope membranes through its N-terminal signal peptide (residues 1-30). Besides the signal peptide, amino acids 31-90 of Nrf1 also negatively regulate the CNC-bZIP factor. In this paper we have tested the hypothesis that amino acids 31-90 of Nrf1, and the overlapping N-terminal homology box 2 (NHB2, residues 82-106), inhibit Nrf1 because they control its topology within membranes. This region contains three amphipathic α-helical regions comprising amino acids 31-50 [called the signal peptide-associated sequence (SAS)], 55-82 [called the cholesterol recognition amino acid consensus sequences (CRACs)], and 89-106 (part of NHB2). We present experimental data showing that the signal peptide of Nrf1 contains a transmembrane 1 region (TM1, residues 7-24) that is orientated across the ER membrane in an N_{cyt}/C_{lum} fashion with its N-terminus facing the cytoplasm and its C-terminus positioned in the lumen of the ER. Once Nrf1 is anchored to the ER membrane through TM1, the remaining portion of the N-terminal domain (residues 1-124) is transiently translocated into the ER lumen. Thereafter, Nrf1 adopts a topology in which the SAS is inserted into the membrane, the CRACs are probably repartitioned to the cytoplasmic side of the ER membrane, and NHB2 may serve as an anchor switch, either lying on the luminal surface of the ER or traversing the membrane with an N_{cyt}/C_{lum} orientation. Thus, Nrf1 can adopt several topologies within membranes that are determined by its N-terminal domain.

INTRODUCTION

Nuclear factor-erythroid 2 (NF-E2)-related factors 1 (Nrf1), 2 (Nrf2) and 3 (Nrf3) are cap'n'collar (CNC) basic-region leucine zipper (bZIP) transcription factors [1-4]. All three CNC-bZIP proteins regulate cytoprotective genes that contain an antioxidant response element (ARE, with minimal consensus sequence 5′-$^{A}/_{G}$TGA$^{C}/_{G}$nnnGC$^{A}/_{G}$-3′) in their regulatory regions [4-8]. Such genes include those encoding the glutamate-cysteine ligase catalytic and modifier subunits, class Alpha and Mu glutathione *S*-transferase (GST) isoenzymes, NAD(P)H:quinone oxidoreductase 1, heme oxygenase 1, ferritin and metallothioneins [9-15]. Nrf1 and Nrf2 are widely expressed in mouse and human tissues [1-3, 12, 16]. By contrast, Nrf3 is particularly abundant in the placenta, and whilst it is present at modest levels in the liver, it is expressed to a much lesser extent in other tissues [4, 17, 18].

Gene-targeting experiments have revealed that Nrf1, Nrf2 and Nrf3 possess distinct biological functions. Global knockout of Nrf1 is embryonically lethal to the mouse, with the livers of foetal mice exhibiting increased levels of apoptosis, resulting from high levels of oxidative stress [19-21]. Furthermore, liver-specific disruption of *Nrf1* in neonatal mice results in the development of non-alcoholic steatohepatitis and ultimately hepatoma [22]. Thus, Nrf1 appears to be necessary for the expression of ARE-driven genes that are critical for liver development. Whilst Nrf1 is essential for foetal development, global knockout of Nrf2 in the mouse yields viable animals that develop normally [23]. However, *Nrf2*$^{-/-}$ mice are intolerant of many environmental stressors, and are susceptible to acute and chronic injuries, inflammation and chemical carcinogenesis [24-28]. Lastly, *Nrf3*$^{-/-}$ mice display no obvious

phenotype [29], and it is possible that Nrf1 and/or Nrf2 can compensate the loss of Nrf3. Collectively, these studies indicate that amongst the CNC-bZIP factors, Nrf1 fulfils an indispensable and unique function in maintaining the redox status and integrity of the liver. This conclusion has prompted us to examine the activity of Nrf1 in greater detail.

Transcription factor Nrf1 exists as a number of isoforms that arise from alterative RNA splicing, translation from several internal ATG codons [2, 6, 16, 30-32], and various post-synthetic modifications including glycosylation and possibly proteolytic cleavage [33, 34]. We have reported previously that full-length Nrf1 is tethered to the endoplasmic reticulum (ER) through a non-cleavable N-terminal signal anchor sequence, and that it is glycosylated in this organelle through its Asn/Ser/Thr-rich (NST) domain [33]. Further studies have shown that Nrf1 is an integral membrane protein which can be sorted from the ER to the nuclear envelope [35]. Upon reaching the inner nuclear membrane, Nrf1 is presumably subject to retrotranslocation, whereupon it is recruited to the promoters of ARE-driven genes through its CNC-bZIP domain and gains access to the transcriptional machinery through its two acidic activation domains (i.e., AD1 and AD2).

The N-terminal domain (NTD, residues 1-124) of Nrf1 was originally described as a region that negatively controls the CNC-bZIP protein [36]. In particular, it has been shown that progressive sequential deletion of residues 2-90 in Nrf1 produces a gradual incremental step-wise increase in its transactivation activity, suggesting that a substantial portion of the NTD contributes to its negative control. The NTD of Nrf1 contains an N-terminal homology box 1 (NHB1, residues 11-30) and an N-terminal homology box 2 (NHB2, residues 82-106), subdomains that were so named because they are also represented in the NTD of Nrf3 [8]. In Nrf1, NHB1 represents the signal anchor sequence [33], whereas in mouse Nrf3 it is cleaved [8]. In Nrf1, less is known about the function of NHB2 than NHB1. However, the NHB2 subdomain of Nrf3 is required for its activity as it contributes to the sorting of the factor within the ER [8]. It is therefore not known why the region between amino acids 31-90 of Nrf1 inhibits its activity, though it is probable that NHB2 contributes to repression as it overlaps with the sequence.

Knowledge of the topology of Nrf1 within the membrane is necessary in order to gain a better understanding of how the CNC-bZIP protein fulfils its biological function. We have previously proposed that full-length Nrf1 possesses three transmembrane (TM) regions, TM1, TMi and TMc [35]. The predicted TM1 α-helix is formed by amino acids 7-24, and whilst it predominantly determines the topology of Nrf1 within the ER membrane, its orientation within the membrane is uncertain. In the present study we have determined the orientation of TM1 in ER membranes. We have also examined whether amino acids 31-90, as well as NHB2, might form amphipathic helices that contribute to the negative regulation of Nrf1.

EXPERIMENTAL

Chemicals, antibodies and other reagents

These were all of the highest quality available and were readily available commercially. The ER extraction kit and all chemicals were purchased from Sigma-Aldrich. Peptide:N-glycosidase (PNGase) F, endoglycosidase (Endo) H and proteinase K (PK) were obtained from New England Biolabs. Rabbit polyclonal antibodies against calreticulin (CRT) and green fluorescent protein (GFP) were bought from Calbiochem and Abcam PLC, respectively. A mouse monoclonal antibody against the V5 epitope and rabbit polyclonal antibodies against DsRed (a *Discosoma sp*. red fluorescent protein) were from Invitrogen Ltd. Antisera against Nrf1 were produced in rabbits using a polypeptide covering amino acids 292-741 (called Nrf1β) as an antigen that was purified in our laboratory. All oligonucleotide primers in this study were synthesized by MWG Biotech.

Expression constructs

Expression constructs for full-length mouse Nrf1 and its mutants, along with other expression constructs for mouse Nrf2, GFP and their chimearic proteins fused with various portions of Nrf1, have been described previously [35, 36]. These mutants and others used in this study were created by PCR-directed deletion mutagenesis within the NTD of Nrf1, as described elsewhere [37]. Two sandwiched fusion proteins, called DsRed/NTD/GFP and DsRed/N65/GFP, were engineered by inserting nucleotide sequences encoding the NTD of Nrf1 or its N-terminal 65 amino acids (N65) between the cDNAs for DsRed2 and GFP within the pDsRed2-GFP vector through either the SalI/KpnI or the HindIII/KpnI multiple cloning sites [35]. The fidelity of all cDNA products was confirmed by sequencing.

Cell culture, transfection, luciferase reporter assays and statistical analysis

Monkey kidney COS-1 cells (3×10^5) were seeded in a 6-well plate and grown for 24 h in Dulbecco's modified Eagle's medium. After reaching 70% confluence, the cells were transfected with a Lipofectamine 2000 (Invitrogen) mixture that contained an expression construct for wild-type Nrf1 or a mutant, along with either a P_{SV40}*GSTA2*-6×ARE-Luc (which contains six copies of the core ARE consensus sequence from rat GSTA2) or P_{TK}*nqo1*-ARE-Luc reporter plasmid (which contains only one copy of the ARE from mouse *nqo1*), together with pcDNA4/HisMax/*lacZ*; the latter encodes β-galactosidase (β-gal) and was used to control for transfection efficiency [35, 36, 38]. Luciferase activity was measured at approximately 36 h after transfection. The data were calculated as a fold change (mean ± S.D) of the activity obtained following transfection with an expression vector for the CNC-bZIP factor, when compared with that obtained following transfection with an empty pcDNA3.1/V5 His B vector. The data presented are each representatives of at least 3 different independent experiments undertaken on separate occasions that were performed in triplicate. The significance of differences in the luciferase activity was determined using the Student's *t* test and is shown as a *p* value.

Immunocytochemistry and confocal and microscopy

These experiments were performed as described elsewhere [33, 36].

Subcellular fractionations and membrane proteinase protection reactions

The intact ER-rich membrane and nuclear fractions were prepared from COS-1 cells, as described previously [33, 35]. These fractions were resuspended in 100 μl of 1 × isotonic buffer [10 mM Hepes, pH 7.8, containing 250 mM sucrose, 1 mM EGTA, 1 mM EDTA and 25 mM KCl]. Subsequently, membrane proteinase protection reactions were performed for 15, 30 or 60 min on ice in an aliquot (50 μg of protein) of the membrane-containing preparation with PK at a final concentration of 50 or 100 μg protein/ml in either the presence or absence of 1% (v/v) Triton X-100 (TX). The reactions were stopped by incubation at 90°C for 10 min following the addition of 1 mM PMSF.

Gycosylation mapping, deglycosylation reactions and Western blotting

The Nrf1$^{1\text{-}7N/Q}$ mutant was constructed by site-directed mutation of all seven endogenous glycosylation consensus asparagines (Asn−X−Ser/Thr, where X is any amino acids except proline) into glutamines, thereby ensuring it would not be glycated in the ER lumen [35]. Using the cDNA for Nrf1$^{1\text{-}7N/Q}$ as a template, a series of N-linked glycosylation asparagine acceptor sites were introduced into its NTD. It was anticipated that, if the engineered glycosylation sites were translocated into the ER lumen, the mutant Nrf1 protein would be glycated by *in vivo* addition of a glycan precursor Glc3Man9GlcNAc2; this technique is called glycosylation mapping mutagenesis [39]. Subsequently, this modification was detected using *in vitro* deglycosylation reactions, followed by Western blotting [33, 35, 40]. On some occasions, nitrocellulose membranes that had already been blotted with an antibody were washed for 30 min

with stripping buffer [7M guanidine hydrochloride, 50 mM glycine, 0.05 mM EDTA, 0.1 M KCl and 20 mM 2-mercaptoethanol at pH 10.8] before being re-probed with additional primary antibodies [41]. The intensity of some Western blots was calculated using the Quantity One® software developed at the Bio-Rad Laboratories.

Bioinformatic analysis

The topology of Nrf1 around the membranes was predicted firstly using several bioinformatic algorithms, including the TopPred (http://mobyle.pasteur.fr/cgi-bin/portal.py?form=toppred), TMpro (http://flan.blm.cs.cmu.edu/tmpro/), HeliQuest (http://heliquest.ipmc.cnrs.fr/), and AmphipaSeek (http://npsa-pbil.ibcp.fr/cgi-bin/npsaautomat.pl?page=/NPSA/npsa_amphipaseek.html) software programmes. Thereafter, the predicted topologies were evaluated using molecular and biochemical experiments.

RESULTS

Identification of potential topological elements in the NTD of Nrf1

Several bioinformatic algorithms have predicted that the TM1 α-helix in Nrf1 is formed by amino acids 7-24 [35]. In the present study we have addressed the question of whether inhibition of Nrf1 activity by its NTD is due in part to the presence of other putative α-helical amphipathic regions within the domain, besides TM1, that are capable of interacting with membranes (Figure 1 and supplemental Figure S1). The first of these helices lies between amino acids 31-50, and we have designated it the signal peptide-associated sequence (SAS). It is apparent that SAS shares 55% similarity with the TM1 sequence, which suggests it may interact with membranes. Moreover, as shown in Figure 1B, SAS also shows conservation with the core N-terminal TM sequence of nucleus-vacuole junction protein 1 (Nvj1), an integral membrane polypeptide anchored in the inner nuclear membrane of *Saccharomyces cerevisiae* [42]. The second predicted α-helical amphipathic region possesses a net basic charge and lies between amino acids 55-82. It overlaps with two potential cholesterol recognition amino acid consensus (CRAC) motifs, L/V-$X_{(1-5)}$-Y-$X_{(1-5)}$-R/K, located between amino acids 62-70 and 74-82 (Figure 1C); we therefore called it the CRACs subdomain. The third possible α-helical amphipathic region has a net acidic charge and lies between amino acids 89-106 (supplemental Figure S1). As shown in Figure 1D, it resides within NHB2. In addition, the NHB2 sequence shares 36% identity and 68% similarity with an N-terminal region of ADP ribosylation factor (Arf)-GTPase-activating protein 1 (ArfGAP1), a membrane-bound protein that regulates membrane traffic and protein transport [43-45].

Consequences of loss of the SAS and CRACs subdomains on Nrf1 activity

We have previously reported that NHB1 negatively controls Nrf1 by targeting it to the ER membrane [33]. Similarly, we have reported that Nrf1 is also repressed by both NHB2 [35] and the region located between NHB1 and NHB2 that includes the SAS and CRACs motifs [36], but the reason why this subdomain within the NTD negatively controls the CNC-bZIP factor is not known. To test the functional significance of SAS and CRACs, we transfected COS-1 cells with an expression construct for wild-type Nrf1 or Nrf1 mutants that lack either SAS and/or CRACs, along with a P_{SV40}*GSTA2*-6×ARE-Luc reporter plasmid. Importantly, internal deletion of the SAS (residues 31-50), CRACs (residues 55-80) or both (residues 31-80) subdomains from Nrf1 gave rise to differential increases in the transactivation activity (Figure 2B). This result indicates that both the SAS and CRACs negatively regulate Nrf1.

Western blotting experiments were performed to determine whether the higher activity of $Nrf1^{\Delta 31\text{-}50}$ and $Nrf1^{\Delta 55\text{-}80}$ is accompanied by the appearance of novel forms of the factor. We have reported previously that wild-type Nrf1 migrates as both a non-glycated ~95-kDa band and a glycated ~120-kDa band during LDS/NuPAGE [33]. As shown

in Figure 2B (*right panel*), Nrf1$^{\Delta 31\text{-}50}$ and Nrf1$^{\Delta 55\text{-}80}$ both ran during LDS/NuPAGE as ~95-kDa and ~120-kDa bands; in both instances, digestion with PNGase F prior to electrophoresis eliminated the ~120-kDa band, thereby providing evidence that it was glycosylated. A non-glycated unstable ~85-kDa band was observed following electrophoresis of Nrf1$^{\Delta 31\text{-}50}$ but not the Nrf1$^{\Delta 55\text{-}80}$ mutant. These data indicate that the high activity of Nrf1 mutants lacking the SAS or CRACs subdomains is not associated with glycosylation status of the protein.

Although removal of the SAS subdomain in Nrf1$^{\Delta 31\text{-}50}$ caused an obvious increase in its transactivation activity, confocal imaging revealed that this mutant protein was located primarily in the extranuclear ER and the nuclear envelope, but not within the nucleus (Figure 2C). By contrast, the Nrf1$^{\Delta 55\text{-}80}$ mutant protein exhibited a relative increase in staining of the nucleus, but this was not accompanied by a decrease in its extranuclear staining when compared with wild-type Nrf1 (also see supplemental Figures S2 and S3).

The SAS and CRACs subdomains regulate the association of Nrf1 with membranes

Comparison of confocal images indicated that loss of SAS had no effect on the extranuclear localization of Nrf1 because both Nrf1$^{\Delta 2\text{-}50}$ and Nrf1$^{\Delta 2\text{-}30}$ mutant proteins gave similarly low levels (~20%) of extranuclear staining in COS-1 cells (supplemental Figure S3). Rather, subcellular fractionation revealed an increased recovery of Nrf1$^{\Delta 2\text{-}50}$ in the microsome-rich fraction when compared with the recovery of Nrf1$^{\Delta 2\text{-}30}$. The recovery of Nrf1$^{\Delta 2\text{-}50}$ was either decreased by removal of CRACs, to give Nrf1$^{\Delta 2\text{-}80}$, or diminished by deletion of both CRACs and NHB2, in the Nrf1$^{\Delta 2\text{-}120}$ mutant (supplemental Figure S4A). Subsequently, membrane proteinase protection reactions showed that a portion of Nrf1$^{\Delta 2\text{-}50}$ was located on the cytoplasmic side of the intact ER/microsome-rich membranes, because this mutant protein was not protected by membranes during proteolysis (supplemental Figure S4B). These data indicate that CRACs and NHB2, but not SAS, are required by Nrf1 to bind membranes.

SAS, CRACs and NHB2 are transiently translocated into the ER lumen

The above experiments suggested that transactivation activity of Nrf1 may be modulated by its topology within membranes. To explore whether the NTD of Nrf1 resides within the lumen of the ER following its targeting to the organelle, we performed glycosylation-mapping mutagenesis using Nrf1$^{1\text{-}7N/Q}$, a non-glycated mutant as the reference protein [35]. Nine novel glycosylation consensus sites were introduced at different indicated positions within the NTD of the Nrf1$^{1\text{-}7N/Q}$ mutant (Figure 3A). Amongst these engineered asparagines (eN), the two that were located closest to the N-terminus or the C-terminus of TM1 (i.e., Nrf1^{eN8} and Nrf1^{eN28}) were not glycosylated because no distinguishable change in their electrophoretic mobility was observed following PNGase F catalyzed deglycosylation (Figure 3B). This finding suggests that after ER-targeting of Nrf1, TM1 is inserted within the ER membrane but is not translocated into the lumen of this organelle. An alternative explanation is that, even if the C-terminus of TM1 were translocated into the lumen, it is associated so closely with the ER membrane that the eN^{28} residue is unable to reach the luminal active site of oligosaccharyl transferase (OST), a membrane-bound enzyme that catalyzes the addition of a glycan precursor Glc3Man9GlcNAc2 to the amino group of the asparagines [39]. Similarly, the eN^{33} residue located in the N-terminus of SAS that adjoins TM1 was not glycated because the electrophoretic mobility of Nrf1^{eN33} was not obviously altered following digestion of this mutant protein with either PNGase F or Endo H (Figure 3B and C).

Further glycosylation scanning mutagenesis revealed that all of the eN residues that had been introduced at positions 51, 60, 78, 97, 112 and 125 in the NTD of Nrf1 were glycated in the ER lumen as their mobilities during NuPAGE differed following PNGase F digestion (Figure 3B). These observations indicate that amino acids 51–125 around and within the SAS, CRACs and NHB2 are translocated transiently into the ER lumen. However, close examination of their electrophoretic mobilities revealed that the Nrf1^{eN97} and Nrf1^{eN112} mutant glycoproteins were

deglycated by Endo H (Figure 3C), but they was not extensively deglycated by PNGase F as the latter digestion did not alter their migration during NuPAGE (Figure 3B). These two different glycosidase reactions suggest that amino acids 97 and 112 located within and around NHB2 are probably embedded within a tightly folded conformation.

ARE-driven reporter luciferase assays showed a marked increase in transactivation by glycated Nrf1^{eN51}, in which eN^{51} was located in a possible loop between SAS and CRACs (Figure 3D). Other glycated proteins Nrf1^{eN60}, Nrf1^{eN78} and Nrf1^{eN112}, but not Nrf1^{eN97}, possessed a modest increase in their transactivation activity. Interestingly, non-glycated Nrf1^{eN28} and Nrf1^{eN33} also had a similar transactivation activity as those obtained from Nrf1^{eN112} or Nrf1^{eN60}, respectively. These data suggest that engineered asparagines at the indicated position and their glycosylation may have an effect on topological orientation of Nrf1 and its subsequent retrotanslation from the ER lumen to the cytoplasm and/or the nucleoplasm.

TM1 is orientated with its N-terminus facing the cytoplasm while its C-terminus resides in the ER lumen

It is known that an ER membrane-bound signal peptidase (SPase) possesses a luminal active site that catalyzes the cleavage of signal peptides immediately after small amino acids within the peptide c-region [46]. To examine whether the C-terminal part of the TM1 region of Nrf1 is positioned in the ER, we introduced two putative SPase cleavage consensus sites (e.g., Ala/Gly−X−Ala/Gly) within the NHB1 to yield the Nrf1$^{L12/13A}$ and Nrf1$^{L19/21A}$ mutants (Figure 4A). Unfortunately, it was difficult to detect by LDS/NuPAGE the loss of only 13 or 21 amino acid residues from full-length Nrf1 proteins (Figure 4B). To resolve this problem, we further examined the electrophoretic properties of N65/GFP×2 and NTD/GFP×2 fusion proteins and their point mutants because they were considerably smaller than Nrf1 and did not contain a functional glycosylation consensus site. Immunoblots of COS-1 cell lysates expressing these fusion proteins revealed that both N65/GFP×2 and NTD/GFP×2 were clearly resolved from GFP×2 by LDS/NuPAGE (Figure 4C); during electrophoresis, N65/GFP×2 and NTD/GFP×2 each gave a clear single band with estimated molecular masses of 65 kDa and 73 kDa, respectively, whilst GFP×2 migrated with an apparent mass of 58 kDa. The relative mobilities of N65/GFP×2 and NTD/GFP×2 were consistent with them containing, respectively, 65 and 125 amino acids more than GFP×2. Importantly, a distinct cleaved polypeptide of ~62.5 kDa was detected in cell lysates expressing N65$^{L19/21A}$/GFP×2 (Figure 4C). Similarly, a cleaved polypeptide of ~71 kDa with a mobility similar to the full-length ~73-kDa protein was also observed in the cells expressing NTD$^{L19/21A}$/GFP×2. These findings demonstrate that the pentapeptide ^{19}Ala−Ser−Ala−Ile−Gly^{23} in N56$^{L19/21A}$/GFP×2 and in NTD$^{L19/21A}$/GFP×2 is located around the membrane luminal interface in close vicinity to the active site of SPase. By contrast, neither N65$^{L12/13A}$/GFP×2 nor NTD$^{L12/13A}$/GFP×2 yielded a cleaved polypeptide, suggesting that the ^{11}Gly−Ala−Ala^{13} tripeptide is positioned in the membrane hydrocarbon center or close to the membrane cytoplasmic interface. Therefore, we deduce that the TM1 α-helix is inserted in membranes with an orientation of N_{cyt}/C_{lum} (i.e., with its N-terminus in the cytoplasm and its C-terminus within the ER lumen).

Whilst the Nrf1$^{L19/21A}$ mutant migrated primarily as a 120-kDa glycoprotein, it also yielded three shorter glycopolypeptides with molecular masses of ~36 kDa, ~52 kDa, and ~65 kDa (Figure 4B). However, the appearance of these additional Nrf1$^{L19/21A}$ products did not result in obvious changes in either its transactivation activity (Figure 4D) or subcellular distribution (Figure 4E) when compared to wild-type Nrf1. The finding that Nrf1$^{L19/21A}$ yields multiple forms is consistent with amino acids at positions either 21−22 or 23−24 being proteolytically cleaved by a luminal SPase, followed possibly by sequential proteolytic cleavage around and within its NST domain. By contrast, the Nrf1$^{L12/13A}$ mutant was expressed exclusively as a nuclear non-glycated ~95-kDa protein (Figure 4B and E), and exhibited increased activity (Figure 4D). This result indicates that the leucine residues 12 and 13 in the TM1 region

are situated in the hydrocarbon center of the membrane lipid bilayer and their hydrophobic properties are essential for insertion of Nrf1 into the ER membranes.

No proteolytical cleavage occurs within the N-terminal 170 amino acids of Nrf1

To further characterize the effect of the N_{cyt}/C_{lum} orientation of TM1 on its post-insertion processes, we created three chimaeric proteins, in which the NTD of Nrf1, the N-terminal 156 amino acids of Nrf1 (N156), and the N-terminal 170 amino acids of Nrf1 (N170) were attached to the N-terminus of Nrf2, giving NTD/Nrf2, N156/Nrf2 and N170/Nrf2. Western blotting of total cell lysates expressing these chimaeric proteins showed that they gave an apparent single electrophoretic band and its migration in the NuPAGE gel increased following PNGase F digestion (Figure 5). No cleaved polypeptide was found in these chimaeric proteins. These results suggest that the TM1-containing signal peptide of Nrf1 anchored the Nrf2-containing chimaeric proteins to the ER, where their remaining C-terminal portions were translocated into the lumen and glycosylated. We conclude that the NTD, N156 and N170 portions of the fusion proteins are not proteolytically cleaved by a SPase, or other proteases, under normal homeostatic conditions.

The TM1-connecting portions of NTD are repositioned from the ER lumen into the extra-luminal compartments

Our above experiments suggested that during the topogenesis of Nrf1 its TM1 region is inserted with an N_{cyt}/C_{lum} orientation within ER membranes, and its associated SAS, CRACs and NHB2 subdomains are translocated into the lumen of the ER. However, it is not known whether any portion of Nrf1 that lies C-terminal to TM1 is repartitioned into the membrane cytoplasmic side of the ER. To address this question, we performed time-lapse membrane PK protection reactions with the intact ER and nuclear fractions purified from COS-1 cells that had been transfected with an expression construct for either DsRed/NTD/GFP or DsRed/N65/GFP. Immunoblotting using either anti-DsRed or anti-GFP antibodies revealed that the full-length DsRed/NTD/GFP sandwiched fusion protein of ~68 kDa was almost totally digested following incubation of the intact ER fractions with PK for 30 min or 60 min (Figure 6A). By contrast, the nuclear ~68 kDa DsRed/NTD/GFP fusion protein appeared to be partially digested by the proteinase (Figure 6A). A cleaved polypeptide ladder of between 10 kDa and 60 kDa was detected by immunoblotting that reacted strongly with anti-GFP antibodies (Figure 6A, *middle panel*) and weekly with anti-DsRed antibodies (*upper panel*). Two PK-digested bands of ~32 kDa and ~45 kDa cross-reacted with antibodies against GFP (*middle panel*) rather than DsRed (*upper panel*). Based on their molecular masses, the immunoreactive bands appear to represent two small GFP fusion proteins of ~32 kDa and ~45 kDa that contain a short C-terminal segment of the NTD of Nrf1. Further examination showed that, as the proteinase reaction time was extended from 5 min to 60 min, the abundance of the ~32 kDa and ~45 kDa polypeptides gradually decreased to a low level, but they were not completely abolished, even in the presence of TX. A similar time course response for ~32-kDa polypeptide was observed in the nuclear membrane PK digestions. Interestingly, no band of between ~33 kDa and ~44 kDa was observed in the nuclear membranes. These data suggest that, during topogenesis of Nrf1, TM1 is orientated in an N_{cyt}/C_{lum} fashion within membranes and SAS, CRACs and NHB2 are reintegrated and repartitioned from the lumen into the extra-luminal compartments around the ER and nuclear envelope membranes (Figure 6C, *on the left-hand side*).

Similar proteinase protection of ectopic DsRed/N65/GFP against PK digestion revealed that a full-length fusion polypeptide of ~62 kDa was recovered in the ER and nuclear fractions (Figure 6B). Incubation of the ER fraction with PK for 5 min to 30 min resulted in loss of most of the ~62-kDa DsRed/N65/GFP protein. In a 60 min digest, PK almost completely eliminated the ~62-kDa fusion protein in the ER, but not in the nuclear fraction. An apparent PK-digested band of ~32-kDa cross-reacted with antibodies against GFP, but not against DsRed, and its abundance was

modestly increased when the enzyme reaction time was extended. An additional weak PK-digested band of ~38-kDa was also detected by immunoblotting with anti-GFP antibodies (*middle panel*). When compared to the ~28 kDa GFP protein, the relative mobilities of the PK-digested ~32-kDa and ~38-kDa bands are consistent with their identification as two GFP fusion polypeptides that contain the entire N65 amino acids or its C-terminal portion, respectively. These data indicate that, after the N-terminal 65 amino acids of Nrf1 are integrated in the membrane through TM1, its flanking N-terminal portion is partitioned in the cytoplasmic or nucleoplasmic sides, and its adjacent SAS subdomain can be repartitioned from the lumen into the extra-luminal environments closely associated with the membranes (Figure 6C, *on the right-hand side*). Furthermore, these two PK-digested polypeptides of ~32-kDa and ~38-kDa were not completely destroyed after the ER membrane was disrupted by 1% TX, suggesting that the N-terminal 65 amino acid portion of Nrf1 may diffuse from the TX detergent-resistant membrane microdomain (DRM) to the detergent-sensitive membrane regions.

SAS, but not CRACs, is required to repartition Nrf1 within the ER

To identify roles for SAS and CRACs in determining the topology of full-length Nrf1 protein, we transfected COS-1 cells with expression constructs for wild-type Nrf1 and three internal deletion mutants of the CNC-bZIP protein, followed by subcellular fractionation and time-lapse membrane PK protection reactions. Figure 7A shows that, following incubation of ER membranes with PK for 5 to 60 min, the abundance of the ~120-kDa wild-type Nrf1 glycoprotein was reduced by proteolytic digestion from ~60% to ~10% of the level observed following incubation in the absence of PK (see supplemental Figure S5A). The nuclear membrane protease protection reactions with PK caused between ~75% and 85% of the ~120-kDa Nrf1 glycoprotein to be digested. Upon addition of 1% TX into these membrane reactions, all remaining Nrf1 was destroyed by PK; it should be noted that no PK-digested polypeptides of between 6 kDa and 94 kDa were observed by immunoblotting with antibodies against either Nrf1 or its C-terminal V5 tag (also see supplementary Figure S6). By contrast, the ER luminal protein CRT was protected by the membranes from PK digestion (Figure 7A). Together, these data suggest that a large C-terminal portion of Nrf1 is dynamically retrotranslocated from the ER lumen to the cytoplasmic and/or nucleoplasmic sides. In addition, all Nrf1 isoforms of less than 120-kDa were recovered primarily in the nuclear fractions and rapidly destroyed by digestion with PK (Figure 7A), suggesting that these products resided in the extra-luminal compartments and thus are not protected by membranes.

When compared with wild-type Nrf1, the Nrf1$^{\Delta 31\text{-}50}$ mutant protein was preferentially retained in the ER lumen following PK digestion (Figure 7B). In the intact ER membrane, ~50% of Nrf1$^{\Delta 31\text{-}50}$ resisted PK digestion for 15 to 60 min. In the nuclear membrane PK reactions, digestion of Nrf1$^{\Delta 31\text{-}50}$ did not differ significantly from the wild-type protein. Similar results were obtained from the membrane PK protection reactions with the Nrf1$^{\Delta 31\text{-}80}$ mutant protein (supplemental Figure S5A and C). Further experiments showed that PK digestion of the Nrf1$^{\Delta 55\text{-}80}$ mutant, lacking CRACs alone (Figure 7C), gave a similar result to those obtained for the wild-type Nrf1 (supplemental Figure S5A). These results indicate that SAS, but not CRACs, is required for integration of Nrf1 into the ER and its repartitioning from the lumen to the membrane cytoplasmic side. This conclusion was also supported by results obtained from specific anti-Nrf1 antibodies in immunoblots of Nrf1$^{\Delta 31\text{-}50}$ and Nrf1$^{\Delta 55\text{-}80}$ after membrane protection reactions (supplemental Figure S6B).

A portion of the NHB2 resides on the luminal side of the ER membrane

Bioinformatic analyses suggested that the C-terminal 70% portion of NHB2 can fold into an amphipathic β-strand, but upon interaction with the membrane lipids it might also fold into an acidic hydrophobic amphipathic α-helix

structure that either lies on membranes or spans across membranes (Figure 8). To test whether the NHB2 influences the orientation of Nrf1 within the ER, we performed time-lapse membrane PK protection reactions with the ER and nuclear fractions purified from COS-1 cells expressing the Nrf1$^{\Delta 96\text{-}106}$ mutant protein. As shown in Figure 7D, digestion of Nrf1$^{\Delta 96\text{-}106}$ by PK at a final concentration of 100 μg protein/ml resulted in its almost complete elimination from the digest over 60 min; at 5 min 46% of the Nrf1$^{\Delta 96\text{-}106}$ input was observed, but at 60 min only 0.2% remained (also see supplemental Figure S5A). In the nuclear membrane PK reactions, the ~120-kDa Nrf1$^{\Delta 96\text{-}106}$ glycoprotein was not totally destroyed (Figure 7D). When larger amounts of protein digests were analyzed we observed a ~38-kDa proteolysis product of Nrf1$^{\Delta 96\text{-}106}$ (Figure 7D), but it was not detected when lesser amounts were analyzed (left panel in supplemental Figure S6B). These results, together with the data for DsRed/NTD/GFP (Figure 6A), suggest that the C-terminal 70% portion of NHB2 allows Nrf1 to be tethered to the membrane luminal side and/or be reintegrated within the membrane.

In common with Nrf1$^{\Delta 96\text{-}106}$, the abundance of the ~120-kDa Nrf1$^{\Delta 81\text{-}106}$ mutant glycoprotein lacking the entire NHB2 sequence was rapidly digested by PK (supplemental Figure S5A and D). Indeed, when the digestion time was extended from 30 to 60 min, this mutant protein almost completely disappeared. However, a membrane protease protection reaction revealed no obvious difference in the amount of the Nrf1$^{\Delta 81\text{-}106}$ mutant protein and the wild-type protein that remained after digestion (right panel in supplemental Figure S6B). These results suggest that the N-terminal 30% of NHB2 has an innate propensity to be repartitioned into the extra-luminal compartments, with an orientation that is distinctive from that of its C-terminal 70% portion.

DISCUSSION

The function of membrane proteins clearly depends on their dynamic topology and distribution within membranes because this defines an important halfway house between their amino acid sequence and their fully folded three-dimensional structures [47-49]. In the present study, we have examined the NHB1, SAS, CRACs and NHB2 subdomains within the NTD of Nrf1 in order to identify topological elements that might help explain how the CNC-bZIP protein is regulated within membranes.

Mechanisms that control the topology of Nrf1 within membranes

Our results demonstrate that Nrf1 exists in several topologies that enable it to partition into different compartments around membranes ([35] and in the reference therein). We propose that the different topologies of Nrf1 could arise through co-translational and/or post-translational mechanisms. According to our present understanding [50-53], the initial topology of Nrf1 is determined by the amino acid composition of the nascent polypeptide as it is synthesized. This CNC-bZIP protein contains several topogenic signals (e.g. hydrophobic segments and the positive/negative charge distribution of associated residues) and glycosylation recognition motifs (e.g. Asn–X–Ser/Thr) in its primary structure. In ribosome-polypeptide-translocon Sec61 complexes, the ribosome serves as a platform for the co-translational processing of newly-synthesized Nrf1 polypeptides, whereas the translocon Sec61 provides a permissive environment required for the concurrent insertion of hydrophobic regions of Nrf1 into the rough ER membranes whilst simultaneously positioning hydrophilic amino acid flanking regions on either side of the membrane according to the positive-inside and/or charge difference rules. We therefore envisage that the NHB1 signal peptide of Nrf1 directs its TM1 to enter the ER membrane-bound translocon Sec61 complex, whereupon it is inserted into membranes with an N_{cyt}/C_{lum} orientation (Figure 8A). During subsequent topogenesis, this translocon complex will continue to partition the remaining regions of Nrf1 into the lumen or extra-luminal environments around and within the ER membrane.

In addition to the above co-translational topologenesis, Nrf1 may be subject to a number of post-translational rearrangements even whilst the nascent CNC-bZIP protein remains engaged with the ribosome and the translocon Sec61. Our glycosylation mapping mutagenesis study and membrane proteinase protection reactions suggest that Nrf1 is reorganized within the ER membrane to possess intermediate topologies; they probably arise through post-translational modifications (e.g. glycosylation and deglycosylation) and other folding processes within the ER. These post-translational mechanisms enable the extra-TM1 regions in the NTD, SAS, CRACs and NHB2, to be either reinserted into membranes or repositioned from the lumen into the cytoplasm and/or nucleoplasm (Figure 8B and C). It should be noted that this topological reorganization is likely to be directed by the combined effects of protein lipid interactions, the most important of which entails the amphipathic lipid membrane and several amphipathic amino acid α-helical regions within Nrf1. These amphipathic α-helical regions within Nrf1 may lie flat on the plane of membrane lipid biolayers, with its hydrophobic surface situated in the core hydrocarbon interior and its hydrophilic surface positioned on either the cytoplasmic, or in the luminal lipid-water interfaces. Furthermore, membrane lipid compositions may also modulate the topology of Nrf1 through ferrying it in the lipid plane between cholesterol-rich DRM and non-DRM membranes. The lipid composition of membranes may cause Nrf1 to flip-flop across the bilayers as the lipids move between the inner and outer leaflets. In addition, it is possible that the topology of Nrf1 is also modulated by internal interactions of TM1 with other amphipathic α-helical regions within the CNC-bZIP protein. Although this hypothesis requires to be confirmed, we predict that these interactions enable Nrf1 to form a local α-helical hairpin or bundle structure across membranes.

Membrane topology of Nrf1 is determined by its NTD

Nrf1 is an integral polytopic protein with a hydrophobic TM1 region, and several amphipathic α-helical regions that seem to bi-span, mono-span or half-span the membrane lipid bilayer (Figure 8B and C). We have presented evidence suggesting that the membrane topology of Nrf1 is determined by its NTD. Within this domain, the TM1 α-helix represents a dominant membrane-interactive region with a hydrophobic surface and an amphipathic surface (see supplemental Figure S1). The polar amino acids Thr^{9}, Gln^{14}, Glu^{10}, Thr^{16} and Ser^{20} in TM1 may regulate its N_{cyt}/C_{lum} orientation within the membrane and the extent to which it interacts with both the amphipathic membrane lipids and other amphipathic amino acid α-helix regions. The SAS α-helix (residues 31-50) is likely to be integrated into the membranes in cooperation with TM1, but its orientation may also be modulated by a putative Pro^{37}-Pro^{38} kink within this region. We postulate that the CRACs subdomain (residues 62-82) interacts with cholesterol-rich DRM membranes, and thus the activity of Nrf1 may be influenced by the lipid contents of membranes or it may even be involved in lipid sorting and membrane trafficking, as has been described for caveolin-1 [54]. Topological orientation of Nrf1 around the membrane is also controlled by its NHB2 region (residues 82-106) as it could serve as a dynamic switch anchoring the CNC-bZIP protein either on the membrane luminal leaflets or across membranes (Figure 8). In addition, the membrane topology of Nrf1 will also be modulated by other amphipathic subdomains (e.g., TMi and TMc) and by glycosylation of the NST domain [35]. It should be noted that membrane topology of Nrf1 is established within and around the ER before it is transported into the inner nuclear envelope membrane, because we observed significant differences in the time course of digestion of Nrf1 proteins in PK protection reactions with intact ER fractions, but not with entire nuclear membrane fractions (supplemental Figure S5).

The NTD of Nrf1 may confer a role in control of lipid homeostasis

The ability of Nrf1 to induce ARE-driven gene expression is a consequence of its CNC-bZIP DNA-binding domain and its AD1 and AD2 transactivation domains [9, 10, 35]. Gene knockout experiments have shown that Nrf1 is

required for mouse development and growth, as well as the maintenance of liver function [19, 20, 22]. The markedly different phenotypes resulting from knockout of *Nrf1* and *Nrf2* indicate that the former transcription factor fulfills a unique role that the latter cannot perform. Amongst the many possible explanations for the distinct phenotypes, one possibility is that the NTD of Nrf1, which is not represented in Nrf2, endows Nrf1 with a unique function. In hepatocytes from liver-specific *Nrf1*$^{-/-}$ knockout mice, a marked increase in the number of lipid vesicles and proliferation of smooth ER has been reported [15, 22]. By contrast, the livers of *Nrf2*$^{-/-}$ mice show no phenotype under normal laboratory conditions [24]. It has been speculated that the abnormalities in the liver-specific Nrf1 null mice result from enhanced expression of the cytochrome P450 4A enzymes responsible for ω-oxidation of fatty acids [22]. We now hypothesize that the NTD of Nrf1 may be involved in controlling lipid homeostasis and membrane trafficking of the factor, because it shares conservation with the N-terminus of ArfGAP1, a Golgi-associated protein that contains an amphipathic α-helix lipid-packing sensor (ALPS) for membrane curvature [55]. Furthermore, like other membrane proteins [48, 55-57], the amphipathic regions in Nrf1 could be organized to accommodate both polar and apolar microenvironments of membrane lipid bilayers. Once these amphipathic regions are integrated as TM α-helices, they might in turn cause a disordering of membrane lipid organization, primarily by acting as a simple barrier and restricting the diffusion of neighbouring lipids. Thus, we propose that some membrane-interactive amphipathic regions within NTD and/or other domains of Nrf1 may enable the CNC-bZIP protein to play a role in membrane-dependent biological processes.

ACKNOWLEDGMENTS

This work was supported by the Association for International Cancer Research (grants 06-015 and 09-0254). We thank Dr Changjiang Dong (at Centre for Biomolecular Science, University, St. Andrews) for his expert advice about the structural biology of membrane proteins. We thank Professor Ronald T Hay (at College of Life Sciences, University of Dundee) for his help and interest in this work.

REFERENCES

1 Chan, J. Y., Han, X. L. and Kan, Y. W. (1993) Cloning of Nrf1, an NF-E2-related transcription factor, by genetic selection in yeast. Proc. Natl. Acad. Sci. USA **90**, 11371-11375
2 Luna, L., Johnsen, O., Skartlien, A. H., Pedeutour, F., Turc-Carel, C., Prydz, H. and Kolsto, A. B. (1994) Molecular cloning of a putative novel human bZIP transcription factor on chromosome 17q22. Genomics **22**, 553-562
3 Moi, P., Chan, K., Asunis, I., Cao, A. and 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 β-globin locus control region. Proc. Natl. Acad. Sci. USA **91**, 9926-9930
4 Kobayashi, A., Ito, E., Toki, T., Kogame, K., Takahashi, S., Igarashi, K., Hayashi, N. and Yamamoto, M. (1999) Molecular cloning and functional characterization of a new Cap'n' collar family transcription factor Nrf3. J. Biol. Chem. **274**, 6443-6452
5 Venugopal, R. and Jaiswal, A. K. (1996) Nrf1 and Nrf2 positively and c-Fos and Fra1 negatively regulate the human antioxidant response element-mediated expression of NAD(P)H:quinone oxidoreductase1 gene. Proc. Natl. Acad. Sci. USA **93**, 14960-14965
6 Johnsen, O., Murphy, P., Prydz, H. and Kolsto, A. B. (1998) Interaction of the CNC-bZIP factor TCF11/LCR-F1/Nrf1 with MafG: binding-site selection and regulation of transcription. Nucleic Acids Res. **26**, 512-520
7 Nioi, P. and Hayes, J. D. (2004) Contribution of NAD(P)H:quinone oxidoreductase 1 to protection against carcinogenesis, and regulation of its gene by the Nrf2 basic-region leucine zipper and the arylhydrocarbon receptor basic helix-loop-helix transcription factors. Mutat. Res. **555**, 149-171
8 Zhang, Y., Kobayashi, A., Yamamoto, M. and Hayes, J. D. (2009) The Nrf3 transcription factor is a membrane-bound glycoprotein targeted to the endoplasmic reticulum through its N-terminal homology box 1 sequence. J. Biol. Chem. **284**, 3195-3210
9 Chen, L., Kwong, M., Lu, R., Ginzinger, D., Lee, C., Leung, L. and Chan, J. Y. (2003) Nrf1 is critical for redox balance and survival of liver cells during development. Mol. Cell. Biol. **23**, 4673-4686
10 Leung, L., Kwong, M., Hou, S., Lee, C. and Chan, J. Y. (2003) Deficiency of the Nrf1 and Nrf2 transcription factors results in early embryonic lethality and severe oxidative stress. J. Biol. Chem. **278**, 48021-48029
11 Yang, H., Magilnick, N., Lee, C., Kalmaz, D., Ou, X., Chan, J. Y. and Lu, S. C. (2005) Nrf1 and Nrf2 regulate rat glutamate-cysteine ligase catalytic subunit transcription indirectly via NF-κB and AP-1. Mol. Cell. Biol. **25**, 5933-5946
12 McMahon, M., Itoh, K., Yamamoto, M., Chanas, S. A., Henderson, C. J., McLellan, L. I., Wolf, C. R., Cavin, C. and Hayes, J. D. (2001) The Cap 'n' Collar basic leucine zipper transcription factor Nrf2 (NF-E2 p45-related factor 2) controls both constitutive and inducible expression of intestinal detoxification and glutathione biosynthetic enzymes. Cancer Res. **61**, 3299-3307

13 Chanas, S. A., Jiang, Q., McMahon, M., McWalter, G. K., McLellan, L. I., Elcombe, C. R., Henderson, C. J., Wolf, C. R., Moffat, G. J., Itoh, K., Yamamoto, M. and Hayes, J. D. (2002) Loss of the Nrf2 transcription factor causes a marked reduction in constitutive and inducible expression of the glutathione S-transferase *Gsta1*, *Gsta2*, *Gstm1*, *Gstm2*, *Gstm3* and *Gstm4* genes in the livers of male and female mice. Biochem. J. **365**, 405-416
14 Kensler, T. W., Wakabayashi, N. and Biswal, S. (2007) Cell survival responses to environmental stresses via the Keap1-Nrf2-ARE pathway. Annu. Rev. Pharmacol. Toxicol. **47**, 89-116
15 Ohtsuji, M., Katsuoka, F., Kobayashi, A., Aburatani, H., Hayes, J. D. and Yamamoto, M. (2008) Nrf1 and Nrf2 play distinct roles in activation of antioxidant response element-dependent genes. J. Biol. Chem. **283**, 33554-33562
16 Murphy, P. and Kolsto, A. (2000) Expression of the bZIP transcription factor TCF11 and its potential dimerization partners during development. Mech. Dev. **97**, 141-148
17 Kobayashi, A., Ohta, T. and Yamamoto, M. (2004) Unique function of the Nrf2-Keap1 pathway in the inducible expression of antioxidant and detoxifying enzymes. Methods Enzymol. **378**, 273-286
18 Etchevers, H. C. (2005) The cap 'n' collar family member NF-E2-related factor 3 (Nrf3) is expressed in mesodermal derivatives of the avian embryo. Int. J. Dev. Biol. **49**, 363-367
19 Farmer, S. C., Sun, C. W., Winnier, G. E., Hogan, B. L. and Townes, T. M. (1997) The bZIP transcription factor LCR-F1 is essential for mesoderm formation in mouse development. Genes Dev. **11**, 786-798
20 Chan, J. Y., Kwong, M., Lu, R., Chang, J., Wang, B., Yen, T. S. and Kan, Y. W. (1998) Targeted disruption of the ubiquitous CNC-bZIP transcription factor, Nrf-1, results in anemia and embryonic lethality in mice. EMBO J. **17**, 1779-1787
21 Kwong, M., Kan, Y. W. and Chan, J. Y. (1999) The CNC basic leucine zipper factor, Nrf1, is essential for cell survival in response to oxidative stress-inducing agents. Role for Nrf1 in *γ-gcs(l)* and *gss* expression in mouse fibroblasts. J. Biol. Chem. **274**, 37491-37498
22 Xu, Z., Chen, L., Leung, L., Yen, T. S., Lee, C. and Chan, J. Y. (2005) Liver-specific inactivation of the Nrf1 gene in adult mouse leads to nonalcoholic steatohepatitis and hepatic neoplasia. Proc. Natl. Acad. Sci. USA **102**, 4120-4125
23 Chan, K., Lu, R., Chang, J. C. and Kan, Y. W. (1996) NRF2, a member of the NFE2 family of transcription factors, is not essential for murine erythropoiesis, growth, and development. Proc. Natl. Acad. Sci. USA **93**, 13943-13948
24 Chan, K. and Kan, Y. W. (1999) Nrf2 is essential for protection against acute pulmonary injury in mice. Proc. Natl. Acad. Sci. USA **96**, 12731-12736
25 Li, J., Stein, T. D. and Johnson, J. A. (2004) Genetic dissection of systemic autoimmune disease in Nrf2-deficient mice. Physiol. Genomics **18**, 261-272
26 Rangasamy, T., Guo, J., Mitzner, W. A., Roman, J., Singh, A., Fryer, A. D., Yamamoto, M., Kensler, T. W., Tuder, R. M., Georas, S. N. and Biswal, S. (2005) Disruption of Nrf2 enhances susceptibility to severe airway inflammation and asthma in mice. J. Exp. Med. **202**, 47-59
27 Ramos-Gomez, M., Kwak, M. K., Dolan, P. M., Itoh, K., Yamamoto, M., Talalay, P. and 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. USA **98**, 3410-3415
28 Chowdhry, S., Nazmy, M. H., Meakin, P. J., Dinkova-Kostova, A. T., Walsh, S. V., Tsujita, T., Dillon, J. F., Ashford, M. L. and Hayes, J. D. (2010) Loss of Nrf2 markedly exacerbates nonalcoholic steatohepatitis. Free Radic Biol. Med. **48**:357-371
29 Derjuga, A., Gourley, T. S., Holm, T. M., Heng, H. H., Shivdasani, R. A., Ahmed, R., Andrews, N. C. and Blank, V. (2004) Complexity of CNC transcription factors as revealed by gene targeting of the Nrf3 locus. Mol. Cell Biol. **24**, 3286-3294
30 Luna, L., Skammelsrud, N., Johnsen, O., Abel, K. J., Weber, B. L., Prydz, H. and Kolsto, A. B. (1995) Structural organization and mapping of the human TCF11 gene. Genomics **27**, 237-244
31 Caterina, J. J., Donze, D., Sun, C. W., Ciavatta, D. J. and Townes, T. M. (1994) Cloning and functional characterization of LCR-F1: a bZIP transcription factor that activates erythroid-specific, human globin gene expression. Nucleic Acids Res. **22**, 2383-2391
32 Husberg, C., Murphy, P., Martin, E. and Kolsto, A. B. (2001) Two domains of the human bZIP transcription factor TCF11 are necessary for transactivation. J. Biol. Chem. **276**, 17641-17652
33 Zhang, Y., Lucocq, J. M., Yamamoto, M. and Hayes, J. D. (2007) The NHB1 (N-terminal homology box 1) sequence in transcription factor Nrf1 is required to anchor it to the endoplasmic reticulum and also to enable its asparagine-glycosylation. Biochem. J. **408**, 161-172
34 Zhang, Y. (2009) Molecular and cellular control of the Nrf1 transcription factor: An integral membrane glycoprotein. Vdm Verlag Dr. Müller Publishing House Germany**.** The first edition, pp1-264
35 Zhang, Y., Lucocq, J. M. and Hayes, J. D. (2009) The Nrf1 CNC/bZIP protein is a nuclear envelope-bound transcription factor that is activated by t-butyl hydroquinone but not by endoplasmic reticulum stressors. Biochem. J. **418**, 293-310
36 Zhang, Y., Crouch, D. H., Yamamoto, M. and Hayes, J. D. (2006) Negative regulation of the Nrf1 transcription factor by its N-terminal domain is independent of Keap1: Nrf1, but not Nrf2, is targeted to the endoplasmic reticulum. Biochem. J. **399**, 373-385
37 Zhang, Y., Cho, Y. Y., Petersen, B. L., Bode, A. M., Zhu, F. and Dong, Z. (2003) Ataxia telangiectasia mutated proteins, MAPKs, and RSK2 are involved in the phosphorylation of STAT3. J. Biol. Chem. **278**, 12650-12659
38 Wang, X. J., Hayes, J. D. and Wolf, C. R. (2006) Generation of a stable antioxidant response element-driven reporter gene cell line and its use to show redox-dependent activation of nrf2 by cancer chemotherapeutic agents. Cancer Res. **66**, 10983-10994
39 van Geest, M. and Lolkema, J. S. (2000) Membrane topology and insertion of membrane proteins: search for topogenic signals. Microbiol. Mol. Biol. Rev. **64**, 13-33
40 Zhang, Y., Zhong, S., Dong, Z., Chen, N., Bode, A. M., Ma, W. and Dong, Z. (2001) UVA induces Ser381 phosphorylation of p90RSK/MAPKAP-K1 via ERK and JNK pathways. J. Biol. Chem. **276**, 14572-14580
41 Zhang, Y., Mattjus, P., Schmid, P. C., Dong, Z., Zhong, S., Ma, W. Y., Brown, R. E., Bode, A. M., Schmid, H. H. and Dong, Z. (2001) Involvement of the acid sphingomyelinase pathway in UVA-induced apoptosis. J. Biol. Chem. **276**, 11775-11782
42 Millen, J. I., Pierson, J., Kvam, E., Olsen, L. J. and Goldfarb, D. S. (2008) The luminal N-terminus of yeast Nvj1 is an inner nuclear membrane anchor. Traffic **9**, 1653-1664
43 Kahn, R. A. (2009) Toward a model for Arf GTPases as regulators of traffic at the Golgi. FEBS Lett **583**, 3872-3879

44 Inoue, H. and Randazzo, P. A. (2007) Arf GAPs and their interacting proteins. Traffic **8**, 1465-1475
45 Nie, Z. and Randazzo, P. A. (2006) Arf GAPs and membrane traffic. J. Cell Sci. **119**, 1203-1211
46 Paetzel, M., Karla, A., Strynadka, N. C. and Dalbey, R. E. (2002) Signal peptidases. Chem. Rev. **102**, 4549-4580
47 Ott, C. M. and Lingappa, V. R. (2002) Integral membrane protein biosynthesis: why topology is hard to predict. J. Cell Sci. **115**, 2003-2009
48 McMahon, H. T. and Gallop, J. L. (2005) Membrane curvature and mechanisms of dynamic cell membrane remodelling. Nature **438**, 590-596
49 Bowie, J. U. (2005) Solving the membrane protein folding problem. Nature **438**, 581-589
50 von Heijne, G. (2006) Membrane-protein topology. Nat. Rev. Mol. Cell Biol. **7**, 909-918
51 Dowhan, W. and Bogdanov, M. (2009) Lipid-dependent membrane protein topogenesis. Annu. Rev. Biochem. **78**, 515-540
52 Johnson, A. E. (2009) The structural and functional coupling of two molecular machines, the ribosome and the translocon. J. Cell Biol. **185**, 765-767
53 Skach, W. R. (2009) Cellular mechanisms of membrane protein folding. Nat. Struct. Mol. Biol. **16**, 606-612
54 Epand, R. M., Sayer, B. G. and Epand, R. F. (2005) Caveolin scaffolding region and cholesterol-rich domains in membranes. J. Mol. Biol. **345**, 339-350
55 Drin, G., Casella, J. F., Gautier, R., Boehmer, T., Schwartz, T. U. and Antonny, B. (2007) A general amphipathic alpha-helical motif for sensing membrane curvature. Nat. Struct. Mol. Biol. **14**, 138-146
56 Phoenix, D. A., Harris, F., Daman, O. A. and Wallace, J. (2002) The prediction of amphiphilic alpha-helices. Curr. Protein Pept. Sci. **3**, 201-221
57 Mesmin, B., Drin, G., Levi, S., Rawet, M., Cassel, D., Bigay, J. and Antonny, B. (2007) Two lipid-packing sensor motifs contribute to the sensitivity of ArfGAP1 to membrane curvature. Biochemistry **46**, 1779-1790

FOOTNOTES

[§]Correspondence should be addressed to Yiguo Zhang (email y.z.zhang@dundee.ac.uk).

Abbreviations used: AD, acidic domain; ALPS, amphipathic α-helix lipid-packing sensor; ARE, antioxidant response element; ArfGAP1, ADP ribosylation factor (Arf)-GTPase-activating protein 1; bZIP, basic-region leucine zipper; CNC, cap 'n' collar; CRACs, cholestrol recognition amino acid consensus sequences; CRT, calreticulin; DRM, Triton X-100 detergent-resistant membrane microdomain; DsRed, *Discosoma sp*. red fluorescent protein; Endo, endoglycosidase; ER, endoplasmic reticulum; eN, an engineered asparagine introduced into Nrf1; β-gal, β-galactosidase; GFP, green fluorescent protein; GST, glutathione S-transferase; LDS, lithium dodecyl sulfate; N65, N156 and N170, N-terminal 65, 156 and170 amino acid residues of Nrf1, respectively; N_{cyt}/C_{lum}, the cytoplasmic N-terminus and the luminal C-terminus; Neh3L, Neh5L and Neh6L, Nrf2-ECH homology (Neh) 3-, 5- and 6-like domain; NHB, N-terminal homology box; Nrf, nuclear factor-erythroid 2 p45 subunit-related factor; NST, Asn/Ser/Thr-rich domain; NTD, N-terminal domain; NuPAGE, Novex® PAGE; Nvj1, nucleus-vacuole junction protein 1; PK, proteinase K; PNGase, peptide N-glycosidase; SAS, signal peptide-associated sequence; SPase, signal peptidase; SR, serine repeats; TM, transmembrane; TX, Triton X-100; and WN, whole nucleus.

FIGURE LEGENDS

Figure 1 Topological elements within the NTD of Nrf1

(**A**) The upper cartoon shows a schematic representation of topological elements contained within the NTD. The lower panel shows nine structural domains in mouse Nrf1 protein. The abbreviations employed are explained in the text and the footnotes. (**B**) A multiple alignment of amino acids 31–53 of Nrf1 showing similarity with amino acids 1–22 of Nvj1, along with other membrane-bound proteins including ALG8 (dolichyl pyrophosphate GlcMan9GlcNAc2 Alpha-1,3-glucosyltransferase 8), ATC1 (calcium-transporting ATPase sarcoplasmic/endoplasmic reticulum type 1), ATRN (attractin), ENTP2 (ectonucleoside triphosphate diphosphohydro-lase 2), GFRA3 (GDNF family receptor alpha-3), KLK5 (kallikrein-5), OR6K3 (olfactory receptor 6K3), PGP (phosphoglycolate phosphatase) and Y4312 (LRR receptor-like serine/threonine-protein kinase At4g31250). Identical amino acid residues are presented as white letters on a black background, whereas amino acids with similar physico-chemical properties are placed on a grey background. Amongst them, those amino acids on black and grey backgrounds comprise a conserved membrane-associated SAS motif. (**C**) A multiple alignment of amino acids 62–82 of Nrf1 showing similarity with membrane lipid-associated proteins Cav1 (caveolin 1), Cav2 (caveolin 2), Cav3 (caveolin 3), Plin (perilipin), ADRP (adipose differentiation-related protein), Tip47 (tail-interacting protein of 47 kDa, also called perilipin-3), Pat1 (perilipin, ADRP and Tip47-related protein 1), Scd1 (stearoyl-coenzyme A desaturase 1), and Scp2 (sterol carrier protein 2 in liver). Some non-conserved amino acids in these proteins have been omitted from the comparison and these are represented by the dotted lines in the sequence alignment. (**D**) An alignment of amino acids 74–124 of Nrf1 is shown with amino acids 1–49 of ArfGAP1.

Figure 2 Consequences of deletion of topological elements within the NTD of Nrf1

(**A**) Diagrammatic representation of the topological elements within NTD and their deletion mutants. (**B**) In the left-hand panel,

an expression construct (1.2 μg of DNA) for wild-type Nrf1 or its indicated mutants, together with 0.6 μg of P_{SV40}*GSTA2*-6×ARE-Luc and 0.2 μg of the β-gal reporter plasmids, were cotransfected into COS-1 cells over 6 h, and thereafter the cells were transferred to fresh complete media for an additional 24 h. Subsequently, cell lysates were prepared and luciferase reporter activity was measured. The data were calculated as a fold change (mean ± S.D.) of the ARE-driven luciferase gene activity obtained following transfection with each of the Nrf1 expression constructs, when compared with reporter gene activity obtained following transfection of an empty pcDNA3.1/V5 His B vector; in all cases the background luciferase activity observed following transfection of an empty pGL3-promoter reporter gene plasmid and an empty pcDNA3.1/V5 His B vector was subtracted after it was first normalized for transfection efficiency on the basis of β-gal activity. The results presented are typical of those obtained from at least three independent experiments, each of which was performed in triplicate. The significance of the differences was determined by the unpaired *t*-test and is shown as a *P*-value (*p < 0.05, **p < 0.01, and ***p < 0.001) when compared with wild-type Nrf1 activity. In the right-hand panel, each of the above expression constructs (1.2 μg), along with 0.2 μg of the β-gal reporter plasmid, was cotransfected into COS-1 cells. Approximately 24 h after transfection, the cells were harvested in RIPA buffer and β-gal activity was measured. Following denaturation, the total cell lysates were (+), or were not (-), incubated for 1 h with 500 units of PNGase F, before they were resolved using 4-12% LDS/NuPAGE, in Tris-Bis running buffer, and immunoblotted with antibodies against the V5 epitope. The amount of protein applied to each polyacrymide sample well was adjusted to ensure equal loading of β-gal activity. The arrows indicate the position of Nrf1 isoforms estimated to be 120, 95 and 85 kDa. (**C**) COS-1 cells were cotransfected with 1.2 μg of an expression construct for wild-type Nrf1 or mutant (as indicated), along with 0.6 μg of an ER/DsRed expression construct. After the cells had been left to recover from transfection for 24 h, the subcellular location of Nrf1 and ER/DsRed proteins was examined by immuocytochemistry, followed by confocal imaging. FITC-labelled second antibody was used to locate V5-tagged Nrf1 protein. Nuclear DNA was stained by DAPI (4', 6-diamiduno-2-phenylindole) (blue). The ER/DsRed protein gave a red image in the ER (as a positive control). The merge signal represents the results obtained when the three images were superimposed. On some occasions, the intensity of staining was analyzed by the Profile program (see supplemental Figures S2 and S3). Bars = 20 μm.

Figure 3 The SAS, CRACs and NHB2 subdomains are transiently translocated into the ER lumen

(**A**) The upper cartoon shows the structural subdomains in the NTD of Nrf1$^{1\text{-}7xN/Q}$ mutant, in which seven endogenous glycosylation consensus sites in Nrf1 had been removed by mutation of the relevant asparagines into glutamines [35]. The lower amino acid sequence shows the positions of eight new glycosylation sites that were introduced into the NTD of Nrf1$^{1\text{-}7xN/Q}$ protein. The new engineered asparagine (eN) glycosylation sites are represented as white letters on a black background. (**B**) The constructs for Nrf1 and its asparagine mutants were cotransfected, along with the β-gal reporter plasmid, into COS-1 cells. Approximately 24 h after transfection, cell lysates (30 μg of protein) were prepared and subjected to *in vitro* deglycosylation using 500 units of PNGase F for 1.5 h at 37°C. The resulting samples were resolved using LDS/NuPAGE containing 4-12% polyacrylamide in Tris-Bis buffer, followed by western blotting. (**C**) The above protein samples were also subjected to a deglycosylation reaction with 500 units of Endo H for 1.5 h at 37°C. The digests were resolved in a 7% LDS/NuPAGE Tris-Acetate gel and Nrf1 protein identified by western blotting. (**D**) COS-1 cells were cotransfected with each of the above expression constructs together with P_{SV40}*GSTA2*-6×ARE-Luc and β-gal reporter plasmids, before luciferase reporter assays were carried out 24 h later as described in Figure 2.

Figure 4 The TM1 is inserted with an N_{cyt}/C_{lum} orientation within the ER membranes

(**A**) Shows amino acids 1–50 of wild-type Nrf1 and the point mutations that were introduced to create potential GAA and ASA cleavage sites for SPase. (**B**) Total lysates prepared from COS-1 cells that had been transfected with an expression construct for Nrf1, Nrf1$^{L12/13A}$ or Nrf1$^{L19/21A}$ were subject to *in vitro* PNGase F deglycosylation reactions, followed by western blotting with anti-V5 antibodies. (**C**) Total lysates from COS-1 cells that had been transfected with an expression construct for NTD/GFP×2, N65/GFP×2 or their point mutants were resolved by 7% LDS/NuPAGE using the Tris-Acetate running gel buffer system, before being visualized by immunoblotting with anti-GFP antibodies. The arrows indicate a proteolytically cleaved band with an electrophoretic mobility faster than its equivalent full-length protein. (**D**) An expression construct (1.2 μg of DNA) for Nrf1, Nrf1$^{L12/13A}$ or Nrf1$^{L19/21A}$, together with 0.6 μg of P_{TK} *nqo1*-ARE-Luc and 0.2 μg of the β-gal reporter plasmids, was transfected

into COS-1 cells over 6 h, and then the cells were transferred to the fresh complete media for an additional 24 h. Subsequently, luciferase reporter activity was measured. The data were calculated as a fold change (mean ± S.D) of the activity observed following transfection with the reporter plasmid and an empty pcDNA3.1/V5 His B expression vector. The results presented are typical of those obtained from at least three independent experiments, each of which was performed in triplicate. The significance of the differences was determined by the unpaired *t*-test and is shown as a *P*-value. (**E**) COS-1 cells were transfected with 1.2 μg of each of the above expression constructs for Nrf1 and its mutants, along with 0.6 μg of an ER/DsRed construct. Approximately 24 h following transfection, the subcellular location of Nrf1 was examined by immuocytochemistry, followed by confocal imaging as described in Figure 2B, with the exception that nuclear DNA was stained blue by DAPI. Bars indicate 20 μm.

Figure 5 The N-terminal 170 amino acid region of Nrf1 does not contain a cleavage site

Total lysates of COS-1 cells expressing wild-type Nrf2 or chimaeric Nrf2 fused N-terminally to the NTD, the N156 or the N170 of Nrf1 were prepared and subjected to deglycosylation using PNGase F. The digests were identified by immunoblotting. The resulting nitrocellulose membrane was exposed to X-ray film for 30 sec (upper panel) and 5 sec (lower panel).

Figure 6 The extra-TM1 regions of NTD is reintegrated from the lumen into the extra-luminal environment

COS-1 cells were transfected with an expression construct for either DsRed/NTD/GFP (**A**) or DsRed/N65/GFP (**B**), two sandwiched fusion proteins containing either the NTD or the N65 flanked N-terminally by DsRed and C-terminally by GFP. The intact ER and whole nuclei (WN) fractions were purified and equal amounts (50 μg of protein in 40 μl of 1× isotonic buffer) were examined. Membrane proteinase protection reactions were carried out by incubating (on ice) either ER or WN fractions with 100 μg/ml (+) of PK in the presence (+) or absence (−) of 1% TX for 5, 15, 30 and 60 min. The reaction products were resolved in a 4-12% LDS/NuPAGE Tris-Bis gel and identified by western blotting with rabbit polyclonal antibodies against DsRed (upper panel) or GFP (middle panel). The DsRed antibody-blotted nitrocellulose membrane was stripped and re-probed with antibodies against GFP and CRT. (**C**) The cartoon shows two proposed topologies of DsRed/NTD/GFP and DsRed/N65/GFP. The scissors indicate potential PK cleavage sites that are not protected by membranes.

Figure 7 Repartitioning of the NTD of Nrf1 is modulated by SAS, CRACs and NHB2 subdomains

COS-1 cells were grown in a 100-mm dish and transfected with 6 μg of an expression construct for wild-type Nrf1 (**A**), Nrf1$^{\Delta 31\text{-}50}$ (**B**), Nrf1$^{\Delta 55\text{-}80}$ (**C**), or Nrf1$^{\Delta 96\text{-}106}$ (**D**). Approximately 24 h following transfection, the cells were subjected to subcellular fractionation, followed immediately by membrane proteinase protection reactions that were carried out for 5, 15, 30, or 60 min in 20 μl of 1× isotonic buffer containing 100 μg/ml (+) of PK plus 1% TX (+) or not (−), together with 100 μg of proteins from the intact ER or WN fractions. The reaction products (10 μg of proteins diluted in 10 μl of 1× loading buffer) were resolved in a 4-12% LDS/NuPAGE Tris-Bis gel and Nrf1 protein identified by western blotting with anti-V5 antibodies. The antibody-blotted nitrocellulose membrane was stripped and re-probed with antibodies against CRT. The intensity of these Western blots was calculated using the Quantity One® software. The relative amount of Nrf1 after PK digestion was estimated by dividing its immunoreactive band intensity with that obtained for CRT, and then the resulting values were normalized to the total non-digested amounts and shown as percentages (%) at the bottom. In addition, these results are shown graphically in the supplementary Figure S5.

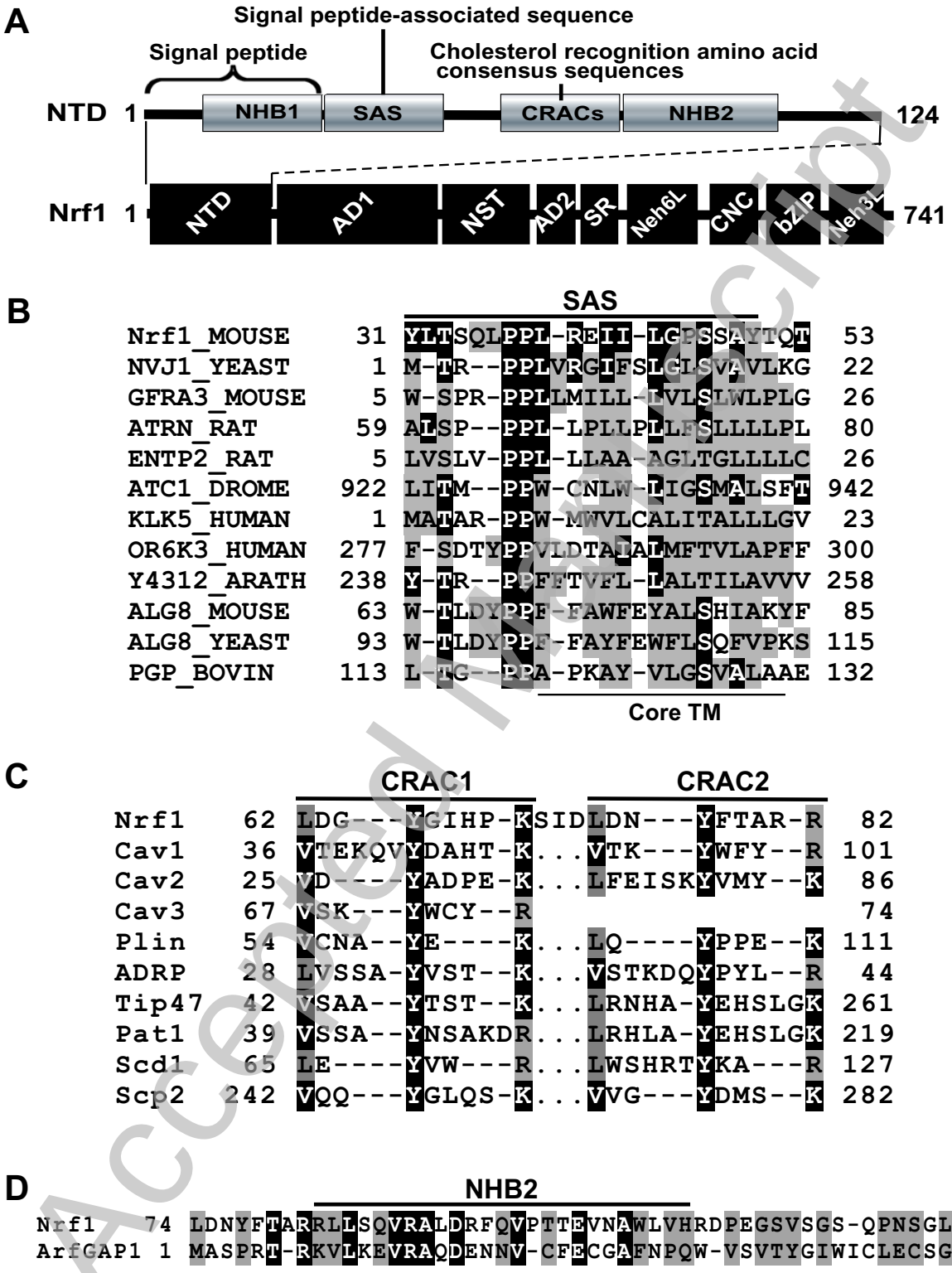
Figure 8 A model to describe a role for the NTD in the topology of Nrf1

After Nrf1 is targeted to the ER through its non-cleaved NHB1 signal peptide, TM1 is inserted into the membrane with an N_{cyt}/C_{lum} orientation. The remaining regions (e.g., SAS, CRACs and NHB2) of the NTD continue to be translocated transiently into the lumen (**A**). These amphipathic regions are subsequently reintegrated from the lumen into the extra-luminal environment around and within the ER membrane. Upon interaction with membrane lipids, the NTD can fold into amphipathic α-helix and/or β-strand structures with distinctive physico-chemical properties that allows them to lie on the plane of membranes or span across the membranes (**B** and **C**). The SAS is predicted to fold into a kinked amphipathic α-helix; it can be reinserted into the membrane in cooperation with TM1, because the latter has an amphipathic surface and is closely connected with SAS through a short six amino acid loop. Although no direct evidence has been provided that CRACs bind to membrane cholesterol, it is postulated that the net basic charged sequence could be essential for the localization of Nrf1 in lipid-ordered DRM membranes

and for the repartitioning of the NTD into the cytoplasmic and/or nucleoplasmic sides of membranes. This repositioning may also require a basic-rich N-terminal portion of NHB2, whilst its C-terminal 70% portion could fold into amphipathic α-helix and/or β-strand structures that are dependent on local membrane microdomains. We propose that NHB2 may serve as an anchor switch that places the NTD in the lumen of membranes or across membranes.

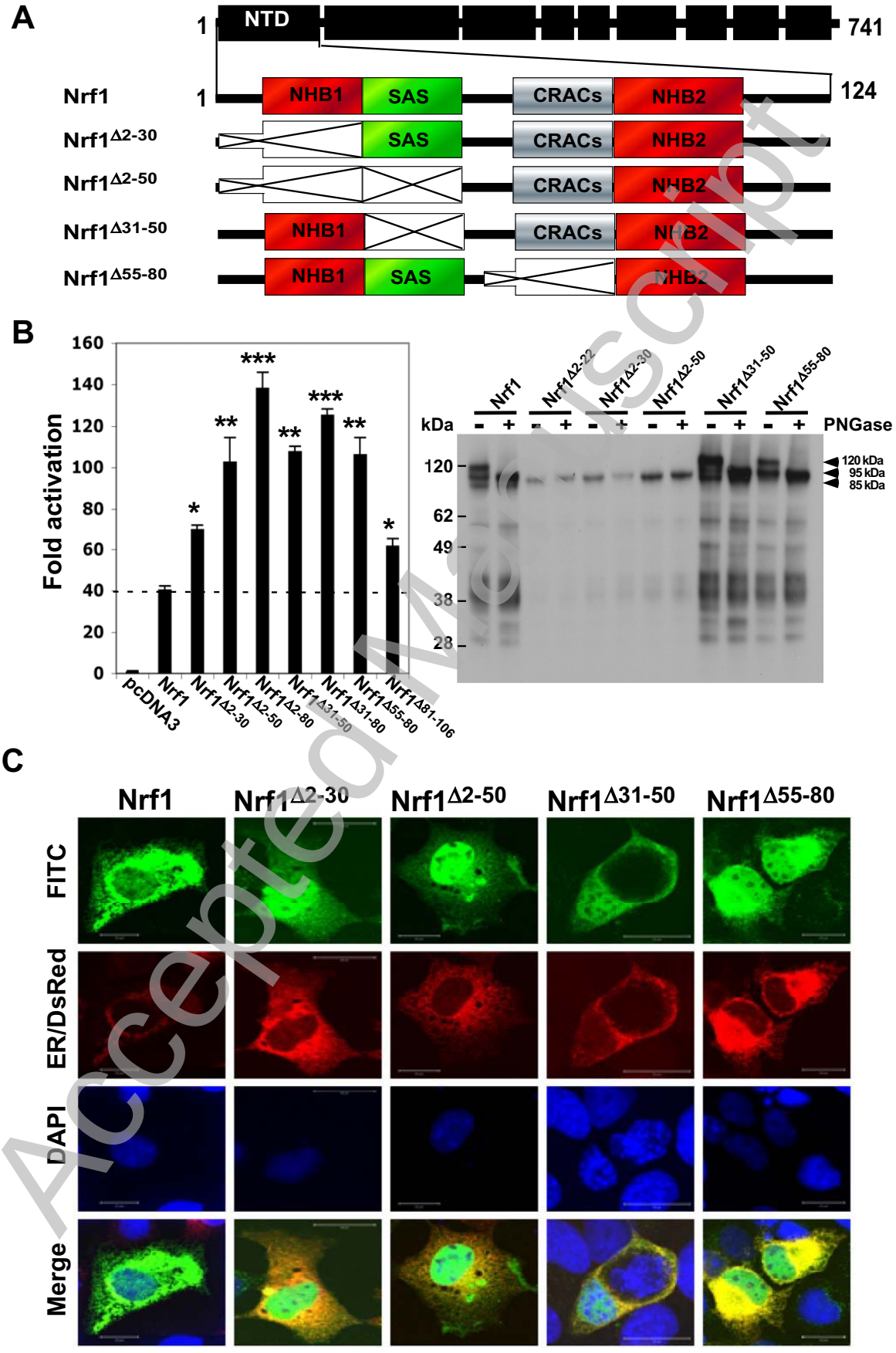
Biochemical Journal Immediate Publication. Published on 14 Jul 2010 as manuscript BJ20100471

Fig 1



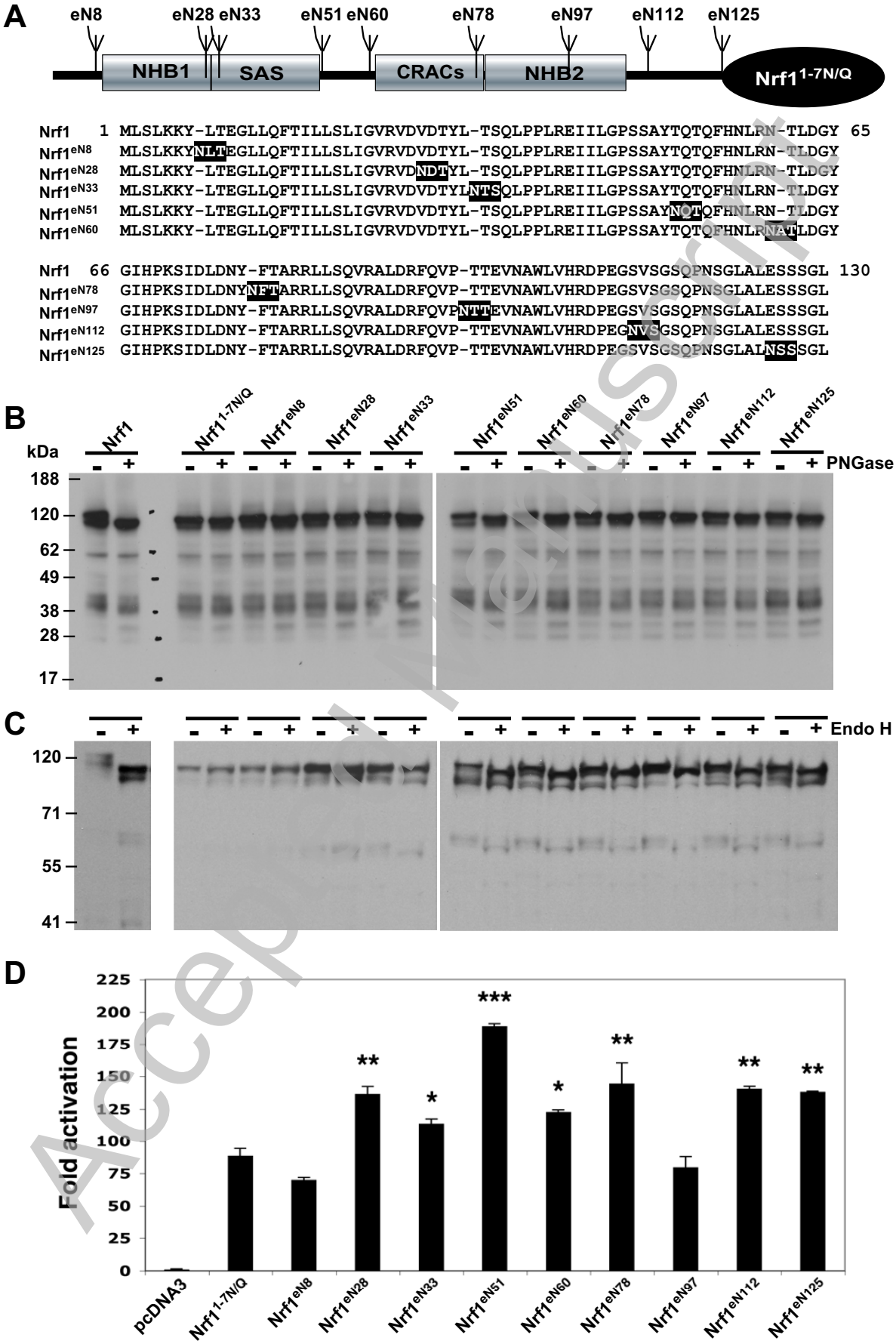
Biochemical Journal Immediate Publication. Published on 14 Jul 2010 as manuscript BJ20100471

Fig 2



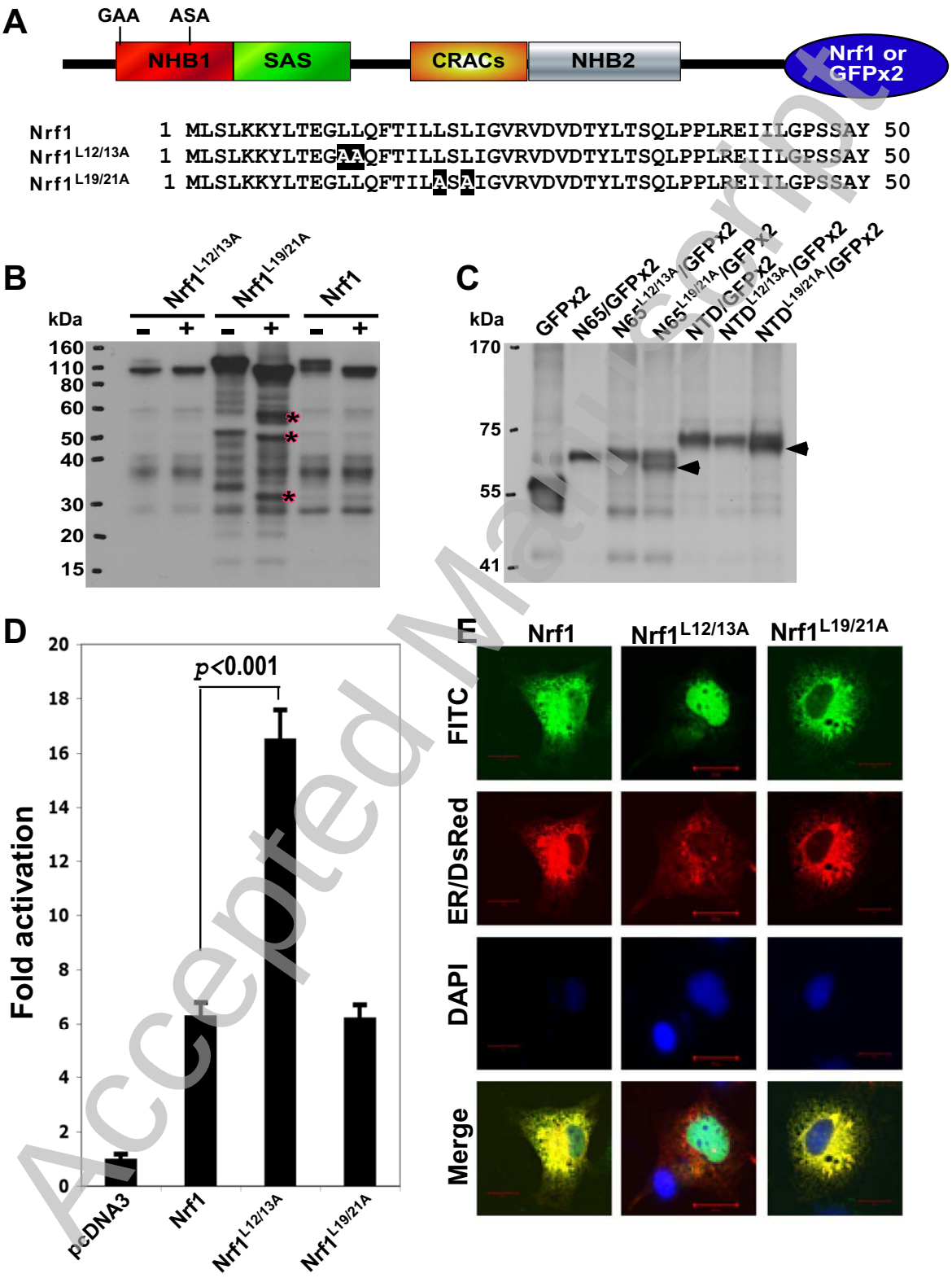
Biochemical Journal Immediate Publication. Published on 14 Jul 2010 as manuscript BJ20100471

Fig 3



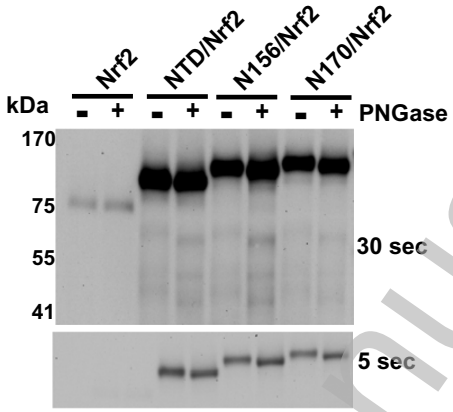
Biochemical Journal Immediate Publication. Published on 14 Jul 2010 as manuscript BJ20100471

Fig 4



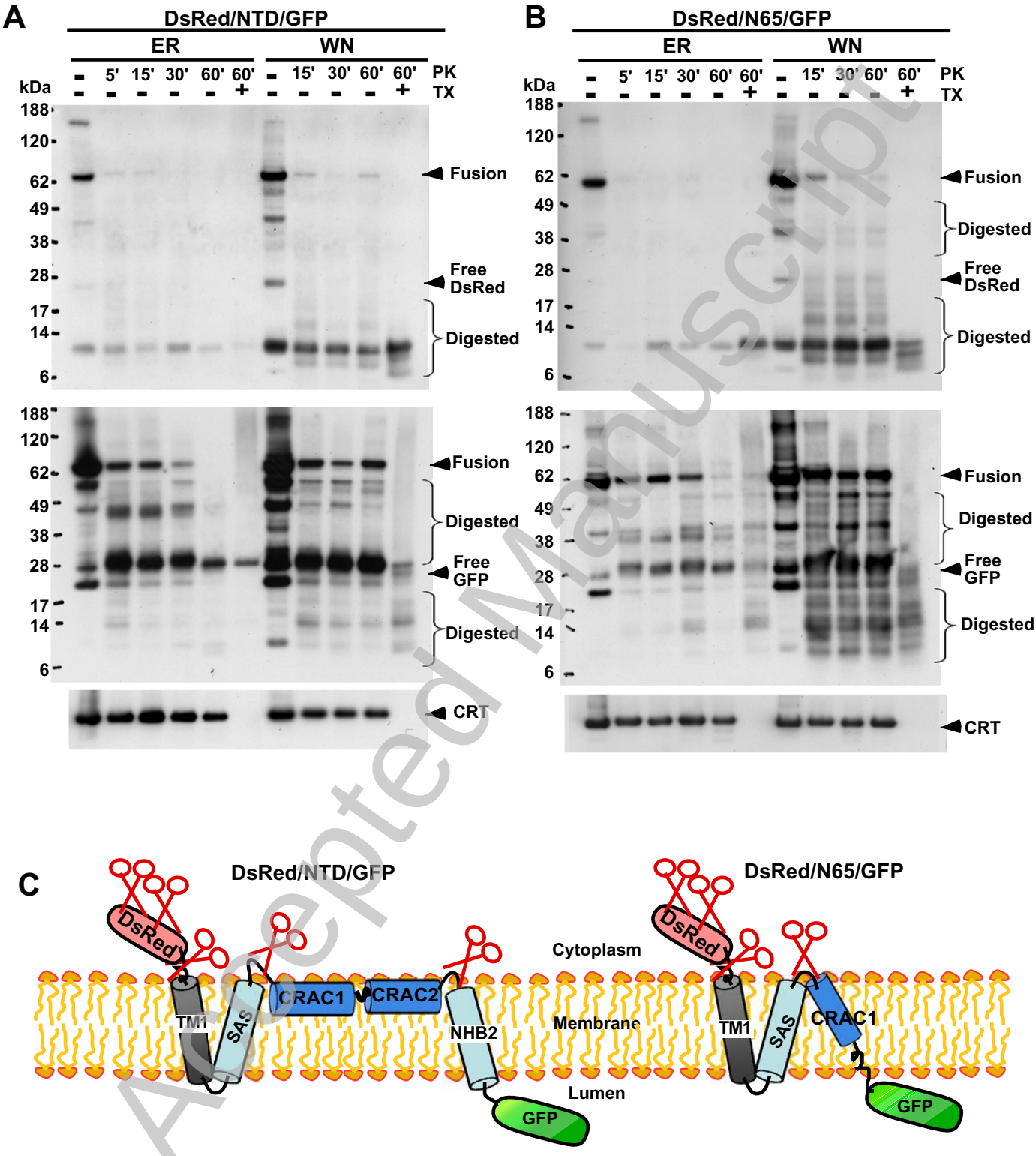
Biochemical Journal Immediate Publication. Published on 14 Jul 2010 as manuscript BJ20100471

Fig 5



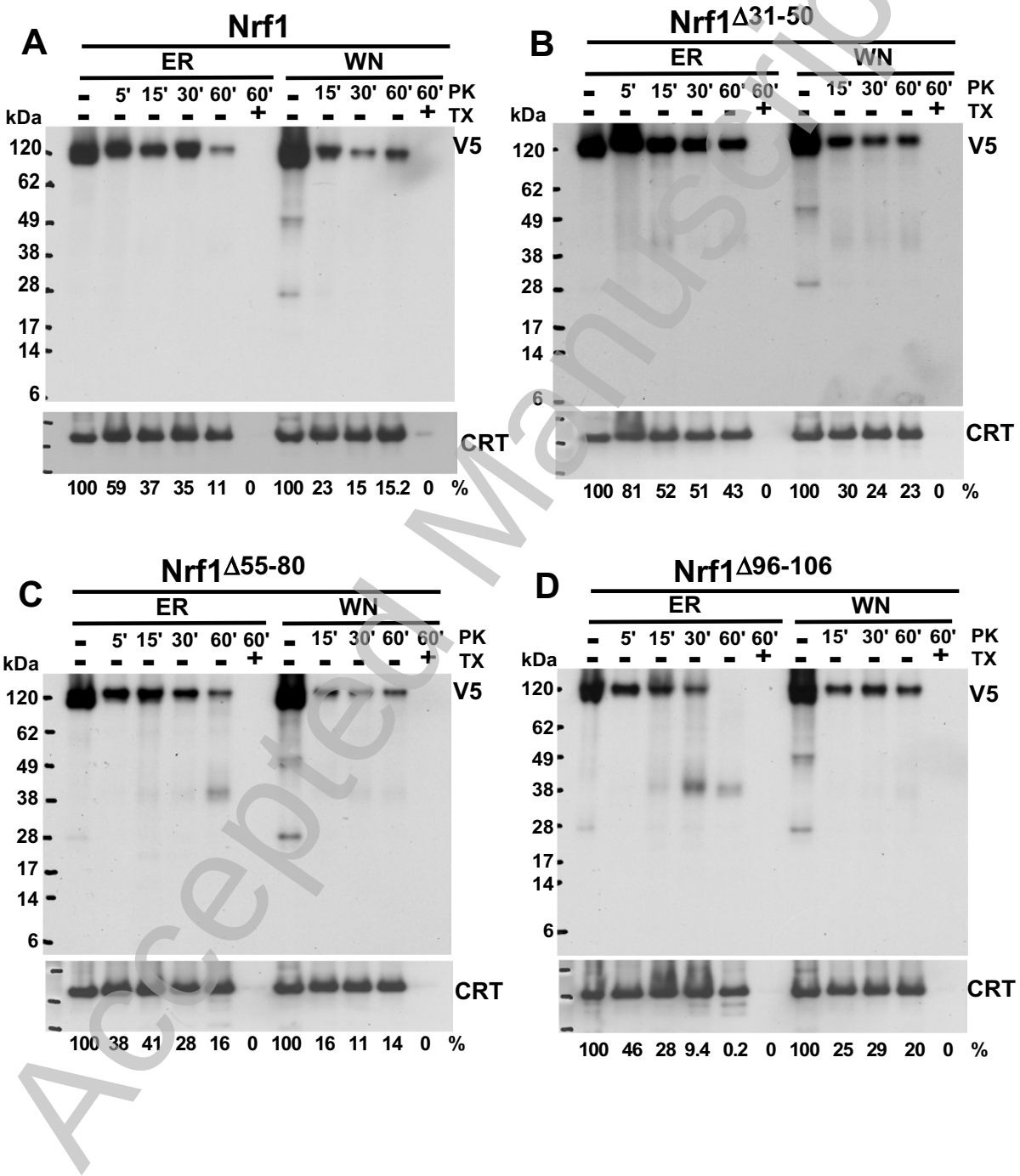
Biochemical Journal Immediate Publication. Published on 14 Jul 2010 as manuscript BJ20100471

Fig 6



Biochemical Journal Immediate Publication. Published on 14 Jul 2010 as manuscript BJ20100471

Fig 7



Biochemical Journal Immediate Publication. Published on 14 Jul 2010 as manuscript BJ20100471

Fig 8