Oncogenic CALR mutant C-terminus mediates dual binding to the thrombopoietin receptor triggering complex dimerization and activation

Nicolas Papadopoulos, Audrey Nédélec, Allison Derenne, Teodor Asvadur Şulea, Christian Pecquet, Ilyas Chachoua, Gaëlle Vertenoeil, Thomas Tilmant, Andrei-Jose Petrescu, Gabriel Mazzucchelli, et al.

nature communications

Article

https://doi.org/10.1038/s41467-023-37277-3

Oncogenic CALR mutant C-terminus mediates dual binding to the thrombopoietin receptor triggering complex dimerization and activation

Received: 11 May 2022

Accepted: 4 March 2023

Published online: 05 April 2023

Check for updates

Nicolas Papadopoulos[1,2], Audrey Nédélec[1,2], Allison Derenne[3], Teodor Asvadur Şulea[4], Christian Pecquet[1,2], Ilyas Chachoua[1,2,5], Gaëlle Vertenoeil[1,2], Thomas Tilmant[6], Andrei-Jose Petrescu[4], Gabriel Mazzucchelli[6], Bogdan I. Iorga[7], Didier Vertommen[2,8] & Stefan N. Constantinescu[1,2,9,10]

Calreticulin (CALR) frameshift mutations represent the second cause of myeloproliferative neoplasms (MPN). In healthy cells, CALR transiently and non-specifically interacts with immature N-glycosylated proteins through its N-terminal domain. Conversely, CALR frameshift mutants turn into rogue cytokines by stably and specifically interacting with the Thrombopoietin Receptor (TpoR), inducing its constitutive activation. Here, we identify the basis of the acquired specificity of CALR mutants for TpoR and define the mechanisms by which complex formation triggers TpoR dimerization and activation. Our work reveals that CALR mutant C-terminus unmasks CALR N-terminal domain, rendering it more accessible to bind immature N-glycans on TpoR. We further find that the basic mutant C-terminus is partially α-helical and define how its α-helical segment concomitantly binds acidic patches of TpoR extracellular domain and induces dimerization of both CALR mutant and TpoR. Finally, we propose a model of the tetrameric TpoR-CALR mutant complex and identify potentially targetable sites.

Myeloproliferative neoplasms (MPNs) are blood malignancies driven by the acquisition of somatic mutations in hematopoietic stem cells[1]. Frameshift mutations in the endoplasmic reticulum (ER) resident chaperone calreticulin (CALR) are the main cause of $JAK2^{V617F}$ negative MPNs and are responsible for ~25% of Essential Thrombocythemia (ET) and myelofibrosis cases[2,3]. The most common CALR frameshift mutations in MPN are a 52-bp deletion denoted CALR del52 (type 1) and a 5-bp insertion called CALR ins5 (type 2), but all referenced mutations lead to the replacement of the wild-type C-terminus and KDEL ER-retention motif by a new sequence rich in methionine and positively charged residues[2,3]. These CALR mutants acquire the ability to specifically bind and activate the thrombopoietin receptor

[1]Ludwig Institute for Cancer Research Brussels, Brussels, Belgium. [2]Université catholique de Louvain and de Duve Institute, Brussels, Belgium. [3]Spectralys Biotech SRL, rue Auguste Piccard 48, 6041 Gosselies, Belgium. [4]Department of Bioinformatics and Structural Biochemistry, Institute of Biochemistry of the Romanian Academy, Splaiul Independentei 296, Bucharest 060031, Romania. [5]Department of Molecular Biology and Genetics, Bilkent University, Ankara, Turkey. [6]Mass Spectrometry Laboratory, MolSys Research Unit, Universiy of Liège, 4000 Liège, Belgium. [7]Université Paris-Saclay, CNRS, Institut de Chimie des Substances Naturelles, UPR 2301 Gif-sur-Yvette, France. [8]de Duve Institute and MASSPROT platform, Brussels, Belgium. [9]Walloon Excelence in Life Sciences and Biotechnology, WELBIO, avenue Pasteur, 6, 1300 Wavre, Belgium. [10]Ludwig Institute for Cancer Research, Nuffield Department of Medicine, Oxford University, Oxford, UK. e-mail: Stefan.constantinescu@bru.licr.org

(TpoR), resulting in constitutive activation of the JAK-STAT pathway[4–8]. The binding of CALR mutants to TpoR relies notably on the interaction between CALR N-domain and immature N-glycans on TpoR[8,9]. Yet, this type of interaction is not specific to TpoR as wild-type CALR naturally associates via its N-domain with thousands of immature N-glycosylated proteins in a cycle of attachment and liberation that ends once the protein achieves proper folding[10]. In addition, the deletion of CALR mutant C-terminus results in reduced binding to TpoR and loss of activation of the JAK-STAT pathway[4,11,12]. These observations suggest that the interaction between mutant CALR and TpoR derives from novel properties acquired by CALR frameshift mutants that go beyond the canonical interaction between CALR N-domain and immature N-glycosylated proteins[10]. In this study, we used a multidisciplinary approach to uncover the basis for the specific and stable interaction between CALR mutants and TpoR and unveil how this interaction leads to productive dimerization and activation of the TpoR. Understanding how frameshift mutations in a master chaperone result in novel binding capacities is of deep interest both conceptually and therapeutically as the detailed characterization of binding and activation mechanisms is required for the development of therapeutic inhibitors. This study further provides a complete model of the CALR mutant-TpoR complex, paving the way for the development of therapeutic avenues.

Results

Frameshift mutations in CALR C-terminus unmask its N-glycan binding domain

The stability of the interaction between CALR N-domain and immature N-glycans on TpoR[4,5,8,11], our observation that CALR del52 exhibits a lower thermal stability than CALR WT[13] (see also Supplementary Fig. 1a) and the fact that both CALR wild-type N- and mutant C-domains are required for TpoR binding[4,11,12] together suggest that CALR frameshift mutations affect the structure of the whole protein. We therefore sought to compare the conformational footprints of CALR WT and mutants using hydrogen-deuterium exchange mass spectrometry (HDx-MS), a technique which uses mass shifts in peptides from a protein after hydrogen-deuterium (H-D) exchange in backbone amide positions to provide a readout of residue accessibility and protein conformation. CALR del52 (type 1) was chosen as the representative of CALR mutants that all acquire a very similar novel C-terminus[2,3]. To delineate the effect of the mutant C-terminus, we created CALR ΔC-tail. This variant contains the complete N-domain (residues 18–197), the proline-rich P-domain (residues 198–308) and part of the C-domain (residues 309–366) but not the C-terminal fragment (C-tail) that differs between CALR WT and CALR del52 (Fig. 1a, b). Similar deletions are frequent in a variety of solid-tumor cancers and are associated with immunosuppressive activity[14]. The three CALR variants (CALR WT, CALR del52 and CALR ΔC-tail) were produced as recombinant proteins and their purity and proper folding were validated by thermal shift, Coomassie blue staining and chromatography (Supplementary Fig. 1a, b; Supplementary Fig. 2). The HDx-MS footprints of CALR WT and variants were acquired with a sequence coverage of 92.6% (Supplementary Fig. 3a, b). Our analysis revealed that the absence of the last 50 amino acids of CALR (as in CALR ΔC-tail) did not drastically alter the conformation and accessibility of the rest of the protein. Only a small fragment of CALR N-domain was less protected in CALR ΔC-tail compared to CALR WT (Fig. 1c; Supplementary Fig. 3c). In sharp contrast, CALR del52 exhibited a globally more accessible conformation, except for fragments of the P-domain that were less accessible after 0.25 min incubation in deuterium but not at longer time points (Fig. 1d; Supplementary Fig. 3d). Critically, the N-domain of CALR del52 was the most affected by the addition of the mutant C-terminus and displayed a strong increase in H-D exchange compared to CALR WT. This sharp increase in accessibility was notably observed in residues involved in direct interactions with immature N-glycans such as C105 and W319[8,15] (Fig. 1d), indicating that the region involved in binding immature N-glycans is unmasked due to the presence of the CALR del52 C-terminus.

CALR mutant C-terminus contains two segments with distinct secondary structure

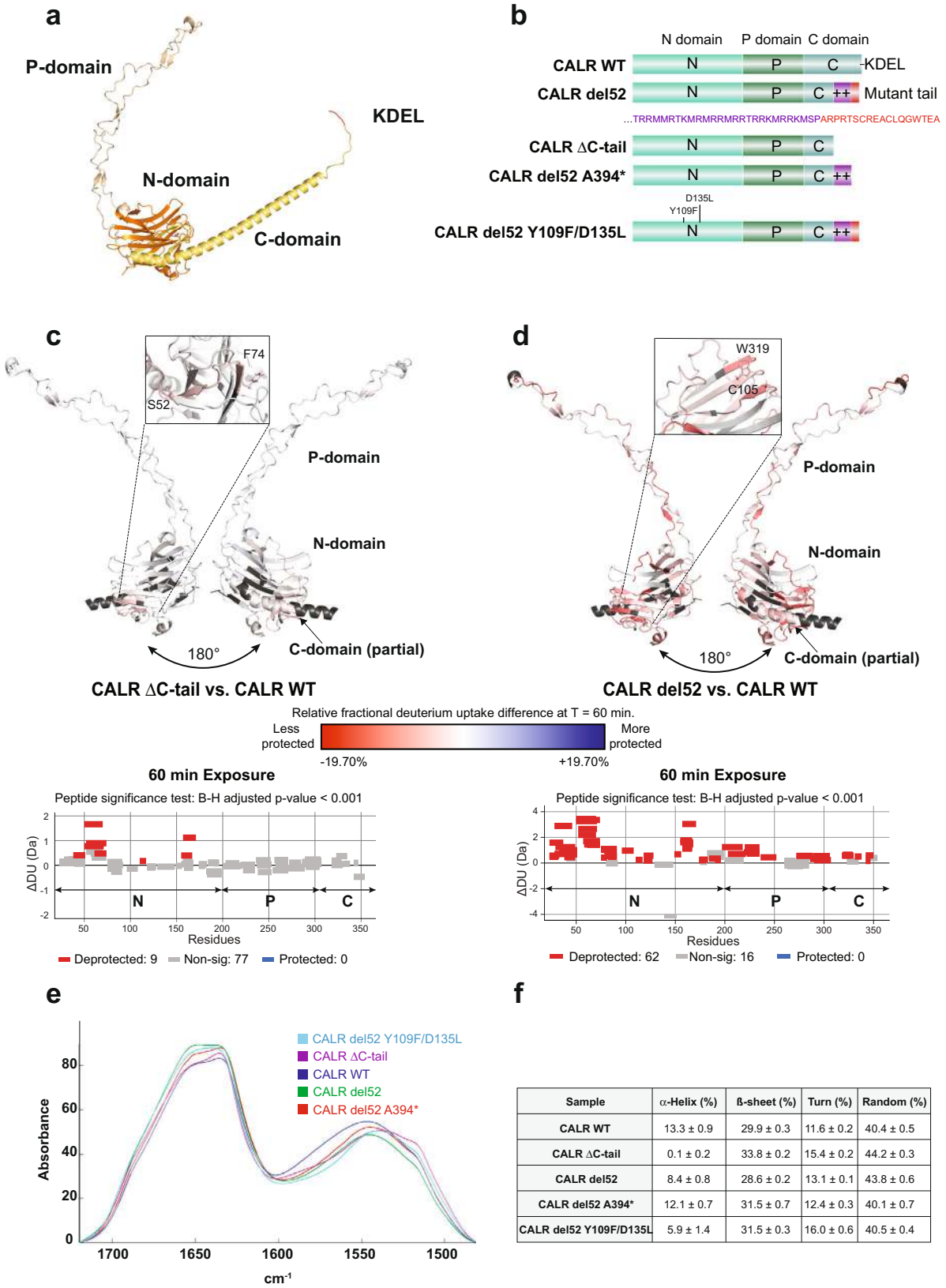
Next, we used Fourier transform infrared spectroscopy (FTIR) to define how the frameshift mutation in CALR del52 influences its secondary structure (Supplementary Figs. 4 and 5). Analysis of the amide I vibration revealed that the protein has high ß-sheet and random coil content, which is abundant in the N- and P-domains[16,17] respectively, and that α-helical content was higher in CALR WT (13.3%) compared with CALR del52 (8.4%) ($p < 0.001$). The FTIR spectra of CALR ΔC-tail revealed that deletion of the last 50 residues resulted in an almost complete loss of α-helical content (0.1%), indicating that helicity is concentrated in the C-terminus of both wild-type and mutant CALR (Fig. 1e, f). The C-terminus of CALR mutant can be further separated in two segments based on amino acid composition. The proximal segment is rich in hydrophobic (Met) and basic (Arg, Lys) residues (Fig. 1b, purple) while the C-terminal segment starting at A394 (Fig. 1b, red) has a more heterogenous amino acid composition. FTIR spectra indicated that deletion of the last 18 residues (as in CALR del52 A394*) resulted in relative increase in α-helix content (12.1%), revealing that helicity was concentrated in the proximal segment of the mutant C-terminus rich in Arg and Met, in line with in silico prediction[18] (Supplementary Fig. 1d). Because ß-sheets are concentrated in the N-domain[17] (Fig. 1a) and α-helices present only in the C-terminus, these changes coupled to our HDx-MS data indicate that the acquisition of frameshift mutations in the C-terminus of CALR del52 disturbs its secondary structure and increases accessibility of CALR del52 N-domain.

Remarkably, introduction of two point mutations, (Y109F/D135L) in the N-domain of CALR del52 that abolish immature N-glycan binding[8,15] also led to decreased helicity (from 8.4 to 5.9%) and of random coil percentage with a compensatory increase in turns and ß-sheets (Fig. 1e, f), suggesting that the different domains of CALR are conformationally linked.

CALR mutant C-terminus directly interacts with a mature form of TpoR extracellular domain

Although the higher accessibility of CALR del52 N-domain could explain its capacity to form a stable interaction with immature N-glycans of TpoR, the specificity of CALR mutant for TpoR versus other cytokine receptors[8] suggested that other binding mechanisms could be at play. To probe this assumption, we produced and purified recombinant TpoR extracellular domain (ECD) labeled TpoR D1-D4 (Supplementary Figs. 2g and 6a, b) that contains mature N-glycans[8] to avoid any generic interaction between CALR N-domain and immature N-glycans. The binding affinity between the mature TpoR ECD and CALR del52 was evaluated using microscale thermophoresis to ~104 nM (Supplementary Fig. 6c). Then, we set up a HDx-MS experiment to determine which regions of CALR del52 interacted with TpoR ECD in absence of immature N-glycans. Considering the level of affinity between the two partners, the proteins were incubated at 1:1 molar ratio (at 20 µM each) prior to experiments. At this concentration, the percentage of complex at steady state is of ~93% given a ~104 nM affinity. The HDx-MS footprint of CALR del52 was acquired with a sequence coverage of 89.6% (Supplementary Fig. 6d). Comparison of the H-D exchange between CALR del52 alone or in presence of TpoR ECD revealed significant ($p < 0.01$) H-D exchange differential in different peptides containing CALR del52 C-terminus (Fig. 2a and Supplementary Fig. 6e, f), indicative of a direct interaction with the mature TpoR ECD. This differential exchange was not observed in the C-terminal extremity of the mutant C-terminus encompassing residues ^{406}QGWTEA^{411} (Supplementary Fig. 6e, f) but only in the α-helical segment containing positively charged residues.

https://doi.org/10.1038/s41467-023-37277-3



Sample	α-Helix (%)	ß-sheet (%)	Turn (%)	Random (%)
CALR WT	13.3 ± 0.9	29.9 ± 0.3	11.6 ± 0.2	40.4 ± 0.5
CALR ΔC-tail	0.1 ± 0.2	33.8 ± 0.2	15.4 ± 0.2	44.2 ± 0.3
CALR del52	8.4 ± 0.8	28.6 ± 0.2	13.1 ± 0.1	43.8 ± 0.6
CALR del52 A394*	12.1 ± 0.7	31.5 ± 0.7	12.4 ± 0.3	40.1 ± 0.7
CALR del52 Y109F/D135L	5.9 ± 1.4	31.5 ± 0.3	16.0 ± 0.6	40.5 ± 0.4

Fragments of the N-domain predicted to be conformationally close to the C-domain also exhibited significant, albeit smaller, decreased accessibility in presence of the mature TpoR ECD (Supplementary Fig. 6e). Thus, CALR mutant C-terminus interacts with TpoR in the absence of the interaction between the N-terminus of CALR del52 and immature N-glycans. Importantly, in this work we show later that the interaction is maintained also when the N-domain interacts with immature N-glycans.

To validate this interaction in living cells, we used Nano-bioluminescence energy transfer (NanoBRET) where close proximity (<10 nm) between a bioluminescent energy donor (NanoLuc) and a fluorescent energy acceptor (HaloTag) results in energy

Fig. 1 | Structural changes induced by CALR frameshift mutation. a Structure of full length CALR WT predicted using AlphaFold 2.0[18]. The N-domain is shown in orange, the P-domain in wheat, the C-domain in yellow and the KDEL in red. **b** Representation of the domains and C-terminal sequences of CALR WT and CALR del52 or variants thereof used in FTIR spectroscopy and HDx-MS experiments. **c**, **d** Top: Structure model (AlphaFold 2.0[18]) of the common region between CALR WT and CALR del52 (corresponding to CALR ΔC-tail). Colors represent the difference in relative fractional uptake (ΔRFU) between CALR ΔC-tail and CALR WT (**c**) or between CALR del52 and CALR WT (**d**) at 1 h incubation in deuterium. Regions in red and blue are respectively less and more protected in CALR ΔC-tail (**c**) or CALR del52 (**d**) compared to CALR WT. The scale from red to blue is proportional to the ΔRFU between indicated CALR species with dark red and dark blue corresponding to highest differential. Dark gray represents regions without peptide coverage for the given time point. The N-glycans binding domain of CALR which is more exposed in CALR del52 compared to CALR WT is highlighted. Raw data are provided in the source file. Bottom: Wood's plots generated with Deuteros 2.0[35]. Each bar (wood) represents the H-D exchange differential for a single peptide between indicated CALR species at 1 h incubation in deuterium. Peptides in red (deprotected) or blue (protected) have significant differential H-D exchange ($p < 0.001$) with the peptide-level significance testing as described[35] ($n = 3$). The N-, P- and C-domains of CALR are indicated on the plots by letters N, P and C, respectively. Source data are provided as a Source data file. **e** Comparison of the mean spectra recorded for each sample to analyze the protein secondary structure. These spectra have been baseline-corrected and normalized. Each sample is identified by a unique color indicated in the legend. The unprocessed spectra are provided in Supplementary Fig. 4a. **f** Secondary structure predictions using the method developed on our in-house database (see Supplementary Methods). 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 ($n = 5$). For the present predictions, the standard error of prediction in cross-validation is 5.7% for the α-helix and 6.7% for the ß-sheet, 3.2% for turns and 8% for random coil.

transfer measured by the BRET ratio. We measured the BRET ratios between fragments of CALR del52-HaloTag and NanoLuc-TpoR as we did before for full length CALR del52[8]. We deleted the N-domain or both N- and P-domains of CALR del52 to create the constructs labeled P-C and C-domain, respectively (Fig. 2b), which do not contain the N-domain required to bind immature N-glycans[15,19]. Both truncated forms of CALR del52 retained significant interaction with TpoR in living cells as measured by the BRET ratio (Fig. 2c; Supplementary Fig. 7). This interaction was further validated by co-immunoprecipitation between FLAG-tagged N-terminal truncations of CALR del52 and HA-tagged TpoR (Fig. 2d). Then, we questioned whether CALR del52 devoid of N-glycan binding domain remained able to induce TpoR activation. We used a luciferase assay[20] to measure the STAT5-dependent transcriptional activity in cells co-expressing TpoR and CALR del52 P-C domain. Remarkably, the latter conserved the ability to induce a significant induction of STAT5 transcriptional activity in presence of TpoR (Fig. 2e) but not of the erythropoietin receptor (EpoR) (Supplementary Fig. 6g), indicating that the deletion of CALR mutant N-domain does not completely inhibit its ability to specifically bind and activate TpoR.

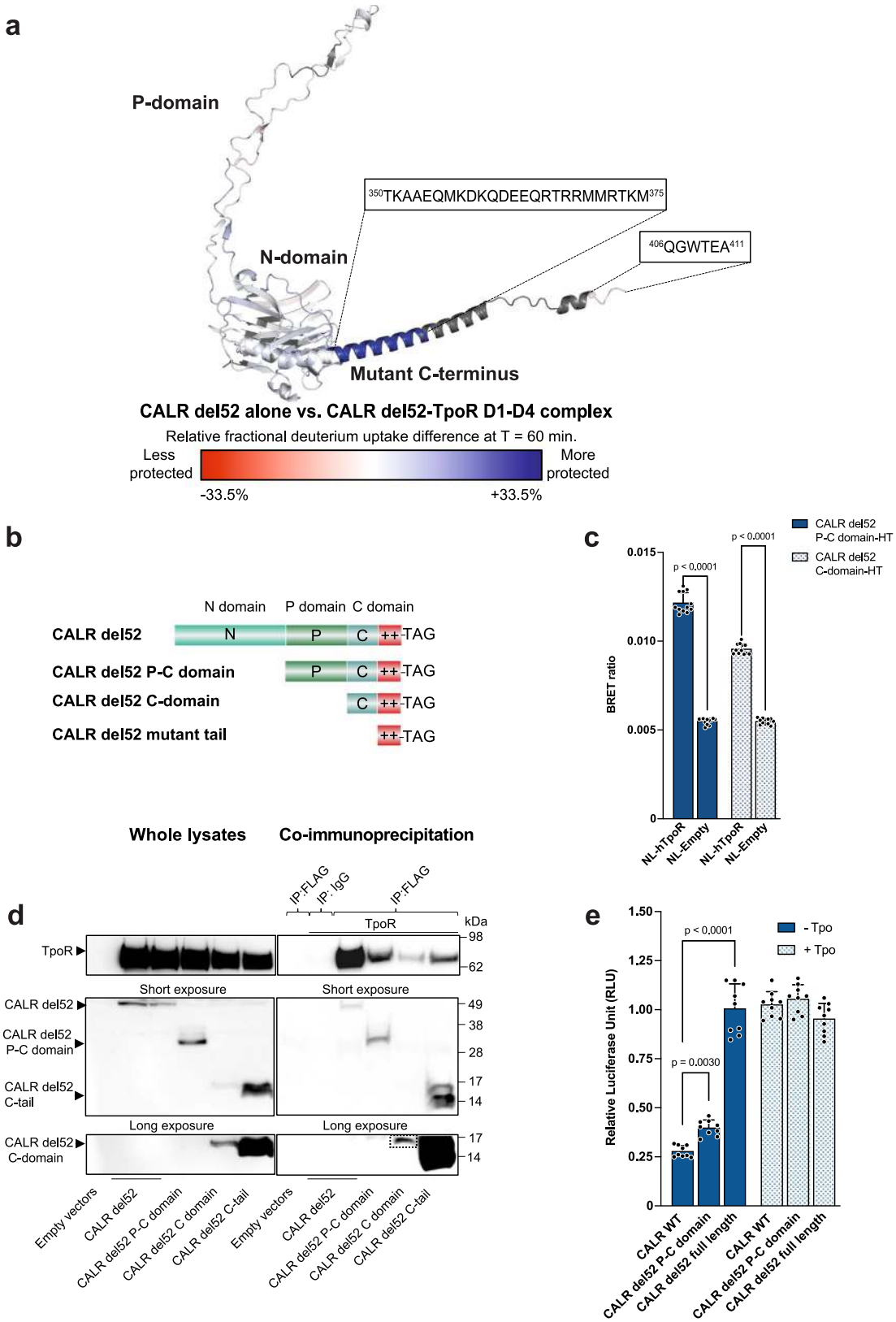
Mapping of interactions in the TpoR-CALR mutant complex

Having established that the C-terminus of CALR mutant directly interacts with TpoR ECD, we sought to identify the region of TpoR involved in this interaction. The TpoR ECD is composed of four subdomains labeled from D1 to D4 starting from the N-terminus. To determine which of these subdomains is involved in binding to CALR mutant, we first used co-immunoprecipitation between CALR del52 and progressive truncations of TpoR ECD starting from the C-terminal D4 fragment. In these conditions where TpoR fragments and CALR del52 are expressed in the same cell, TpoR fragments retain immature N-glycans[8]. Deletions of the D3D4 domains of TpoR did not reduce interaction with CALR del52 while deletion of D2 resulted in a modest decrease in co-immunoprecipitation compared to D1 alone (Fig. 3a, b), indicating that binding of TpoR to CALR del52 occurs essentially via the D1 domain. This is reminiscent of our previous finding that immature N-glycans attached to Asn117 of D1 are the major site of CALR del52 binding via the N-domain[5,8]. Given our result that CALR mutant C-terminus binds TpoR also in presence of mature N-glycans (Fig. 2), we additionally probed binding of TpoR ECD fragments to the CALR del52 Y109F/D135L double mutant that is deficient for binding immature N-glycans[8,15]. Expectedly, loss of N-glycan-dependent interaction led to a sharp decrease in co-immunoprecipitation ratios (Fig. 3b). However, unlike with non-mutated CALR del52, the interaction of CALR del52 Y109F/D135L was similar between D1 and D1D2 ECD fragments (Fig. 3a, b), suggesting that interaction between CALR mutant C-terminus and TpoR occurs essentially via the D1 domain. Noteworthy, the TpoR D1D2 species that co-immunoprecipitated with CALR del52 Y109F/D135L had a smaller apparent molecular size than that interacting with non-mutated CALR del52 (Fig. 3a). This small size shift, visible only for TpoR D1D2 due to better resolution in this part of the gel, correlates with the fact that the TpoR D1D2 that interacts with non-mutated CALR del52 retains immature N-glycans[8], unlike the TpoR D1D2 mature species that binds the CALR del52 Y109F/D135L double mutant that is deficient for N-glycan binding. To confirm this in live cells, we used our NanoBRET assay as in Fig. 2c to measure binding between CALR del52 P-C or C-domain-HaloTag and fragments of NanoLuc-TpoR ECD. Confirming results from co-immunoprecipitations, the interaction between all fragments of TpoR ECD and CALR del52 P-C and C-domain was conserved (Fig. 3c). Strikingly, the interaction between CALR del52 P-C or C-domain and TpoR ECD was even increased in D1 and D1D2 compared to the full TpoR ECD, possibly indicating that removing the C-terminal segments of TpoR ECD improves accessibility of the D1 domain to CALR mutant C-terminus.

CALR mutant interacts with TpoR via two major domains

In the physiologically relevant CALR mutant-TpoR complex, TpoR retains immature N-glycans attached to Asn117[8]. We thus sought to assess whether the presence of immature N-glycans on TpoR affected the H-D exchange profile of CALR del52 in presence of TpoR ECD and to identify the binding sites of CALR del52 with immature N-glycans on TpoR. The recombinant CALR del52-TpoR D1D2 complex was produced in S2 cells, and its purity was verified by thermal shift assay, Coomassie blue staining and size-exclusion chromatography (Fig. 3d; Supplementary Figs. 2f and 8a, b). In this complex, immature N-glycans are attached to Asn117[8]. Comparison of the H-D exchange profile between CALR del52 alone or in complex with TpoR D1D2 with immature N-glycans (sequence coverage of ~90%, Supplementary Fig. 8c) indicated that CALR del52 interacted with TpoR via two major domains. The strongest H-D differential was observed in the putative N-glycans binding site of CALR (Fig. 3e; Supplementary Fig. 8d). This region included notably C105, Y109, D315 and W319 that were reported to be key for binding of immature N-glycans[8,19,21,22] and were all more protected in presence of TpoR. Importantly, this region is not involved in binding to mature TpoR (Fig. 2a), indicating that these residues are specifically involved in the interaction with immature N-glycans. The second major H-D differential was present in peptides containing the mutant C-terminus of CALR del52 which exhibited strong protection in the TpoR-CALR mutant complex compared to CALR mutant alone (Fig. 3e; Supplementary Fig. 8d, e). These peptides were similar to the ones exhibiting differential H-D uptake between CALR del52 alone or in complex with mature TpoR (Fig. 2; Supplementary Fig. 6e, f), indicating that CALR mutant C-terminus interacts with both mature and immature forms of TpoR ECD. Interestingly, fragments of the P-domain were significantly more accessible in the CALR mutant-TpoR complex (Figs. 2a, 3e and

https://doi.org/10.1038/s41467-023-37277-3



Supplementary Figs. 6e, 8d). This observation suggests that in absence of TpoR, the P-domain may interact with CALR N- or C-domain and that this interaction is destabilized upon binding to TpoR. This hypothesis is also supported by previous reports that the deletion of CALR mutant P-domain improves the binding of CALR mutant to TpoR[4].

CALR mutant C-terminus interacts with acidic patches on TpoR D1 domain

To study with more precision the residues of TpoR ECD that may interact with CALR mutant C-terminus, we used HDx-MS with the same set-up as in Fig. 2a, where we showed that CALR del52 interacts with mature TpoR exclusively through the mutant C-terminus. H-D

Fig. 2 | Interaction between CALR mutant C-terminus and TpoR independently of N-glycans. a Structure model (AlphaFold 2.0[18]) of CALR del52. Colors represent the difference in relative fractional uptake (ΔRFU) between CALR del52 alone and CALR del52 in complex with TpoR D1-D4 with mature N-glycans. Regions in red and blue are respectively less and more protected in the CALR del52-TpoR D1-D4 complex compared to CALR del52 alone. The scale from red to blue is proportional to the ΔRFU at 1 h incubation in deuterium between indicated species (mean of $n = 3$ experiments), with dark red and dark blue corresponding to highest differential. Dark gray represents regions without peptide coverage for the given time point. Sequences of the CALR mutant C-terminus are highlighted. Source data are provided as a Source data file. **b** Representation of N-terminal truncations of CALR del52 fused to either a FLAG tag or a HaloTag at the C-terminus. **c** NanoBRET between NanoLuc-TpoR (NL-TpoR) or NanoLuc not fused to any protein (NL-Empty) and CALR del52 P-C or C-domain-HaloTag. Data represent mean ± SD ($n = 12$ biologically independent samples over 4 independent experiments). Data were analyzed by two-ways ANOVA followed by Sidak multiple comparison test. Source data are provided as a Source data file. **d** Representative co-immunoprecipitation (from 3 independent experiments) of HA-TpoR with CALR del52-FLAG full length or N-terminal truncations as indicated. Source data are provided as a Source data file. **e** STAT5 transcriptional activity induced by indicated CALR truncations in presence of TpoR. HEK293T were transiently transfected with vectors coding for human TpoR and CALR del52 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. Data represent mean ± SD ($n = 9$ biologically independent samples over 3 independent experiments). Data were analyzed with two-ways ANOVA followed by Sidak multiple comparison test. Source data are provided as a Source data file.

exchange on TpoR ECD was acquired with a lower sequence coverage (~50%) than for CALR del52 due to the presence of 4 N-glycans on TpoR ECD that complicated pepsin digestion (Supplementary Fig. 9a). Amongst the covered region, by far the strongest interaction was observed with the ^{41}FSRTFEDL^{48} motif of TpoR S1 region (Fig. 4a), for which differential H-D uptake between TpoR ECD alone or in presence of CALR del52 was significant ($p < 0.001$) for all incubation time points (Fig. 4a; Supplementary Fig. 9b). Remarkably, the same peptide of TpoR ECD also exhibited strong differential H-D uptake when compared to the CALR del52-TpoR produced as a complex with immature N-glycans (Supplementary Fig. 9c). Consistently, mutations of the ^{44}TFED^{47} motif to alanine prevented TpoR activation by CALR del52 in our STAT5 transcriptional luciferase assay (Fig. 4b). In addition, lower but significant differential H-D exchange was also detected for the ^{52}WDEEEAAPSGT^{62} peptide (Supplementary Fig. 9d).

Next, we turned to molecular dynamics (MD) simulations to study the configurations in which CALR mutant C-terminus could interact with TpoR ECD. We first generated the structure of TpoR D1D2 and CALR del52 C-terminus (Fig. 4c; Supplementary Fig. 10a). Based on previous characterization of the TpoR N-glycans composition in the TpoR D1D2-CALR del52 complex[8], immature (high-mannose) and mature N-glycans were attached to Asn117 and Asn178 of TpoR, respectively. The final model was similar to the one generated with AlphaFold 2.0[18]. 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 (Fig. 4c). Since CALR mutant C-terminus is strongly positively charged and that the two peptides identified by HDx-MS are rich in negatively charged residues, we hypothesized that electrostatic interactions could mediate binding between CALR mutant C-terminus and TpoR ECD. To challenge this assumption, we generated mutants of CALR del52 where all hydrophobic (Met) or basic amino acids (Arg/Lys) of the mutant C-terminus are replaced by either Gly or Asn residues. Using our STAT5 transcriptional assay, we observed that mutations of basic but not of hydrophobic residues to either Gly or Asn resulted in complete loss of TpoR activation by CALR del52 (Supplementary Fig. 9e, f), in line with previous reports[12]. Taking into consideration the localization of the negatively charged acidic residues on TpoR, three main start configurations (poses) of the complex were chosen for assessing complex formation and were used as inputs in HADDOCK 2.4[23] for complex optimization searches. The poses predicted interaction either with the extended S1 region of TpoR (pose 1 and pose 2) or via the more restricted S2 region (pose 3) (Fig. 4d). Starting from the poses generated by HADDOCK (Supplementary Fig. 10b), each of the three poses were subjected to triplicate 500 ns unconstrained MD stimulations (Supplementary Fig. 10c–f). The interaction of CALR mutant C-terminus with the Asp and Glu residues of the ^{44}TFED^{47} motif was consistent with this observed in silico with pose 2 (Fig. 4c, d) and this interaction was conserved after 500 ns of unconstrained MD simulations in all three replicates (Supplementary Fig. 11). Similarly, binding to the ^{52}WDEE^{55} motif was compatible with pose 3 of our model (Fig. 4c).The free energy (ΔG) of the three poses was then estimated by both a knowledge-based method (using the PRODIGY server[24]) and a physical MD estimation (using the MM-GBSA method[25]) at 150 mM salt concentration (Supplementary Fig. 12 and Supplementary Methods). Both methods indicated that in all three poses CALR del52 displays a very high affinity for TpoR D1D2 ($\Delta G < -9$ kcal/mol) with the same order of magnitude for poses P1, P2 and P3. Moreover, MD simulations identified in each pose multiple microstates of the complex (Supplementary Fig. 13 and Supplementary Methods) that can target both the continuous area found mainly on D1 (and partly on D2) labeled S1, but also the small acidic patch in the N-terminal region of D1 (S2). Taken together, these results suggest that, in absence of stabilization by immature N-glycans, multiple micro-configurations of CALR del52-TpoR ECD can co-exist but that the S2 patch centered on ^{44}TFED^{47} plays a central role in binding.

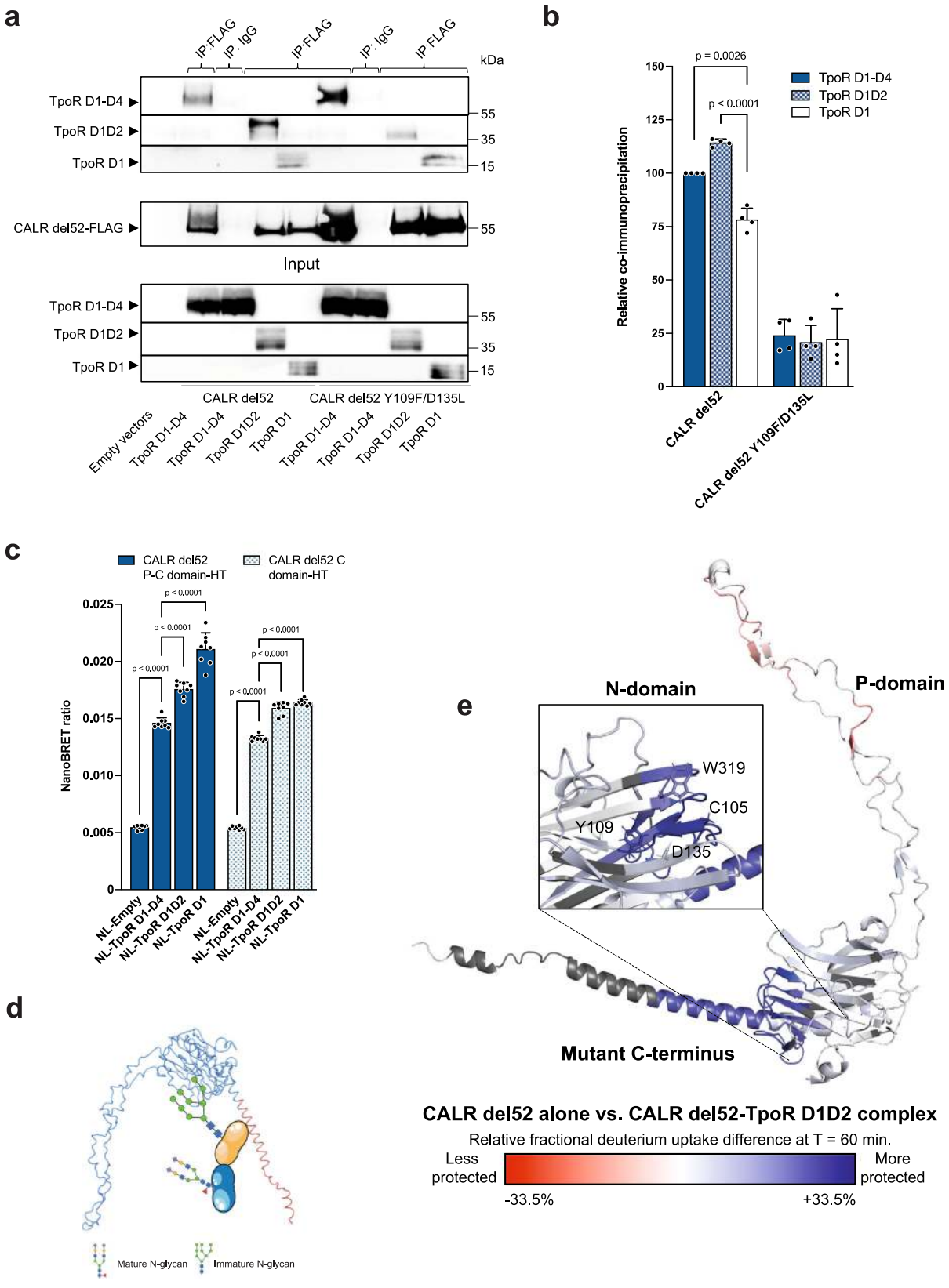
Dimerization of TpoR and of CALR del52 is mediated by the α-helical segment of CALR mutant C-terminus

Our data indicated that both CALR mutant C-terminus and N-domain are involved in direct interaction with TpoR. Previous studies provided evidence that CALR mutant C-terminus is also indispensable for activation of the receptor[4,12] which occurs after homodimerization of mutant CALR[26]. Yet, exactly how activation is achieved remained unclear.

To close this gap, we sought to determine the exact region of CALR mutant C-terminus required to induce TpoR activation and dimerization. We measured autonomous proliferation of cytokine-dependent hematopoietic cells (Ba/F3) stably expressing TpoR together with progressive truncations of the C-terminus of CALR del52 (Fig. 5a). The deletion of the non α-helical segment of CALR mutant C-terminus (as in CALR del52 A394*) (Fig. 1e, f) did not prevent TpoR-dependent proliferation of Ba/F3 cells. In contrast, further deletions in the α-helical segment of CALR del52 either reduced (M387*) or prevented (M377* and M371*) CALR del52 mediated activation of TpoR (Fig. 5b). Similarly, CALR del52 Y109F/D135L mutant, which is deficient for N-glycan binding and disturbs the helicity of the mutant C-terminus (Fig. 1e, f), was not able to induce Ba/F3 autonomous proliferation in presence of TpoR.

Since activation of homodimeric cytokine receptors occurs upon ligand-induced dimerization[27] in a productive conformation[28,29], we then assessed whether the same truncations of CALR del52 C-terminus precluded homodimerization of TpoR in a live-cell cysteine crosslinking assay. The L508C point mutation, homologous to murine L501C, was introduced in human TpoR, placing the cysteine residue at a position facing inward the transmembrane α-helix in the active conformation of TpoR dimer[28]. The specificity of the crosslinking was achieved by using a truncated form of the TpoR devoid of intracellular cysteines which remains active[28]

https://doi.org/10.1038/s41467-023-37277-3



and by preventing crosslinking of free cysteines of the ECD by pre-incubation with N-ethyl-maleimide which blocks free extracellular cysteines (Fig. 5c). In agreement with above results, truncations until M387 allowed CALR del52 induced dimerization of TpoR while further truncations precluded the formations of homodimers (Fig. 5d).

Because oligomerization of CALR mutants themselves precedes TpoR activation[26,30], the same set of CALR del52 C-terminal deletions (Fig. 5a) was used to probe the role of the α-helical segment of CALR mutant C-terminus in CALR homodimerization. By co-immunoprecipitating HA-tagged full length CALR del52 with FLAG-tagged CALR del52 truncations, we observed that truncations beyond

Fig. 3 | N-glycan dependent and independent interactions of CALR mutant with TpoR. a Representative co-immunoprecipitation of HA-TpoR ECD domains by CALR del52-FLAG or CALR del52 Y109F/D135L-FLAG using an anti-FLAG antibody for capture and anti-HA antibody or anti-CALR mutant C-terminus antibody (SAT602) for detection of HA-TpoR and CALR del52-FLAG (and Y109F/D135L mutant), respectively. Source data is provided as a Source data file. **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 ($n = 4$). Data were analyzed by two-ways ANOVA followed by SIDAK multiple comparison test. Source data are provided as a Source data file. **c** NanoBRET between NanoLuc-TpoR subdomains (or NL-Empty) and CALR del52-HaloTag truncated from the N-terminus. Data represent mean + SD ($n = 8$ biologically independent samples from 4 independent experiments). Source data are provided as a Source data file. **d** Cartoon representing the complex between CALR del52 and TpoR D1D2 domain containing immature N-glycans produced in Schneider (S2) cells. Items in this figure were created with BioRender. **e** Structure model (AlphaFold 2.0[18]) of CALR del52. Colors represent the difference in relative fractional uptake (ΔRFU) at 1 h incubation in deuterium (mean of $n = 3$ experiments) between CALR del52 alone and CALR del52-TpoR D1D2 complex with immature N-glycans. Regions in red and blue are respectively less and more protected in the CALR del52-TpoR D1D2 complex compared to CALR del52 alone. The scale from red to blue is proportional to the ΔRFU between indicated species with dark red and dark blue corresponding to highest differential. The same scale as Fig. 2a was used for comparison. Raw data are provided in the source file. Dark gray represents regions without peptide coverage for the given time point. The N-glycan binding domain of CALR is highlighted. Source data are provided as a Source data file.

the non α-helical segment strongly decreased CALR del52 oligomerization, indicating that the same 28 α-helical residues of the mutant C-terminus required for TpoR activation and dimerization also mediate CALR oligomerization (Fig. 5e). Consistently, the oligomeric profile in native conditions was similar between recombinant CALR del52 and CALR del52 A394* with or without reducing agent (indicating that C-terminal cysteines are not required for homo-multimerization) while CALR ΔC-tail did not form any oligomer (Supplementary Fig. 14).

Finally, we used RosettaDock[31] to model CALR del52 dimer formation using the monomeric structure predicted using AlphaFold 2.0[18]. The top 10 models predicted dimerization through the mutant C-terminus via residues prior A394 and the two C-terminal cysteines. The best scoring prediction is depicted in Fig. 5f and shows dimerization of CALR del52 via the mutant C-terminus which forms a coiled-coil like structure with interactions involving Arg (dark blue), Met (orange) and Thr (purple), but not the cysteines at the extremity of CALR mutant C-terminus.

Comprehensive model of the TpoR-CALR mutant complex

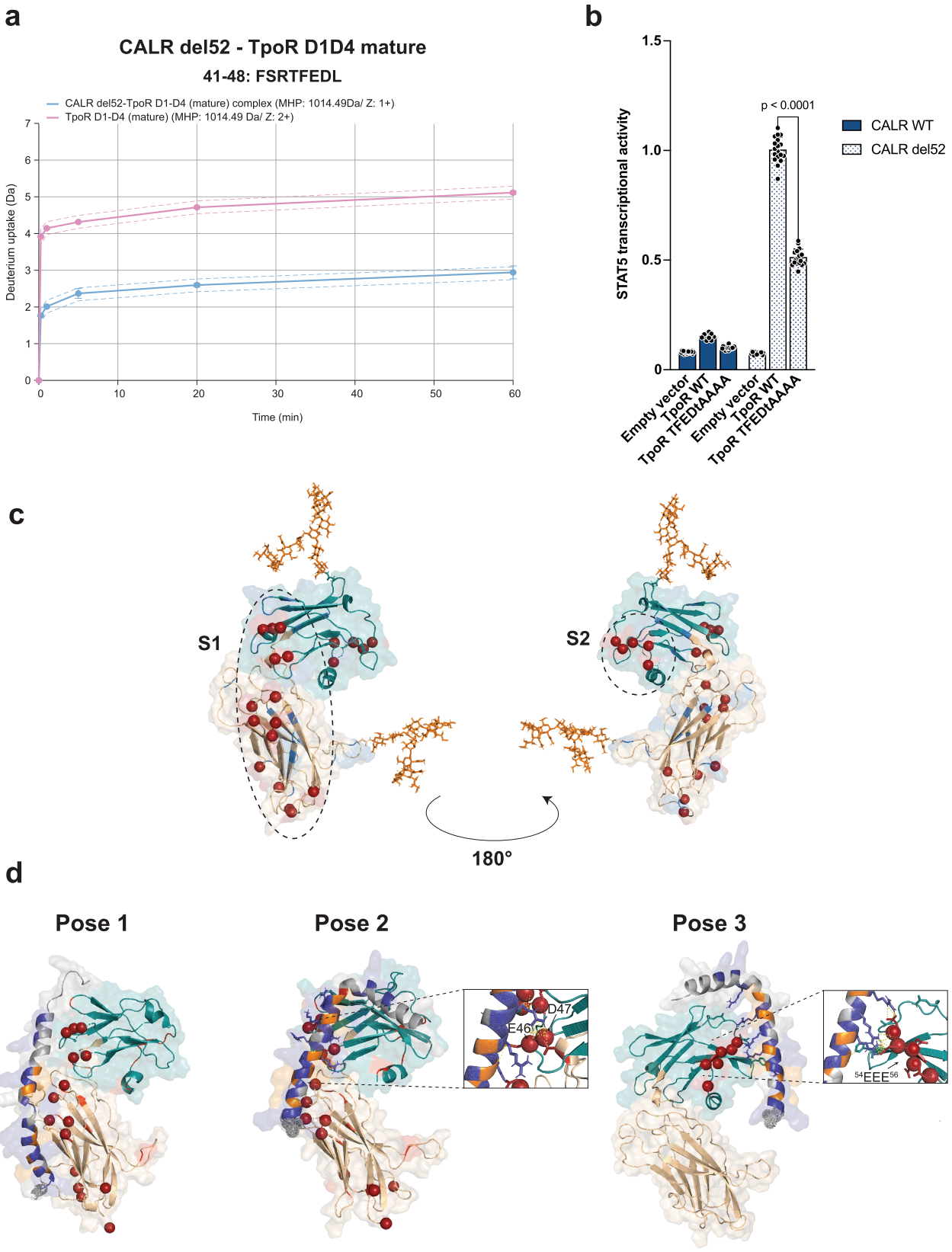
On the basis of our experimental data, we generated an atomistic model of the complete TpoR-CALR mutant tetrameric complex. Our data indicated that binding of CALR del52 C-terminus alone to TpoR could occur in a variety of micro-configurations but that the extended S1 acidic region centered on ^{44}TFED^{47} was key for interaction. Yet, binding of TpoR to full length CALR del52 also involves strong interaction between specific residues of the N-domain and immature N-glycans on Asn117[8] (Fig. 3e). We used AlphaFold 2.0[18] to complete our modeling of TpoR and generate the full extracellular domain and transmembrane 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 (Fig. 5c, d). CALR del52 dimer (Fig. 5f) was docked to the dimer of TpoR taking into consideration our experimental data indicating that binding occurs concomitantly between immature N-glycans on Asn117 of TpoR and residues of CALR N-domain and between TpoR S1 acidic region and CALR mutant C-terminus. The final structure places the mutant C-terminus in a configuration where the main interacting sites are located around the ^{44}TFED^{47} motif, in line with above results. Likewise, immature N-glycans on Asn117 of TpoR interact with the N-domain pocket containing key residues involved in N-glycan binding including C105, Y109 and W319 (Fig. 6). This glycoproteic tetramer was embedded in a POPC lipid bilayer and a water box (comprising a total of ~1 million atoms) and subjected to triplicate all-atom molecular dynamics simulations for 100 ns. The complex remained stable during this timeframe, except for the very flexible P-domain (Supplementary Fig. 15). Most contacts identified during the simulations relied on basic-acidic interactions and occurred both in *cis* and in *trans*, thereby further stabilizing TpoR dimers. They involved the ^{44}TFED^{47} motif but also other negative patches including ^{96}PDQEE^{100} and ^{154}WEEP^{157} of the extended S1 negative patch (Supplementary Data 1). To assess whether our model was also compatible with CALR ins5 (CALR type 2 mutant), which harbors a similar but longer C-terminus, we also generated the CALR ins5-TpoR tetrameric complex following the same procedure as for CALR del52 and subjected the complex to all-atom MD simulations in triplicate (Supplementary Fig. 16). Like for CALR del52, the complex remained stable over the 100 ns timeframe. Analysis of interacting residues over the simulation timeframe revealed that the ^{44}TFED^{47} motif, ^{96}PDQEE^{100} and ^{154}WEEP^{157} motifs were conserved in the CALR ins5-TpoR tetrameric complex (Supplementary Data 2).

Our experimental data and atomistic simulations indicate that CALR mutants interact through two regions of TpoR essentially on the D1 domain. First, the mutant C-terminus 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 C-terminus 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 specific configuration. When this interaction occurs in the context of immature TpoR, strong interactions between CALR N-domain and immature N-glycans mainly on Asn117 of TpoR stabilize the complex. Thus, the mutant C-terminus provides the specificity and stability of the tetrameric complex.

Discussion

Our work unveils the molecular basis for the recognition and activation of the thrombopoietin receptor by frameshift mutants of calreticulin that are at the origin of myeloproliferative neoplasm[2,3]. These findings provide mechanistic insights into the mechanisms leading to a switch from the transient, N-glycan-only based interaction between wild-type calreticulin and thousands of proteins to a specific and stable interaction between CALR frameshift mutants and TpoR. Our results indicate that this specificity relies on two complementary mechanisms. First, the presence of CALR mutant C-terminus induces a conformational overhaul of CALR N-domain, resulting in increased accessibility of the N-glycan binding pocket that interacts with immature N-glycans on Asn117 of TpoR[5,8]. Remarkably, this N-glycan binding pocket 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 TpoR[9]. Increased accessibility of this pocket in CALR del52 explains the partial specificity of hematoxylin to target mutant cells[9]. Secondly, our results demonstrate a direct interaction between the positively charged CALR mutant C-terminus and negatively charged residues on TpoR D1 domain. We posit that this second interaction is the basis for the specificity of CALR mutant for TpoR versus other N-glycosylated proteins. In addition, our HDx-MS results highlight several regions that exhibit different accessibility between CALR WT and CALR del52, suggesting potential targetable sites. Of interest, mutations in CALR C-terminus are also observed in a variety of non-hematological cancers and their exposure at the cell surface leads to immunosuppressive activity[14]. Conceptually, it is

https://doi.org/10.1038/s41467-023-37277-3



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.

Most recent work in the field elegantly demonstrated using single molecule fluorescence tracking that TpoR is a monomer at the basal state and that dimerization is induced upon ligand binding to the ECD[27,32]. Several dimer conformations can be induced by synthesized ligands such as diabodies that can tune the receptor for subtly different effects that all require induced dimerization[32]. Previous studies demonstrated that CALR del52, like Tpo, can act as a cytokine to induce dimerization of TpoR[8,26]. Here, we define the basis for this pathological ligand-induced dimerization. We find that the first segment of CALR

Fig. 4 | CALR mutant C-terminus interacts with acidic patches on TpoR D1 domain. a Deuterium uptake (Da) of the FSRTFEDL peptide from CALR del52 alone or in complex with TpoR D1D4 (with mature N-glycans) at 5 different exchange time points. The dotted lines represent standard deviation (SD), the full line represents the mean of triplicates ($n = 3$). Source data are provided as a Source data file. **b** STAT5 transcriptional activity induced by CALR del52 in presence of empty vector, TpoR WT or TFEDtAAAA mutant. Data represent mean ± SD ($n = 12$ biologically independent samples over 4 independent experiments). Data were analyzed with one-way ANOVA followed by Sidak multiple comparison test. Source data are provided as a Source data file. **c** Prediction of TpoR D1D2 domain (described in Supplementary Methods). The model shows one extensive (S1) and one more restricted (S2) patch of acidic residues represented by red spheres. Basic residues are shown in blue, acidic residues are shown in red. TpoR D1 domain is shown in blue/green and TpoR D2 is shown in wheat. Complex N-glycans are attached to position Asn117 and Asn178 and are shown in orange. **d** Pose 1 (left), pose 2 (middle) and pose 3 (right) generated by HADDOCK between CALR del52 mutant C-terminus and TpoR D1D2. The best docking complexes were chosen, as ranked by the HADDOCK score. Highlighted are the strong interactions between Arg of CALR mutant C-terminus and E46 and D47 of TpoR D1 domain for pose 2 and the interactions between Arg of CALR mutant C-terminus and the ^{54}EEE^{56} motif on TpoR D1 for pose 3. The basic (Arg/Lys) and hydrophobic (Met) residues of CALR del52 mutant C-terminus are shown in dark blue and orange, respectively. Other residues are shown in gray.

mutant C-terminus that is α-helical is involved in CALR oligomerization, TpoR binding and TpoR dimerization. This observation suggests that one face of the α-helix forms a coiled-coil like structure at the core of the CALR dimer while the other face interacts with negative patches of TpoR D1 domain (Fig. 6). These results contradict previous reports claiming that CALR mutant dimerization occurs through disulfide bonds involving cysteines of the C-terminal extremities of CALR mutant, and that the same cysteines of a fusion protein containing the mutant C-domain are involved in binding TpoR[33]. However, this was inferred from comparison of western blots in denaturing conditions with or without reducing agent, thus not reproducing the native conformation of the protein. While disulfide bonds may form after the Cys-independent interaction and activation, our live-cell cysteine cross-linking assays, oligomerization analysis in native conditions and co-immunoprecipitation studies all indicate that CALR mutant C-terminal cysteines are not required for CALR oligomerization nor for TpoR dimerization. Noteworthy, these results confirm our observation and those from others[11,12] that C-terminal cysteines of CALR mutant are not involved in TpoR activation.

Finally, our experimentally validated in silico analyses identify specific residues of the TpoR D1 domain that interact with CALR mutant C-terminus. Identification of these interaction sites is of major importance for the development of inhibitors targeting the TpoR-CALR mutant C-terminus interaction. Based on our HDx-MS and functional results, we further propose a complete model of the tetrameric CALR mutant-TpoR complex for both type 1 (CALR del52) and type 2 (CALR ins5) frameshift mutants and highlight key interacting residues. We posit that the interaction between immature N-glycans of TpoR and CALR N-domain occurs concomitantly to binding of negatively charged patches of TpoR D1 domain to CALR mutant C-terminus, the latter providing specificity and additional stability for the interaction. Previous studies suggested that homo-multimerization of CALR mutant proteins preceded binding to TpoR[26] in the ER and Golgi apparatus[8,34]. This is compatible with our structural models of CALR del52 and CALR ins5 where different faces of the α-helical C-terminus are involved in CALR mutant homodimerization and TpoR interaction. Interestingly, the 19 amino acid insertion in ins5 which increases the distance between the two sites by more than 25Ä does not seem to have a major effect on the complex formation. 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 CALR frameshift mutants and the TpoR ECD, providing mechanistic understanding of the specificity of this interaction and paving the way for inhibitory therapeutic avenues.

Methods

Production and purification of recombinant proteins

Recombinant proteins for human TpoR (hTpoR) D1–D4 and CALR del52-hTpoR D1D2 complex were produced as described previously in Schneider S2 cells[8] (see also Supplementary Methods). The amino acid sequence of hTpoR D1–D4 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 A411.

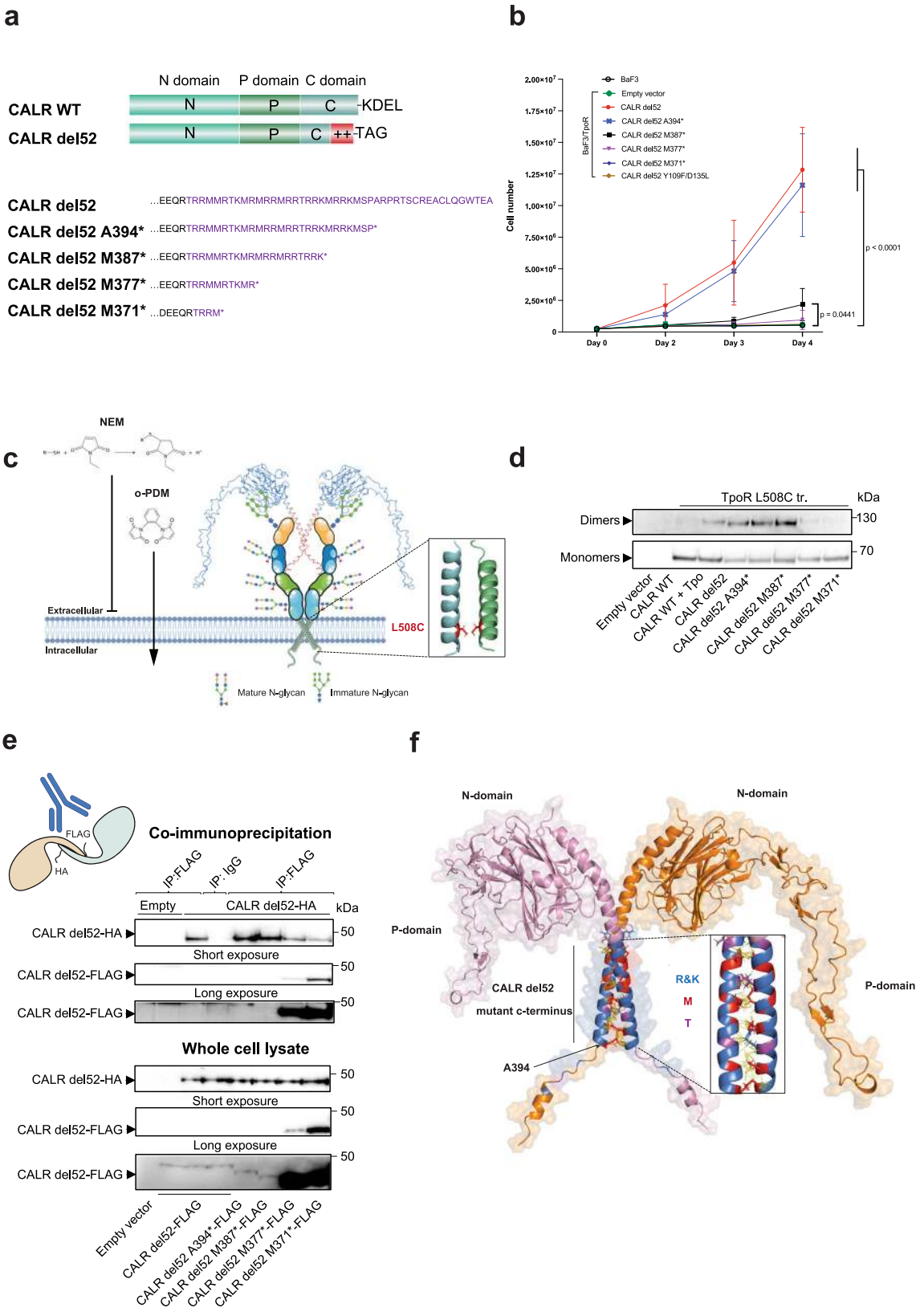
Recombinant human CALR wild-type, CALR del52 and its derivatives contain a N-terminal His tag sequence (MGSHHHHHHGSSG) that replaces the CALR signal peptide sequence (aa1–17). In addition, the cysteine 163 was mutated to serine. CALR proteins were produced in *Escherichia coli* (see also Supplementary Methods). The purity and folding of each protein were verified by SDS-PAGE, thermal stability, and chromatography (Supplementary Figs. 1b, c, 2a–g, 6a, b, 8a, b).

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 K_2HPO_4, 5 mM KH_2PO_4 dissolved in H_2O, pH 7). For interaction analysis between recombinant mature TpoR D1-D4 and CALR del52, proteins were first mixed together at a 1:1 molarity for 30 min at room temperature followed by 3 h at 4 °C. Proteins were then kept at 0 °C. Labeling was performed with a 20-fold dilution of samples in labeling buffer (5 mM K_2HPO_4, 5 mM KH_2PO_4 dissolved in D_2O, pD 7) for 6 different incubation times (0, 0.25, 1, 5, 20 and 60 min) at 20 °C in a randomized order. Final D_2O concentration was 95% during labeling reaction (3 µL proteins/57 µL labeling buffer). After incubation, the reaction was quenched using a 1:1 dilution in the quench buffer (0.05 M K_2HPO_4, 0.05 M KH_2PO_4 with 30 mM TCEP, pH 2.3) prior to injection into a pepsin column (Enzymate BEH Pepsin 2.1 × 30 column, Waters CAT. 186007233) with dynamic flowrate of 150–75 µL/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 1.7 µM VANGUARD Pre-column (Waters Cat. 186009375), separated on a ACQUITY UPLC C18 1.7 µM 1.0 × 100 mm 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) with a database containing the sequence of all proteins present in the sample and the pepsin used for digestion. The peptides identified were further analyzed with DynamX 3.0 (Waters) using a tolerance of 10 ppm, a maximum length of 35 amino acids, 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 Deuteros 2.0 with peptide-level significance testing, which controls for the false discovery rate[35]. A summary of the HDx-MS experimental set-up and raw data are provided in the source data file following reporting suggestions described by Masson and colleagues[36].

For the representation of HDx-MS data, structures of CALR WT, CALR del52 or CALR ΔC-tail were generated with AlphaFold 2.0[18]. The structures were colored with a gradient from red (less protected,

https://doi.org/10.1038/s41467-023-37277-3



negative value) to blue (more protected, positive values) corresponding to difference in relative fractional uptake (ΔRFU) with dark red and dark blue corresponding to highest H-D differential. The ΔRFU per residue was inserted as b-factor in Pymol 2.4.2. and the structures were colored according to b-factor values with the scale indicated in figure legends. The RFU for a peptide 1 is computed as:

$$\mathrm{RFU}_{\mathrm{peptide\,1}} = \frac{\mathrm{Uptake(Da)}_{\mathrm{peptide\,1}}}{\mathrm{Total\ number\ of\ backbone\ amide\ hydrogen}_{\mathrm{peptide\,1}}} \quad (1)$$

Fig. 5 | Dimerization of TpoR and of CALR del52 is mediated by the α-helical segment of CALR mutant C-terminus. a Representation and sequence of CALR del52 C-terminal truncations used in proliferation assay. **b** Proliferation assay. Ba/F3 cells stably expressing TpoR in pMX-IRES-GFP vector were transduced with indicated CALR variants or an empty vector (pMSCV-IRES-mCherry) and sorted by FACS. 250,000 cells were washed and seeded in 10 mL of complete culture medium without cytokine and counted each day using an automated cell counter. Values represent mean of 3 independent experiments (±SD) ($n = 3$). Source data are provided as a Source data file. Data were analyzed using two-ways ANOVA followed by Sidak multiple comparison test at day 4 using the average of technical triplicate counting from 3 independent experiments ($n = 3$). **c** Cartoon representation of the crosslinking assay to assess homodimerization of TpoR in a productive orientation. o-PDM: ortho-phenylene dimaleimide. NEM: N-ethylmaleimide. Items in this figure were created with BioRender. **d** Crosslinking study of TpoR dimerization in presence of Tpo, CALR del52 full length or C-terminal truncations. Shown is a representative western blot (from 3 independent experiments) in denaturing and reducing conditions showing human TpoR monomers and o-PDM crosslinked dimers in the indicated conditions. Source data are provided as a Source data file. **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 (from 3 independent experiments). Source data are provided as a Source data file. **f** RosettaDock top scoring simulation of CALR del52 dimers. Structure of CALR del52 was modeled using AlphaFold 2.0[18]. CALR del52 monomers are shown in orange and pink. Residues of the CALR mutant C-terminus are shown in purple (Thr), dark blue (Arg) and red (Met). Close interactions (<3Å) are shown by yellow dashed lines.

The ΔRFU for peptide 1 between condition A and B in then computed as:

$$\Delta RFU_{\text{peptide 1[A-B]}} = RFU_{1[A]} - RFU_{1[B]} \quad (2)$$

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 (~22 °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[37,38]. Detailed protocol and unprocessed data are provided in Supplementary Methods and Supplementary Fig. 4a, respectively.

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[39–46]. Closest templates were retrieved with Phyre 2[47]. Modeller 9.21[48], AlphaFold2 and Rosetta Folding[18]. 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[23,49] for TpoR D1D2-CALR del52 mutant tail complex optimizations searches. The top configurations were further optimized using 500 ns molecular dynamics runs performed with OpenMM version 7.4.1[50] using a Monte Carlo Barostat, at 300 K, using a Langevin integrator with 1ps-1 friction coefficient and a 2 fs timestep and the FF14SB Force Field[51] to obtain 3 final poses in which the last residues of the mutant become unfolded. Free energy was estimated by both a knowledge-based method, using PRODIGY server[24] and a physical MD estimation approach based on 3 simulations for each pose, using the MM-GBSA method[25] at 150 mM salt concentration, implemented in AMBER20[52]. Conformational discretization for microstate analysis was performed using Time-Lagged Independent Component Analysis (TICA). The backbone dihedral angles of the CALR del52 mutant C-terminus molecule were used as input coordinates for TICA. TICA and free energy surfaces were computed using the PyEMMA (2.5.11) package[53], resulting plots were generated using the Matplotlib (3.5.1) package[54]. The inflection core state (InfleCS) clustering method[55] was used to cluster the two transformed coordinates with the highest eigenvalues and the associated cluster centers were plotted on the corresponding free energy surface. Clustering was performed using 10 components, re-estimation of the same Gaussian mixture model was done 5 times.

Templates from AlphaFold2 and Rosetta Folding were used to effectively build the tetrameric 3D models and identify the interaction interface between the two CALR mutants. HDx-MS data was used to identify contacts between TpoR and CALR del52 in the formation of the tetramer complex. The ER specific G1M9 glycans of TpoR in contact to CALR were modeled with Glycopack[56] in the configuration consistent with NMR data[57] while the rest are of complex type, built in agreement with SAGS Database (https://sags.biochim.ro/)[58,59]. 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 1907 POPC molecules accommodating the TM region of TpoR and in 478479 TIP3P water molecules, 1328 chloride and 1402 sodium ions describing the solvent region hydrating the rest of the tetramer using the CHARMM-GUI server[60]. This overall system consisting of ~ 1 million atoms was subjected to a mild simulated annealing procedure consisting in a start minimization, heating to 300 K followed by cooling to 0 K and final extended minimization, using NAMD v.2.13[61] CHARMM36 forcefield[62–64]. The same procedure was used to build TpoR-CALR ins5 complex. The TpoR-CALR-del52/ins5 models were further subjected to 3 molecular dynamics runs to explore the configuration sample space. More detailed protocols, including intermediate modeling steps, free energy estimates and detailed TICA analysis are presented in Supplementary Methods.

Transcriptional dual luciferase assay

Transcriptional dual luciferase assays were performed as described[5]. Briefly, HEK293T were transiently transfected with empty vector 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 25 ng/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 48 h post transfection with NP-40 buffer. After pre-clearing, samples

https://doi.org/10.1038/s41467-023-37277-3

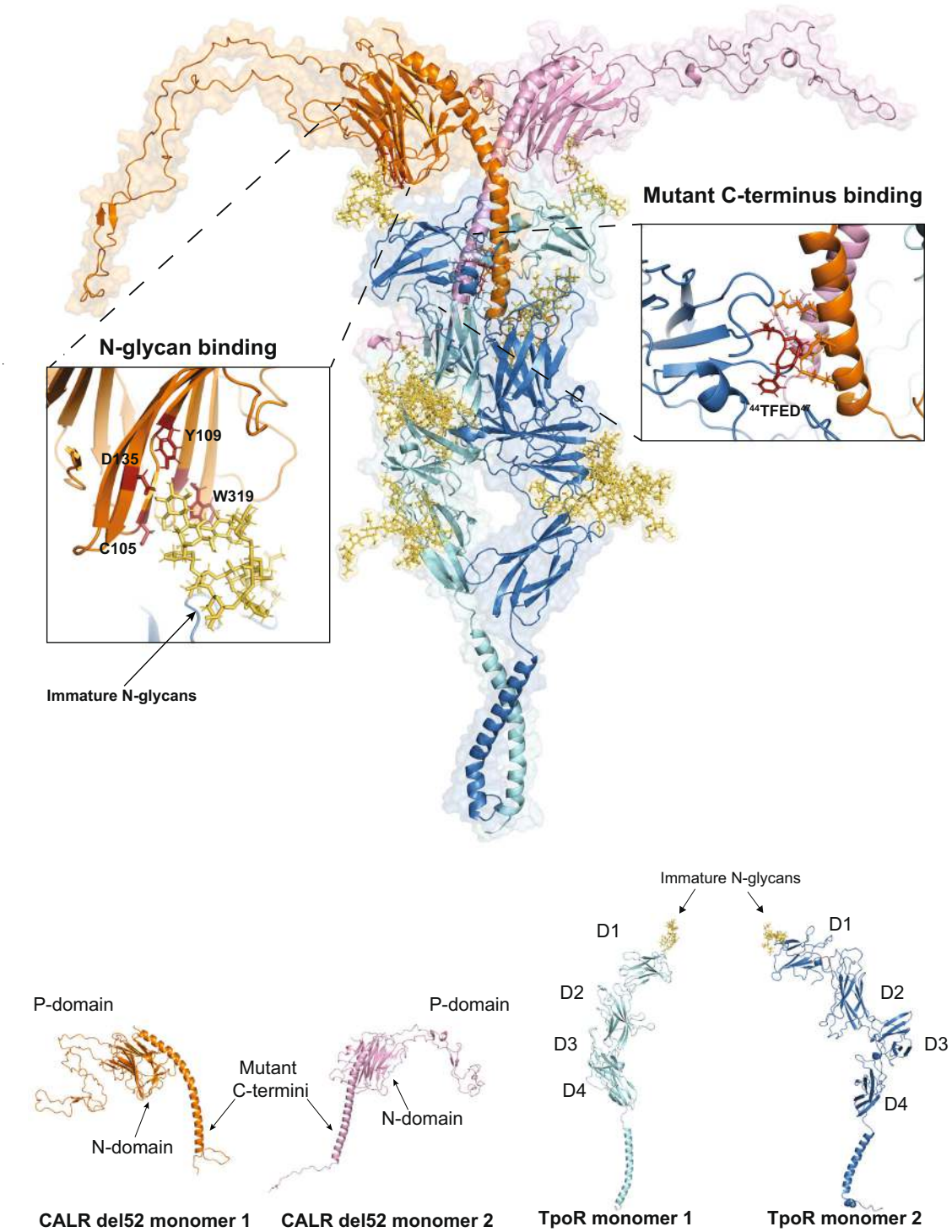


Fig. 6 | Comprehensive model of the TpoR-CALR mutant complex. Molecular dynamics of the CALR del52-TpoR tetrameric complex. The structure represents the last frame of one out of three replicates of 100 ns unconstrained MD of the CALR del52-TpoR tetrameric complex. All replicates are shown in Supplementary Fig. 15. The model illustrates how CALR mutant interacts with TpoR via two distinct regions. One region concerns the interaction between CALR N-domain (N-glycan binding domain) and immature N-glycans attached to Asn117 of TpoR. The second region involves binding of CALR mutant C-terminus to acidic patches on TpoR ECD. The interaction between TpoR negative residues and CALR mutant C-terminus may possibly occur both in *cis* (with the same molecule of CALR mutant with which TpoR interacts via N-glycans) or in *trans*. The binding between specific residues of CALR N-domain and immature N-glycans attached to Asn117 and that of the ^{44}TFED^{47} motif of TpoR to CALR mutant C-termini is illustrated. The contact list of interactions detected during MD triplicate runs is provided as Supplementary data 1. TpoR molecules are shown in cyan and dark blue. CALR del52 molecules are shown in orange and pink. Key residues of CALR del52 N-domain involved in binding of immature N-glycans are shown in dark red. The ^{44}TFED^{47} motif of TpoR D1 domain that interacts with CALR del52 C-termini is shown in red. The different domains of CALR mutant and TpoR are indicated.

were incubated with anti-FLAG antibody (Genscript, Cat. No. A00187) at 2 μg/mL or corresponding isotype control (Genscript Cat. No. A01730) overnight at 4 °C. Bound proteins were pulled down with 40 μL/mL of rProtein G Agarose (ThermoFisher, 20397) for 3 h 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 (Cell Signaling, C29F4) for HA-hTpoR, HA-CALR del52 or anti-CALR mutant tail (Clone SAT602, MyeloPro GmBH).

Antibodies

Anti-FLAG tag (Genscript No. A00187) 3 μg/mL whole cell lysate for immunoprecipitation and 1:500 dilution for western blots. Anti-HA tag clone C29F4 (Cell Signalling #3724) 1:1000 dilution for western blots. Mouse IgG control (Genscript No. A01730) 3 μg/mL whole cell lysate for immunoprecipitation. Anti-mutant CALR (Myelopro, Clone SAT602) 1.3 μg/mL for western Blot. Anti-rabbit IgG, HRP-linked (Cell Signaling #7074) 1:5000 for western blots. Anti-mouse IgG, HRP-linked (Cell Signaling #7076) 1:5000 for western blots.

Nano-bioluminescence energy transfer (BRET)

Nano-bioluminescence resonance energy transfer (BRET) was performed as previously described[8]. The specificity of the interaction was validated following the manufacturer's instruction (Promega), see Supplementary Methods and Supplementary Fig. 7.

Microscale thermophoresis (MST)

Recombinant human CALR del52 was labeled with NHS- chemistry according to the manufacture's instruction (Protein Labeling Kit RED-NHS 2nd Generation, NanoTemper Technology), referred as the "target protein". Briefly, the target protein (at 10 μM) was incubated with the dye solution in the labeling buffer (130 mM $NaHCO_3$, 50 mM NaCl, pH 8.2–8.3) for 1 h on ice. The dye carries a reactive NHS-ester group that reacts with primary amines (lysine residues) to form covalent bonds. The surplus of the dye not bound to the target protein was removed through passage on a resin column prior to elution of the target in the equilibration buffer (Tris-HCL 1 M, pH 7.6). For MST measurement, the CALR del52-NHS was used at 20 nM final concentration in MST buffer (Tris based supplemented with 0.01% of tween 20).

TpoR D1–D4 (the "ligand") remained label free. Serial dilutions (0.15 nM to 5 μM) of the ligand (TpoR D1-D4) were performed to titer the target protein (CALR del52). The measurements were performed on a NanoTemper monolith NT.115 instrument (NanoTemper technologies, Germany) at 40% LED and medium MST-power with a standard 5 s. before, MST-on for 30 s. and 5 s. after MST-off.

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 h post transfection, cells were harvested without trypsinization and washed in PBS. Cells were then re-suspended and incubated for 15 min at room temperature in crosslinking buffer (PBS 1 mM $MgCl_2$, 0.1 mM $CaCl_2$) with 100 μM N-ethylmaleimide (NEM, ThermoFisher Cat.:23030) to avoid non-specific crosslinking of extracellular cysteines. 200 ng/mL of rhTpo was added in the indicated condition. Samples were mixed gently and further incubated for 15 min at room temperature. Cells were then centrifuged for 5 min at 500×*g* and re-suspended in crosslinking buffer with 100 μM o-phenylene dimaleimide (o-PDM, Sigmaaldrich) for 10 min at room temperature. Cells were further centrifuged 5 min at 500×*g* and re-suspended in lysis buffer (NP-40, 2% ß-mercaptoethanol) with protease inhibitor cocktail. Cell lysates were analyzed by SDS-PAGE in denaturing and reducing conditions with anti-HA antibody.

Cell lines and cell culture

HEK293T were obtained from the American Type Culture Collection (ATCC) (CRL-3216™). They were cultivated in Dulbecco's Modified Eagle Medium (Gibco) supplemented with 10% fetal bovine serum (Gibco). Ba/F3 cell lines were obtained previously after isolation of clones with pro-B lymphocyte characteristics[65] and were transferred from laboratory of Prof. Harvey Lodish (Whitehead Institute, MIT) to Ludwig Institute for Cancer Research, Brussels Branch. Ba/F3 were cultured in Roswell Park Memorial Institute (RPMI) medium (Gibco) supplemented with 10% FBS (Gibco) and 0.5 ng/mL of murine IL-3 (RnDsystems). IL-3 was removed by washing the cells at three times with PBS prior to experiments.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

Mass spectrometry data was deposited on the ProteomeXchange repository (https://www.proteomexchange.org) under accession number PXD034131. PDB files of the molecular dynamics simulations were deposited on Figshare (https://figshare.com/s/b4ceb87fdce1f242e469 and https://figshare.com/s/9033970b5a1d3f8d6fa7). The detailed protocols to produce the recombinant proteins used in this study are provided as Supplementary Methods. We can provide plasmids coding for each protein upon request after signing of a material transfer agreement and payment of a fee corresponding to shipment and preparation costs. Source data are provided with this paper. The source data file provides the raw data and reporting of HDx-MS experiments following the guidelines suggested by Masson et al.[36] in addition to source data for all other experiments. Source data are provided with this paper.

References

1. Constantinescu, S. N., Vainchenker, W., Levy, G. & Papadopoulos, N. Functional consequences of mutations in myeloproliferative neoplasms. *Hemasphere* **5**, e578 (2021).
2. Klampfl, T. et al. Somatic mutations of calreticulin in myeloproliferative neoplasms. *N. Engl. J. Med.* **369**, 2379–2390 (2013).
3. Nangalia, J. et al. Somatic CALR mutations in myeloproliferative neoplasms with nonmutated JAK2. *N. Engl. J. Med.* **369**, 2391–2405 (2013).
4. Araki, M. et al. Activation of the thrombopoietin receptor by mutant calreticulin in CALR-mutant myeloproliferative neoplasms. *Blood* **127**, 1307–1316 (2016).
5. Chachoua, I. et al. Thrombopoietin receptor activation by myeloproliferative neoplasm associated calreticulin mutants. *Blood* **127**, 1325–1335 (2016).
6. Marty, C. et al. Calreticulin mutants in mice induce an MPL-dependent thrombocytosis with frequent progression to myelofibrosis. *Blood* **127**, 1317–1324 (2016).

7. Nivarthi, H. et al. Thrombopoietin receptor is required for the oncogenic function of CALR mutants. *Leukemia* **30**, 1759–1763 (2016).
8. Pecquet, C. et al. Calreticulin mutants as oncogenic rogue chaperones for TpoR and traffic-defective pathogenic TpoR mutants. *Blood* **133**, 2669–2681 (2019).
9. Jia, R. et al. Hematoxylin binds to mutant calreticulin and disrupts its abnormal interaction with thrombopoietin receptor. *Blood* **137**, 1920–1931 (2021).
10. Varricchio, L. et al. Calreticulin: challenges posed by the intrinsically disordered nature of calreticulin to the study of its function. *Front. Cell Dev. Biol.* **5**, 96 (2017).
11. Elf, S. et al. Defining the requirements for the pathogenic interaction between mutant calreticulin and MPL in MPN. *Blood* **131**, 782–786 (2018).
12. Elf, S. et al. Mutant calreticulin requires both its mutant c-terminus and the thrombopoietin receptor for oncogenic transformation. *Cancer Discov.* **6**, 368–381 (2016).
13. Pecquet, C. et al. Secreted mutant calreticulins as rogue cytokines trigger thrombopoietin receptor activation specifically in CALR mutated cells: perspectives for MPN therapy. *Blood* **132**, 4–4 (2018).
14. Liu, P. et al. Immunosuppression by mutated calreticulin released from malignant cells. *Mol. Cell* **77**, 748–760.e749 (2020).
15. Kapoor, M. 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* **43**, 97–106 (2004).
16. Ellgaard, L. et al. NMR structure of the calreticulin P-domain. *Proc. Natl Acad. Sci. USA* **98**, 3133–3138 (2001).
17. Chouquet, A. et al. X-ray structure of the human calreticulin globular domain reveals a peptide-binding area and suggests a multimolecular mechanism. *PLoS ONE* **6**, e17886 (2011).
18. Jumper, J. et al. Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021).
19. Thomson, S. P. & Williams, D. B. Delineation of the lectin site of the molecular chaperone calreticulin. *Cell Stress Chaperones* **10**, 242–251 (2005).
20. Wood, T. J. et al. Specificity of transcription enhancement via the STAT responsive element in the serine protease inhibitor 2.1 promoter. *Mol. Cell Endocrinol.* **130**, 69–81 (1997).
21. Gopalakrishnapai, J. et al. Isothermal titration calorimetric study defines the substrate binding residues of calreticulin. *Biochem. Biophys. Res. Commun.* **351**, 14–20 (2006).
22. Kozlov, G. et al. Structural basis of carbohydrate recognition by calreticulin. *J. Biol. Chem.* **285**, 38612–38620 (2010).
23. Dominguez, C., Boelens, R. & Bonvin, A. M. HADDOCK: a protein-protein docking approach based on biochemical or biophysical information. *J. Am. Chem. Soc.* **125**, 1731–1737 (2003).
24. Xue, L. C., Rodrigues, J. P., Kastritis, P. L., Bonvin, A. M. & Vangone, A. PRODIGY: a web server for predicting the binding affinity of protein-protein complexes. *Bioinformatics* **32**, 3676–3678 (2016).
25. Kollman, P. A. et al. Calculating structures and free energies of complex molecules: combining molecular mechanics and continuum models. *Acc. Chem. Res.* **33**, 889–897 (2000).
26. Araki, M. et al. Homomultimerization of mutant calreticulin is a prerequisite for MPL binding and activation. *Leukemia* **33**, 122–131 (2019).
27. Wilmes, S. et al. Mechanism of homodimeric cytokine receptor activation and dysregulation by oncogenic mutations. *Science* **367**, 643–652 (2020).
28. Staerk, J. et al. Orientation-specific signalling by thrombopoietin receptor dimers. *EMBO J.* **30**, 4398–4413 (2011).
29. Seubert, N. et al. Active and inactive orientations of the transmembrane and cytosolic domains of the erythropoietin receptor dimer. *Mol. Cell* **12**, 1239–1250 (2003).
30. Rivera, J. F. et al. Zinc-dependent multimerization of mutant calreticulin is required for MPL binding and MPN pathogenesis. *Blood Adv.* **5**, 1922–1932 (2021).
31. Lyskov, S. & Gray, J. J. The RosettaDock server for local protein-protein docking. *Nucleic Acids Res.* **36**, W233–W238 (2008).
32. Cui, L. et al. Tuning MPL signaling to influence hematopoietic stem cell differentiation and inhibit essential thrombocythemia progenitors. *Proc. Natl Acad. Sci. USA* **118**, e2017849118 (2021).
33. Venkatesan, A. et al. Mechanism of mutant calreticulin-mediated activation of the thrombopoietin receptor in cancers. *J. Cell Biol.* **220**, e202009179 (2021).
34. Masubuchi, N. et al. Mutant calreticulin interacts with MPL in the secretion pathway for activation on the cell surface. *Leukemia* **34**, 499–509 (2020).
35. Lau, A. M., Claesen, J., Hansen, K. & Politis, A. Deuteros 2.0: peptide-level significance testing of data from hydrogen deuterium exchange mass spectrometry. *Bioinformatics* **37**, 270–272 (2021).
36. Masson, G. R. et al. Recommendations for performing, interpreting and reporting hydrogen deuterium exchange mass spectrometry (HDX-MS) experiments. *Nat. Methods* **16**, 595–602 (2019).
37. Oberg, K. A., Ruysschaert, J. M. & Goormaghtigh, E. Rationally selected basis proteins: a new approach to selecting proteins for spectroscopic secondary structure analysis. *Protein Sci.* **12**, 2015–2031 (2003).
38. Goormaghtigh, E., Ruysschaert, J. M. & Raussens, V. Evaluation of the information content in infrared spectra for protein secondary structure determination. *Biophys. J.* **90**, 2946–2957 (2006).
39. Drozdetskiy, A., Cole, C., Procter, J. & Barton, G. J. JPred4: a protein secondary structure prediction server. *Nucleic Acids Res.* **43**, W389–W394 (2015).
40. Jones, D. T. Protein secondary structure prediction based on position-specific scoring matrices. *J. Mol. Biol.* **292**, 195–202 (1999).
41. Yachdav, G. et al. PredictProtein—an open resource for online prediction of protein structural and functional features. *Nucleic Acids Res.* **42**, W337–W343 (2014).
42. Petersen, B., Petersen, T. N., Andersen, P., Nielsen, M. & Lundegaard, C. A generic method for assignment of reliability scores applied to solvent accessibility predictions. *BMC Struct. Biol.* **9**, 51 (2009).
43. Rost, B. & Sander, C. Prediction of protein secondary structure at better than 70% accuracy. *J. Mol. Biol.* **232**, 584–599 (1993).
44. Ishida, T. & Kinoshita, K. PrDOS: prediction of disordered protein regions from amino acid sequence. *Nucleic Acids Res.* **35**, W460–W464 (2007).
45. Romero, P. et al. Sequence complexity of disordered protein. *Proteins* **42**, 38–48 (2001).
46. Barik, A. et al. DEPICTER: intrinsic disorder and disorder function prediction server. *J. Mol. Biol.* **432**, 3379–3387 (2020).
47. Kelley, L. A., Mezulis, S., Yates, C. M., Wass, M. N. & Sternberg, M. J. The Phyre2 web portal for protein modeling, prediction and analysis. *Nat. Protoc.* **10**, 845–858 (2015).
48. Webb, B. & Sali, A. Comparative protein structure modeling using MODELLER. *Curr. Protoc. Bioinforma.* **54**, 5.6.1–5.6.37 (2016).
49. van Zundert, G. C. P. et al. The HADDOCK2.2 web server: user-friendly integrative modeling of biomolecular complexes. *J. Mol. Biol.* **428**, 720–725 (2016).
50. Eastman, P. et al. OpenMM 7: Rapid development of high performance algorithms for molecular dynamics. *PLoS Comput Biol.* **13**, e1005659 (2017).
51. Maier, J. A. et al. ff14SB: improving the accuracy of protein side chain and backbone parameters from ff99SB. *J. Chem. Theory Comput* **11**, 3696–3713 (2015).
52. Case D. A. et al. *AMBER 2020* [computer program] (University of California, San Francisco, 2020).

53. Scherer, M. K. et al. PyEMMA 2: a software package for estimation, validation, and analysis of Markov models. *J. Chem. Theory Comput.* **11**, 5525–5542 (2015).
54. Hunter, J. D. Matplotlib: A 2D graphics environment. *Comput. Sci. Eng.* **9**, 90–95 (2007).
55. Westerlund, A. M. & Delemotte, L. InfleCS: clustering free energy landscapes with gaussian mixtures. *J. Chem. Theory Comput.* **15**, 6752–6759 (2019).
56. Paduraru, C. et al. An N-linked glycan modulates the interaction between the CD1d heavy chain and beta 2-microglobulin. *J. Biol. Chem.* **281**, 40369–40378 (2006).
57. Petrescu, A. J. et al. The solution NMR structure of glucosylated N-glycans involved in the early stages of glycoprotein biosynthesis and folding. *EMBO J.* **16**, 4302–4310 (1997).
58. Petrescu, A. J., Petrescu, S. M., Dwek, R. A. & Wormald, M. R. A statistical analysis of N- and O-glycan linkage conformations from crystallographic data. *Glycobiology* **9**, 343–352 (1999).
59. Petrescu, A. J., Wormald, M. R. & Dwek, R. A. Structural aspects of glycomes with a focus on N-glycosylation and glycoprotein folding. *Curr. Opin. Struct. Biol.* **16**, 600–607 (2006).
60. Jo, S., Kim, T., Iyer, V. G. & Im, W. CHARMM-GUI: A web-based graphical user interface for CHARMM. *J. Comput. Chem.* **29**, 1859–1865 (2008).
61. Phillips, J. C. et al. Scalable molecular dynamics on CPU and GPU architectures with NAMD. *J. Chem. Phys.* **153**, 044130 (2020).
62. Guvench, O. et al. CHARMM additive all-atom force field for carbohydrate derivatives and its utility in polysaccharide and carbohydrate–protein modeling. *J. Chem. Theory Comput.* **7**, 3162–3180 (2011).
63. Huang, J. & MacKerell, A. D. Jr CHARMM36 all-atom additive protein force field: validation based on comparison to NMR data. *J. Comput. Chem.* **34**, 2135–2145 (2013).
64. Pastor, R. W. & MacKerell, A. D. Development of the CHARMM force field for lipids. *J. Phys. Chem. Lett.* **2**, 1526–1532 (2011).
65. Palacios, R. & Steinmetz, M. Il-3-dependent mouse clones that express B-220 surface antigen, contain Ig genes in germ-line configuration, and generate B lymphocytes in vivo. *Cell* **41**, 727–734 (1985).

Acknowledgements

We thank Dr. 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 Raphaël Frédérick from Louvain Drug Research Institute (LDRI) for his guidance and expertise in microscale thermophoresis experiments. We also thank Jean-François Collet and Steve O. Smith for their sound advices in the writing process of the manuscript. Funding to S.N.C. is acknowledged from Ludwig Institute for Cancer Research, Fondation contre le cancer, Salus Sanguinis and Fondation "Les avions de Sébastien", projects Action de recherché concertée (ARC) 16/21-073 and WELBIO F 44/8/5 - MCF/UIG – 10955, avenue Pasteur, 6, 1300 Wavre (Belgium). Funding to G.M. for the HDx-MS platform creation is acknowledged to the FRS-FNS (Appel Grands Equipement 2018, ref:32938497). G.V. has received an Aspirant PhD fellowship from the FRS-FNRS, Belgium. N.P. 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. A.N. performed MST experiments. 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., B.I., T.A.S. and N.P. performed docking predictions, structure refinement and molecular dynamics analyses. N.P., A.N., I.C., C.P. and G.V. created constructs. N.P. A.N. and S.N.C. designed experiments and interpreted data. N.P. and S.N.C wrote the manuscript and supervised the study.

Competing interests

S.N.C. is co-founder of MyeloPro Diagnostics and Research GmbH, Vienna. S.N.C., N.P. and A.N. are co-inventors of a patent application related to the use of inhibitors of the TpoR-mutant CALR interaction (patent application no EP22166303.2). The remaining authors declare no conflict of interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41467-023-37277-3.

Correspondence and requests for materials should be addressed to Stefan N. Constantinescu.

Peer review information *Nature Communications* thanks Vladimir Uversky, and the other, anonymous, reviewers for their contribution to the peer review of this work.

Supplementary Information

Oncogenic CALR Mutant C-terminus Mediates Dual Binding to the Thrombopoietin Receptor Triggering Complex Dimerization and Activation

Nicolas Papadopoulos[1,2], Audrey Nédélec[1,2], Allison Derenne[3], Teodor Asvadur Şulea[4], Christian Pecquet[1,2], Ilyas Chachoua[1,2,5], Gaëlle Vertenoeil[1,2], Thomas Tilmant[6], Andrei-Jose Petrescu[4], Gabriel Mazzucchelli[6], Bogdan I. Iorga[7], Didier Vertommen[2,8], Stefan N. Constantinescu[1,2,9,10]*

*Corresponding Author: Stefan N. Constantinescu

Email: Stefan.constantinescu@bru.licr.org

[1]Ludwig Institute for Cancer Research Brussels, Brussels, Belgium.

[2]Université Catholique de Louvain and de Duve Institute, Brussels, Belgium.

[3] Spectralys Biotech SRL, rue Auguste Piccard 48, 6041 Gosselies, Belgium

[4]Department of Bioinformatics and Structural Biochemistry, Institute of Biochemistry of the Romanian Academy, Splaiul Independentei 296, Bucharest 060031, Romania.

[5]Bilkent University, Department of Molecular Biology and Genetics, Ankara, Turkey

[6]Mass Spectrometry Laboratory, MolSys Research Unit, Universiy of Liège, 4000, Liège, Belgium.

[7]Université Paris-Saclay, CNRS, Institut de Chimie des Substances Naturelles, UPR 2301, Gif-sur-Yvette, France

[8] de Duve Institute and MASSPROT platform, Brussels, Belgium

[9]Walloon Excelence in Life Sciences and Biotechnology, WELBIO, avenue Pasteur, 6, 1300 Wavre, Belgium.

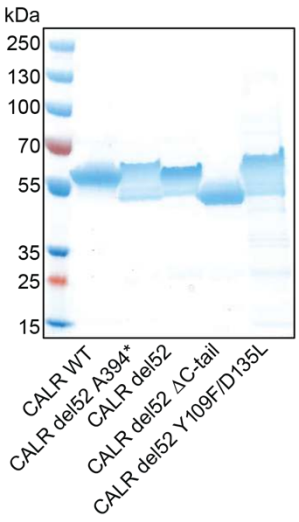
[10]Ludwig Institute for Cancer Research, Nuffield Department of Medicine, Oxford University, Oxford, UK.

Supplementary Figures

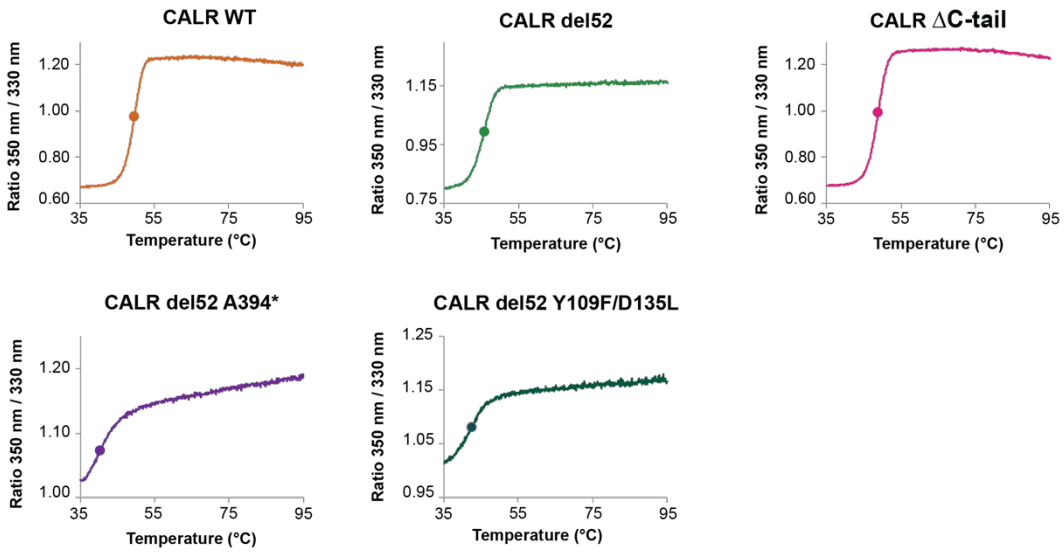
a

CALR species	Theoretical MW	Measured Tm (°C)
CALR WT	47.8 kDa	49.7
CALR del52	47.3 kDa	45.7
CALR del52 A394*	45.3 kDa	40.3
CALR del52 ΔC-tail	41.7 kDa	48.8
CALR del52 Y109F/D135L	47. 3 kDa	42.5

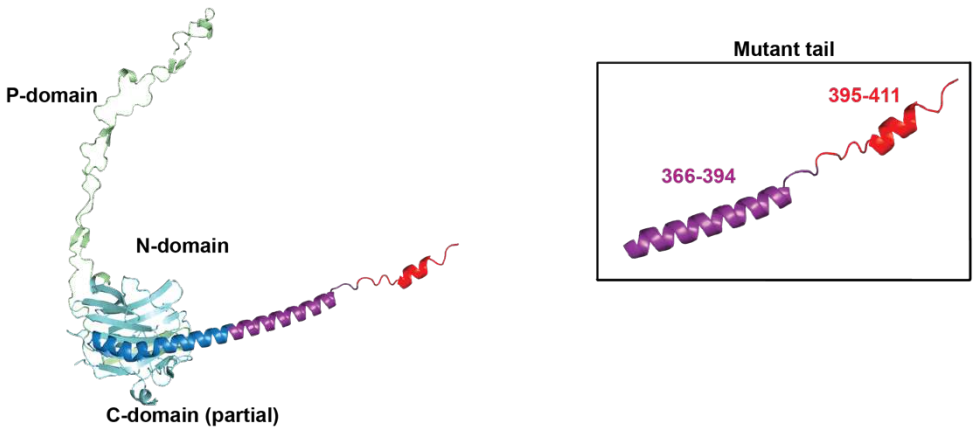
b



c

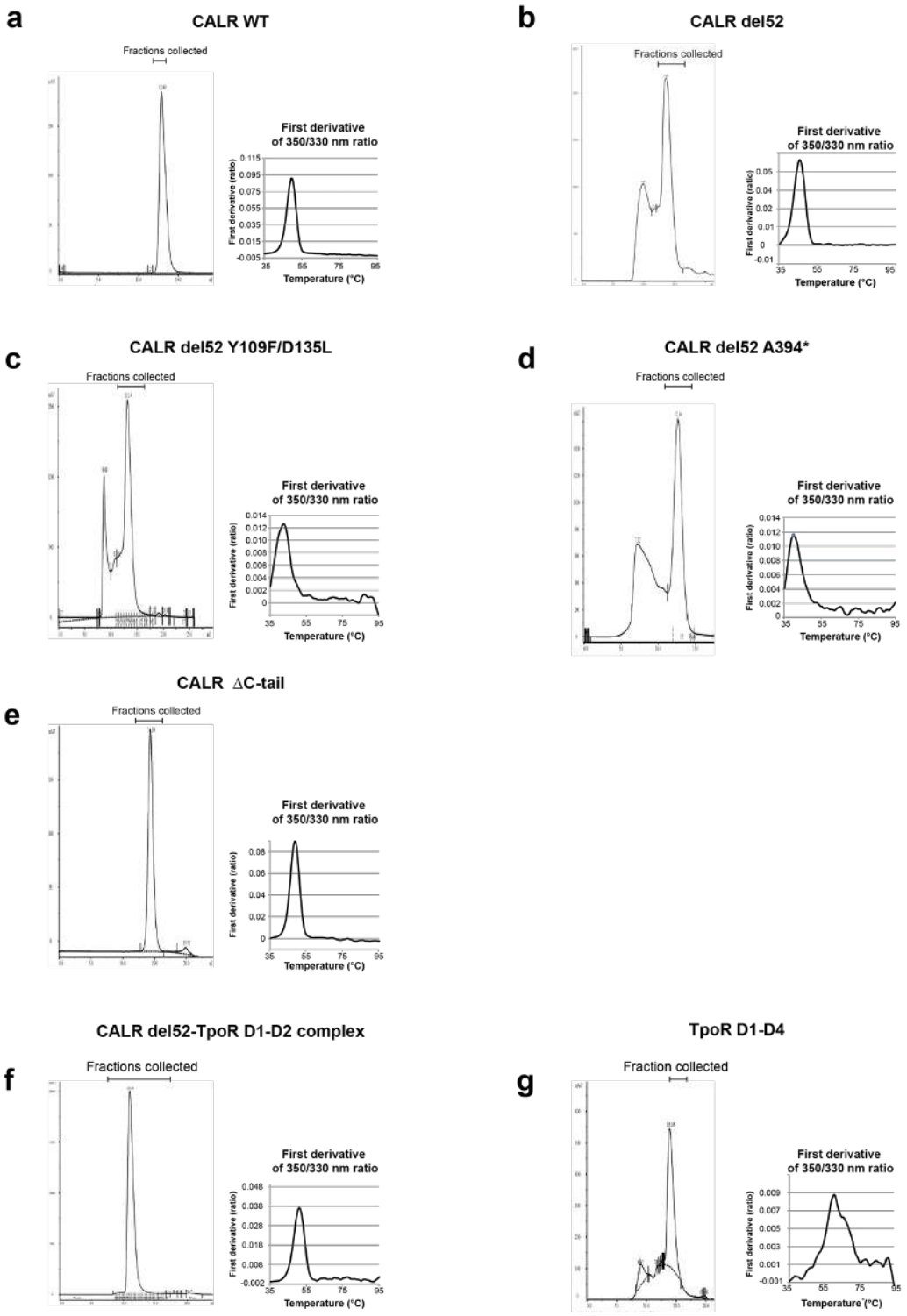


d



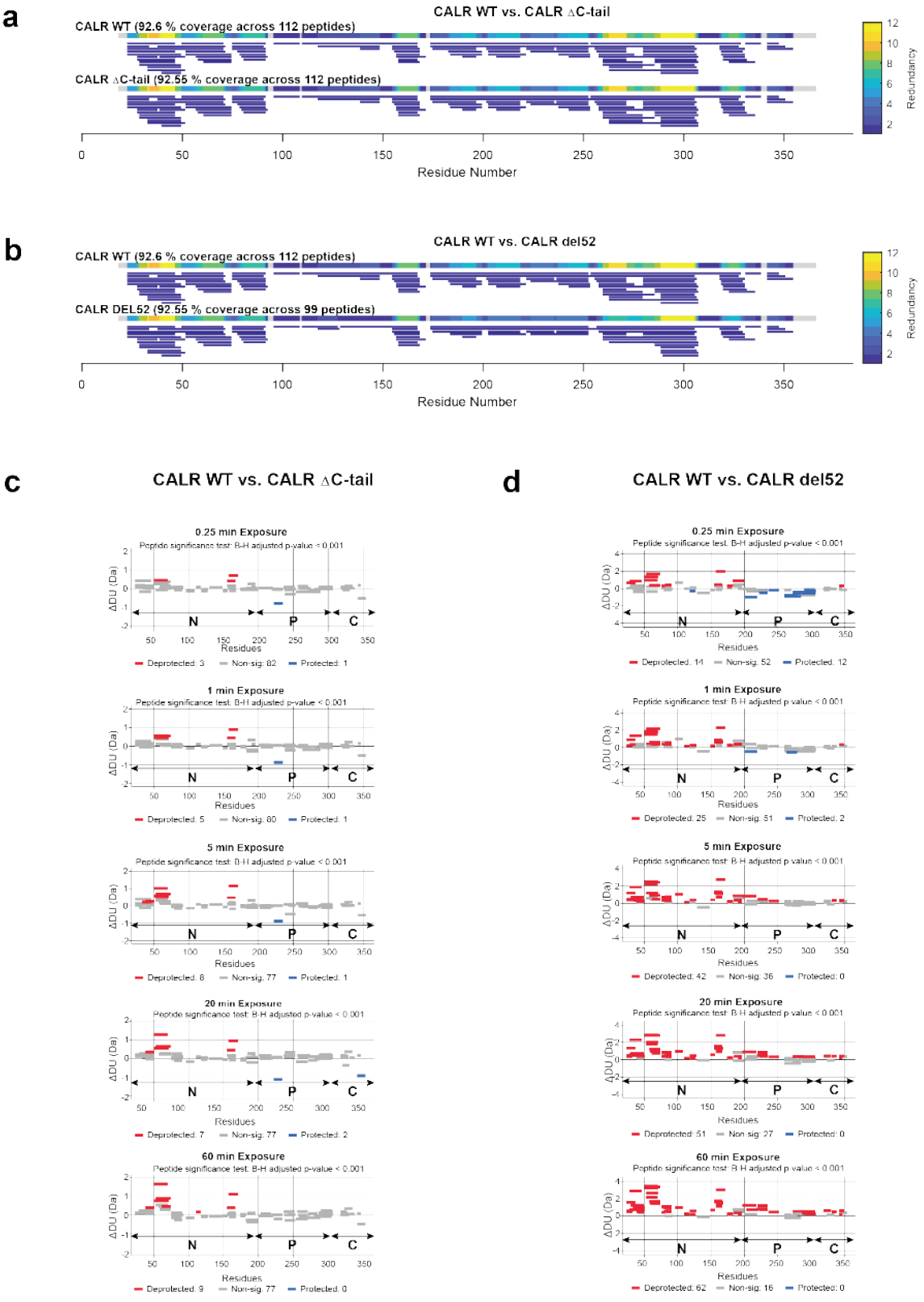
Supplementary Figure 1. CALR species used in this study.

a. Summary of recombinant CALR proteins used in HDx-MS and FTIR experiments. Theoretical molecular weight (MW) and T_m measured with Tycho NT.6 instrument (NanoTemper). **b**. Coomassie Blue staining of CALR WT, CALR del52 and indicated variants. Purified samples were analyzed by SDS-PAGE in denaturing and reducing conditions and stained with Coomassie Blue for total protein detection**.** Representative gel from 3 experiments. **c.** Thermal stability of purified proteins. The graphs represent the 350/330 nm intrinsic fluorescence from Trp and Tyr residues at different temperatures. The S-shaped curve is typical of well-folded proteins as the accessibility of Tyr and Trp residues gradually increases upon temperature-induced protein unfolding. The T_m is computed as the temperature at which half of the proteins in the sample are denatured. Source data are provided as a Source data file. **d.** Structure prediction using AlphaFold 2.0[1] of CALR del52 mutant protein. The mutant C-terminus is highlighted. The N-domain is shown in light blue, the P-domain in green, the partial C-domain of CALR WT in dark blue and the mutant tail in magenta until position 394 and in red from 395 to 411.



Supplementary Figure 2. Size-exclusion chromatography and purity of recombinant proteins.

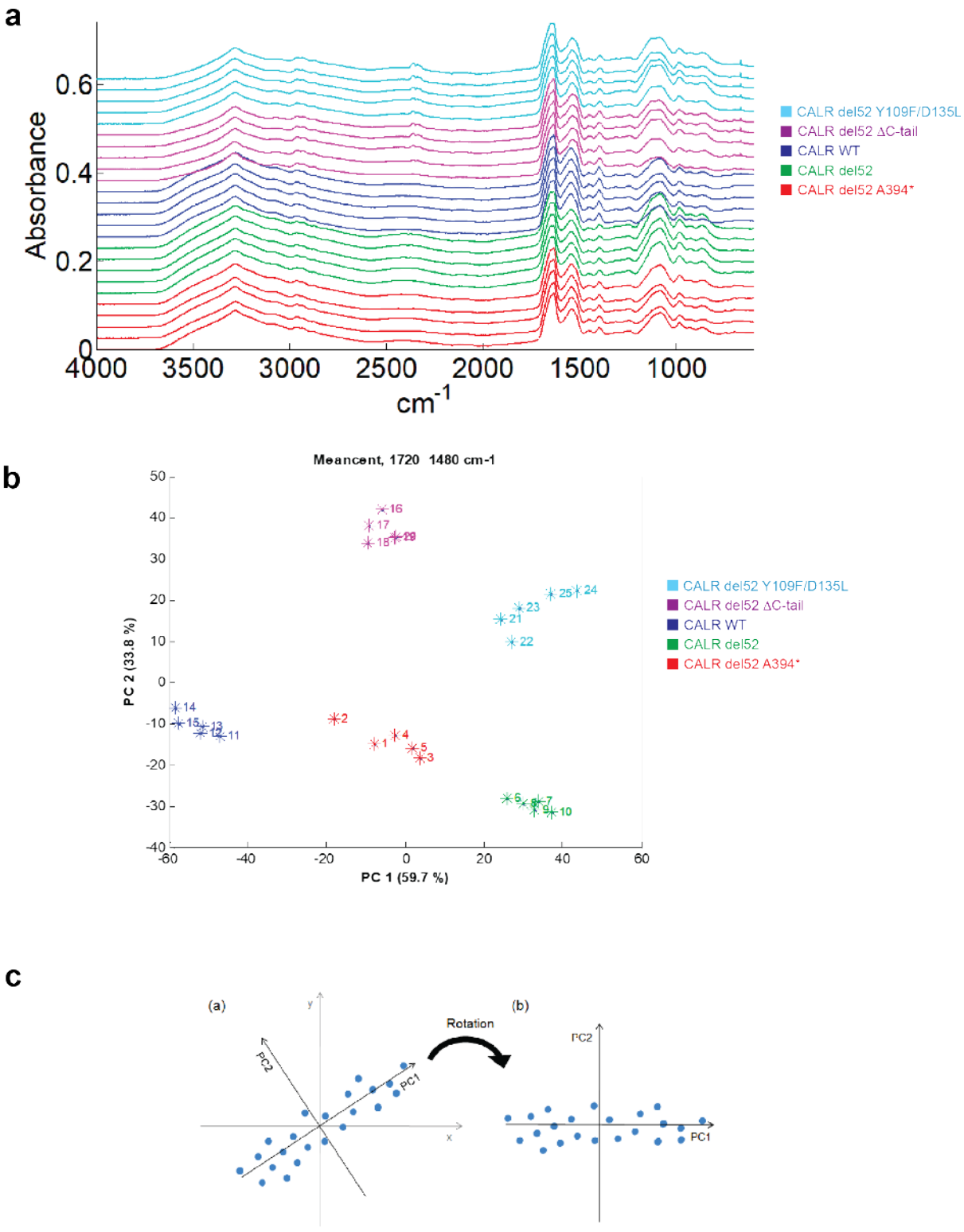
a-g. Left: Size-exclusion chromatograph of indicated proteins in HDx equilibration buffer. The final fraction collected for is indicated. Right: Profile of the first derivative of the 350/330 nm ratio of indicated proteins. The position of the peak on the X axis indicates the Tm of each protein. The shape of the first derivative ratio is used as an additional indicator of protein purity.



Supplementary Figure 3. HDx-MS footprints of CALR WT, CALR del52 and CALR ΔC-tail

a-b. Sequence coverage obtained for the H-D exchange analysis between CALR WT and CALR ΔC-tail (a) and between CALR WT and CALR del52 (b). The percentage of sequence

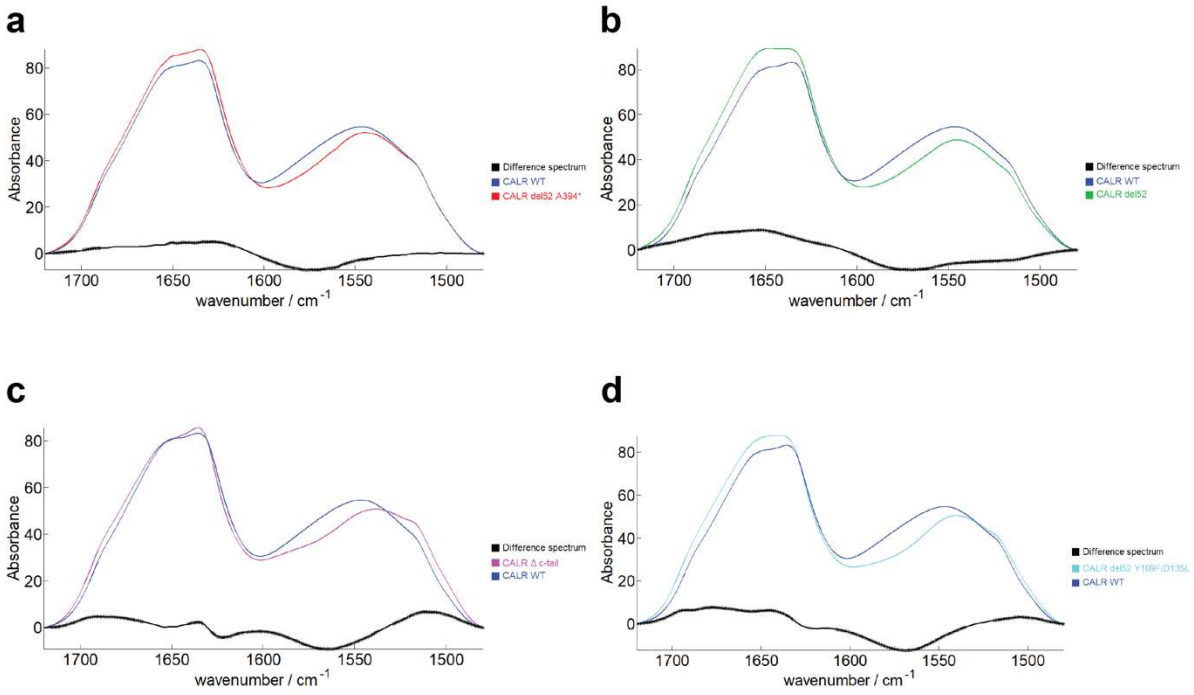
coverage and the number of peptides detected are indicated. **c-d.** Wood's plots linked to Figure 1C-D generated with Deuteros 2.0[2]. Each bar (wood) represents the H-D exchange differential for a single peptide between CALR WT and CALRΔC-tail (c) or between CALR WT and CALR del52 (d) at 0,0.25,1,5,15 or 60 minutes incubation in deuterium. Peptides in red (deprotected) or blue (protected) have significant differential H-D exchange ($p<0.001$) with the peptide-level significance testing ($n = 3$) as described [2]. The N-, P- and C-domains of CALR are indicated on the plots by letters N, P and C, respectively. Source data are provided as a Source data file.



Supplementary Figure 4. FTIR spectra of CALR species.

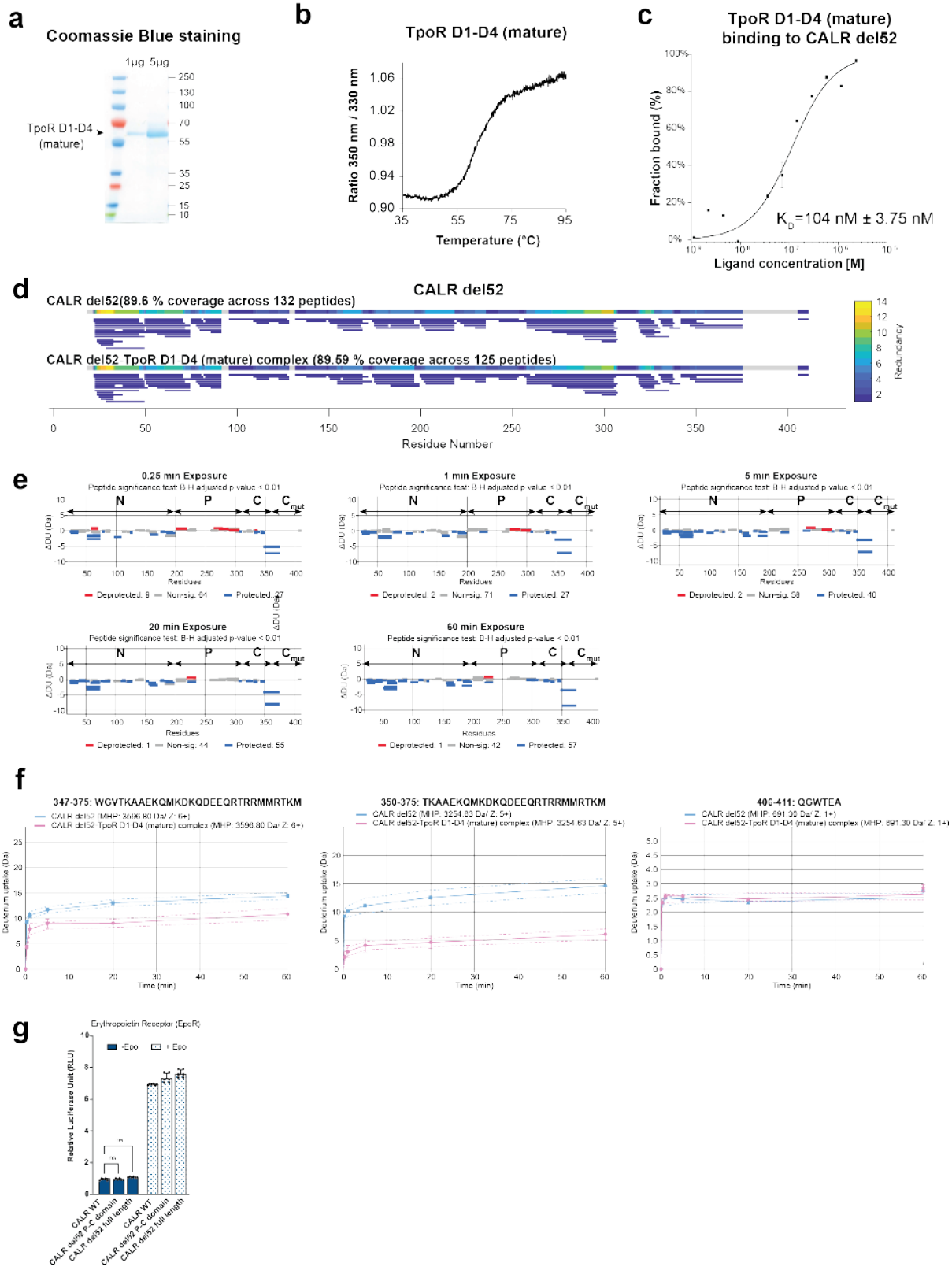
a. Raw data - all FTIR spectra recorded (without any preprocessing step) to analyse the structure. Each sample is identified by a unique color indicated in the legend. For better readability, spectra were offset along the absorbance axis. **b.** PCA score plot depicting the projection of the 25 individual preprocessed FTIR spectra of the WT and mutant proteins in

the PC1-PC2 space performed on the 1720-1480 cm^{-1} spectral region. Each star stands for one spectrum. For the sake of clarity, a colour is associated with each sample but the analysis is completely unsupervised. Percentages on the axis labels indicate the variance described by PC1 (59.7%) and PC2 (33.8%). A mean centering (subtraction of the arithmetic mean from all the spectra) was applied on this set of data. **c.** (a) Representation of the original data in terms of the two axes x and y. Each point represents one IR spectra. (b) As a result of the PCA the axes are rotated and the data will be represented in the two-dimensional principal component space. Thereby PC1 represents the largest variance in the data.



Supplementary Figure 5. Raw FTIR spectra of CALR species.

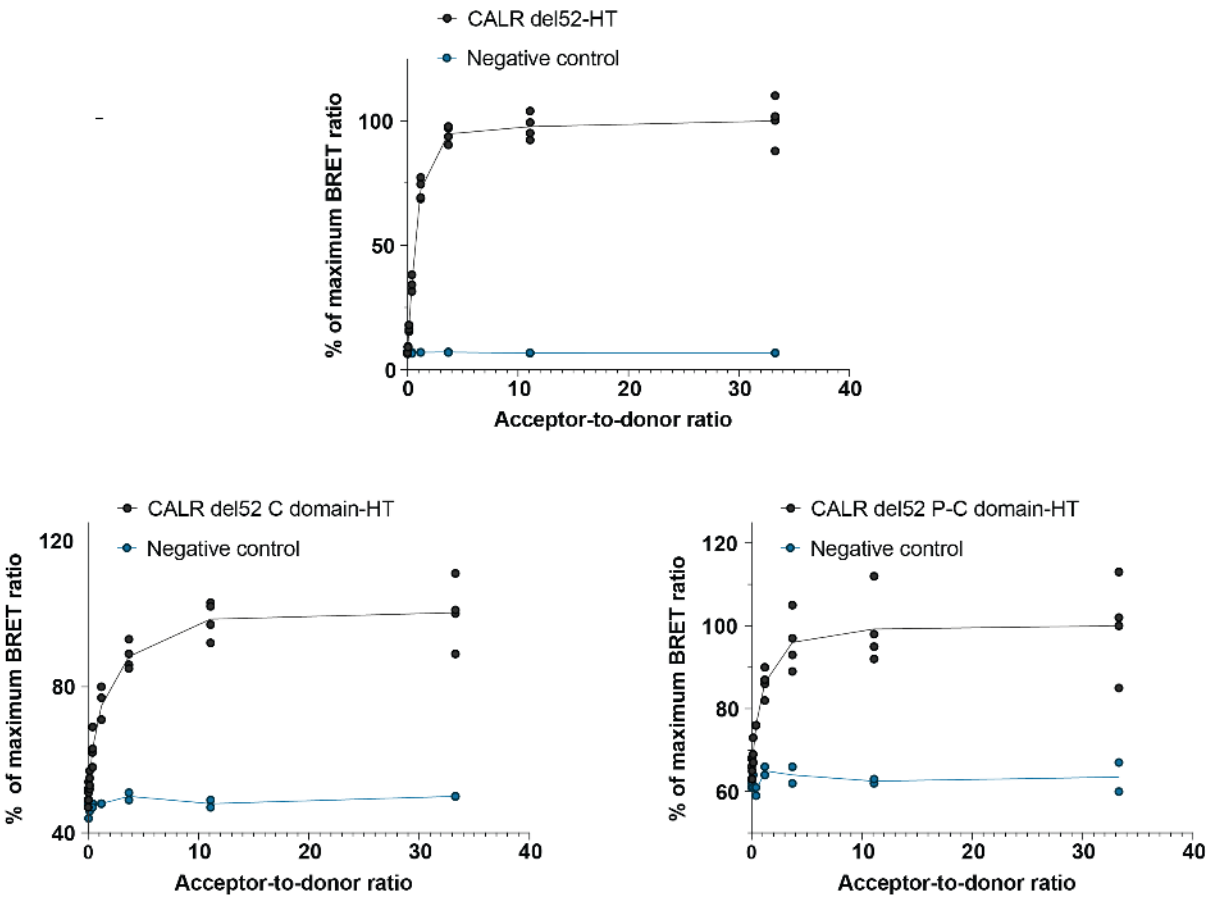
a-d. Comparison of the processed mean spectra recorded. The difference between the two mean spectra indicated is shown in black – zoom on the spectral region related to proteins absorption (1720-1480 cm^{-1}). A Student's t-test was carried out at each wavenumber with a confidence level α=0.1%. The significant spectral differences are revealed with black stars on the difference spectrum. Each spectrum is identified by a unique colour indicated in the legend.



Supplementary Figure 6. CALR mutant C-terminus specifically interacts with TpoR in absence of immature N-glycans.

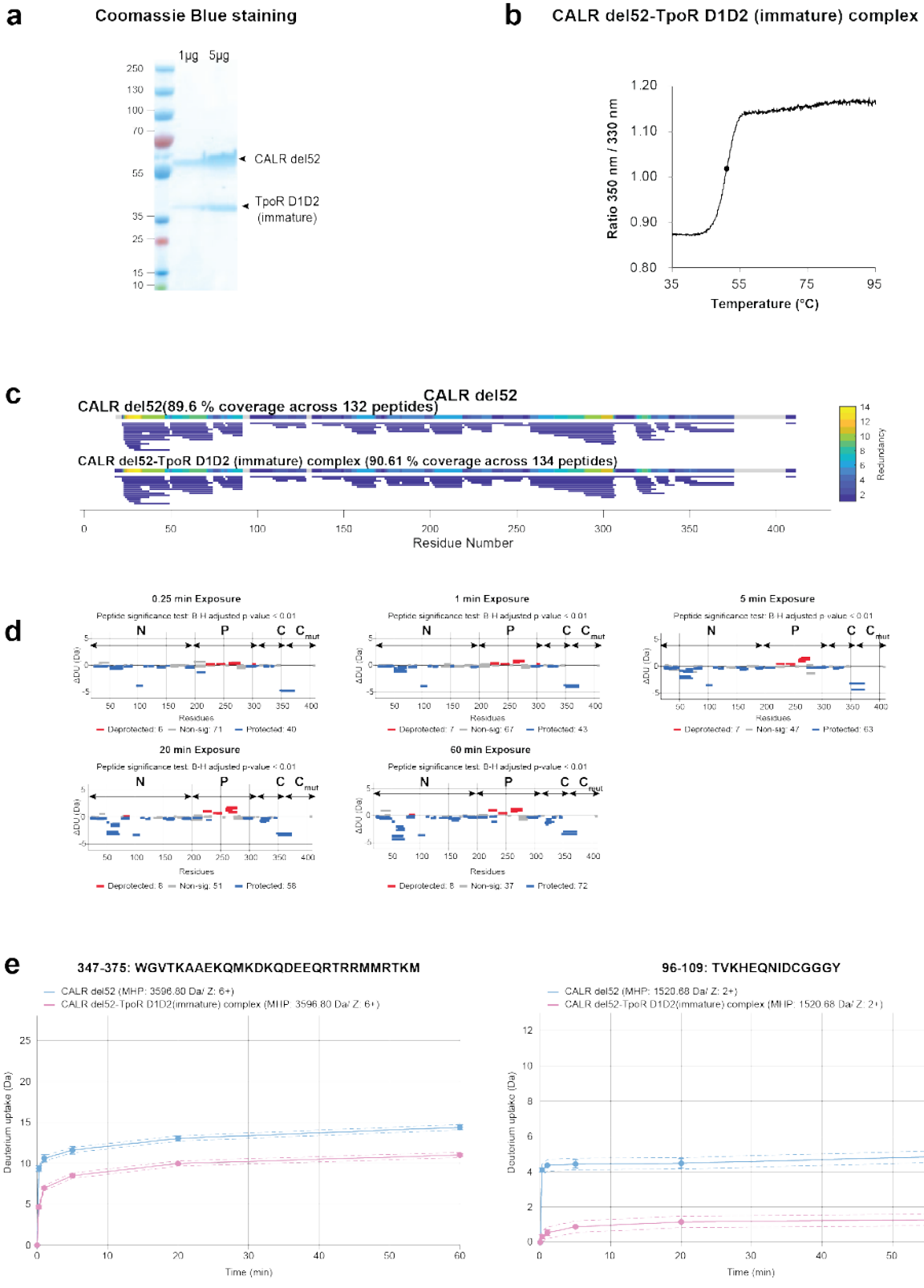
a. Coomassie Blue staining of TpoR D1-D4 with mature N-glycans. Purified TpoR D1-D4 was analyzed by SDS-PAGE in denaturing and reducing conditions and stained with Coomassie

Blue for total protein detection. Representative gel from 3 experiments. **b.** Thermal stability of TpoR D1-D4 with mature N-glycans. The graphs represent the 350/330 nm intrinsic fluorescence from Trp and Tyr residues at different temperatures. The S-shaped curve is typical of well-folded proteins as the accessibility of Tyr and Trp residues gradually increases upon temperature-induced protein unfolding. Source data are provided as a Source data file. **c**. Microscale Thermophoresis between CALR del52 and TpoR D1-D4 with mature N-glycans. CALRdel52 was labeled with RED-NHS 2nd Generation chemistry according to the manufacture's instruction (NanoTemper Technology). The curve corresponds to the mean (± SD) of two independent experiments following the fluorescence of the target protein (CALRdel52-NHS) with titration of the the ligand (TpoR D1-D4). While the concentration of the target is kept constant at 20 nM, the ligand concentration ranges from 5uM and 0,15 nM. The binding curve represents the percentage of bound fraction of CALR del52 to TpoR D1-D4 and yields a K_D of 104 nM ± 3.75 nM. Source data are provided as a Source data file. **d.** Sequence coverage obtained for the H-D exchange analysis between CALR del52 alone and CALR del52 in complex with TpoR D1-D4 with mature N-glycan. The percentage of sequence coverage and the number of peptides detected are indicated. **e.** Wood's plots linked to Figure 2A generated with Deuteros 2.0[2]. Each bar (wood) represents the H-D exchange differential for a single peptide between CALR del52 alone and CALR del52 in complex with TpoR D1-D4 with mature N-glycans at 0,0.25,1,5,15 or 60 minutes incubation in deuterium. Peptides in red (deprotected) or blue (protected) have significant differential H-D exchange ($p<0.001$) with the peptide-level significance testing (n = 3) as described [2]. The N-, P- and WT C and mutant C-domains of CALR are indicated on the plots by letters N, P and C and C_{mut}, respectively. Source data are provided as a Source data file. **f**. Deuterium uptake (Da) of the indicated peptides of the mutant C-terminus from CALR del52 alone or in complex with TpoR D1D4 (with mature N-glycans) at 5 different exchange time points. The dotted lines represent standard deviation (SD), the full line represents average of triplicates. Source data are provided as a Source data file. **g.** STAT5 transcriptional activity in presence of EpoR and indicated CALR truncations. HEK293T were transiently transfected with human EpoR 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. Data represent mean ± SD (n = 6 biologically independent samples over 2 independent experiments). Data were analyzed with two-ways ANOVA followed by Sidak multiple comparison test. ns: non-significant ($p>0.05$). Source data are provided as a Source data file.



Supplementary Figure 7. Donor Saturation Assay between TpoR-NanoLuc and CALR del52-HaloTag.

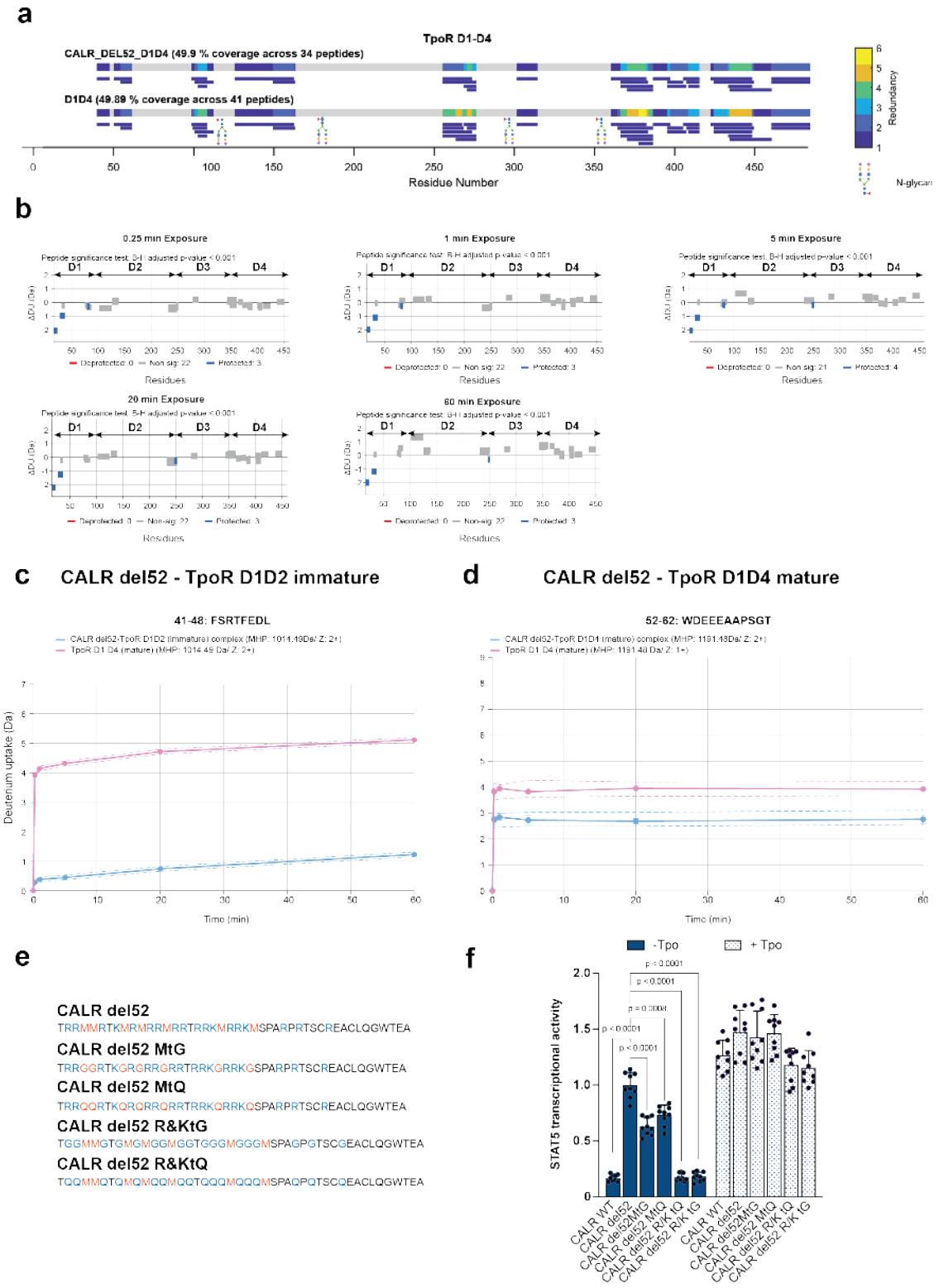
Donor Saturation Assay between TpoR-NanoLuc and indicated CALR del52-HaloTag (HT) construct. HEK293T were co-transfected with fixed amount of donor (TpoR-NanoLuc) and increasing ratios of acceptor (HaloTag fusion proteins). The negative control corresponds to a HaloTag protein non-fused to TpoR. A specific BRET signal will increase in a hyperbolic manner before reaching a plateau. A non-specific interaction will be less intense and increase linearly without reaching a plateau. The shape of the curve in a Donor Saturation Assay provides a control for the specificity of the interaction according to the manufacturer instruction (Promega).



Supplementary Figure 8. Interaction between CALR mutant and TpoR in presence of immature N-glycans.

a. Coomassie Blue staining of the complex CALR del52-TpoR D1D2 with immature N-glycans on Asn117. The purified complex was analyzed by SDS-PAGE in denaturing and reducing

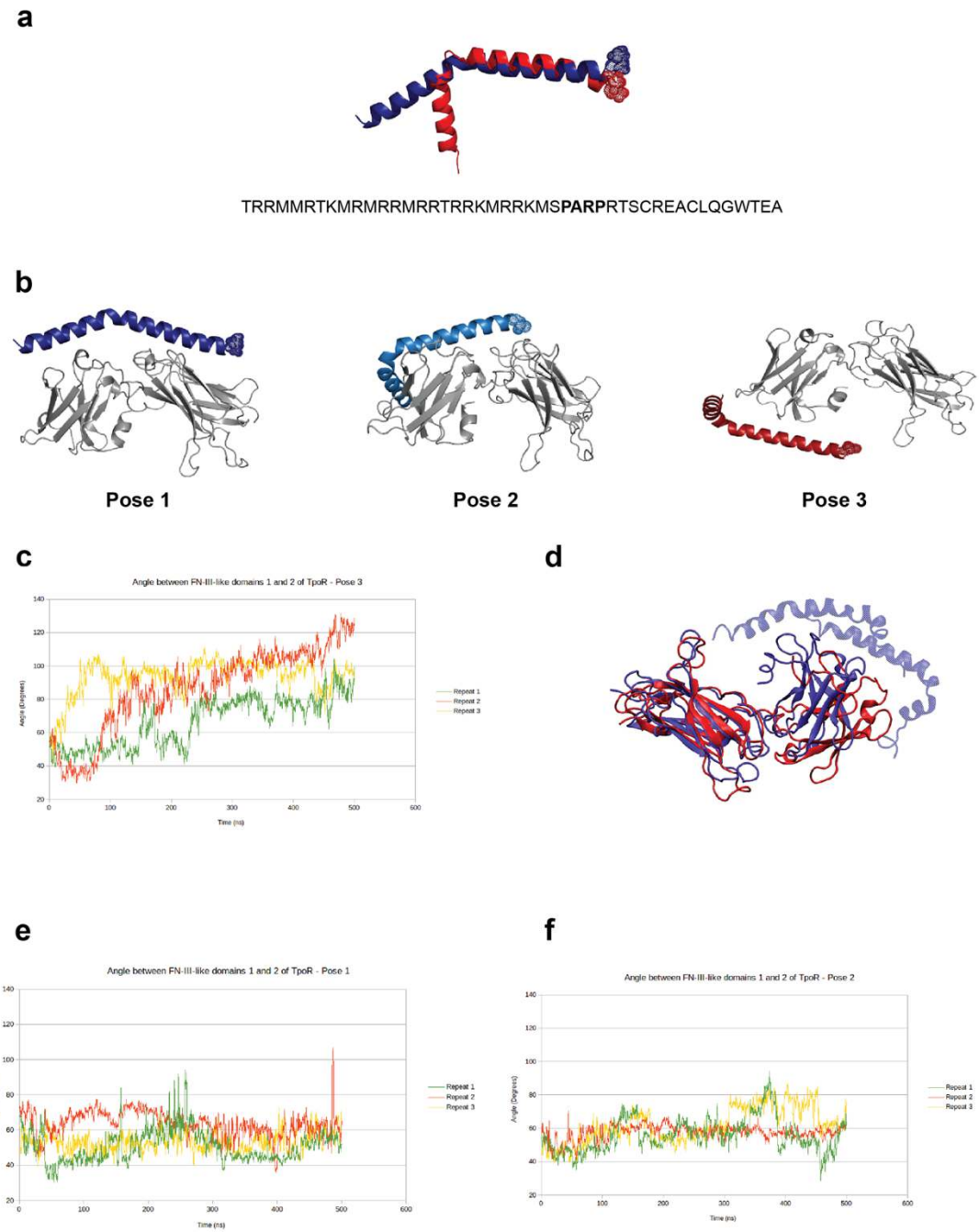
conditions and stained with Coomassie Blue for total protein detection. Representative gel from 3 experiments. **b.** Thermal stability of the CALR del52-TpoR D1D2 complex. The graphs represent the 350/330 nm intrinsic fluorescence from Trp and Tyr residues at different temperatures. The S-shaped curve is typical of well-folded proteins as the accessibility of Tyr and Trp residues gradually increases upon temperature-induced protein unfolding. Source data are provided as a Source data file. **c.** Sequence coverage obtained for the H-D exchange analysis between CALR del52 alone and the CALR del52-TpoR D1D2 complex with immature N-glycans on TpoR Asn117. The percentage of sequence coverage and the number of peptides detected are indicated. **d.** Wood's plots linked to Figure 3E generated with Deuteros 2.0[2]. Each bar (wood) represents the H-D exchange differential for a single peptide between CALR del52 alone and the CALR del52-TpoR D1D2 complex at 0,0.25,1,5,15 or 60 minutes incubation in deuterium. Peptides in red (deprotected) or blue (protected) have significant differential H-D exchange ($p<0.001$) with the peptide-level significance testing ($n = 3$) as described [2]. The N-, P- WT C and mutant C-domains of CALR are indicated on the plots by letters N, P and C and C_{mut}, respectively. Source data are provided as a Source data file. **e**. Deuterium uptake (Da) of the indicated peptides of the mutant C-terminus (left) or N-glycans binding sites (right) from CALR del52 alone or in complex with TpoR D1D2 (with immature N-glycans on Asn117) at 5 different exchange time points. The dotted lines represent standard deviation (SD), the full line represents average of triplicates. Source data are provided as a Source data file.



Supplementary Figure 9. Interaction between CALR mutant C-terminus and acidic patches on TpoR D1 domain.

a. Sequence coverage obtained for the H-D exchange analysis between TpoR D1-D4 alone (with mature N-glycans) and in presence of CALR del52. The percentage of sequence coverage and the number of peptides detected are indicated. The peptide signal sequence is

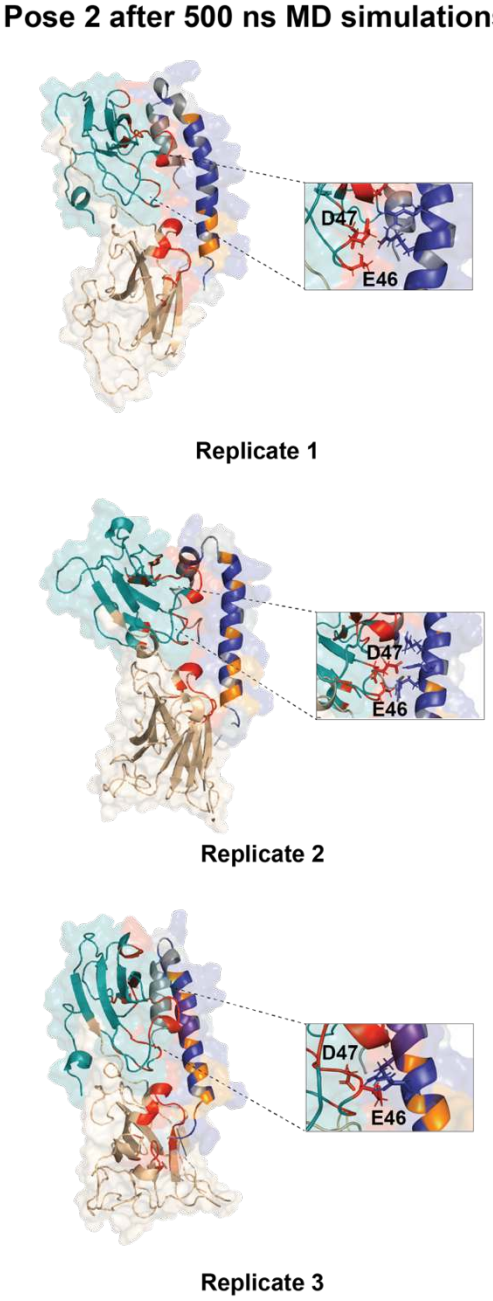
not shown. The sites of N-glycans attachment are indicated. **b.** Wood's plots generated with Deuteros 2.0[2]. Each bar (wood) represents the H-D exchange differential for a single peptide between TpoR D1-D4 alone and the CALR del52-TpoR D1-D4 complex at 0,0.25,1,5,15 or 60 minutes incubation in deuterium. Peptides in red (deprotected) or blue (protected) have significant differential H-D exchange ($p<0.001$) with the peptide-level significance testing as described [2]. The D1, D2,D3 and D4 domains of TpoR are indicated. Source data are provided as a Source data file. **c-d**. Deuterium uptake (Da) of (C) the FSRTFEDL peptide from TpoR D1-D4 alone and from the CALR del52-TpoR D1D2 complex and (D) the WDEEEAAPSGT peptide from TpoR D1-D4 alone and in presence of CALR del52 at 5 different exchange time points. The dotted lines represent standard deviation (SD), the full line represents average of triplicates. Source data are provided as a Source data file. **e.** Sequence of CALR del52 C-terminus and mutants as indicated. **f.** STAT5 transcriptional activity with human TpoR and indicated CALR del52 mutants. HEK293T were transiently transfected with human TpoR and CALR mutants 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. Data represent mean $\pm$ SD (n = 9 biologically independent samples over 3 independent experiments). Data were analyzed with two-ways ANOVA followed by Sidak multiple comparison test. Source data are provided as a Source data file.



Supplementary Figure 10. Molecular Dynamics between CALR mutant C-terminus and TpoR D1D2 domains.

a. Representative conformations of the main 2 clusters of CALR del52 mutant C-terminus with the bending region in bold in sequence. **b.** Poses generated by HADDOCK. The best docking complex was chosen, as ranked by the HADDOCK score. Render done in PyMOL[40] v. 2.2.3 **c.** Timeseries of the angle between the two FN-III-like domains of the membrane distal TpoR domain, moving average with 0.5 ns window. The C_α atoms of Q26, S129 and D283 were used to compute the angle. **d.** Conformation superposition of the TpoR-CALR del52 mutant C-terminus structure before (purple/blue) and after (red) 500 ns of molecular dynamics. **e-f.**

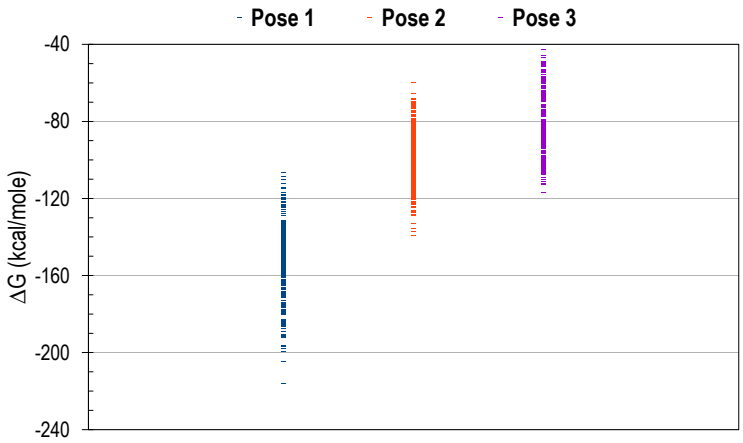
Timeseries of the angle between the two FN-III-like domains of the membrane distal TpoR domain for pose 1 and 2, moving average with 0.5 ns window. The C_α atoms of Q26, S129 and D283 were used to compute the angle.



Supplementary Figure 11. Final frames Molecular Dynamic simulations between CALR mutant C-terminus and TpoR D1D2 domain.

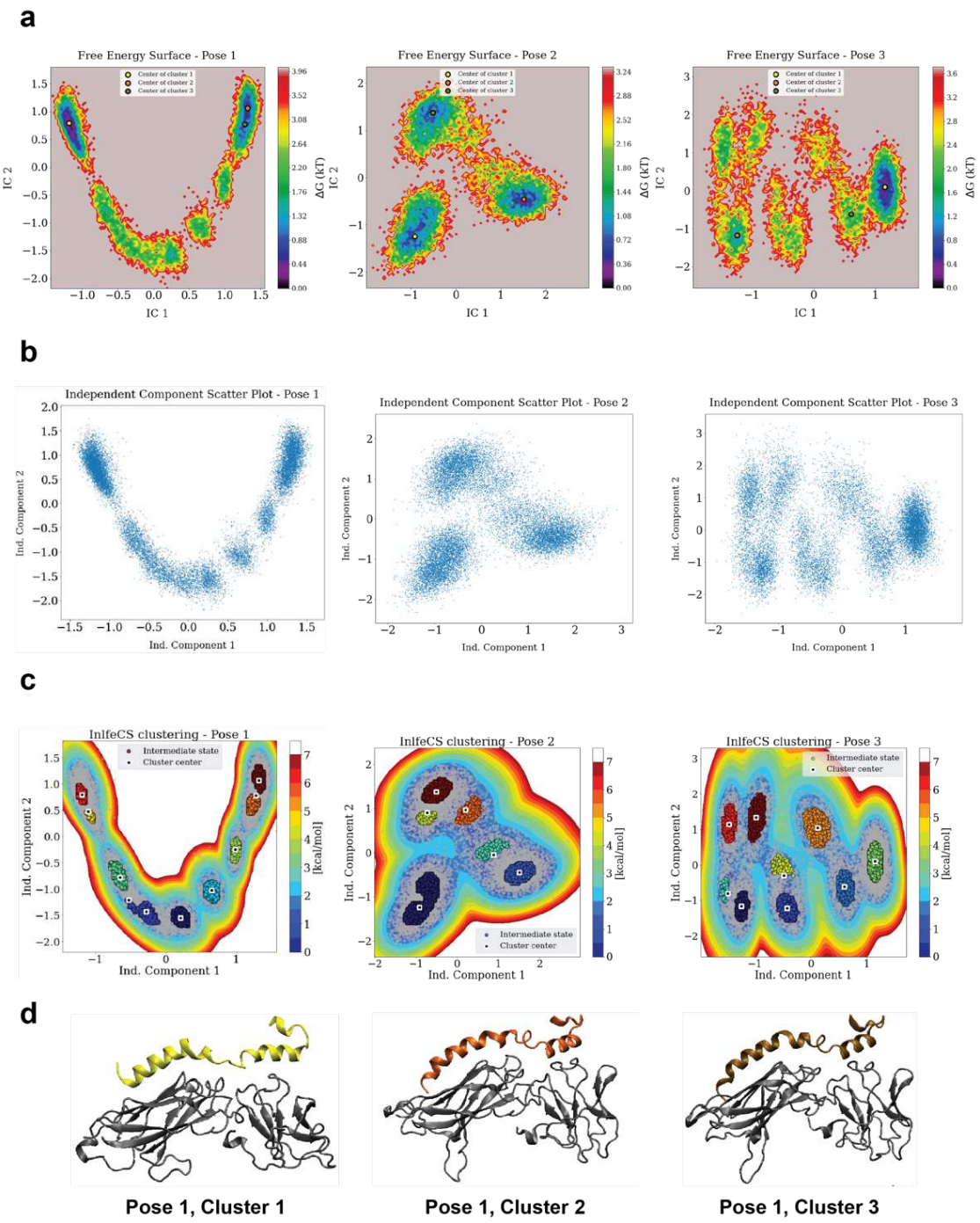
Last frame of triplicates unconstrained molecular dynamics simulations (500 ns) for pose 2. The last frame is shown for each replicate independently. The D1 and D2 domain of TpoR are shown in blue/deepteal and beige, respectively. The Arg and Lys of CALR del52 C-terminus are shown in dark blue, the Met in orange and other residues in grey. Residues of TpoR that with a distance < 3.5 Ä from CALR del52 C-terminus are shown in red. The E and F of the

^{44}TFED^{47} motif that interact with Arg of CALR del52 C-terminus are highlighted. Render done with Pymol 2.4.2.



Supplementary Figure 12. Free energy between CALR mutant C-terminus and TpoR D1D2 domain.

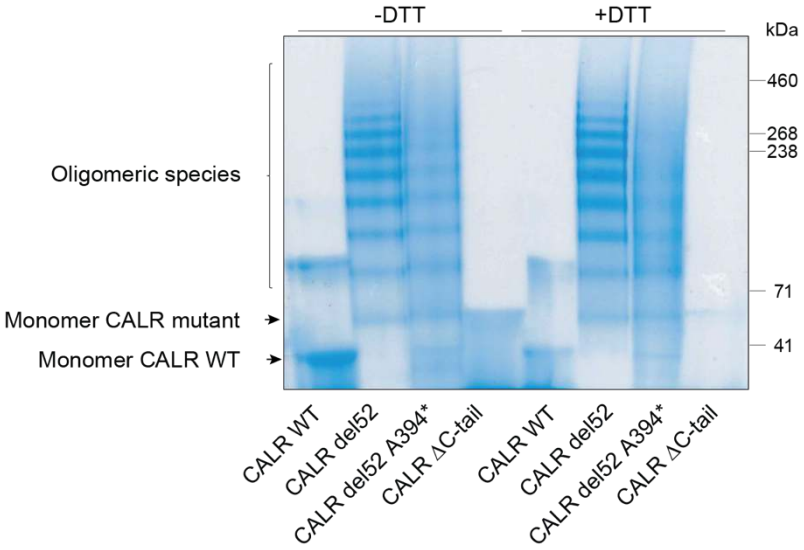
Comparative MM-GBSA free energy differences, and their associated standard deviation (SD). Data represent the mean (± SD) of MM-GBSA free energy computed on 100 simulation frames (N = 100) for each pose. The average is -153.43 (± SD 18.49) ΔG (kcal/mol) for pose 1, -100.45 (± SD 14.58) ΔG (kcal/mol) for pose 2 and -81.94 (± SD 17.01) ΔG (kcal/mol) for pose 3.



Supplementary Figure 13. *In silico* analyses of the CALR mutant C-terminus interaction with TpoR D1D2 domain.

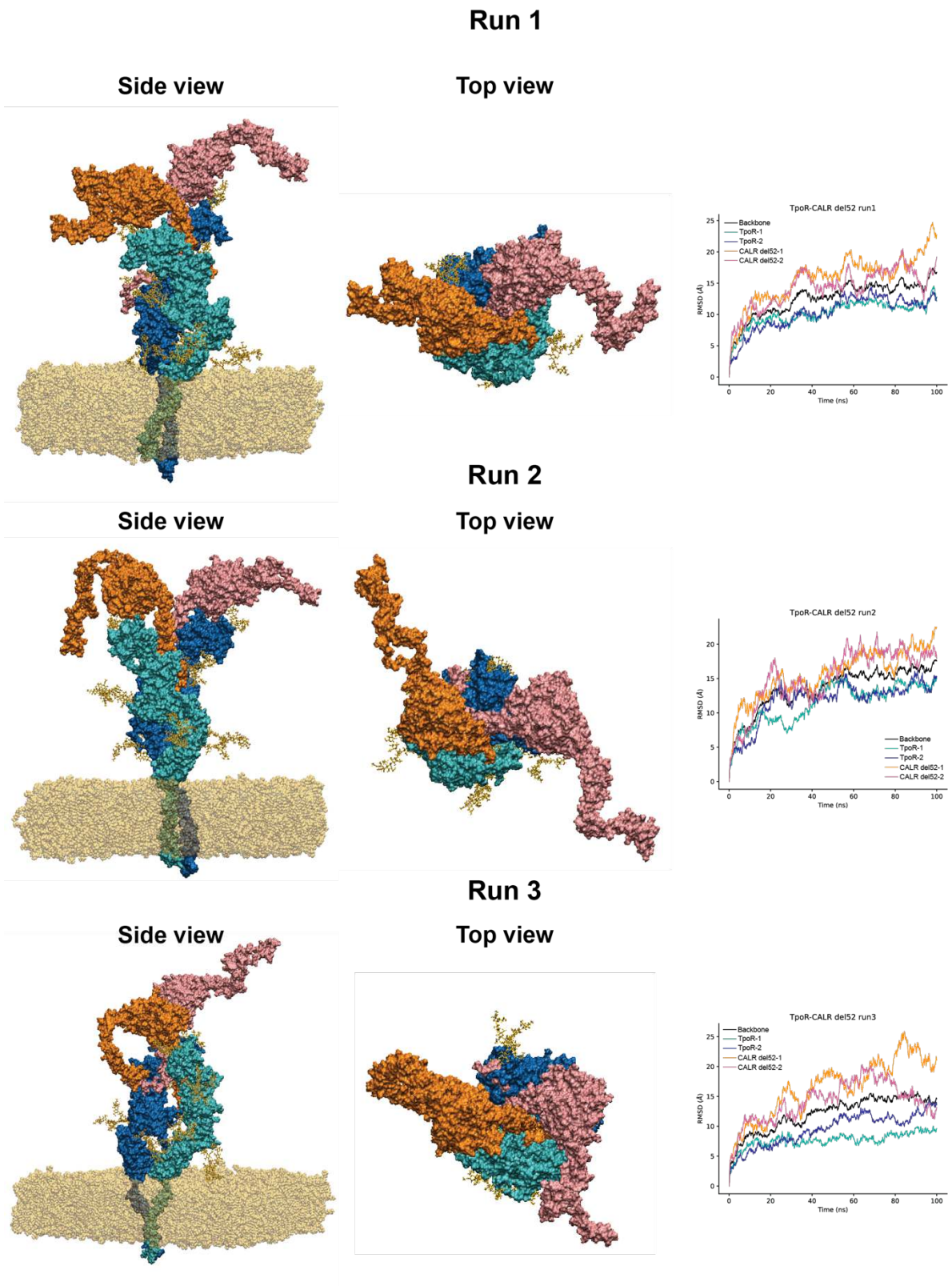
a. Free energy surface of the independent components for pose 1, pose 2 and pose 3. The centers of the top 3 most probable clusters are shown (see also **Supplementary Methods**). **b.** Top two most auto-correlated components generated by TICA for pose 1, pose 2 and pose 3**. c.** InfleCS cluster of the independent components for pose 1, pose 2 and pose 3. The generated clusters are shown in colored circles and their associated centers in squares. **d.** Cluster center conformation of pose 1. Left: cluster 1, middle: cluster 2, right: cluster 3. These

corresponds to clusters 9, 10 and 8 of **Supplementary Table 1**, respectively. Renders done in VMD v1.9.4.



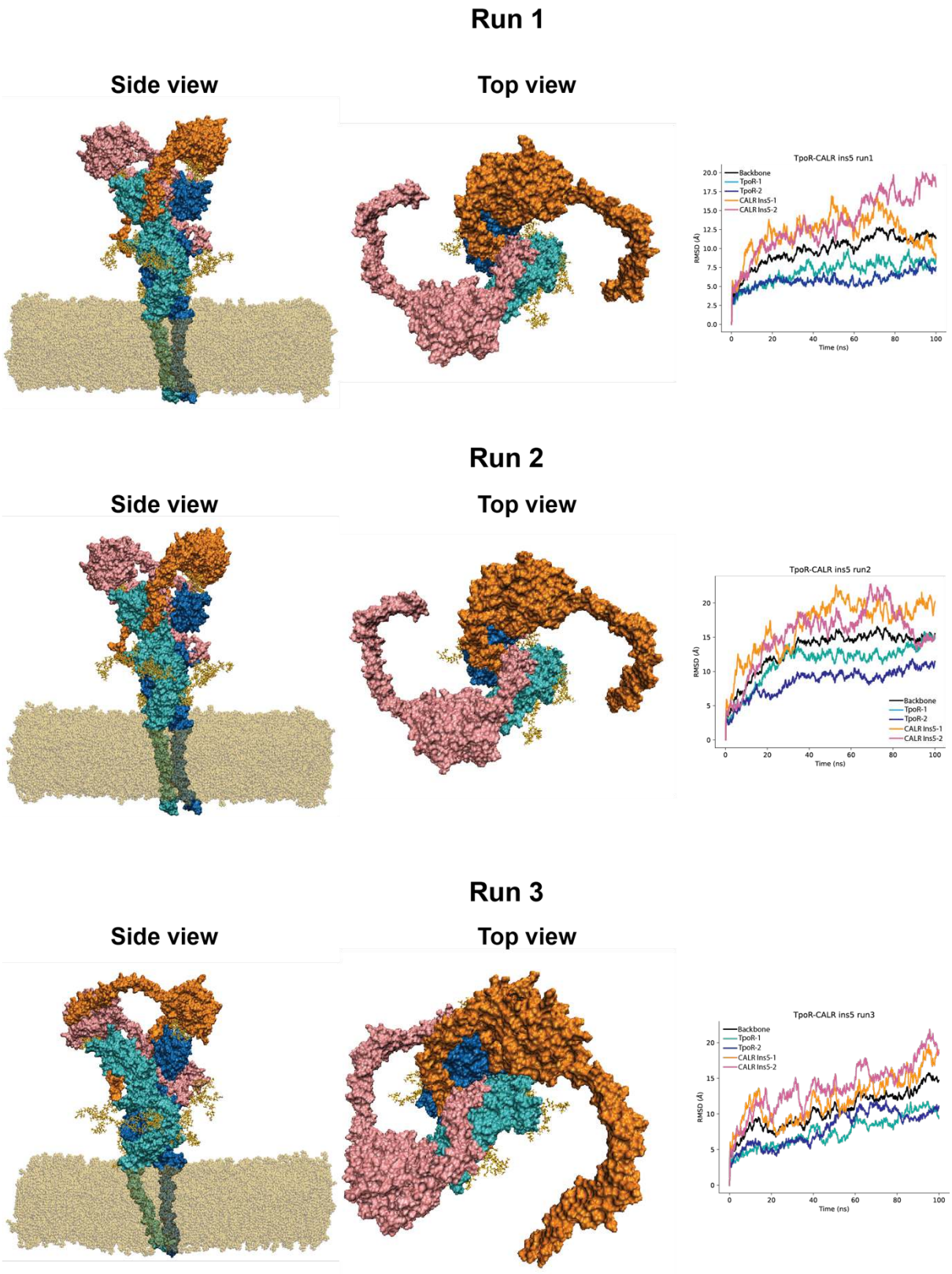
Supplementary Figure 14. CALR mutant oligomerization is independent of C-terminal cysteines.

Native western blot of indicated CALR species, with or without reducing agent (DTT). Staining with Coomassie Blue. Representative gel from 3 experiments.



Supplementary Figure 15. Molecular Dynamic Simulations of CALR del52-TpoR tetrameric complex.

Left: side view of the last frame of unconstrained molecular dynamics simulations (100 ns) of the CALR del52-TpoR tetrameric complex performed in triplicates. CALR del52 molecules are shown in orange and pink. TpoR molecules are shown in cyan and dark blue. N-glycans attached to Asn117 (immature), Asn178 (mature), Asn298 (mature) and Asn358 (mature) are shown as yellow sticks. Middle: Top views of the last frame of triplicates molecular dynamics CALR del52-TpoR tetrameric complex. Right: RMSD plots of each molecular dynamic simulations. RMSD over the simulations courses are shown in black (backbone), cyan/light blue (TpoR-1), dark blue (TpoR-2), orange (CALR del52-1) and pink (CALR del52-2).



Supplementary Figure 16. Molecular Dynamic Simulations of CALR Ins5-TpoR tetrameric complex.

Left: side view of the last frame of unconstrained molecular dynamics simulations (100 ns) of the CALR Ins5-TpoR tetrameric complex performed in triplicates. CALR Ins5 molecules are

shown in orange and pink. TpoR molecules are shown in cyan and dark blue. N-glycans attached to Asn117 (immature), Asn178 (mature), Asn298 (mature) and Asn358 (mature) are shown as yellow sticks. Middle: Top views of the last frame of triplicates molecular dynamics CALR Ins5-TpoR tetrameric complex. Right: RMSD plots of each molecular dynamic simulations. RMSD over the simulations courses are shown in black (backbone), cyan/light blue (TpoR-1), dark blue (TpoR-2), orange (CALR Ins5-1) and pink (CALR Ins5-2).

Supplementary Table

Cluster Number	State Probability (%)	Cluster Population (determined by Voronoi Tesselation)
1	3%	1441
2	3.2%	886
3	0.05%	595
4	2.3%	955
5	3.3%	1102
6	2.2%	991
7	1.8%	1758
8	4.3%	2035
9	7.6%	3245
10	6.7%	1992

Supplementary Table 1. Detailed overview of the generated InfleCS clusters for pose 1 linked to Supplementary Figure 13d.

Supplementary Methods

Production and purification of recombinant proteins

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 (a.a. 1-17). In addition, the cysteine 163 was mutated to serine. The amino acid sequence of human TpoR D1D2D3D4 (TpoR D1-D4) starts at Q26 and ends at T489 and this of human TpoR D1D2 (TpoR D1D2) starts at Q26 and ends at Q290. Both contains a histidine tag at the C-terminus. The amino acid sequence of human CALR WT starts at E18 and ends at L417.

The sequences of all recombinant proteins are provided below.

Protein	Sequence
rhCALRWT	MGSHHHHHHGSSGEPAVYFKEQFLDGDGWTSRWIESKHKSDFGKFVLSSG KFYGDEEKDKGLQTSQDARFYALSASFEPFSNKGQTLVVQFTVKHEQNIDC GGGYVKLFPNSLDQTDMHGDSEYNIMFGPDICGPGTKKVHVIFNYKGKNVLI NKDIRSKDDEFTHLYTLIVRPDNTYEVKIDNSQVESGSLEDDWDFLPPKKIKD PDASKPEDWDERAKIDDPTDSKPEDWDKPEHIPDPDAKKPEDWDEEMDGE WEPPVIQNPEYKGEWKPRQIDNPDYKGTWIHPEIDNPEYSPDPSIYAYDNFG VLGLDLWQVKSGTIFDNFLITNDEAYAEEFGNETWGVTKAAEKQMKDKQDE EQRLKEEEEDKKRKEEEEAEDKEDDEDKDEDEEDEEDKEEDEEEDVPGQA KDEL
rhCALRdel52	MGSHHHHHHGSSGEPAVYFKEQFLDGDGWTSRWIESKHKSDFGKFVLSSG KFYGDEEKDKGLQTSQDARFYALSASFEPFSNKGQTLVVQFTVKHEQNIDC GGGYVKLFPNSLDQTDMHGDSEYNIMFGPDICGPGTKKVHVIFNYKGKNVLI NKDIRSKDDEFTHLYTLIVRPDNTYEVKIDNSQVESGSLEDDWDFLPPKKIKD PDASKPEDWDERAKIDDPTDSKPEDWDKPEHIPDPDAKKPEDWDEEMDGE WEPPVIQNPEYKGEWKPRQIDNPDYKGTWIHPEIDNPEYSPDPSIYAYDNFG VLGLDLWQVKSGTIFDNFLITNDEAYAEEFGNETWGVTKAAEKQMKDKQDE EQRTRRMMRTKMRMRRMRRTRRKMRRKMSPARPRTSCREACLQGWTEA
rhCALRdel52 A394*	MGSHHHHHHGSSGEPAVYFKEQFLDGDGWTSRWIESKHKSDFGKFVLSSG KFYGDEEKDKGLQTSQDARFYALSASFEPFSNKGQTLVVQFTVKHEQNIDC GGGYVKLFPNSLDQTDMHGDSEYNIMFGPDICGPGTKKVHVIFNYKGKNVLI NKDIRSKDDEFTHLYTLIVRPDNTYEVKIDNSQVESGSLEDDWDFLPPKKIKD PDASKPEDWDERAKIDDPTDSKPEDWDKPEHIPDPDAKKPEDWDEEMDGE WEPPVIQNPEYKGEWKPRQIDNPDYKGTWIHPEIDNPEYSPDPSIYAYDNFG

	VLGLDLWQVKSGTIFDNFLITNDEAYAEEFGNETWGVTKAAEKQMKDKQDE EQRTRRMMRTKMRMRRMRRTRRKMRRKMSP
rhCALR **ΔC-tail**	MGSHHHHHHGSSGEPAVYFKEQFLDGDGWTSRWIESKHKSDFGKFVLSSG KFYGDEEKDKGLQTSQDARFYALSASFEPFSNKGQTLVVQFTVKHEQNIDC GGGYVKLFPNSLDQTDMHGDSEYNIMFGPDICGPGTKKVHVIFNYKGKNVLI NKDIRSKDDEFTHLYTLIVRPDNTYEVKIDNSQVESGSLEDDWDFLPPKKIKD PDASKPEDWDERAKIDDPTDSKPEDWDKPEHIPDPDAKKPEDWDEEMDGE WEPPVIQNPEYKGEWKPRQIDNPDYKGTWIHPEIDNPEYSPDPSIYAYDNFG VLGLDLWQVKSGTIFDNFLITNDEAYAEEFGNETWGVTKAAEKQMKDKQDE EQR
rhCALRdel52 **Y109F/D135L**	MGSHHHHHHGSSGEPAVYFKEQFLDGDGWTSRWIESKHKSDFGKFVLSSG KFYGDEEKDKGLQTSQDARFYALSASFEPFSNKGQTLVVQFTVKHEQNIDC GGGFVKLFPNSLDQTDMHGDSEYNIMFGPLICGPGTKKVHVIFNYKGKNVLI NKDIRSKDDEFTHLYTLIVRPDNTYEVKIDNSQVESGSLEDDWDFLPPKKIKD PDASKPEDWDERAKIDDPTDSKPEDWDKPEHIPDPDAKKPEDWDEEMDGE WEPPVIQNPEYKGEWKPRQIDNPDYKGTWIHPEIDNPEYSPDPSIYAYDNFG VLGLDLWQVKSGTIFDNFLITNDEAYAEEFGNETWGVTKAAEKQMKDKQDE EQRTRRMMRTKMRMRRMRRTRRKMRRKMSPARPRTSCREACLQGWTEA
TpoR D1-D4	QDVSLLASDSEPLKCFSRTFEDLTCFWDEEEAAPSGTYQLLYAYPREKPRA CPLSSQSMPHFGTRYVCQFPDQEEVRLFFPLHLWVKNVFLNQTRTQRVLFV DSVGLPAPPSIIKAMGGSQPGELQISWEEPAPEISDFLRYELRYGPRDPKNS TGPTVIQLIATETCCPALQRPHSASALDQSPCAQPTMPWQDGPKQTSPSRE ASALTAEGGSCLISGLQPGNSYWLQLRSEPDGISLGGSWGSWSLPVTVDLP GDAVALGLQCFTLDLKNVTCQWQQQDHASSQGFFYHSRARCCPRDRYPIW ENCEEEEKTNPGLQTPQFSRCHFKSRNDSIIHILVEVTTAPGTVHSYLGSPF WIHQAVRLPTPNLHWREISSGHLELEWQHPSSWAAQETCYQLRYTGEGHQ DWKVLEPPLGARGGTLELRPRSRYRLQLRARLNGPTYQGPWSSWSDPTRV ETATETAWHHHHHH
TpoR D1D2	QDVSLLASDSEPLKCFSRTFEDLTCFWDEEEAAPSGTYQLLYAYPREKPRA CPLSSQSMPHFGTRYVCQFPDQEEVRLFFPLHLWVKNVFLNQTRTQRVLFV DSVGLPAPPSIIKAMGGSQPGELQISWEEPAPEISDFLRYELRYGPRDPKNS TGPTVIQLIATETCCPALQRPHSASALDQSPCAQPTMPWQDGPKQTSPSRE ASALTAEGGSCLISGLQPGNSYWLQLRSEPDGISLGGSWGSWSLPVTVDLP GDAVALGLQHHHHHH

Production of CALR wild-type and mutants

The corresponding DNA sequences of the CALR recombinant proteins were cloned into a derivative of plasmid pET9 for transformation and expression in the *E. Coli* bacterial strain Rosetta2(DE3). For each production, 2 L of E. Coli culture was centrifuged at 4°C and put on ice. The pellet was lysed in 40 mL buffer A using French press. The lysate was centrifuged at 4.500 G for 20 min. at 4°C and the supernatant was further filtered on 0.45 µM. The clarified lysate was frozen at -80°C until use.

The soluble recombinant proteins were captured from total lysate using Histidine-affinity with a Ni2+ purification resin (HiTRAP IMAC 5mL, Roche) equilibrated in Buffer B. The column was washed with 20 columns volume (100mL) of buffer C. Recombinant proteins were then eluted with by increasing the imidazole concentration (50-200mM). Eluted fractions were concentrated by ultrafiltration and further purified by size-exclusion chromatography using a Superdex 200 Increase 10/300 column (GE healthcare, Chalfont St. Giles, United Kingdom). The proteins were kept in buffer D at -80°C. The size-exclusion chromatographs are provided in **Supplementary Figure 2**. The purity was further confirmed by SDS-PAGE and by measuring the computing the first derivative of the 330/350 nm ratio upon thermal unfolding (**Supplementary Figures 1 and 2**). The correct folding of the proteins was validated by thermal unfolding experiments with Tycho NT.6 (**Supplementary Figure 1**).

Buffer	**CALR WT** **CALR del52 A394***	**CALR del52** **CALR del52 Y109F/D135L**	**CALR ΔC-tail**
A	Tween 50mM, NaCl 500mM, CaCl2 5 mM, imidazole 10 mM, CHAPS 1mM, pH 7.5	MES 50mM, NaCl 500mM, CHAPS 1%, EDTA 1mM, pH 6.5	Tween 50mM, NaCl 500mM, CaCl2 5 mM, imidazole 10 mM, CHAPS 1mM, pH 7.5
B	Tween 50mM, NaCl 500mM, Imidazole 10 mM, CHAPS 1%, EDTA 1 mM, pH 7.5	MES 50 mM, NaCl 500mM, CHAPS 1%, pH 6.5	Tween 50mM, NaCl 500mM, Imidazole 10 mM, CHAPS 1%, EDTA 1 mM, pH 7.5
C	Tween 50mM, NaCl 500mM, Imidazole 10 mM, CHAPS 1%, EDTA 1 mM, pH 7.5	MES 50 mM, NaCl 500mM, CHAPS 1%, imidazole 25 mM (CALR del52) or 10mM (CALR del52 Y109F/D135L), pH 6.5	Tween 50mM, NaCl 500mM, Imidazole 10 mM, CHAPS 1%, EDTA 1 mM, pH 7.5
D	HEPES 20mM NaCl 150mM CaCl2 1mM	MES 20mM NaCl 150mM CaCl2 1mM	HEPES 20 mM CaCl2 1mM

Fourrier Transformed Infrared Stepctroscopy (FTIR)

FTIR experiments were performed under the supervision of Allison Derenne, Ph.D., founder and general manager of Spectralys Biotech, Belgium. The full report is provided below.

Sample preparation

In order to avoid any interference with buffers or with the high concentration in salt, a buffer exchange was performed. The buffer was replaced by a phosphate buffer 5 mM at pH 7.01 using Micro-Bio Spin P6 Gel from Bio-Rad. 30µL of the five samples was charged on the columns.

To obtain meaningful comparison, five FTIR spectra were recorded for each sample described here above (after buffer exchange). FTIR spectra (raw data, without any pre-processing step) are provided in **Supplementary Figure 4a.**

Result analysis

To interpret the FTIR data in terms of secondary structure, two types of analysis were performed:

- Comparison of the spectra in the region specific to proteins absorption to evaluate the similarity between the wild type protein and the two mutated forms;
- A prediction of the secondary structure content using an in-house database of FTIR spectra of proteins.

The preprocessing steps described in the methods section are required for meaningful comparison of FTIR spectra. Results shown in main **Figure 1e-f** presents the mean of all preprocessed spectra for each sample with a zoom on the spectral region related to protein absorption.

With the naked eye, clear spectral variations can be noted among the five samples. Regarding the region of the Amide II, it must be noted that the mean spectra of the wild type form is quite different. This would mainly arise from the contribution of the Aspartate and Glutamate residues which are much more abundant in the wild type (55 and 54 respectively) than in CALR del52 (43 and 34), in CALRdel52 A394* (43 and 32), in CALR ΔC-tail (43 and 32) and in CALR del52 D135L/Y109F (42 and 34). These two amino acids have a significant contribution around 1568 cm^{-1} due to the absorption of the carbonyl group in their side chains.

In addition to this first observation, it can be noticed that the CALR WT and the CALRdel52 A394* have a similar shape, especially in the amide I band. The CALR ΔC-tail has a similar shape in the amide I band but exposes very different characteristics in the amide II band. The CALRdel52 and the CALRdel52 D135L/Y109F have each, specific and distinct features in both, the amide I and the amide II bands.

Multivariate analyses can be applied to assess similarities and discrepancies among the samples. Principal component analysis (PCA) is a powerful tool for high-dimension data to identify patterns and to express the data in such a way as to highlight their similarities and differences. This tool reduces the number of variables without much loss of information, taking into account most of the variance. It is an unsupervised analysis: the classification obtained does not suppose any a priori condition on grouping obtained. Technical details concerning PCA are provided in the “Methods” section.

Supplementary Figure 4b displays the PCA score plot. Every point (star) in this plot is the projection of one spectrum in the space defined by the first two principal components (PC). The different samples are identified by a unique colour (indicated in the right caption). Each mutant clusters separately, underlying that each mutant has specific spectral features.

The CALR del52 A394* and the CALR del52 separate from the CALR WT along the first principal component (PC1). The CALR del52 A394* is the closest to the CALR WT. The two other mutants separate from the CALR WT along the PC1 and the PC2.

Secondary structure prediction

As described in the methods section, the estimation was realized using three wavenumbers in the Amide I and II bands. The wavenumbers used for this secondary structure determination are the following:

- α-helix: 1545, 1655 and 1613 cm^{-1}
- β-sheet: 1656, 1635 and 1692 cm^{-1}
- Turn: 1678, 1528 and 1600 cm^{-1}
- Random: 1544, 1627 and 1692 cm^{-1}

The table shown in main **Figure 1f** presents the results of this prediction for each sample.

According to the prediction obtained, samples CALR WT and CALR del52 A394* have a similar α-helix, β-sheet, turn and random structures content. Sample CALR del52 has a lower content in α-helix and a slightly higher content in random structures. The α-helix content further decreases for sample CALRdel52 D135L/Y109F and becomes null for CALR ΔC-tail. In addition, the CALR del52 D135L/Y109F has a higher content in turn and the CALR ΔC-tail has a higher turn and random contents. A slight increase of the β-sheet content is also observed for the CALR ΔC-tail.

For the present predictions, the standard error of prediction in cross-validation obtained using the 50-protein database is 5.7% for the α-helix and 6.7% for the β-sheet, 3.2% for turns and 8% for random. Despite these prediction errors, those results are consistent with the observations made on main **Figure 1e** and **Supplementary Figure 4a.**

Instrument

The IR measurements were performed with a Bruker Tensor 27 FTIR-spectrometer (Bruker Optics GmbH, Ettlingen, Germany) with the software Opus 6.5 (Bruker Optics GmbH, Ettlingen, Germany). The FTIR-spectrometer was equipped with a Mercury-Cadmium-Telluride detector, which was cooled down with liquid nitrogen. The spectra were recorded with the ATR mode by using a Golden GateTM ATR accessory (Specac, Orpington, United Kingdom) with an integrated total reflection element composed of a single reflection diamond. The angle of incidence was 45 degrees.

Methods

FTIR measurement

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 (~22°C). For each sample, at least four spectra were recorded.

Multivariate data analysis

Each wavelength in an IR spectrum is considered as a variable. There are therefore a few thousand wavenumbers at which biological molecules absorb, and several spectra are recorded per sample. Thus, the measured data contains a high number of variables, which are additionally often correlated with each other. Consequently, the analysis and interpretation of this large amount of information is complicated. To extract useful information from the measured data, multivariate data analysis can be applied. Multivariate data analysis can be carried out based on unsupervised or supervised learning procedures. For the unsupervised procedures no a priori knowledge about the training set samples is required.

Principal component analysis (PCA) is an unsupervised multivariate data analysis. This technique permits reducing the dimensionality of the data consisting of many correlated variables. At the same time, most of the variation present in the data set is retained [3]. Therefore the variables are transformed in a new uncorrelated set of variables, the principal components (PC) [3]. The PC are linear combinations of the initial variables representing a maximum of variation present in the data [3] . This transformation simplifies the interpretation and visualisation of the data [4].

Mathematically, each IR spectrum can be represented as a linear combination of p wavenumbers. For n spectra the matrix A [nxp] (samples x variables) can be expressed whereat one spectrum corresponds to one row and each column characterises one wavenumber. Thus, each element of the matrix represents the absorbance of a spectrum at one specific wavenumber. The principal components are determined by calculating the eigenvalues and their corresponding eigenvectors of the covariance matrix A [pxp']. In the next step, the eigenvectors are ordered by eigenvalue, from the highest one to the lowest one, to obtain the principal components in order of their significance. The eigenvector with the largest eigenvalue is defined as the first principal component (PC1) as it represents the direction of the greatest variance. The eigenvector with the second largest eigenvalue (PC2) represents the next highest variance in the orthogonal direction and so forth. Finally, the coordinate system of the original spectral data will be transformed to express the spectral data in

terms of the principal components as new axes. This transformation is demonstrated in **Supplementary Figure 4c**. In fact, the plot of the original data in terms of the axes x and y will be rotated and the principal components form the new axes.

In fact, there are as many principal components as variables in the data. However, the first few principal components represent generally over 99% of the present variance in the data. Thus, PCA permits reducing the dimensionality of the spectral data while retaining the majority of the information. This is simply done by projecting the spectra in the principal components space.

The representation of the composition of all spectra in terms of the PC is called score plot. Each point or star in a score plot represents a spectrum. Thus, a score plot permits visualising similarities and difference between spectra and to determine if the spectra are related with each other by forming groups [5].

Secondary structure prediction

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 [6,7].

Statistical analysis

Supplementary Figure 5 presents the mean preprocessed spectra for each sample, with a zoom on the spectral region related to protein absorption. A statistical pairwise comparison with the wild type was also performed to evidence spectral changes. The difference spectrum (black spectrum) corresponds to the difference between the mean spectra of each sample. The black stars on the difference spectrum refer to significant differences defined by a Student's t-test at each wavenumber. Details on the statistical techniques are provided here below.

<u>Student's t-test</u>

In order to evidence spectral changes between samples, the mean spectrum of one was subtracted from the mean spectrum of another. We thus obtained a "difference

spectrum". All difference spectra were calculated with fully preprocessed spectra (baseline corrected and normalized).

The Student's t-test is a parametrical hypothesis test. It is used to determine whether two populations are significantly different from each other or not by comparing the means of the measurements derived from these two populations. The test is applicable if the measurements follow a normal distribution, and the variance of each population is the same.

Two hypotheses are tested:

- H_0: $\lambda_1 = \lambda_2$ which means that there is no difference between the means of the two populations
- H_1: $\lambda_1 \neq \lambda_2$ which means that the means of the two populations are significantly different

where λ_i is the signal intensity for a given wavelength.

The t-test statistic is calculated as follows:

$$t = \frac{(\bar{x}_1) - (\bar{x}_2)}{\sqrt{\frac{S_P^2}{n_1} + \frac{S_P^2}{n_2}}}$$

where $\bar{x}_i$ is the mean of the sample i, S_P^2 is the estimated common variance for the two samples and n_i the number of the sample in the population i.

The variance S_P^2 is calculated with the following formula:

$$S_P^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$$

where s_i is the standard deviation of the sample i and n_i the number of the samples in the population i.

The test was carried out with a significance level of α = 0.1% ($p<0.001$). This threshold is defined as the probability of rejecting the null hypothesis under the assumption that it is true.

Student's t-tests were computed at every wavenumber and allowed a statistical comparison between the spectra of the two samples. Wavenumbers where a significant difference occurs (with a significance α = 0.1%) are indicated by black stars.

Nano-Bioluminescence Energy Transfer (BRET)

Nano-Bioluminescence Energy Transfer (BRET) is a technique that measures proximity between two proteins in living cells. When the two partners are in close proximity (< 10 nm), bioluminescence energy transfer (BRET) occurs between a donor (NanoLuciferase) and an acceptor (HaloTag ligand. This technique has been used to measure protein-protein interaction in living cells between a wide variety of proteins thanks to its ease of use, reproducibility and specificity [8].

Construct

The Thrombopoietin Receptor (TpoR) was cloned into pNL-N vector (Promega) to generate the N-terminally fused NanoLuc-TpoR construct as described. Extracellular forms of the receptor (TpoR D1-D4, TpoR D1D2 and TpoR D1) were obtained from this initial construct by introducing a stop codon by site-directed mutagenesis.

The CALR del52-HaloTag construct was generated by cloning the cDNA from CALR del52 into the pHT-C vector (Promega) to generation the CALR del52-HaloTag fusion protein[9]. The CALR del52 P-C and C-domain-HaloTag were obtained by truncating the N domain or N and P-domain, respectively using the Q5® Site-Directed Mutagenesis method with KLD (NEB). All constructions were sequenced by Macrogen.

Methods

HEK293T were co-transfected with the NanoLuc and HaloTag fusion proteins with a 1:1 cDNA ratio using TransLT-1 in white opaque 96-wells plate. The NanoBRET 618 ligand was added 8h post-transfection according to the manufacturer instruction (Promega). 24h post-transfection, the medium was removed and 100 µL of luminescence substrate (DMEM/F12 without FBS, without phenol red with 2 µL of NanoGlo reagent) was added to each well. The signal was acquired directly using a GloMax® MicroPlate reader at 618 nm (acceptor) and 460 nm (donor) at 37°C. The NanoBRET ratio was computed as:

$$\frac{\text{Acceptor signal}}{\text{Donor Signal}} = \text{NanoBRET ratio}$$

The specificity of the interaction was further validated by performing a Donor Saturation Assay (DSA) (**Supplementary Figure 7**) following the manufacturer instruction (Promega).

Molecular Modelling, Docking, Molecular Dynamics and Free energy simulations

Generation of TpoR D1D2 and CALR mutant C-terminus

Sequences of TpoR extracellular regions which consists of two Cytokine Receptor Modules (CRM1 and 2) comprising each two Fibronectin III like (FN-III) domains and of CALR Del52 were profiled for secondary structure, intrinsic disorder and accessibility propensities with state-of-the-art predictors [10-20]. Models were raised manually for CRM1 using Modeller 9.21 [17], and automatically for the TpoR-CRM2, TpoR-linker, TpoR-Transmembrane domain and CALR using Alpha Fold 2.0 [1] and Rosetta Folding [1].

When compared to its closest template (EpoR, RCSB code: 1CN4) - TpoR displays a long, 65 amino acids insertion in its second FN-III domain of CRM1 which bears local propensity for two ß-strands. Hence a second template, the 2E8 antibody of LDL receptor (RCSB code: 12E8) from the same CATH 2.60.40.10 superfamily, has had to be used to model the insertion and bring the TpoR-CRM1 architecture from a 3/4 β-sandwich (EpoR) to a 4/5 β-sandwich TpoR model.

In order to computationally investigate the interaction between CRM1 and CALR-del52, a model of the mutant C-terminus of CALR del52 was also generated using tLEaP and the FF14SB [21] protein forcefield. Given the secondary structure propensity of CALR del52 C-terminus, this region was modelled first as a straight, ~72Å long, 12 turn α-helix. This structure was then subjected to a 1 μs MD simulation in order to gather the conformational pool to be used in the selection of several starting CALR del52 mutant C-terminus structures for CALR del52 C-terminus-TpoR D1D2 docking. The simulation was performed with OpenMM v.7.4.1 [22] in explicit solvent, using a Monte Carlo barostat to maintain constant pressure, at a temperature of 300K. The conformation clustering along the trajectory and structure extraction was performed with TASKit, an in-house developed Python module, based on the All-vs-All RMSD matrix for clustering and on MDTraj [23] and SciPy [24] for structure extraction. Using a maximum inter-cluster distance cutoff of 8Å, two major configurational basins were identified, from which structural selections were performed. The structures in these

two basins differ by the bending angle of a break in the α-helix shaping up at the 'PARP' motif of CALR del52 (**Supplementary Figure 10a**).

Docking trials between TpoR D1D2 and CALR del52 mutant C-terminus

Taking into account the location of negatively charged acidic amino acids onto the surface of TpoR, three main start configurations (*poses*) of the complex were chosen for assessing the complex formation: one involving a potential interaction of CALR del52 mutant C-terminus with the large extended acidic area found on the 'dorsal' region of TpoR, a second similar pose - but using the more bent CALR del52 mutant C-terminus configuration, and a third involving the second, minor acidic cluster found in the N-ter region of the TpoR sequence. Several structures from the two configurational basins of CALR-del52 were manually set in the above-mentioned poses of the complex as inputs for HADDOCK 2. 4 [8] for complex optimization searches (**Supplementary Figure 10b**). The positively charged, basic residues on CALR and negatively charged, acidic residues on TpoR were selected as “interacting residues” in the HADDOCK run and no other explicit restraints were used. The generated poses were then filtered by their HADDOCK score and the top three poses were chosen for further analysis.

Molecular dynamics simulations of TpoR D1D2-CALR del52 mutant C-terminus complex and subsequent analysis

Starting from the three poses generated by HADDOCK, we sampled the conformational space of the complex via unconstrained molecular dynamics simulations. Each of the three poses were subjected to triplicate 500ns molecular dynamics simulations in which the last residues of the CALR del52 mutant C-terminus become unfolded, in explicit TIP3P water boxes with a 12Å buffer, using a Monte Carlo barostat, at 300K, a Langevin integrator with $1ps^{-1}$ friction coefficient and a 2fs timestep. OpenMM [22] was used to run all simulations, with the same ff14SB [21] forcefield. PDB files corresponding to the first and last frame of triplicates MD simulations for each pose have been deposited to FigShare (https://figshare.com/s/9033970b5a1d3f8d6fa7).

Trajectory analysis

A strong bending motion was noticed in the TpoR molecule of Pose 3 (lowest binding free energy) between the two FN-III-like domains (**Supplementary Figure 10c-d**). This was not observed in the other poses (**Supplementary Figure 10e-f**) which could

indicate that the presence of the CALR del52 C-terminus has a stabilizing effect on the TpoR. This might also imply that the domain is more flexible when unbound, but the inter domain joint stiffens when bound.

Free energy estimations

Free energy was estimated by both a knowledge based method, using PRODIGY server[25], and a physical MD estimation approach based on 3 simulations for each pose, using the MM-GBSA method [26] at 150mM salt concentration, implemented in AMBER20 [27]. Results using both methods were in line with HADDOCK scores.

Free energy estimation using the PRODIGY: the best free energy is shown by the dark-blue complex (**Supplementary Figure 10b**) with ΔG= –12.6 kcal/mol, the sky-blue complex displays an intermediate estimated affinity with ΔG= –11.2 kcal/mol, while the red complex shows the lowest binding capacity with ΔG= –9.6 kcal/mol.

Free energy estimation using MM-GBSA: The first half (250ns) of simulation time was removed as equilibration and 100 frames were uniformly drawn at every 2.5ns, from each repeat, and retained for MM-GBSA calculation. MM-GBSA calculations were performed using the MM-PBSA module included with AMBER20 [28]. A total of 300 frames were used for each pose. The OBC GB model [29] and a salt concentration of 150 mM was used. The LCPO algorithm [30] was used to calculate the surface area of the solute.

The three poses' binding energy estimates using MM-GBSA show exactly the same trend as initial PRODIGY estimates and provide insight into the relative stabilities of the three poses. The values presented in **Supplementary Figure 12** show that the stretched CALR del52 C-terminus conformation (Pose 1) *binds the strongest* to the membrane-distal TpoR domain and that Pose 3 binds the weakest. This could explain the higher flexibility of the TpoR domain when bound in the Pose 3 configuration.

Conformational discretization:

Microstates were delimited using Time-Lagged Independent Component Analysis (TICA). The backbone dihedral angles of the CALR del52 mutant C-terminus molecule were used as input coordinates for TICA. TICA and free energy surfaces were computed using the PyEMMA (2.5.11) python package[31], and the resulting plots were generated using the Matplotlib (3.5.1) python package [32].

The inflection core state (InfleCS) clustering method [33] was used to cluster the two transformed coordinates with the highest eigenvalues and the associated cluster centers were plotted on the corresponding free energy surface. Clustering was performed using 10 components, and re-estimation of the same model was done 5 times. Bayesian information criterion was used for identifying the model.

Trajectory clustering using TICA

The independent components generated by TICA were sorted by their eigenvalues and the top two most autocorrelated components were plotted (**Supplementary Figure 13b**). The subsequent InfleCS clustering (**Supplementary Figure 13c**, **Supplementary Table 1**) shows the formation of discrete conformational clusters, showing that a range of microstates were sampled. Free energy surfaces were generated and are presented in **Supplementary Figure 13a**, along with the centers of the top 3 most probable clusters for pose 1, for which trajectory frames are also presented (**Supplementary Figure 13d**).

Modelling of the TpoR-CALR del52 and ins5 tetramers

Initial models from AlphaFold 2.0[1] and RosettaDock[34] were used to build the tetrameric 3D TpoR-CALR mutant models and to identify the interaction interface between the two CALR del52 and CALR ins5 mutants. H-D exchange data was used to identify contacts between TpoR and CALR del52 in the formation of the tetramer complex. The ER specific G1M9 glycans of TpoR in contact to CALR mutant were modelled with Glycopack [35] and the Glycam server in the configuration consistent with NMR data [36] while the rest are of complex type, built in agreement with SAGS Database [37,38].

The contacts identified by HDx-MS and the crosslinking data on the TM region configuration of TpoR dimer were used as constraints in generating the overall 2CALR del52-2TpoR model. The glycoproteic tetramer was then gently optimized in a five stage process in implicit solvent: (1) first, this was heated to 300K over 1 ns with harmonic cartesian constraints of K=1 on all backbone atoms predicted to be found in secondary structures and K=0.5 on backbone atoms in predicted coil regions; then (2) the system was subjected to equilibration for 2 ns with harmonic cartesian constraints of K=0.5 on secondary structures and no constraints on predicted coil region; this was followed by (3) a further equilibration of 20ns with distance based harmonic constraints of K=1 on interdomain contact points and hydrogen bonds in predicted secondary

structure regions; then (4) the system was cooled over 2 ns with distance based constraints in place and finally (5) extensively minimized without constraints.

This glycoproteic tetramer was then immersed into a full-atom representation of the environment - consisting of a lipid bilayer of 1162 POPC molecules accommodating the TM region of TpoR and in 263023 TIP3P water molecules, 726 chloride and 789 sodium ions describing the solvent region hydrating the rest of the tetramer using the CHARMM-GUI server [39]. This overall system consisting of ~ 1 million atoms was subjected to further unconstrained extensive minimization to obtain the final model for MD simulation. A similar procedure was used for preparing the system containing the TpoR-CALR ins5 tetramer.

The explicit solvent MD simulations were performed with NAMD v.2.13[36] and the CHARMM36 [37-39] forcefield. at constant pressure (1 atm) in two steps: (1) heating using a 1fs time step in order to ensure an even energy distribution, followed by (2) constant temperature simulation using a 2 fs timestep. All MD simulations used a Langevin integrator, with a coupling coefficient of 1 ps^{-1}.

Production MD simulations of the heterotetrameric TpoR-CALR del52 and ins5 systems were carried out for 100 ns in triplicate. Root-mean-square deviation (RMSD) analysis of these simulation was performed using MDAnalysis [41,42] and the corresponding plots are presented in Supplementary Figure 8 (CALR del52) and Supplementary Figure 9 (CALR Ins5). For the analysis of inter-residues contacts, the first half (50 ns) of each simulation was discarded as equilibration and only the second half (which shows a plateau in the RMSD plot) was used for computing the distances.

Tetramer trajectory analysis

Inter-domain and glycan-CALR contacts were computed using MDTraj[19] and were averaged over three 100ns runs. A threshold of 8 Å was used as contact cutoff between heavy atoms and the data was collected every 0.5 ns starting from 50 ns - to account for a model equilibration period – resulting in 100 frames per run. Shown are only contacts present, on average, in more than 60% of frames.

Thermal unfolding and stability

Thermal unfolding experiments were conducted to measure the T_m of all proteins used in HDx-MS and FTIR experiments and to verify the quality and purity of protein preparations. This technique uses the intrinsic fluorescence of Trp and Tyr residues to verify quality, purity and structural integrity of a protein. The intrinsic fluorescence of Trp and Tyr residues (detected at 350 and 330 nm) are measured over a range of increasing temperature. As temperature ramps up, changes in fluorescence signal inform on the protein folding status. The 350/350 nm ratio provides information about protein integrity and T_m. The first derivative of the 350/330 ratio informs on the purity of the protein preparation. A pure sample will have a single peak, indicating that a single protein specie is present in the sample while samples that are not pure will have multiple peaks. All measurements were performed with Tycho NT.6 (NanoTemper).

Supplementary References

1. Jumper, J. et al. Highly accurate protein structure prediction with AlphaFold. *Nature* (2021).
2. Lau, A.M., Claesen, J., Hansen, K. & Politis, A. Deuteros 2.0: peptide-level significance testing of data from hydrogen deuterium exchange mass spectrometry. *Bioinformatics* **37**, 270-272 (2021).
3. Jolliffe, I.T. *Principal component analysis for special types of data*, (Springer, 2002).
4. Ringnér, M. What is principal component analysis? *Nature Biotechnology* **26**, 303-304 (2008).
5. Boydston-White, S. et al. Cell-cycle-dependent variations in FTIR micro-spectra of single proliferating HeLa cells: principal component and artificial neural network analysis. *Biochim Biophys Acta* **1758**, 908-14 (2006).
6. Goormaghtigh, E., Ruysschaert, J.M. & Raussens, V. Evaluation of the information content in infrared spectra for protein secondary structure determination. *Biophys J* **90**, 2946-57 (2006).
7. Oberg, K.A., Ruysschaert, J.M. & Goormaghtigh, E. Rationally selected basis proteins: a new approach to selecting proteins for spectroscopic secondary structure analysis. *Protein Sci* **12**, 2015-31 (2003).
8. Dale, N.C., Johnstone, E.K.M., White, C.W. & Pfleger, K.D.G. NanoBRET: The Bright Future of Proximity-Based Assays. *Frontiers in Bioengineering and Biotechnology* **7**(2019).
9. Pecquet, C. et al. Calreticulin mutants as oncogenic rogue chaperones for TpoR and traffic-defective pathogenic TpoR mutants. *Blood* **133**, 2669-2681 (2019).
10. Barik, A. et al. DEPICTER: Intrinsic Disorder and Disorder Function Prediction Server. *J Mol Biol* **432**, 3379-3387 (2020).
11. Dominguez, C., Boelens, R. & Bonvin, A.M. HADDOCK: a protein-protein docking approach based on biochemical or biophysical information. *J Am Chem Soc* **125**, 1731-7 (2003).
12. Ishida, T. & Kinoshita, K. PrDOS: prediction of disordered protein regions from amino acid sequence. *Nucleic Acids Res* **35**, W460-4 (2007).
13. Kelley, L.A., Mezulis, S., Yates, C.M., Wass, M.N. & Sternberg, M.J. The Phyre2 web portal for protein modeling, prediction and analysis. *Nat Protoc* **10**, 845-58 (2015).
14. Petersen, B., Petersen, T.N., Andersen, P., Nielsen, M. & Lundegaard, C. A generic method for assignment of reliability scores applied to solvent accessibility predictions. *BMC Structural Biology* **9**, 51 (2009).
15. Romero, P. et al. Sequence complexity of disordered protein. *Proteins* **42**, 38-48 (2001).
16. Rost, B. & Sander, C. Prediction of protein secondary structure at better than 70% accuracy. *J Mol Biol* **232**, 584-99 (1993).
17. Webb, B. & Sali, A. Comparative Protein Structure Modeling Using MODELLER. *Curr Protoc Bioinformatics* **54**, 5.6.1-5.6.37 (2016).
18. Drozdetskiy, A., Cole, C., Procter, J. & Barton, G.J. JPred4: a protein secondary structure prediction server. *Nucleic Acids Research* **43**, W389-W394 (2015).
19. Jones, D.T. Protein secondary structure prediction based on position-specific scoring matrices. *J Mol Biol* **292**, 195-202 (1999).
20. Yachdav, G. et al. PredictProtein—an open resource for online prediction of protein structural and functional features. *Nucleic Acids Research* **42**, W337-W343 (2014).

21. Maier, J.A. et al. ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. *J Chem Theory Comput* **11**, 3696-713 (2015).
22. Eastman, P. et al. OpenMM 7: Rapid development of high performance algorithms for molecular dynamics. *PLoS Comput Biol* **13**, e1005659 (2017).
23. McGibbon, R.T. et al. MDTraj: A Modern Open Library for the Analysis of Molecular Dynamics Trajectories. *Biophys J* **109**, 1528-32 (2015).
24. Virtanen, P. et al. SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nature Methods* **17**, 261-272 (2020).
25. Xue, L.C., Rodrigues, J.P., Kastritis, P.L., Bonvin, A.M. & Vangone, A. PRODIGY: a web server for predicting the binding affinity of protein-protein complexes. *Bioinformatics* **32**, 3676-3678 (2016).
26. Kollman, P.A. et al. Calculating structures and free energies of complex molecules: combining molecular mechanics and continuum models. *Acc Chem Res* **33**, 889-97 (2000).
27. D.A. Case, H.M.A., K. Belfon, I.Y. Ben-Shalom, J.T. Berryman, S.R. Brozell, D.S. Cerutti, T.E. Cheatham, III, G.A. Cisneros, V.W.D. Cruzeiro, T.A. Darden, R.E. Duke, G. Giambasu, M.K. Gilson, H. Gohlke, A.W. Goetz, R. Harris, S. Izadi, S.A. Izmailov, K. Kasavajhala, M.C. Kaymak, E. King, A. Kovalenko, T. Kurtzman, T.S. Lee, S. LeGrand, P. Li, C. Lin, J. Liu, T. Luchko, R. Luo, M. Machado, V. Man, M. Manathunga, K.M. Merz, Y. Miao, O. Mikhailovskii, G. Monard, H. Nguyen, K.A. O'Hearn, A. Onufriev, F. Pan, S. Pantano, R. Qi, A. Rahnamoun, D.R. Roe, A. Roitberg, C. Sagui, S. Schott-Verdugo, A. Shajan, J. Shen, C.L. Simmerling, N.R. Skrynnikov, J. Smith, J. Swails, R.C. Walker, J. Wang, J. Wang, H. Wei, R.M. Wolf, X. Wu, Y. Xiong, Y. Xue, D.M. York, S. Zhao, and P.A. Kollman AMBER 2020. (University of California, San Francisco, 2020).
28. Miller, B.R. et al. MMPBSA.py: An Efficient Program for End-State Free Energy Calculations. *Journal of Chemical Theory and Computation* **8**, 3314-3321 (2012).
29. Onufriev, A., Bashford, D. & Case, D.A. Exploring protein native states and large-scale conformational changes with a modified generalized born model. *Proteins: Structure, Function, and Bioinformatics* **55**, 383-394 (2004).
30. Weiser, J., Shenkin, P.S. & Still, W.C. Approximate atomic surfaces from linear combinations of pairwise overlaps (LCPO). *Journal of Computational Chemistry* **20**, 217-230 (1999).
31. Scherer, M.K. et al. PyEMMA 2: A Software Package for Estimation, Validation, and Analysis of Markov Models. *Journal of Chemical Theory and Computation* **11**, 5525-5542 (2015).
32. Hunter, J.D. Matplotlib: A 2D Graphics Environment. *Computing in Science & Engineering* **9**, 90-95 (2007).
33. Westerlund, A.M. & Delemotte, L. InfleCS: Clustering Free Energy Landscapes with Gaussian Mixtures. *Journal of Chemical Theory and Computation* **15**, 6752-6759 (2019).
34. Lyskov, S. & Gray, J.J. The RosettaDock server for local protein-protein docking. *Nucleic Acids Res* **36**, W233-8 (2008).
35. Paduraru, C. et al. An N-linked glycan modulates the interaction between the CD1d heavy chain and beta 2-microglobulin. *J Biol Chem* **281**, 40369-78 (2006).
36. Petrescu, A.J. et al. The solution NMR structure of glucosylated N-glycans involved in the early stages of glycoprotein biosynthesis and folding. *Embo j* **16**, 4302-10 (1997).

37. Petrescu, A.J., Petrescu, S.M., Dwek, R.A. & Wormald, M.R. A statistical analysis of N- and O-glycan linkage conformations from crystallographic data. *Glycobiology* **9**, 343-52 (1999).
38. Petrescu, A.J., Wormald, M.R. & Dwek, R.A. Structural aspects of glycomes with a focus on N-glycosylation and glycoprotein folding. *Curr Opin Struct Biol* **16**, 600-7 (2006).
39. Jo, S., Kim, T., Iyer, V.G. & Im, W. CHARMM-GUI: A web-based graphical user interface for CHARMM. *Journal of Computational Chemistry* **29**, 1859-1865 (2008).