

Structural Basis of Oncogenesis by Mutants of Calreticulin

Stefan N. Constantinescu (✉ stefan.constantinescu@bru.licr.org)

Ludwig Institute for Cancer Research, de Duve Institute, Université catholique de Louvain, WELBIO
<https://orcid.org/0000-0002-8599-2699>

Nicolas Papadopoulos

Ludwig Institute for Cancer Research, de Duve Institute, Université catholique de Louvain,

Audrey Nédélec

Ludwig Institute for Cancer Research, de Duve Institute, Université catholique de Louvain,

Allison Derenne

Université Libre de Bruxelles

Teodor Asvadur Şulea

Institute of Biochemistry of the Romanian Academy

Christian Pecquet

Ludwig Institute for Cancer Research, de Duve Institute, Université catholique de Louvain,
<https://orcid.org/0000-0002-8623-3483>

Ilyas Chachoua

Department of Oncology-Pathology, Karolinska Institutet,

Gaëlle Vertenoel

Ludwig Institute for Cancer Research, de Duve Institute, Université catholique de Louvain,

Thomas Tilmant

Mass Spectrometry Laboratory, MolSys Research Unit, University of Liège

Andrei-Jose Petrescu

Gabriel Mazuccheli

Mass Spectrometry Laboratory, MolSys Research Unit, University of Liège

Didier Vertommen

UCLouvain, de Duve Institute <https://orcid.org/0000-0001-7648-8282>

Article

Keywords: Myeloproliferative Neoplasms (MPN), Calreticulin, Thrombopoietin Receptor, HDx-MS, Structure

Posted Date: May 19th, 2022

DOI: <https://doi.org/10.21203/rs.3.rs-1605601/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Version of Record: A version of this preprint was published at Nature Communications on April 5th, 2023.

See the published version at <https://doi.org/10.1038/s41467-023-37277-3>.

Abstract

Calreticulin (CALR) is a master lectin chaperone that normally guides the proper folding of integral membrane proteins in the endoplasmic reticulum. In healthy cells, CALR transiently and non-specifically interacts with thousands of immature N-glycosylated proteins through its N-terminal glycan-binding domain. Conversely, frameshift mutants of CALR turn into rogue cytokine by acquiring the ability to stably and specifically interact with the Thrombopoietin Receptor (TpoR). Strikingly, this interaction induces constitutive activation of TpoR leading to myeloproliferative neoplasms, blood cancers resulting from the overproduction of blood cells. Using a multidisciplinary approach, we unveil how the CALR frameshift mutations result in structural overhaul of the entire protein and identify the structural basis of its acquired specificity for TpoR. We further describe the mechanisms by which complex formation triggers TpoR dimerization and activation. In addition, we provide the first complete dynamic conformational footprints of both wild-type and mutant CALR and identify novel potentially targetable sites.

Significance Statement

Molecular chaperones generally assist the conformational unfolding and re-folding of proteins often by recognizing exposed hydrophobic patches on misfolded proteins. Calreticulin is a master chaperone that targets the carbohydrate groups on a wide array of N-glycosylated proteins to guide their proper folding in the endoplasmic reticulum. In a series of frameshift mutants of calreticulin this normal house-keeping function leads to Myeloproliferative Neoplasms (MPNs) which are blood cancers caused by the constitutive activation of the thrombopoietin receptor (TpoR). Here we used a combination of biophysical and biochemical techniques along with functional assays to uncover the mechanism underlying TpoR activation. We discover that frameshift mutants of CALR adopt a different dynamic structure than wild-type CALR. We show that CALR mutants bind in a non-continuous way to both the glycans and several negatively charged patches in TpoR. This interaction leads to tight complex formation and charge-driven interactions, which provide the specificity of mutant CALRs for TpoR among the very many N-glycosylated proteins that interact with CALR physiologically. We map the sites of direct interactions, and we identify the structural changes occurring upon binding. The structural measurements along with molecular dynamics simulations allow us to build a model that precisely depicts how the binding of mutant CALRs to the TpoR extracellular domain via its D1 domain leads to complex formation, dimerization of the receptor in the membrane and constitutive activation. These results pinpoint specific residues to inhibit interaction and provide novel therapeutic avenues both for mutant CALRs and also for wild type CALR, which is involved in apoptosis and regulation of immune response.

Introduction

Upon their exit from ribosomes into the endoplasmic reticulum (ER), a large proportion of proteins are N-glycosylated¹. In their early life, immature N-glycosylated proteins are recognized by an ER resident chaperone called calreticulin (CALR) that helps nascent proteins achieve mature folding. In addition,

CALR also acts as a key regulator of calcium homeostasis through its C-terminal Ca^{2+} binding domain ^{2,3}. In some instances, CALR acquires frameshift mutations that lead to the loss of the wild-type C-terminus and replacement of the KDEL ER retention signal with a sequence rich in methionine and basic amino acids. Unexpectedly, these frameshift mutants of CALR cause myeloproliferative neoplasms (MPNs), a myeloid blood cancer ^{4,5}. MPNs are blood malignancies driven by the acquisition of somatic mutations in hematopoietic stem cells (HSCs) leading to cytokine-independent activation of the JAK-STAT pathway and uncontrolled proliferation of myeloid hematopoietic progenitors ⁶.

We and others reported that CALR mutants act by constitutively activating the thrombopoietin receptor (TpoR), leading to aberrant hematopoiesis ⁷⁻¹¹. In healthy cells, the N-terminal lectin domain of CALR transiently interacts with immature N-glycans in a cycle of attachment and liberation that ends once the protein achieves proper folding. In previous studies, we showed that the interaction between CALR and TpoR also requires the presence of immature glycans on Asn117 of TpoR ^{12,13}. Strikingly though, this normally transient interaction becomes stable between CALR mutants and TpoR via a yet unknown mechanism. Even more astonishingly, this switch from a transient to a stable interaction between CALR mutants and immature TpoR glycans has been described only for TpoR amongst thousands of other N-glycosylated proteins. In the present study, we used a multidisciplinary approach to uncover the structural mechanisms leading to the specific and stable interaction between CALR mutants and TpoR and unveil how this interaction leads to productive dimerization and activation of TpoR. In addition, this study generates the first structural footprints of wild-type and mutant CALR and provides a complete model of the CALR-TpoR complex, paving the way for novel therapeutic avenues.

Results

CALR mutant C-terminus opens up the wild-type lectin domain

Full length CALR has three domains (**Figure 1A**). The N-terminal domain (residues 18-197) interacts with immature N-glycans, the P-domain (residues 198-308) is associated with the chaperone functions and the C-domain is involved in calcium buffering and contains the KDEL ER retention motif. In agreement with crystallographic structure of the N-domain of human CALR ¹⁴ and NMR structure of rat P-domain ¹⁵, the AlphaFold prediction of full length calreticulin ¹⁶ reveals a compact N-domain rich in β -sheet and a central P-domain forming a long hairpin structure comprising several anti-parallel β -sheet segments. Although it has not been solved experimentally, the structure of the C-terminus is predicted to be essentially α -helical with the exception of the last 50 residues which appear to be unstructured (**Figure 1A**).

The CALR mutants involved in MPNs were previously reported to interact with TpoR through generic interaction between CALR lectin or N-domain and immature N-glycans on TpoR ^{12,13}. However, this interaction is normally only transient ³ and cannot explain the stable binding observed between TpoR and CALR mutants.

To understand how CALR mutants, represented by CALR del52, acquired the ability to stably interact with TpoR, we dissected the structural changes in full length CALR del52, which result from a frameshift that replaces the C-terminal 45 amino acids with a sequence that is rich in methionine and basic amino acids, but lacks the KDEL retention signal (**Figure 1B**).

To assess how the frameshift mutation in del52 influences the secondary structure and overall packing of CALR, we undertook Fourier transform infrared spectroscopy (FTIR) and hydrogen-deuterium exchange mass spectrometry (HDx-MS). We produced recombinant full length human CALR WT together with CALR del52 and different variants thereof as described in **Figure 1A and Supplementary Figure 1A-B**. FTIR spectroscopy reports on protein secondary structure and an analysis of the amide I vibration revealed that the overall structure of the protein was similar to that expected from the structural data of isolated domains or truncated proteins. The protein has high β -sheet and random coil content, which is likely due to the N- and P-domains, respectively. The α -helical content was higher in CALR wild-type (13.3%) compared CALR del52 (8.4%). To address the origin of the α -helical content, we deleted the last 45 residues (as in CALRc-tail, leaving only the common regions between CALR WT and CALR del52) (**Figure 1B**). Strikingly this deletion resulted in an almost complete loss of α -helical content, demonstrating that helicity is concentrated in the C-terminus of both wild-type and mutant CALR (**Figure 1C-D**). To define more precisely the region of the mutant tail that is α -helical, we truncated the last segment of CALR del52 tail (labelled as CALR del52 A394*) (**Figure 1B**). The mutant tail of CALR mutant can indeed be separated in two segments based on amino acid composition. The first region is rich in hydrophobic (Met) and basic (Arg, Lys) residues (**Figure 1B, purple**) while the second region starting at A394 (**Figure 1B, red**) has a more heterogenous amino acid composition. Analysis of FTIR spectra showed that CALR del52 A394* harbored increased relative α -helix content, demonstrating that helicity was concentrated in the first segment of the mutant tail rich in Arg and Met. This was further confirmed by *in silico* prediction using AlphaFold 2.0¹⁶ (**Supplementary Figure 1C**).

Next, we sought to assess whether introduction of two key point mutations (Y109F/D135L) in CALR del52 N-domain led to changes in its secondary structure. These mutations are in the N-domain (**Figure 1B**) and abolish the binding of CALR to the immature N-glycans on TpoR^{12,17}, Strikingly, this mutant showed a clear decrease in C-terminal helicity and of random coil percentages with corresponding increases in turns and β -sheets (**Figure 1C-D**). Because β -sheets are concentrated in the N-domain¹⁸ (**Figure 1B**) and α -helix present only in the C-terminus, these changes suggest conformational coupling between the N-domain and mutant tail of CALR del52.

To further dissect changes induced by the tail of CALR mutant, we acquired structural footprints of CALR WT and mutants using HDx-MS. With this technique, a mass shift in peptides from a protein after deuterium (D) incorporation in backbone amide positions provides a readout of residue accessibility and the overall conformation of the protein. We acquired the first complete HDx-MS structural footprint of CALR WT which was in line with a compact N-domain rich in β -sheets¹⁴ and an elongated P-domain that is highly accessible¹⁵ (**Supplementary Figure 2A-B**). The last segment of the P-domain (from residue

281) was less accessible, in line with *in silico* prediction¹⁹. The accessibility of the C-domain was consistent with an α -helical secondary structure as shown by our FTIR data. However, lower accessibility was observed in the KDEL ER retention signal than expected from current models (**Supplementary Figure 2A-B**).

We then compared the structural footprint of CALR WT with that of CALR del52 and CALR c-tail. Strikingly, the removal of the last 50 amino acids of CALR WT-did not significantly affect the conformation of the rest of the protein. Only a small fragment of the N-domain, represented by residues 52SSKGGKFGDEEKDKGLQTSQDARF⁷⁴ in contact with the C-domain based on AlphaFold prediction (**Figure 1E**) was less protected in CALR c-tail compared to CALR WT, suggesting that the C-terminus closes up the structure, but that it involves a small region of the N-terminus. This finding also suggests that mutations that involve deletion of the C-terminus without frameshift lead to enhanced accessibility of the N-terminus, making it a potential therapeutic target. Such mutations are frequent in a variety of solid-tumor cancer and are associated with immunosuppressive activity²⁰. In sharp contrast, CALR del52 was globally more amenable to deuterium exchange than CALR WT, demonstrating a more flexible conformation (**Figure 1F**). The N-domain was mostly affected by the addition of the novel mutant tail, suggesting again that the mutant tail directly interacts with and destabilizes the N-domain (**Figure 1F and Supplementary Figure 2C**), in line with our FTIR results. Remarkably, key residues involved in interactions with immature glycans such as C105 and W319^{12,17} were more accessible in CALR del52 compared to CALR WT (**Figure 1F and Supplementary Figure 2C**). This finding that the lectin binding residues are more available in CALR mutant provided a first explanation for the more stable interaction between CALR mutant and immature glycans on TpoR. Identifications of such key regions that are more accessible in CALR mutant than in CALR WT also unveils potential sites for specific therapeutic targeting.

CALR mutant tail directly interacts with TpoR extracellular domain

The results above indicated that the strong interaction between CALR mutant and TpoR could be at least partially explained by increased accessibility of key residues in the CALR-N-domain able to bind immature glycans. Yet, this did not solve the specificity of CALR del52 for TpoR versus thousands of other N-glycosylated proteins.

To investigate the basis for this specificity, we sought to assess whether other regions of CALR mutants directly interacted with TpoR independently of the presence of immature glycans on TpoR. Therefore, we used recombinant extracellular domain (ECD) of TpoR, labelled as D1D4, that we showed contains mature glycans¹². In a mature glycosylated form, TpoR ECD can only bind via other residues than those involved in binding of immature TpoR to CALR. The two proteins were incubated at 1:1 molar ratio before HDx-MS analysis. Analysis revealed significant ($p < 0.001$) hydrogen-deuterium exchange differential in multiple peptides containing the mutant tail (**Figure 2A and Supplementary Figure 3A-C**), demonstrating a direct interaction with the mature TpoR extracellular domain. This differential exchange was not observed

in the very last residues of the mutant tail encompassing residues ⁴⁰⁵QWGTEA⁴¹¹ (**Supplementary Figure 3A-C**), showing that this last segment is not involved in the interaction. Different fragments of the N-domain predicted to be conformationally close to the C-domain, also showed significant, albeit smaller, increase in accessibility in presence of mature TpoR ECD (**Supplementary Figure 3B-C**). This observation suggests that binding of the mutant tail to the receptor leads to concomitant loss of contact between the new tail and the N-domain of CALR del52, thus further opening up the mutant CALR.

To assess whether interaction between CALR mutant tail and TpoR could be observed in living cells, we created N-terminal truncations of CALR del52 (**Figure 2B**) lacking the N-domain and thus unable to interact with immature glycans on TpoR. We used Bioluminescence Energy Transfer (BRET) to measure direct interaction in living cells between TpoR N-terminally fused with a NanoLuciferase (NL) and N-terminal truncations of CALR del52 fused to a HaloTag at their C-terminus as we did previously for full length CALR del52 ¹². In line with our structural findings, complete deletion of the N-domain (denoted CALR del52 P-C domain) or both the N and P-domain of CALR del52 (CALR del52 C-domain) still allowed strong interaction with the receptor in living cells (**Figure 2C and Supplementary Figure 3D**). This was further validated by co-immunoprecipitation of FLAG-tagged N-terminal truncations of CALR del52 and HA-tagged TpoR (**Figure 2D**). Thus, CALR del52 interacts with TpoR both through its lectin domain and its mutant tail. Finally, we used a STAT5-dependent transcriptional luciferase assay to assess whether CALR del52 could induce activation of TpoR independently of the lectin binding. Unexpectedly, we found that the P-C domain alone of CALR del52 was able to induce a small but significant STAT5 transcriptional activity via TpoR, but not the erythropoietin receptor (EpoR) (**Figure 2E and Supplementary Figure 3E**). Together these results demonstrated that direct binding of CALR del52 to TpoR is mediated both by the mutant tail and N-domain interaction with immature sugars, the former providing specificity for TpoR against other N-glycosylated proteins.

The α -helical segment of CALR mutant tail induces TpoR dimerization and CALR oligomerization

Our data indicated that the CALR mutant tail provides specificity for TpoR through direct interaction. It is also known that the mutant tail is indispensable for activation of the receptor ^{7,9}. Yet, exactly how activation is achieved remains unclear.

To close this gap, we sought to assess how the mutant tail could at the same time bind TpoR and induce its activation. We first determined the exact region of the mutant tail required for TpoR activation by progressively truncating the C-terminus of CALR del52 (**Figure 3A**). We probed the ability of these truncations to induce activation of the TpoR by measuring autonomous proliferation of cytokine-dependent hematopoietic cell lines (Ba/F3) stably expressing TpoR and either of the CALR truncations. Truncations until position 394, thus removing the non-helical segment of the mutant tail, allowed autonomous growth similar to that of full length CALR del52. In sharp contrast, further truncations led to decrease (CALR del52 M387*) or complete abolition of the activity of the mutant CALR. Thus, it is the -

helical region of the mutant tail that is required for activation of TpoR. Importantly, the CALR del52 Y109F/D135L mutant, which abolishes glycan-dependent binding and disturbs the helicity of the mutant tail, did not allow Ba/F3 autonomous proliferation (**Figure 3B**). Then, we used cysteine crosslinking in live cells to study whether the same α -helical segment of CALR mutant tail was key for TpoR homo-dimerization. We used the L508C mutant of human TpoR, a mutation homologous to murine TpoR L501C, putting the cysteine residue in a dimeric orientation that we previously established to reflect the active state of TpoR²¹. The specificity of the crosslinking was achieved by using a truncated form of the TpoR devoid of intracellular cysteines which remains active²¹ and by preventing crosslinking of cysteines in the ECD thanks to N-ethyl-maleimide which blocks free extracellular cysteines (**Figure 3C**). In line with our functional assays, we observed that truncations until M387 still enabled dimerization of the TpoR transmembrane domain in an active conformation while further truncations did not allow the formation of dimers (**Figure 3D**). Because oligomerisation of CALR mutants themselves is key for TpoR activation^{22,23}, we then used the same set of deletions to probe directly CALR oligomerization by co-immunoprecipitation of HA-tagged full length CALR del52 with FLAG-tagged CALR del52 truncations. Comparably to their effect in TpoR dimerization and activation, truncations of the mutant tail beyond its non- α -helical segment led to sharp decrease in CALR mutant oligomerization. Thus, the first 28 α -helical residues of the mutant tail are required for TpoR activation and dimerization of both CALR mutant and TpoR (**Figure 3E**). We confirmed these results by showing that recombinant CALR c-tail prevented the formation of oligomers while truncations of the non- α -helical segment of the mutant tail, like in CALR del52 394*, did not change the oligomeric profile in presence or not of a reducing agent (DTT) in native conditions (**Supplementary Figure 3F**).

Finally, we used RosettaDock²⁴ to model CALR del52 dimers formation with monomeric structure predicted by AlphaFold 2.0¹⁶. The top 10 models predicted dimerization through the mutant tail through residues prior A394 and the two C-terminal cysteines. The best scoring prediction is depicted in **Figure 3F** and shows dimerization of CALR del52 mutant tail forming a coiled-coil like structure with interactions involving Arg (dark blue), Met (orange) and Thr (purple), but not the cysteines at the C-terminus of the mutant tail. Together with above results, this demonstrates that one face of the α -helical mutant C-terminus provides specificity through direct interaction with a precise region of TpoR while the other face induces dimerization of CALR, imposing an active dimeric conformation to TpoR.

Mapping interactions in the TpoR-CALR mutant complex

Having established that the tail of CALR mutant directly interacted with the TpoR ECD, we sought to identify the regions of TpoR involved in this interaction. The TpoR ECD is composed of four sub-domains labelled D1 to D4. Our previous data showed that glycosylation of Asn117 in D1, and to a lesser extent, Asn178 in D2, was critical for the formation of a productive TpoR-CALR mutant complex^{12,13}.

To investigate whether the same domains of TpoR were also sufficient to mediate binding via the mutant tail, we first used co-immunoprecipitation of C-terminal truncations of the TpoR ECD with CALR del52. Truncations of the D3 and D4 domains did not impact binding while truncations of D2 led to a small but significant reduction of co-immunoprecipitation compared to the D1 domain alone in line with a minor role of glycans linked to Asn178 (**Figure 4A-B**). Since we showed that only the mutant tail interacted with TpoR in presence of mature glycans, we used the CALR del52 Y109F/D135L deficient for binding immature glycans^{12,17}. Expectedly, loss of glycan-dependent binding led to a sharp decrease in co-immunoprecipitation ratios. Remarkably, though, deletions of D3D4 or even D2D3D4 allowed similar binding to CALR del52 Y109F/D135L as D1 alone (**Figure 4A-B**), suggesting that the mutant tail interacted essentially with TpoR D1 domain. To confirm that interaction between the mutant tail occurred through D1, we measured BRET in living cells between CALR del52 P-C and C-domain and subdomains of TpoR ECD. In line with co-immunoprecipitation, BRET showed similar or even slightly higher interaction between the CALR mutant devoid of the lectin binding domain and TpoR D1 compared to D1D2 or D1D2D3D4 (**Figure 4C**).

Next, we sought to assess whether the presence of immature glycans on TpoR affected the interaction with the mutant tail and identify the binding sites of CALR del52 with immature glycans on TpoR. Knowing that TpoR D1D2 was sufficient to mediate full binding to CALR del52 through both glycans and the mutant tail, we produced the CALR del52-TpoR D1D2 complex in S2 cells (**Figure 4D and Supplementary Figure 4A-B**). In this complex, immature glycans are attached to Asn117. Using our HDX-MS set up, we mapped all sites of CALR del52 interaction with TpoR D1D2. This revealed that mutant calreticulin interacted with TpoR via two major domains. First, the strongest interaction was observed in the putative glycan binding site of calreticulin (**Figure 4E-F and Supplementary Figure 4C-E**). This region included notably C105, Y109, D135 and W319, all reported to be key for binding of sugar moieties^{12,25-27}. Importantly, this region was not involved in binding of mature TpoR (**Figure 2B**), confirming that it is specifically involved in interaction with immature glycans. Secondly, multiple peptides containing CALR del52 mutant tail equally showed a high degree of differential uptake between the CALR del52-D1D2 complex and CALR del52 alone (**Figure 4E-F and Supplementary Figure 4C-E**); these were similar as the ones exhibiting differential uptake also when in complex with the mature TpoR. This confirmed that the mutant tail is involved not only in binding to mature forms of the receptors, but also to immature TpoR, the interaction being further consolidated by strong interactions between the CALR glycan binding pocket and immature sugar moieties on Asn117 of TpoR D1D2¹³.

CALR mutant tail interacts with acidic patches on TpoR D1 domain

To achieve a more thorough understanding of the CALR-TpoR interaction, we then aimed to identify specific residues on TpoR involved in binding the mutant tail. We first used molecular dynamics (MD) to study possible sites of interactions between CALR mutant tail and TpoR ECD. We generated the structure of TpoR D1D2 (see methods) to which glycosylation sites were attached at Asn117 (immature) and

Asn178 (mature) (**Supplementary Figure 5A**). The final model was in agreement to the one recently published by AlphaFold 2.0¹⁶. Sequence analysis indicated that TpoR D1D2 exhibits an unbalanced charge composition with an excess of 11 negatively charged acidic amino acids with one extensive (S1) and a second more localized (S2) negatively charged region (**Supplementary Figure 5A**), both particularly prone for interaction with the highly basic C-terminus tail of mutant CALR. The role of the acidic-basic interactions was confirmed by showing that mutations of hydrophobic residues to either Gly or Asn led to a small but significant decrease in activation of the TpoR in a STAT5 transcriptional assay but that mutations of basic residues to Asn or Gly completely abolished CALR del52 dependent activation of TpoR (**Supplementary Figure 5B-C**), in agreement with previous reports⁷. Then, we used *in silico* simulation to probe for the most stable interacting sites of the mutant tail alone to TpoR D1D2 domains (see Methods). Three final poses were obtained after 500 ns Molecular Dynamics runs (**Supplementary Video 1-3**). Amongst them, two were in line with our experimental data, predicting interactions occurring essentially through either the extended S1 region of TpoR D1 (centered on ⁴⁶ED⁴⁷) or the more restricted S2 regions (centered on ⁵³EEE⁵⁵) and basic Arg on CALR mutant tail (**Figure 5A**). We further estimated the free energy (ΔG) using the Prodigy^{28,29}. There were very large number of microstates of the complex showing very large negative values of the free energy ($\Delta G < -9$ kcal/mol), consistent with a high binding affinity of CALR del52 mutant tail for TpoR D1D2. The computational work indicated that CALR del52 mutant tail has the ability to engage TpoR with very high affinity in a very large number of micro-configurations that target both the continuous acidic area found mainly on D1 (and partly on D2) labelled as S1, but also the small acidic patch in the N-terminal region of D1 (S2).

To challenge our MD simulation experimentally, we used HDx-MS with the same set-up as in **Figure 2A**, where we showed that CALR mutants interact with mature TpoR exclusively through the mutant tail. Amongst the covered region, by far the strongest interaction was observed with the ⁴¹FSRTFEDL⁴⁸ motif of TpoR D1 domain, which was statistically significant ($p < 0.001$) for all incubation time points (**Figure 5B-D** and **Supplementary Figure 5D**). Remarkably, the very same peptide could also be detected from the CALR del52-TpoR produced as a complex containing immature glycans as depicted in **Figure 4D** where it was even less accessible than in the glycan-independent interaction (**Figure 5D**). This demonstrates that this region is involved not only in binding the mutant tail in absence of immature glycans but in complex formed in living cells with immature TpoR. Among other noticeable peptides, the ⁵¹WDEE⁵⁴ from the S2 patch and ²⁷²WSLPVT²⁷⁶ close to the extended S1 region showed significant, albeit lower, decrease in accessibility for most time points after incubation of mature TpoR D1D4 with CALR del52 (**Figure 5B-C** and **Supplementary Figure 5E**). This suggests that, in absence of immature glycans, multiple micro-configurations of TpoR-CALR del52 can co-exist but that the S1 patch centered on ⁴⁴TFED⁴⁷ plays a central role in binding. Consistently, mutating this motif to alanine in full length TpoR led to complete loss of TpoR activation by both CALR del52 full length and P-C domain, as assessed by a STAT5 luciferase transcriptional assay (**Figure 5E**). Thus, interaction between CALR mutant tail and TpoR occurs through an extended patch of negative charges on TpoR D1 domain centered on residues ⁴⁶ED⁴⁷.

Comprehensive model of the TpoR-CALR mutant complex

HDx-MS and MD simulations using only the mutant tail showed that the latter was able to bind the receptor in a large variety of micro-configurations and identified the extended S1 acidic region centered on ⁴⁶ED⁴⁷ of TpoR as critical for the mutant tail interaction with TpoR. Yet, binding of immature TpoR to full length CALR del52 also involves strong interaction between specific residues of the N-domain and immature glycans on Asn117 (**Figure 4E**). We sought to create a complete model of the tetrameric TpoR-CALR mutant complex using our structural data. We used AlphaFold 2.0 ¹⁶ to complete our modeling of TpoR to obtain the full extracellular domain of the receptor. TpoR monomers were dimerized through their TM domain with residue L508 in the interface as in the active configuration in presence of CALR del52 (**Figure 3C-D**). CALR del52 dimers (**Figure 3F**) were docked to dimers of TpoR through glycan binding domains and mutant tail based on experimental data and energy minimization. The final structure places the mutant tail in a configuration where the main interacting sites are located around the ⁴⁶ED⁴⁷ motif, in line with our HDx-MS and functional results. Likewise, immature glycans on Asn117 of the receptor interacting with the N-domain pocket containing key residues involved in glycans binding including C105, Y109, D135 and W319 (**Figure 6**). This glycoproteic tetramer was immersed into a full-atom representation of the environment, including a POPC lipid bilayer. This overall system consisting of ~1.5 million atoms was subjected to a complete molecular dynamics cycle at 300K for 20 ns (see methods). The complex remained very stable during this timeframe, with the exception of the very flexible P-domain (**Supplementary Video S4**). Contacts between TpoR and CALR molecules were analyzed over the course of the simulation. Most contacts relied on basic-acidic interactions and occurred both in *cis* and in *trans*, thereby further stabilizing TpoR dimers (**Supplementary File 1**). In conclusion, we provide here a model where CALR del52 interacts through two regions with TpoR D1 domain. First, the mutant tail directly interacts with multiple negatively charged residues on the inner/lateral face of TpoR D1 domain represented by the S1 negative patch. Given the ability of the mutant tail to interact with multiple acidic residues on TpoR D1 domain, it is likely that different micro-configurations co-exist in living cells, especially in absence of immature glycans to stabilize one particular configuration. When this interaction occurs between immature TpoR and CALR del52, it is further stabilized by strong interactions between CALR glycan binding domain and immature sugar moieties mainly on Asn117 of TpoR with the mutant tail providing strong specificity and stability for interaction between CALR mutants and TpoR.

Discussion

Our work reveals how frameshift mutants of the archetype of a lectin chaperone, calreticulin, lead to a conformational overhaul and lead to specific interaction and activation of the TpoR to induce myeloproliferative neoplasms. These findings provide mechanistic insights into the mechanisms leading to a switch from the transient, lectin-only based interaction between wild-type calreticulin and thousands of proteins to a specific and stable interaction between mutant CALR and TpoR causing its constitutive dimerization and activation. Unexpectedly, we find that conformational coupling exists between the C-terminus and the lectin domain of CALR mutants. Acquisition of the first structural footprint of full length CALR WT and mutants revealed that this conformational coupling leads to an opening up of the glycan binding domain of calreticulin and increased accessibility of residues involved in direct interaction with

immature glycans. In addition, this provides unprecedented structural data on wild-type human calreticulin. Remarkably, the identified glycan binding pocket of CALR mutants is reminiscent of the one that we and our collaborators recently identified as the hematoxylin binding site via which hematoxylin acts as an inhibitor of CALR del52 binding to TpoR³⁰. Our finding that this pocket is more accessible in CALR mutants compared to CALR wild-type explains the partial specificity of hematoxylin to target mutant cells. In addition, we identify several novel regions conserved between CALR wild-type and mutant that exhibit different accessibility, thereby paving the way for novel therapeutically targetable sites. Besides, our study unequivocally demonstrates a direct interaction between a specific region of the mutant tail of CALR del52 that we show is α -helical and the N-terminus of the TpoR.

Aside to binding to TpoR, we previously showed that CALR del52 induced dimerization of TpoR¹² and of JAK2 (Chachoua et al. 2016) and others reported that multimerization of mutant CALR was a prerequisite for TpoR activation²². The finding of the present study demonstrates that it is in fact the α -helical region of mutant tail which is directly involved not only in binding TpoR, but also in dimerization of mutant CALR proteins and of TpoR chains. Importantly, using a multidisciplinary approach we identify that the binding of CALR mutant tail to TpoR occurs through negative patches of the N-terminal D1 extracellular domain of the receptor. Identification of these sites is of major importance for the development of inhibitors targeting the TpoR-CALR mutant tail interaction. Finally, we provide the first complete model of the tetrameric CALR-TpoR complex based on experimental data and Molecular Dynamics simulation at 300K.

Of interest, mutations in CALR C-terminus are also observed in a variety of non-hematological cancer and their exposure at the cell surface leads to immunosuppressive activity²⁰. At the conceptual level, it is tempting to speculate that the enhanced accessibility of CALR N-terminus that our work identifies upon deletion of the C-terminus could be used for therapeutical targeting.

In conclusion, the work presented herein identifies how frameshift mutations in a master chaperone leads to its conformational remodeling resulting in oncogenic properties. It characterizes specific interactions between the mutant CALR and the TpoR ECD, providing mechanistic understanding of the specificity of this interaction and opening the road for inhibitory therapeutic avenues. Finally, we propose a first complete model of the CALR del52-TpoR tetrameric complex where the mutant CALR tail acts like a specific scratch for negative patches of TpoR extracellular domain, and where the glycan mediated interaction provides the necessary glue to stabilize the interaction. On a more general aspect, our work provides the first dynamic structural footprint of both full length CALR WT, CALR C-terminus deletion and CALR del52. These findings provide important advances for the study of calreticulin biology and anti-cancer therapy.

Materials And Methods

Production and purification of recombinant proteins

Recombinant proteins for human TpoR (hTpoR) D1D2D3D4 and CALR del52-hTpoR D1D2 complex were produced as described previously¹². The amino acid sequence of hTpoR D1D2D3D4 starts at Q26 and ends at T489 and this of hTpoR D1D2 starts at Q26 and ends at Q290. Both contains a histidine tag at the C-terminus. The amino acid sequence of hCALR del52 starts at E18 and ends at L417.

Recombinant human CALR wild-type, CALR del52 and its derivatives contain a N-terminal His tag sequence (MGSHHHHHHGSSG) that replaces the CALR signal peptide sequence (aa1-17). In addition, the cysteine 163 was mutated to serine. The corresponding DNA sequences of the recombinant proteins were cloned into a derivative of plasmid pET9 for transformation and expression in the *E. Coli* bacterial strain Rosetta2(DE3). The soluble recombinant proteins were extracted from the cells and purified using Histidine-affinity with a Ni²⁺ + purification resin. The eluted proteins were then concentrated by ultrafiltration and further purified by size-exclusion chromatography by loading on a Superdex 200 Increase 10/300 column (GE healthcare, Chalfont St. Giles, United Kingdom). The fractions were analyzed and purity was confirmed by SDS-PAGE (Supplementary Fig. 1A).

Hydrogen-Deuterium Exchange Mass Spectrometry (HDx-MS)

Hydrogen-Deuterium Exchange Mass Spectrometry was performed with a Waters nanoAcquity UPLC with HDx technology coupled with Synapt G2-Si. All purified recombinant proteins were used at 20 µM concentration in equilibration buffer (5 mM K₂HPO₄, 5 mM KH₂PO₄ dissolved in H₂O, pH 7). For interaction analysis between recombinant mature D1D2D3D4 and CALR del52, proteins were first mixed together at a 1:1 molarity for 30 minutes at room temperature followed by 3 hours at 4°C. Proteins were then kept at 0°C. Labelling was performed with a 20-fold dilution of samples in labelling buffer (5 mM K₂HPO₄, 5 mM KH₂PO₄ dissolved in D₂O, pH 7) for 6 different incubation times (0, 0.25, 1, 5, 20 and 60 minutes) at 20°C. After incubation, the reaction was quenched using a 1:1 dilution in the quench buffer (0.05 M K₂HPO₄, 0.05 M KH₂PO₄ with 30 mM TCEP, pH 2.3) prior to injection into a pepsin column (Enzymate BEH Pepsin 2.1x30 Column, Waters CAT. 186007233) with dynamic flowrate of 150 – 75µL/min. All mixes were performed automatically by a PAL-RTC robot station. Peptides resulting from the pepsin digestion were captured on a ACQUITY BEH C18 1.7µM VANGUARD Pre-column (Waters Cat. 186009375), separated on a ACQUITY UPLC C18 1.7µM 1.0x100mm column (Waters Cat. 186002346) and electrosprayed into the Waters SYNAPT G2-Si quadrupole time-of-flight mass spectrometer. Measurements were performed in HDMSe mode with ion mobility. Lock mass correction was performed with infusion of leucine-enkephalin (m/z = 556.277). The peptides were identified from triplicates using the PLGS3.0 software (Waters). The peptides identified were further analyzed with DynamX 3.0 (Waters) using a tolerance of 10 ppm, a maximum length of 35 a.a., a minimum products per amino acid of 0.2 and requiring that each peptide was identified in 3 out of 3 replicates. All peptides were visually validated based on retention time, drift time and isotopic m/z. Data was statistically analyzed using Deuterios 2.0 with peptide-level significance testing³¹.

Fourier Transformed Infrared (FTIR) Spectroscopy

0.5 μL of sample was loaded on the diamond crystal of the ATR device of the FTIR spectrometer and quickly dried with a constant, gentle nitrogen flow: elimination of the water molecules prevents overlapping of the large water absorption peaks with the sample's absorption spectrum. After each spectrum, the crystal was cleaned with water. A background was recorded with a clean crystal before the start of the measurement and before every new sample. FTIR spectra were recorded between 4000 and 600 cm^{-1} at a resolution of 2 cm^{-1} . Each spectrum was obtained by taking an average of 128 scans. The FTIR measurements were carried out at room temperature ($\sim 22^\circ\text{C}$). For each sample, at least four spectra were recorded. All the spectra were preprocessed as follows. The water vapor contribution was subtracted with 1956 – 1935 cm^{-1} as reference peak. All spectra were then baseline-corrected and normalized as follows. Straight lines were interpolated between the following frequencies: 3700 3000 2800 1720 1480 1204 980 cm^{-1} . Then, they were subtracted from the spectrum. Normalization for equal area was applied between 1720 and 1480 cm^{-1} . Using a database of 50 protein containing as little fold redundancy as possible, an ascending stepwise method was applied to determine the protein secondary structure. It was demonstrated that three wavenumbers contain all the nonredundant information related to the secondary structure content. The standard error of prediction in cross-validation obtained using the 50-protein database was 5,7% for the α -helix and 6,7% for the β -sheet, 3,2% for turns and 8% for random ^{32,33}.

Molecular Dynamics and docking simulations

Sequences of TpoR extracellular and transmembrane regions and of CALR Del52 were profiled for secondary structure, intrinsic disorder and accessibility propensities with state-of-the-art predictors ^{34–41}. Closest templates were retrieved with Phyre 2 ⁴², Modeller 9.21 ⁴³, Alpha-Fold2 and Rosetta Folding ¹⁶. For the study of interactions between CALR mutant tail and TpoR D1D2, docking trials were performed using three main start configurations of the complex based on the acidic areas of TpoR set as inputs in HADDOCK ^{44,45} for TpoR D1D2-CALR del52 mutant tail complex optimizations searches. The top configurations were further optimized using 500 ns Molecular Dynamics runs to obtain 3 final poses in which the last residues of the mutant become unfolded. Templates from AlphaFold2 and Rosetta Folding were used to effectively build the tetrameric 3D models and identify the interaction interface between the two CALR Del52 mutants. HDx-MS data was used to identify contacts between TpoR and CALR Del52 in the formation of the tetramer complex. The ER specific G3M9 glycans of TpoR in contact to CALR were modeled with Glycopack ⁴⁶ in the configuration consistent with NMR data ⁴⁷ while the rest are of complex type, built in agreement with SAGS Database ^{48,49}. The HDx-MS identified contacts and the solid-NMR data on the TM region configuration of TpoR dimer were used as constraints in generating the overall 2CALR-2TpoR model. This glycoproteic tetramer was immersed into a full-atom representation of the environment - consisting in a lipid bilayer of 1607 POPC molecules accommodating the TM region of TpoR and in 409966 TIP3P water molecules and 63 Sodium ions describing the solvent region hydrating the rest of the tetramer. This overall system consisting of ~ 1.5 million atoms was subjected to a complete molecular dynamics cycle using OpenMM ⁵⁰. First the system was minimized using L-BFG algorithm till a threshold of 1kJ/mole. Heating was then performed slowly over a 13ns period, from 0 to 300K in an NPT ensemble at 1 bar pressure and a membrane surface tension of 340 bar·nm using a

Langevin integrator of 2fs timestep, with hydrogen bonds constrained in order to achieve this high timestep. Production was then continued for 20ns. Visualization & Analysis was performed using VMD 1.9.3 and in-house Python 3.6 scripts, which use the MDTraj⁵¹ library.

Transcriptional Dual Luciferase Assay

Transcriptional Dual luciferase assays were performed as described¹³. Briefly, HEK293T or γ 2a were transfected or human TpoR WT with indicated CALR species. In both cases, SpiLuc reporter was used as a readout of STAT5 transcriptional activation and pRLTK was used as an internal control (Promega). Cells were stimulated, or not, with 25ng/ml of rhTpo (Milteneyi Biotec) as indicated.

Western Blotting and co-immunoprecipitation

HEK293T were plated in 10 cm dishes and transiently transfected with cDNA coding for the indicated constructs. Confluent cells were lysed 48h post-transfection with NP-40 buffer. After pre-clearing, samples were incubated with anti-FLAG antibody (Genscript, Cat. No. A00187) at 2 μ g/ml or corresponding isotype control (Genscript Cat. No. A01730) overnight at 4°C. Bound proteins were pulled down with 40 μ L/ml of rProtein G Agarose (ThermoFisher, 20397) for 3 hours at 4°C. Samples were then centrifuged, washed three times and immunoprecipitated proteins were analyzed by SDS-PAGE followed by revelation with an anti-HA antibody (Roche, 3F10) for HA-hTpoR, HA-CALR del52 or anti-CALR mutant tail (SAT602).

Nano-Bioluminescence Energy Transfer (BRET)

Nano-bioluminescence resonance energy transfer (BRET) was performed as previously described¹².

Proliferation assay

Ba/F3 were transduced with human TpoR in pMX-IRES-GFP and CALR variants or an empty vector (pMSCV-IRES-mCherry) and sorted by FACS for similar levels of GFP and mCherry. 250,000 cells were washed and seeded in 10 ml RPMI, 10% FBS without cytokine and counted each day using a Coulter automated cell counter in triplicates. The experiments were performed in three different biological replicates (N = 3).

Mutagenesis

All mutants were made alternatively using the QuickChange (Agilent), the KLD enzyme mix (NEB) procedure following the manufacturer instruction or purchased from Genscript. All constructs were verified by sequencing.

Crosslinking

HEK293T were plated in 6 wells plates and co-transfected with indicated constructs. TpoR L508C was truncated after Box 2 to avoid non-specific crosslinking of intracellular cysteines. 48 hours post transfection, cells were harvested without trypsinization and washed in PBS. Cells were then re-suspended and incubated for 15 minutes at room temperature in crosslinking buffer (PBS 1mM MgCl₂, 0.1mM CaCl₂) with 100 μ M N-ethylmaleimide (NEM) to avoid non-specific crosslinking of extracellular

cysteines. 200ng/ml of rhTpo was added in the indicated condition. Samples were mixed gently and further incubated for 15 minutes at room temperature. Cells were then centrifuged for 5 minutes at 500g and re-suspended in crosslinking buffer with 100 μ M o-phenylene dimaleimide (o-PDM) for 10 minutes at room temperature. Cells were further centrifuged 5 minutes at 500g and re-suspended in lysis buffer (NP-40, 2% β -mercaptoethanol) with protease inhibitor cocktail. Cell lysates were analyzed by SDS-PAGE in denaturing and reducing conditions with anti-HA antibody.

Antibodies

Antibody	Reference	Assay
Anti-FLAG tag	Cat. No. A00187 (Genscript)	Co-immunoprecipitation, Flow Cytometry
Anti-HA tag	C29F4 mAB #3724 (Cell Signalling)	Western blot
Mouse IgG control	Cat. No. A01730 (Genscript)	Co-immunoprecipitation Flow Cytometry
anti-CALR antibody	FMC75, Abcam	ELISA
Anti-mutant CALR	SAT602 (Myeloproliferative)	Western Blot ELISA
anti-mouse IgG, coupled to APC	Clone Poly4053 (Biolegend)	Flow Cytometry
Anti-rabbit IgG, HRP-linked	#7074 (Cell Signalling)	Western Blot
Anti-mouse IgG, HRP-linked	Southern Biotech	ELISA

Declarations

Acknowledgments

We thank Didier Colau for his outstanding work in the production and purification of recombinant proteins used in this study, Lidvine Genet and Céline Mouton for expert technical support, and Dr. Nicolas Dauguet for flow cytometry assistance. We also thank Jean-François Collet and Steve O. Smith for their sound advices in the writing process of the manuscript. Funding to SNC is acknowledged from Ludwig Institute for Cancer Research, Fondation contre le cancer, Salus Sanguinis and Fondation “Les avions de Sébastien”, projects Action de recherche concertée (ARC) 16/21-073 and WelBio F 44/8/5 - MCF/UIG – 10955. NP has received an FSR PhD Fellowship from Université catholique de Louvain and an Aspirant PhD Fellowship from the FRS-FNRS, Belgium.

Author Contributions: N.P. and A.N. performed functional and biochemical experiments, analyzed and interpreted data. N.P. G.M., D.V. and T.T. performed HDx-MS experiments and analyzed the data. A.D.

performed FTIR experiments and analyzed data. A.-J.P. and T.A.S. and N.P. performed docking predictions, structure refinement and Molecular Dynamics analysis. N.P., A.N., I.C., C.P. and G.V. created constructs. S.N.C., N.P., A.N. designed experiments and interpreted data. S.N.C. and N.P. wrote the manuscript.

Competing Interest Statement: SNC is co-founder of MyeloPro Diagnostics and Research GmbH, Vienna.

References

1. Van den Steen P, Rudd PM, Dwek RA, Opdenakker G. Concepts and principles of O-linked glycosylation. *Crit Rev Biochem Mol Biol.* 1998;33(3):151–208.
2. Caramelo JJ, Parodi AJ. Getting in and out from calnexin/calreticulin cycles. *J Biol Chem.* 2008;283(16):10221–10225.
3. Varricchio L, Falchi M, Dall'Ora M, et al. Calreticulin: Challenges Posed by the Intrinsically Disordered Nature of Calreticulin to the Study of Its Function. *Front Cell Dev Biol.* 2017;5:96.
4. Nangalia J, Massie CE, Baxter EJ, et al. Somatic CALR mutations in myeloproliferative neoplasms with nonmutated JAK2. *N Engl J Med.* 2013;369(25):2391–2405.
5. Klampfl T, Gisslinger H, Harutyunyan AS, et al. Somatic Mutations of Calreticulin in Myeloproliferative Neoplasms. *N Engl J Med.* 2013;369(25):2379–2390.
6. Constantinescu SN, Vainchenker W, Levy G, Papadopoulos N. Functional Consequences of Mutations in Myeloproliferative Neoplasms. *Hemasphere.* 2021;5(6):e578.
7. Elf S, Abdelfattah NS, Chen E, et al. Mutant Calreticulin Requires Both Its Mutant C-terminus and the Thrombopoietin Receptor for Oncogenic Transformation. *Cancer Discov.* 2016;6(4):368–381.
8. Elf S, Abdelfattah NS, Baral AJ, et al. Defining the requirements for the pathogenic interaction between mutant calreticulin and MPL in MPN. *Blood.* 2018;131(7):782–786.
9. Araki M, Yang Y, Masubuchi N, et al. Activation of the thrombopoietin receptor by mutant calreticulin in CALR-mutant myeloproliferative neoplasms. *Blood.* 2016;127(10):1307–1316.
10. Nivarthi H, Chen D, Cleary C, et al. Thrombopoietin receptor is required for the oncogenic function of CALR mutants. *Leukemia.* 2016;30(8):1759–1763.
11. Marty C, Pecquet C, Nivarthi H, et al. Calreticulin mutants in mice induce an MPL-dependent thrombocytosis with frequent progression to myelofibrosis. *Blood.* 2016;127(10):1317–1324.
12. Pecquet C, Chachoua I, Roy A, et al. Calreticulin mutants as oncogenic rogue chaperones for TpoR and traffic-defective pathogenic TpoR mutants. *Blood.* 2019;133(25):2669–2681.
13. Chachoua I, Pecquet C, El-Khoury M, et al. Thrombopoietin receptor activation by myeloproliferative neoplasm associated calreticulin mutants. *Blood.* 2016;127(10):1325–1335.
14. Chouquet A, Païdassi H, Ling WL, et al. X-ray structure of the human calreticulin globular domain reveals a peptide-binding area and suggests a multi-molecular mechanism. *PLoS One.* 2011;6(3):e17886.

15. Ellgaard L, Riek R, Herrmann T, et al. NMR structure of the calreticulin P-domain. *Proceedings of the National Academy of Sciences*. 2001;98(6):3133–3138.
16. Jumper J, Evans R, Pritzel A, et al. Highly accurate protein structure prediction with AlphaFold. *Nature*. 2021.
17. Kapoor M, Ellgaard L, Gopalakrishnapai J, et al. Mutational analysis provides molecular insight into the carbohydrate-binding region of calreticulin: pivotal roles of tyrosine-109 and aspartate-135 in carbohydrate recognition. *Biochemistry*. 2004;43(1):97–106.
18. Chouquet A, Païdassi H, Ling WL, et al. X-ray structure of the human calreticulin globular domain reveals a peptide-binding area and suggests a multi-molecular mechanism. *PloS one*. 2011;6(3):e17886-e17886.
19. Migliaccio AR, Uversky VN. Dissecting physical structure of calreticulin, an intrinsically disordered Ca(2+)-buffering chaperone from endoplasmic reticulum. *J Biomol Struct Dyn*. 2018;36(6):1617–1636.
20. Liu P, Zhao L, Loos F, et al. Immunosuppression by Mutated Calreticulin Released from Malignant Cells. *Mol Cell*. 2020;77(4):748–760.e749.
21. Staerk J, Defour JP, Pecquet C, et al. Orientation-specific signalling by thrombopoietin receptor dimers. *Embo j*. 2011;30(21):4398–4413.
22. Araki M, Yang Y, Imai M, et al. Homomultimerization of mutant calreticulin is a prerequisite for MPL binding and activation. *Leukemia*. 2019;33(1):122–131.
23. Rivera JF, Baral AJ, Nadat F, et al. Zinc-dependent multimerization of mutant calreticulin is required for MPL binding and MPN pathogenesis. *Blood Adv*. 2021;5(7):1922–1932.
24. Lyskov S, Gray JJ. The RosettaDock server for local protein-protein docking. *Nucleic Acids Res*. 2008;36(Web Server issue):W233-238.
25. Thomson SP, Williams DB. Delineation of the lectin site of the molecular chaperone calreticulin. *Cell Stress Chaperones*. 2005;10(3):242–251.
26. Gopalakrishnapai J, Gupta G, Karthikeyan T, et al. Isothermal titration calorimetric study defines the substrate binding residues of calreticulin. *Biochem Biophys Res Commun*. 2006;351(1):14–20.
27. Kozlov G, Pocanschi CL, Rosenauer A, et al. Structural basis of carbohydrate recognition by calreticulin. *J Biol Chem*. 2010;285(49):38612–38620.
28. Vangone A, Bonvin AM. Contacts-based prediction of binding affinity in protein-protein complexes. *Elife*. 2015;4:e07454.
29. Xue LC, Rodrigues JP, Kastritis PL, Bonvin AM, Vangone A. PRODIGY: a web server for predicting the binding affinity of protein-protein complexes. *Bioinformatics*. 2016;32(23):3676–3678.
30. Jia R, Balligand T, Atamanyuk V, et al. Hematoxylin binds to mutant calreticulin and disrupts its abnormal interaction with thrombopoietin receptor. *Blood*. 2021;137(14):1920–1931.
31. Lau AM, Claesen J, Hansen K, Politis A. Deuteros 2.0: peptide-level significance testing of data from hydrogen deuterium exchange mass spectrometry. *Bioinformatics*. 2021;37(2):270–272.

32. Oberg KA, Ruyschaert JM, Goormaghtigh E. Rationally selected basis proteins: a new approach to selecting proteins for spectroscopic secondary structure analysis. *Protein Sci.* 2003;12(9):2015–2031.
33. Goormaghtigh E, Ruyschaert JM, Raussens V. Evaluation of the information content in infrared spectra for protein secondary structure determination. *Biophys J.* 2006;90(8):2946–2957.
34. Drozdetskiy A, Cole C, Procter J, Barton GJ. JPred4: a protein secondary structure prediction server. *Nucleic Acids Research.* 2015;43(W1):W389-W394.
35. Jones DT. Protein secondary structure prediction based on position-specific scoring matrices. *J Mol Biol.* 1999;292(2):195–202.
36. Yachdav G, Kloppmann E, Kajan L, et al. PredictProtein—an open resource for online prediction of protein structural and functional features. *Nucleic Acids Research.* 2014;42(W1):W337-W343.
37. Petersen B, Petersen TN, Andersen P, Nielsen M, Lundegaard C. A generic method for assignment of reliability scores applied to solvent accessibility predictions. *BMC Structural Biology.* 2009;9(1):51.
38. Rost B, Sander C. Prediction of protein secondary structure at better than 70% accuracy. *J Mol Biol.* 1993;232(2):584–599.
39. Ishida T, Kinoshita K. PrDOS: prediction of disordered protein regions from amino acid sequence. *Nucleic Acids Res.* 2007;35(Web Server issue):W460-464.
40. Romero P, Obradovic Z, Li X, Garner EC, Brown CJ, Dunker AK. Sequence complexity of disordered protein. *Proteins.* 2001;42(1):38–48.
41. Barik A, Katuwawala A, Hanson J, Paliwal K, Zhou Y, Kurgan L. DEPICTER: Intrinsic Disorder and Disorder Function Prediction Server. *J Mol Biol.* 2020;432(11):3379–3387.
42. Kelley LA, Mezulis S, Yates CM, Wass MN, Sternberg MJ. The Phyre2 web portal for protein modeling, prediction and analysis. *Nat Protoc.* 2015;10(6):845–858.
43. Webb B, Sali A. Comparative Protein Structure Modeling Using MODELLER. *Curr Protoc Bioinformatics.* 2016;54:5.6.1–5.6.37.
44. Dominguez C, Boelens R, Bonvin AM. HADDOCK: a protein-protein docking approach based on biochemical or biophysical information. *J Am Chem Soc.* 2003;125(7):1731–1737.
45. van Zundert GCP, Rodrigues J, Trellet M, et al. The HADDOCK2.2 Web Server: User-Friendly Integrative Modeling of Biomolecular Complexes. *J Mol Biol.* 2016;428(4):720–725.
46. Paduraru C, Spiridon L, Yuan W, et al. An N-linked glycan modulates the interaction between the CD1d heavy chain and beta 2-microglobulin. *J Biol Chem.* 2006;281(52):40369–40378.
47. Petrescu AJ, Butters TD, Reinkensmeier G, et al. The solution NMR structure of glucosylated N-glycans involved in the early stages of glycoprotein biosynthesis and folding. *Embo j.* 1997;16(14):4302–4310.
48. Petrescu AJ, Petrescu SM, Dwek RA, Wormald MR. A statistical analysis of N- and O-glycan linkage conformations from crystallographic data. *Glycobiology.* 1999;9(4):343–352.

49. Petrescu AJ, Wormald MR, Dwek RA. Structural aspects of glycomes with a focus on N-glycosylation and glycoprotein folding. *Curr Opin Struct Biol.* 2006;16(5):600–607.
50. Eastman P, Swails J, Chodera JD, et al. OpenMM 7: Rapid development of high performance algorithms for molecular dynamics. *PLoS Comput Biol.* 2017;13(7):e1005659.
51. McGibbon RT, Beauchamp KA, Harrigan MP, et al. MDTraj: A Modern Open Library for the Analysis of Molecular Dynamics Trajectories. *Biophys J.* 2015;109(8):1528–1532.

Figures

Figure 1

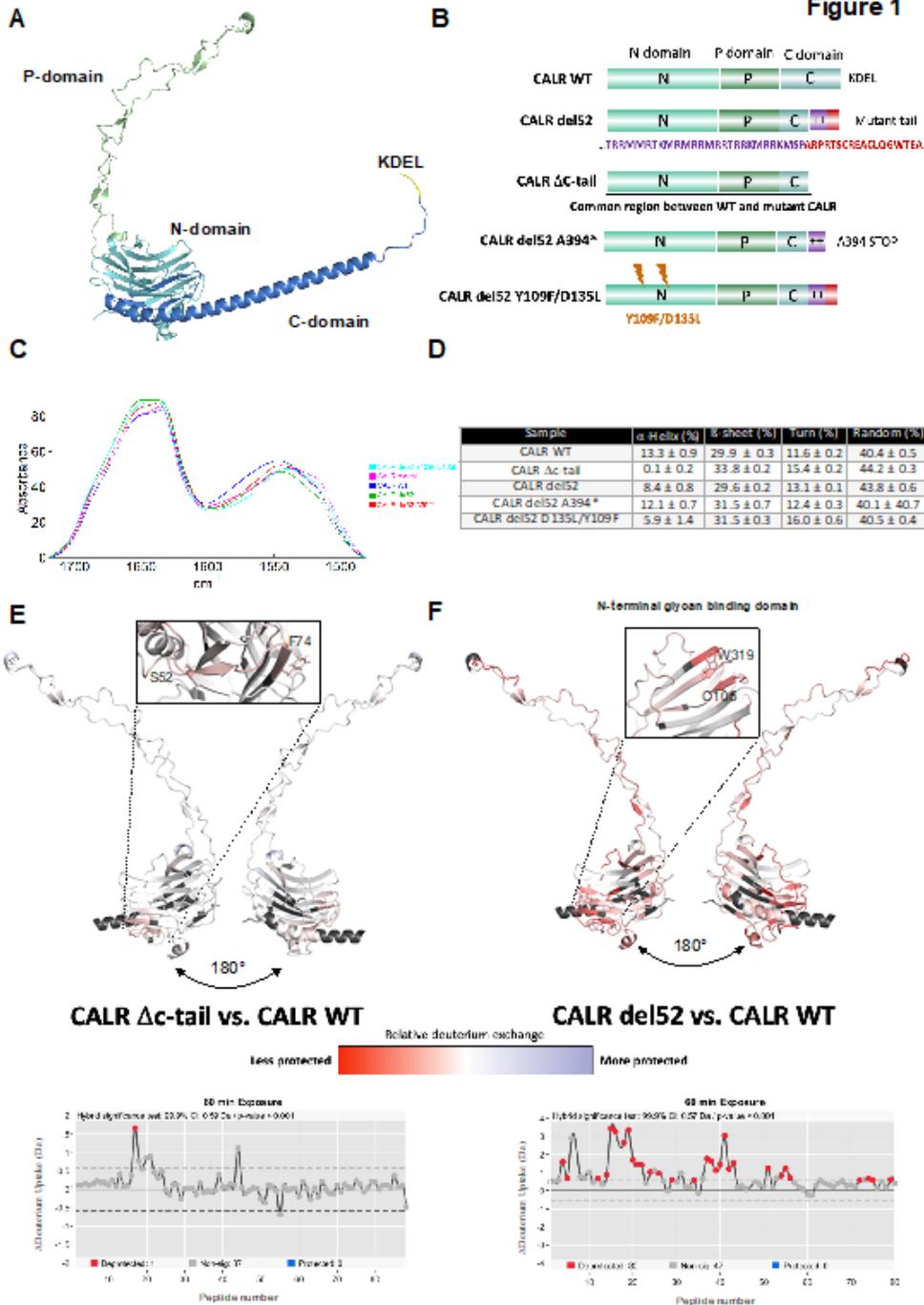


Figure 1

A. Structure of full length CALR WT as predicted by AlphaFold 2.0¹⁶. The N-domain is shown in light blue, the P-domain in green and the C-domain in dark blue. **B.** Representation of the domains and C-terminal sequence of the mutants of CALR del52 used in FTIR spectroscopy and HDx-MS experiments. **C.** Comparison of the mean spectra recorded for each sample to analyse the protein secondary structure. Each sample is identified by a unique color indicated in the legend. **D.** Secondary structure predictions

using the method developed on our in-house database. The prediction is realized on each individual FTIR spectrum. The average and the standard deviation for the 5 spectra recorded for each sample is shown in this table. **E-F.** Top: Relative fractional deuterium uptake differential between the indicated species. The scale represents the relative differential of deuterium incorporation between the left and right species indicated. Regions in dark grey represents regions not covered by detected peptides at the given time point. Shown are the common domains between CALR WT and CALR del52 at 1 hour incubation in deuterium (mean of triplicate experiments). The lectin binding domain of CALR which is more exposed in CALR del52 compared to CALR WT is highlighted. The structure was computed using the AlphaFold 2.0¹⁶ running on our in-house servers. Bottom: Relative deuterium uptake analysis between the indicated species at 1 hour incubation in deuterium. Dots indicated in red (deprotected) or blue (protected) show significant differential deuterium intake ($p < 0.001$) with the peptide-level significance testing (hybrid mode) as described³¹.

Figure 2

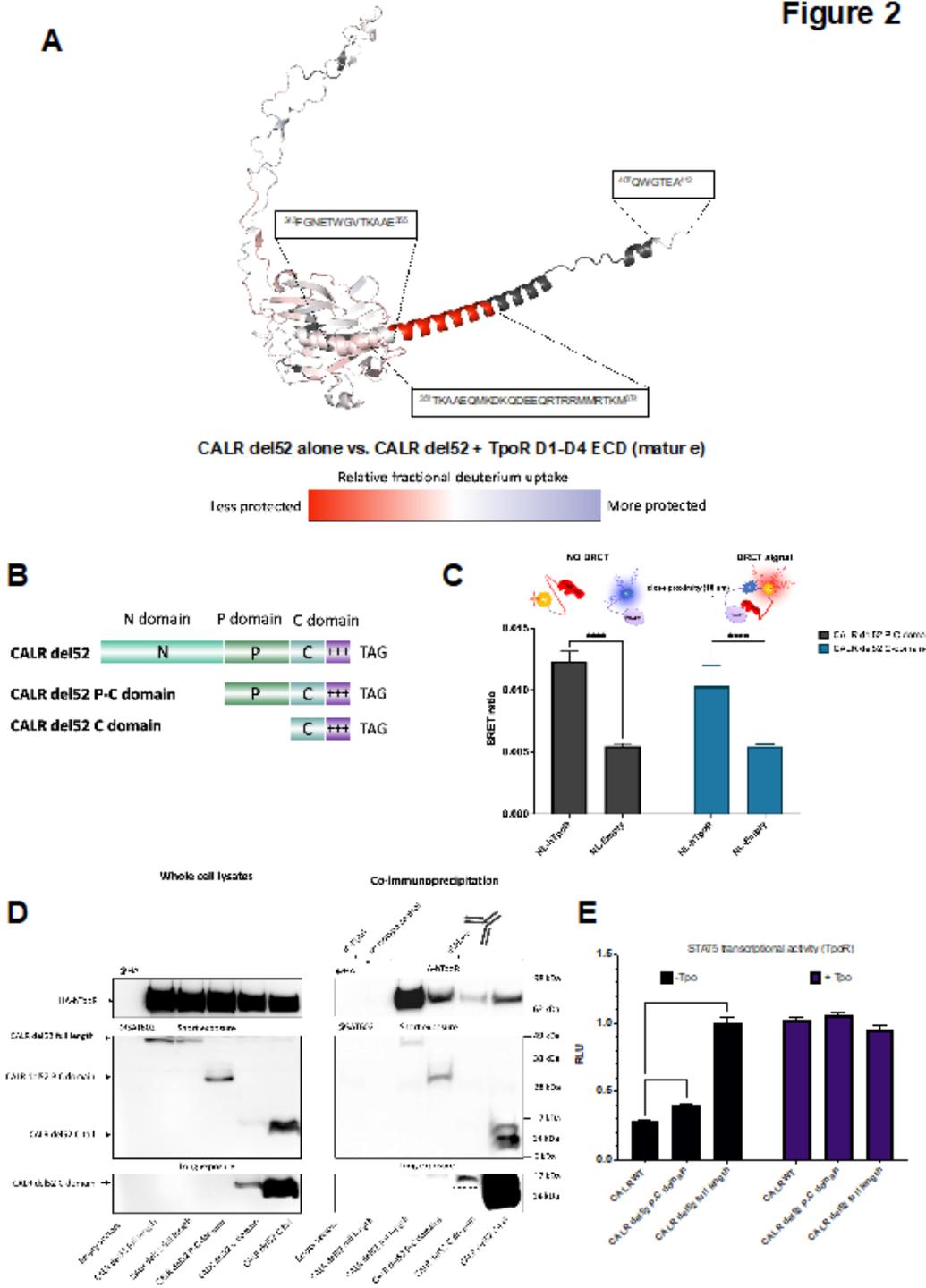


Figure 2

A. Relative fractional deuterium uptake differential between CALR del52 alone or after incubation with TpoR extracellular domain at 1:1 molar ratio. The scale represents the relative differential of deuterium incorporation between CALR del52 alone and CALR del52-TpoR ECD complex. The structure was computed using the AlphaFold 2.0¹⁶. Shown is the differential exchange at 1 hour incubation in deuterium (mean of triplicate experiments). Regions in dark grey represents regions not covered by

detected peptides at the given time point. **B.** Representation of N-terminal truncations of CALR del52 fused to either a FLAG tag or a HaloTag. **C.** NanoBRET between NanoLuc-TpoR and CALR del52-HaloTag truncated from the N-terminus. Shown are average 5 independent experiments (

SD) (N = 5, n = 20). Two-ways ANOVA followed by SIDAK multiple comparison test. ****: $p < 0.0001$, ns: non-significant. **D.** Representative co-immunoprecipitation of HA-hTpoR with CALR del52-FLAG full length or N-terminal truncations as indicated. **E.** STAT5 transcriptional activity with hTpoR and indicated CALR truncations. HEK293T were transiently transfected with HA-hTpoR and CALR truncations along with cDNAs coding for STAT5, JAK2 and SpiLuc Firefly luciferase reporter reflecting STAT5 transcriptional activity and normalized with a control reporter (pRLTK) containing Renilla luciferase. Shown are average 3 independent experiments (

SD) performed in triplicates (N = 3, n = 9). Two-ways ANOVA followed by SIDAK multiple comparison test. ***: $p < 0.001$, ****: $p < 0.0001$.

Figure 3

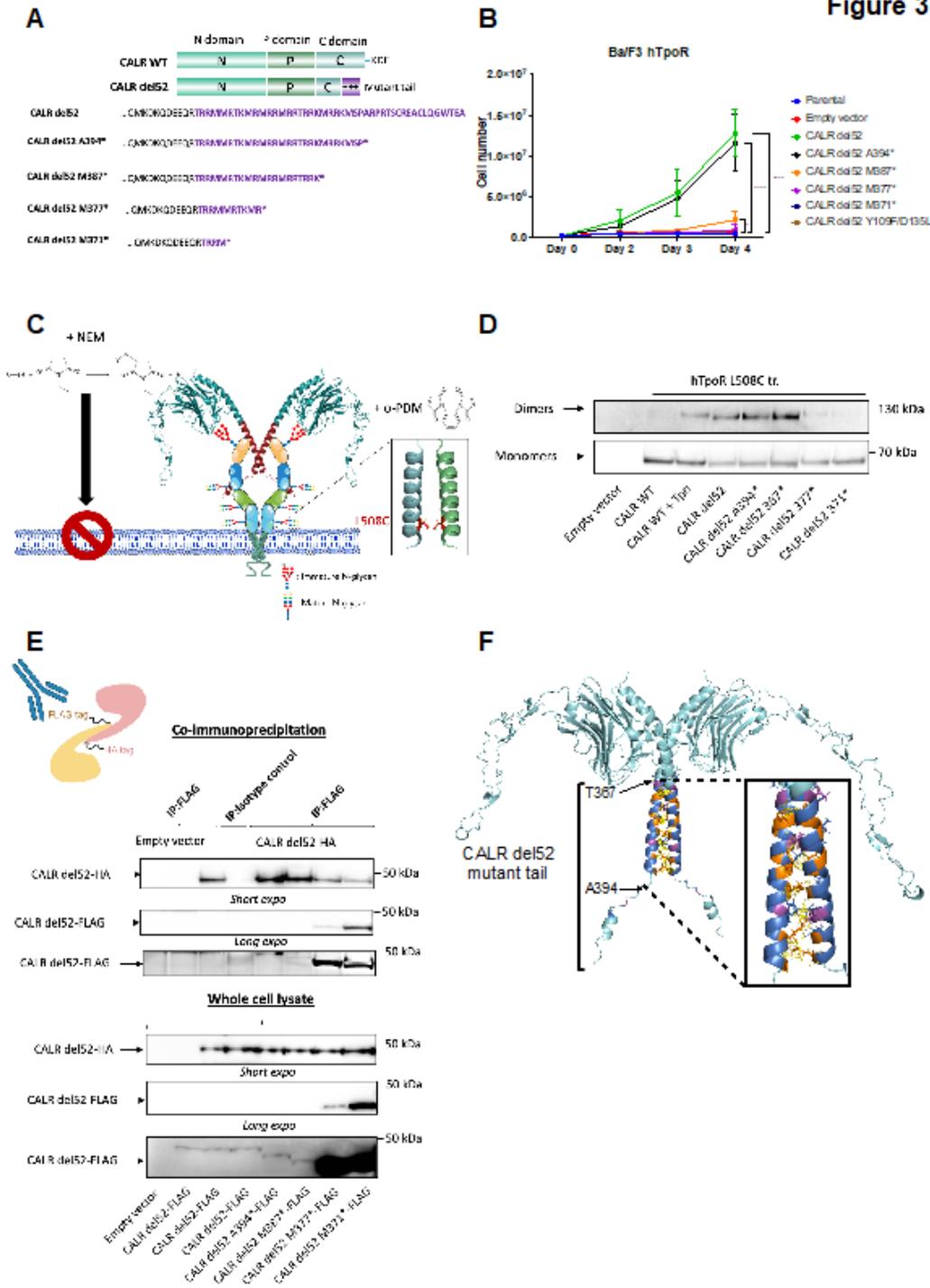


Figure 3

A. Representation and sequence of CALR del52 C-terminal truncations used in proliferation assay (Figure 1B). **B.** Proliferation assay. BaF3 cells stably expressing hTpoR in pMX-IRES-GFP were infected with indicated CALR variants or an empty vector (pMSCV-IRES-mCherry) and sorted by FACS. 250,000 were washed and seeded in 10ml of complete culture medium without cytokine and counted each day using an automated cell counter (n = 3). Values represent average of 3 independent experiments (

SD) with technical triplicates (n = 6-9). **C.** Cartoon representation of the crosslinking assay to assess homo-dimerization of TpoR in a productive orientation. o-PDM: ortho-phenylene dimaleimide. NEM: N-ethylmaleimide. **D.** Crosslinking study of hTpoR dimerisation in presence of Tpo, CALR del52 full length of C-terminal truncations. Shown is a representative western blot in denaturing and reducing conditions showing hTpoR monomers and o-PDM crosslinked dimers in the indicated conditions. **E.** Co-immunoprecipitation of CALR del52-HA full length by CALR del52-FLAG full length or truncated to assess dimerization. Shown are representative western blots in denaturing conditions. **F.** RosettaDock top scoring simulation of CALR del52 dimers. Structure of CALR del52 was modelled by AlphaFold 2.0. Residues of the mutant tail are shown in purple (Thr), dark blue (Arg) and orange (Met). Interactions < 3Å are shown by yellow dashed lines.

Figure 4

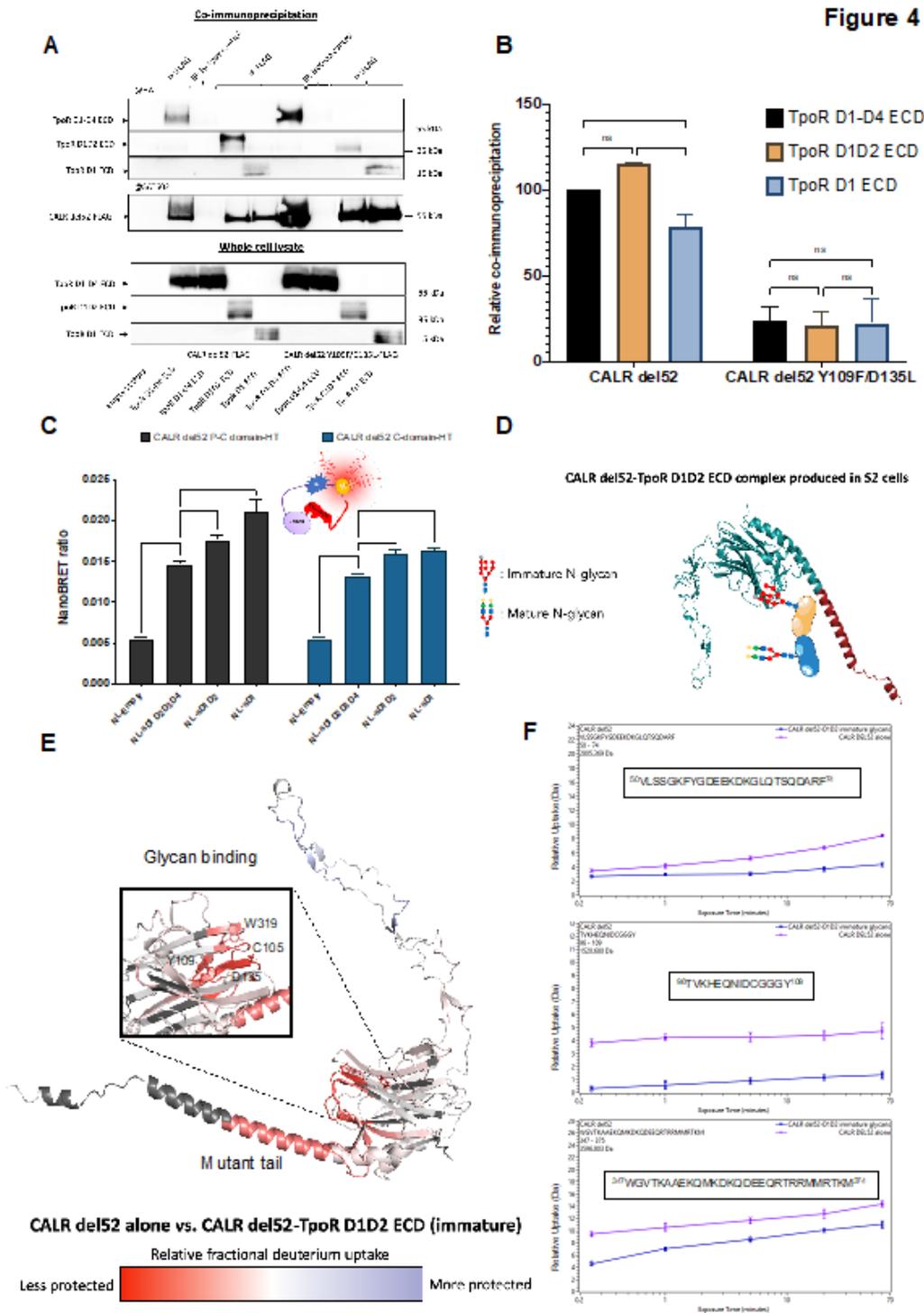


Figure 4

A. Representative co-immunoprecipitation of HA-hTpoR ECD domains with CALR del52-FLAG or CALR del52 Y109F/D135L-FLAG as indicated. **B.** Quantification of relative co-immunoprecipitation of TpoR species by CALR del52 (mutated or not). Western blot quantification performed with ImageJ. Shown are the ratios (+ SD) of TpoR species on CALR del52 normalized for TpoR species expression in whole lysates (N = 2, n = 6). Two-ways ANOVA followed by SIDAK multiple comparison test. ***: $p < 0.001$, *: $p < 0.05$ ns:

non-significant. **C.** NanoBRET between NanoLuc-TpoR subdomains and CALR del52-HaloTag truncated from the N-terminus. Shown are average 3 independent experiments (

SD) (N = 3, n = 12). **D.** Cartoon representing the complex between CALR del52 and TpoR D1D2 domain containing immature glycans produced in Schneider (S2) cells. **E.** Relative fractional deuterium uptake differential between CALR del52 alone and CALR del52-D1D2 complex. The scale represents the relative differential of deuterium incorporation between CALR del52 alone and CALR del52-D1D2 complex. The structure was computed using the AlphaFold 2.0 webserver¹⁶. Shown is the differential exchange at 1 hour incubation in deuterium (mean of triplicate experiments). Regions in dark grey represents regions not covered by detected peptides at the given time point. **F.** Deuterium uptake of representative peptides from CALR del52 alone or CALR del52-D1D2 complex corresponding to key regions described in the text. The first peptide corresponds to the amino acids 50-74 of the N-domain. The second peptide contains residues C105 and Y109 involved in glycan binding. The last peptide contains the mutant c-terminus.

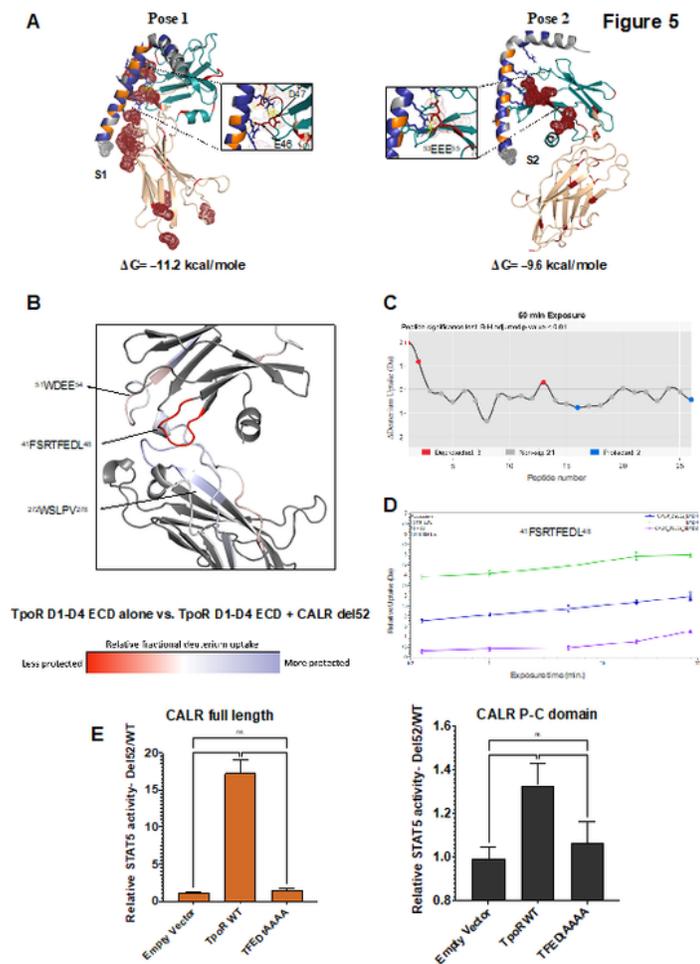


Figure 5

A. Left: Pose labelled as pose 1 generated by HADDOCK between CALR del52 mutant tail and hTpoR D1D2 domains. The best docking complexes were chosen, as ranked by the HADDOCK score. Highlighted is the strong interaction between Arg of CALR mutant tail and E46 and D47 of TpoR D1 domain. Right: Pose labelled as pose 1 generated by HADDOCK between CALR del52 mutant tail and hTpoR D1D2 domains. Highlighted is the interaction between Arg of CALR mutant Tail and the ⁵³EEE⁵⁶ motif on TpoR

D1. Left and right: TpoR D1 is shown is light blue and TpoR D2 in light orange. Acidic residues of TpoR are shown in red and the S1 region is shown with red dots. CALR del52 mutant basic residues (Arg) are shown in dark blue, hydrophobic residues (Met) are shown in orange. The computed free energy for both configurations are indicated. **B.** Relative fractional deuterium uptake differential between the indicated species. The scale represents the relative differential of deuterium incorporation between the left and right species. Regions in dark grey represents regions not covered by detected peptides at the given time point. **C.** Relative deuterium uptake analysis between TpoR full ECD with mature glycans in absence or presence of CALR del52 at 1 hour incubation in deuterium. Dots indicated in blue (protected) show significant differential deuterium intake ($p < 0.001$) with the peptide-level significance testing (hybrid mode) as described ³¹. **D.** Deuterium uptake of the FSRTFEDL peptide peptides from mature TpoR full ECD (D1D4) alone, the CALR-TpoR D1D4 (mature) complex (CALR_del52_D1D4) or the CALR del52-TpoR D1D2 (immature) complex (CALR del52_D1D2) at 5 different exchange time points. **E.** STAT5 transcriptional activity with empty vector, TpoR WT or TFEDtAAAA mutant. The scale represents the fold change between the STAT5 activity of indicated TpoR species in presence of CALR del52 (full length or P-C domain) and CALR WT (full length or P-C domain). and indicated CALR truncations Shown are average of 2-3 independent experiments (

SD) (N = 2-3, n = 6-9). One-way ANOVA followed by SIDAK multiple comparison test. ns: non-significant. ****: $p < 0.0001$.

Figure 6

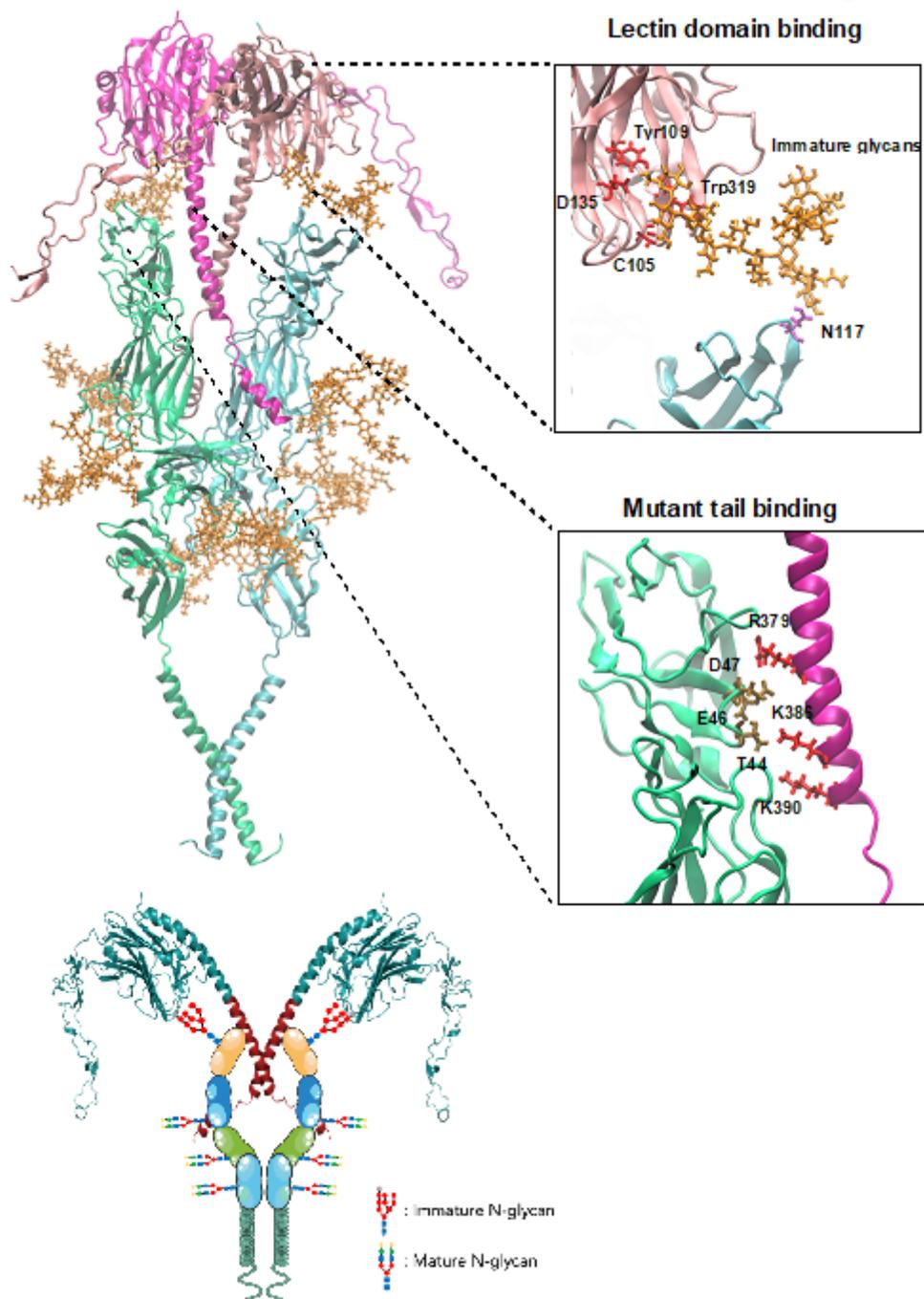


Figure 6

Model of the CALR del52-TpoR tetrameric complex. Top: The structure of CALR del52 monomer was computed with AlphaFold 2.0¹⁶. TpoR extracellular domain was modelled as described in methods and was overall in good agreement with the one generated with AlphaFold 2.0. Mature glycans were added to Asn178, Asn298 and Asn358, and immature glycans whose composition was previously determined¹² were added to Asn117. TpoR monomers were put into POPC lipidic bilayers and dimers of CALR del52 were further docked to TpoR with binding sites determined by HDx-MS (see methods). The figure shows the lectin binding domain with key residues involved in glycans interactions, and the mutant tail binding

region involving the ⁴⁴TFED⁴⁷ motif as determined by HDx-MS. Bottom: Cartoon representation of the tetrameric complex between CALR del52 and TpoR containing immature glycans linked to Asn117.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [Supplementaryfile1.xlsx](#)
- [VideoS1.mp4](#)
- [VideoS2.mp4](#)
- [VideoS3.mp4](#)
- [VideoS4.mpg](#)
- [SupplementaryFigures.docx](#)