Article

Structural and Dynamic Differences between Calreticulin Mutants Associated with Essential Thrombocythemia

Ragousandirane Radjasandirane and Alexandre G. de Brevern *

Université Paris Cité and Université de la Réunion and Université des Antilles, INSERM, BIGR, DSIMB Bioinformatics Team, F-75014 Paris, France
* Correspondence: alexandre.debrevern@univ-paris-diderot.fr; Tel.: +33-1-44493000

Abstract: Essential thrombocythemia (ET) is a blood cancer. ET is characterized by an overproduction of platelets that can lead to thrombosis formation. Platelet overproduction occurs in megakaryocytes through a signaling pathway that could involve JAK2, MPL, or CALR proteins. CALR mutations are associated with 25–30% of ET patients; CALR variants must be dimerized to induce ET. We classified these variants into five classes named A to E; classes A and B are the most frequent classes in patients with ET. The dynamic properties of these five classes using structural models of CALR's C-domain were analyzed using molecular dynamics simulations. Classes A, B, and C are associated with frameshifts in the C-domain. Their dimers can be stable only if a disulfide bond is formed; otherwise, the two monomers repulse each other. Classes D and E cannot be stable as dimers due to the absence of disulfide bonds. Class E and wild-type CALR have similar dynamic properties. These results suggest that the disulfide bond newly formed in classes A, B, and C may be essential for the pathogenicity of these variants. They also underline that class E cannot be directly related to ET but corresponds to human polymorphisms.

1. Introduction

Essential thrombocythemia (ET) is one of the principal chronic myeloproliferative neoplasms (MPNs). It progresses when megakaryocytes in the bone marrow produce an excess number of platelets [1,2]. ET has an incidence of 1.2 to 3.0 per 100,000 population per year [3]. ET has chronic courses complicated by thrombohemorrhagic events, systemic symptoms, splenomegaly, and progression to myelofibrosis and/or blast phase [4–9]. Various drugs are prescribed to normalize the platelet count in patients at high risk of ET. However, the superiority of either interferon-α or hydroxyurea has not yet been established [10,11].

ET-specific driver mutations were recently discovered. The first localized protein is a famous tyrosine kinase enzyme, the Janus kinase 2 (JAK2) protein, and, more specifically, it is JAK2V617F [12]. It is also the major player in another MPN, polycythemia vera [13], characterized by excessive red blood cell levels. JAK2 is involved in the signal transduction of erythropoietin and thrombopoietin; JAK2V617F is represented in 60–65% of ET patients [14,15]. The second protein is a transmembrane receptor interacting with cytosolic JAK2, namely, the thrombopoietin receptor protein or MPL [16]. MPL is involved in 4 to 10% of cases [17]; its main variant is MPLW515K, which is located at the border of the membrane [18].

A third protein was characterized in 2013 [19,20], namely, calreticulin (CALR) [21–24]. CALR is an endoplasmic reticulum (ER) Ca^{2+}-binding chaperone that functions in the folding and assembly of glycoproteins [25–30]. CALR binds half of the ER Ca^{2+} [31–33] and has specificity for monoglycosylated N-glycans on substrate proteins, which are transiently

acquired during glycoprotein maturation in the ER. CALR contains three domains: (i) a lectin-like N-terminal domain (N-domain) containing glycan-binding and high-affinity calcium-binding sites [34], (ii) a central elongated hairpin-like P-domain containing a co-chaperone-binding site [27], and (iii) an acidic C-terminal domain (C-domain), which contains multiple low-affinity Ca^{2+}-binding sites [32]. The three domains have been crystallized separately, and a cryo-EM structure has been proposed as a global model with the end of the C-domain missing and the beginning of the C-domain mainly composed of helical residues [35]. CALR mutations (CALRms) associated with ET always affect the ninth exon [36]. This region corresponds to the third and last structural domains of CALR (named the C-domain). CALRms consist of deletions and/or insertions; they often have a one-base-pair shift, leading to significant changes in the C-domain [37,38]. From a physicochemical point of view, CALRms become more basic. The C-domain also loses the ER retention motif KDEL; it does not stay in the ER but is exported outside the megakaryocyte and then interacts with MPL. This binding leads to platelet proliferation [7,39].

CALRm accounts for 25 to 30% of ET patients and ∼20% of MPN patients [40]. The two most common CALRms account for 80–85% of cases; they are c.1092_1143del, p.L367fs*46, designated as the classical type 1, and c.1154_1155insTTGTC, p.K385fs*47, designated as the classical type 2. Different studies have shown a correlation between the CALR mutation type and the prognosis value for the stratification of thrombotic risk in ET patients [41–44]. The complications originally described only for the JAK2V617F mutation are increasingly being described for CALRm patients, such as stroke [45] or splenic mass [46].

To have a better outlook on CALRm, we have proposed a specific database named CALR-ETdb (https://www.dsimb.inserm.fr/CALR-ET, accessed 20 February 2023) [47] that aggregates all CALRms from the COSMIC database [48] and bibliographic searches. Originally composed of 155 CALRms, it has now been updated to 169. CALR-ETdb is searchable and provides information not only in terms of references and types but also of structural properties with secondary structures and three-dimensional structural models.

Thanks to this large number of CALRms, we propose a new classification of CALRm sequences defining five different classes and some surprising underlying properties. Hence, class A aggregates 96% of the former type 1 (and those that are 1-like), while class B comprises only 64% of the former type 2 (and those that are 2-like), i.e., 1/3 of them belong to other classes. Class C is mainly composed of the former other type. The unexpected cases were (i) class D with 4% of type 1, 8% of type 2, and 5.6% of other and which is associated with a particularly short length, and (ii) class E with no type 1, 28% of type 2, and 3.7% of the other type. Both classes have only point mutations that do not induce any frame change, and for the last of these, even the ER retention signal is maintained.

With loss of the retention signal in the ER (KDEL), CALRms cannot remain. However, another essential point is that CALRms must be in dimeric form to bind to MPL at the membrane surface of megakaryocytes [49–51]. It has been shown that the homomultimerization of mutant calreticulin is an obligatory prerequisite for MPL binding and activation. Recently, Raghavan's group investigated the relevance of disulfide-bond-mediated interactions in CALRm multimerization in primary patient platelets and human cell lines expressing recombinant mutant CALRms [52,53]. They showed that new cysteine residues in the C-terminal ends of CALRm mutants contributed to CALRm (type 1) and MPL interactions by forming disulfide bridges, both in primary patient platelets and also in human cell lines expressing recombinant CALRm mutants.

In the present research, we have extended this analysis to underline if the ability of these different classes to dimerize is found for our different CALRm classes. Indeed, different dynamic properties could be related to oncological impacts. For this purpose, molecular dynamics simulations were performed for the reference CALR and five CALRm classes. We tested the monomer and dimer for CALRm and for the wild-type CALR. The results clearly show that (i) the protocol was correct, (ii) the dynamics of classes A, B, and C are slightly different and are stable as dimers, and (iii) that classes D and E cannot be stable

as dimers. These last two seem to not be associated directly with MPL binding. Therefore, they also may not be directly linked to ET and platelet proliferation.

2. Materials and Methods

2.1. Dataset

The CALR sequence was retrieved from UniProt (UniProt ID: P27797) [54]. For clarity, the wild-type CALR was named CALRwt. ET impacts the C-terminus of the CALR protein, i.e., the C-domain corresponding to the third (and last) domain; it is the ninth exon. The C-domain extends from alanine 352 to leucine 417 of CALRwt. The 5 other sequences of CALRms were taken from CALR-ETdb. Each one corresponds to one of the five classes described in [47], with class A being p.L367Tfs*46 (and corresponding to classical CALR type 1), CALR-ETdb variant 20; class B being p.K385Nfs*47 (and corresponding to classical CALR type 2), CALR-ETdb variant 10; class C being p.E380Dfs*51, CALR-ETdb variant 5; class D being p.E389*, CALR-ETdb variant 16; and class E being p.L367I, CALR-ETdb variant 3. Figure 1 summarizes all the different studied systems.

```
               * * * * * * * * * * * * * *                  * * * * * * * * * * * * * * * * * * * * * * * * *
Class A    AAEKQMKDKQDEEQ-------------------RTRRMMRTKMRMRRMRRTRRKMRRKMS 392(41)
Class B    AAEKQMKDKQDEEQRLKEEEEDKKRKEEEEAEDNCRRMMRTKMRMRRMRRTRRKMRRKMS 411(60)
Class C    AAEKQMKDKQDEEQRLKEEEEDKKRKEED-RRQRTRRMMRTKMRMRRMRRTRRKMRRKMS 410(59)
                          I        II        III
               * * * * * * * * * * * * * *
Class D    AAEKQMKDKQDEEQRLKEEEEDKKRKEEEEAEDKEDD--------------------- 388(37)
Class E    AAEKQMKDKQDEEQRIKEEEEDKKRKEEEEAEDKEDDEDKDEDEEDEEDKEEDEEEDVPG 411(60)
CALRwt     AAEKQMKDKQDEEQRLKEEEEDKKRKEEEEAEDKEDDEDKDEDEEDEEDKEEDEEEDVPG 411(60)
               * * * * * * * * * * * * * *  * * * * * * * * * * * * * * * * * * *oooooooooooooooooooooooooo

               * * * * * * * * * * * * * * * * * * *
Class A    PARPRTSCREACLQGWTEA    411(60)
Class B    PARPRTSCREACLQGWTEA    430(79)
Class C    PARPRTSCREACLQGWTEA    429(78)
Class D    -------------------    388(37)
Class E    QAKDEL-------------    417(66)
CALRwt     QAKDEL-------------    417(66)
               oooooo
```

Figure 1. The 6 different CALR C-domains used in the study. The sequences of CALR wild-type (CALRwt) and the 5 CALRms (class A: p.L367Tfs*46, CALR-ETdb variant 20; B: p.K385Nfs*47, variant 10; C: p.E380Dfs*51, variant 5; D: p.E389*, variant 16; and E: p.L367I, variant 3) were aligned with CLUSTAL Omega (and modified manually) and are shown with red stars to indicate the positions always found, with black stars, if only found for one of the two sub-clusters, and with circles, if only found between class E and CALRwt. The residues that differ between these two are in italics and underlined. The positions of the three stretches are also shown (I: found in all classes, II: found only in classes B and C, and III: complete only in class B). CALR C-domain begins with alanine 352; length of each CALR is provided at the end of each line.

2.2. Structural Models

The structural templates to model CALRwt were selected thanks to PSI-BLAST software [55] used with the Protein Data Bank [56]. It makes it possible to detect proteins with sequences close to the query sequence and with structures that can be used for modeling, namely, structural modeling. The mining of the different databases was performed with PSI-BLAST with an e-value greater than 0.005.

CALRwt was based on the calreticulin partial structure (PDB ID 6ENY chain E structure [35] from the Protein Data Bank [56]) and completed (please notice that the C-domain is identical to the one from the PDB, ID 5LK5 [57]). The CALRm structural models were taken from our database CALR-ETdb (https://www.dsimb.inserm.fr/CALR-ET accessed on 17 December 2022) [47]. The modeling of the 3D models was performed with MOD-

ELLER v.10.0 [58]. The selection of the best structural model was mainly based on the renowned DOPE approach [59]. The three-dimensional structure and structural models were visualized using PyMOL (the PyMOL Molecular Graphics System, Version 1.7.2.2 Schrödinger, LLC) [60]. Other modeling approaches were also used, i.e., I-TASSER [61], Robetta [62], AlphaFold2 [63], and Phyre2 [64], but without satisfying results compared with a human-supervised approach using MODELLER (see the Discussion section). Figure 2 shows all the modeled systems (they are all provided in Data S1).
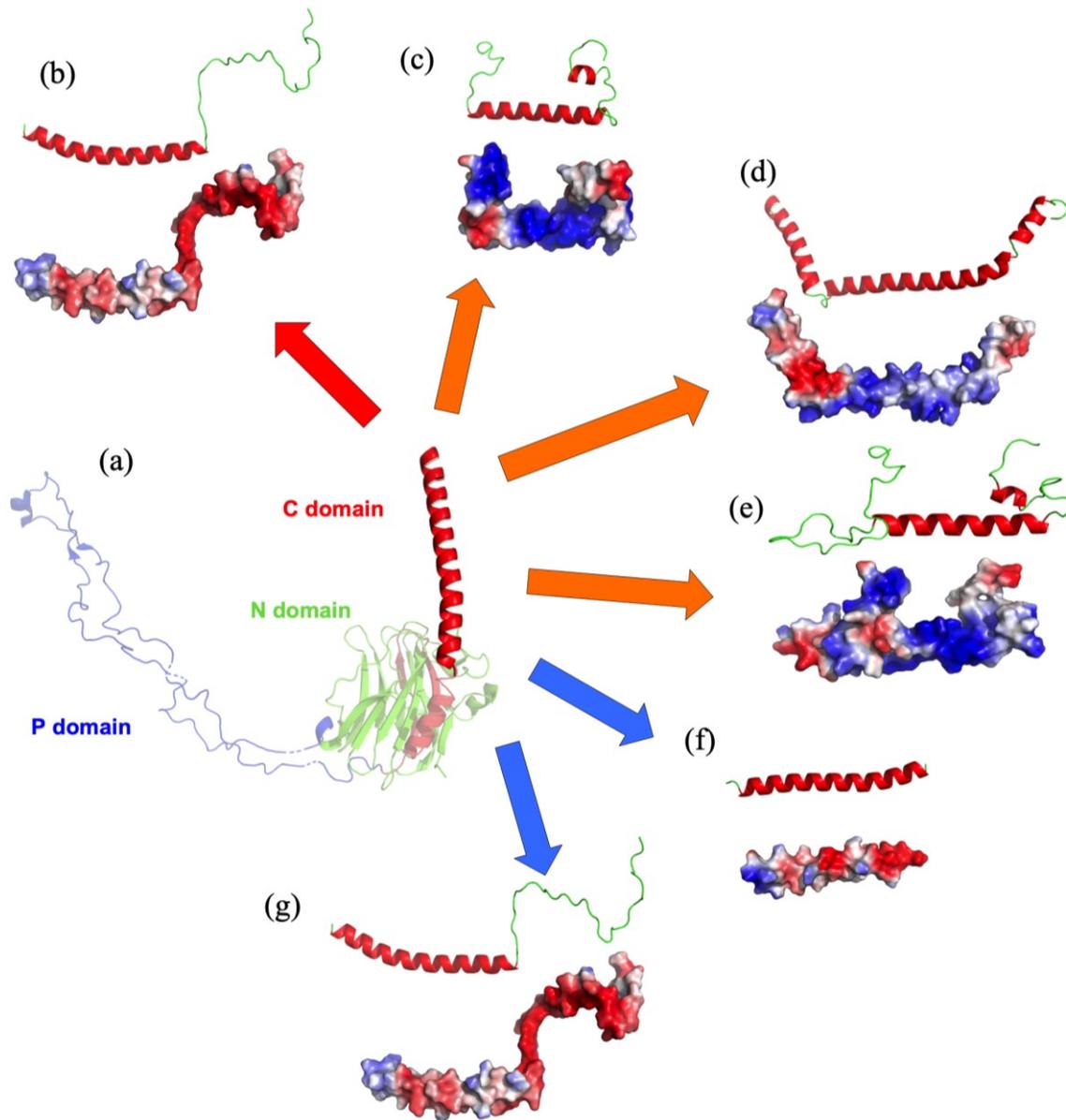
Figure 2. The 6 different structural models of CALR C-domains. (**a**) Calreticulin structure from PDB, ID 6ENY a chain E structure [35] with the three determined domains (N-, P-, and C-domains, the last of which is incomplete); and (**b**–**g**) provide the different modeled C-domains with (up) secondary structure and (down) electrostatics visualization: (**b**) CALRwt, (**c**) class A, (**d**) class B, (**e**) class C, (**f**) class D, and (**g**) class E.

2.3. Dimer Structural Models

Following the same principle, dimers of every system were built. For classes A, B, and C, CALRms with two novel cysteines appeared in the sequences due to the frameshift. They have been shown to have a determinant role in CALR dimerization and MPL receptor

activation through the CALRm dimers [52]. Thus, the dimeric form of CALRms is based on two disulfide bonds between chains using the two cysteines within the CREAC motif at the end of the sequence. MODELLER [58] was used with two monomer chains and a specific constraint between cysteines to create the disulfide bridges. For CALRwt, and classes D and E, their sequences have no cysteines to form disulfide bonds. The dimers were built using distance constraints to keep each chain close enough together to be comparable with other CALRm dimers with disulfide bonds. For all these models, the best DOPE score [59] was used to select the best model (they are all provided in Data S1). The two chains will be named chain 1 and chain 2, respectively, in the text, as they often did not share the same dynamics.

2.4. Molecular Dynamics

Molecular Dynamic (MD) simulations were performed using GROMACS 2020.4 software [65] with the CHARMM-36 forcefield [66]. Before starting any simulation experiments, each structural variant was energy-minimized for 500 steps of the steepest descent and 500 steps of the conjugate gradient optimized with the SHAKE algorithm, using the GROMACS suite. CALRwt and 5 different CALRms, both in monomeric and dimeric forms, were soaked in a rhombic dodecahedral simulation box with TIP3P water molecules. A total of 13 calcium ions (Ca^{2+}) were added as positive counter-ions into the system, in order to neutralize it, for CALRwt and CALRm class E because these two systems are negatively charged. The other variants (classes A, B, and C) had chloride ions (Cl^-) as counter-ions because they are positively charged.

The MD protocol was standardized through our previous works [67,68]. After 1 nanosecond (nsec) of equilibration (with position restraints on the proteins), each system was simulated through multiple classical independent production runs ranging from a total of 100 to 500 nanoseconds, as in [67]. The equilibration consists of one step with an NVT system and three more with an NPT system (with a gradual decrease in position constraints). During the first step in an NVT system and the second step in an NPT system, the protein is totally constrained and cannot move, and the equilibration affects the water molecules. In the third and fourth steps in an NPT system, the constraints on the protein are slowly released. Molecular conformations were saved every 100 picoseconds for downstream analysis. The first 5 nsec of each MD simulation was discarded as the residues at the extremities could not be taken into calculations. Trajectory analyses were carried out with GROMACS software and in-house Python and R scripts. Root-mean-square deviations (RMSDs) and root-mean-square fluctuations (RMSFs) were calculated for Cα atoms only.

2.5. Molecular Dynamics Analysis

The analysis of MD was carried out using classic tools, for example, the RMSD and RMSF, and other more innovative ones such as PBxplore [69], a tool developed within the team. PBxplore makes it possible to analyze protein blocks throughout the MD. MD analysis [70], a Python library, was used to compute distance calculations between Ca^{2+} ions and their binding sites during the simulations.

The RMSD (root-mean-square deviation) makes it possible to see how a structure varies during the dynamics by comparing it with the starting structure (reference). For each frame, an average of the differences between the reference positions and the positions of the current frame is performed in order to obtain an RMSD value per unit of time.

The RMSF (root-mean-square fluctuation) is similar to the RMSD. It makes it possible to determine the fluctuation of each residue following the same principle as for the RMSD, i.e., a comparison with a reference that is the average position of each residue and, thus, measure the difference between the current position and the average position in order to have a flexibility value per position.

Protein Blocks (PBs) are a structural alphabet composed of 16 local prototypes [71]. Each specific PB is characterized by the φ, ψ dihedral angles of five consecutive residues, with each PB assignment focused on the central residue. Obtained through an unsupervised

training approach and performed on a representative non-redundant databank, PBs give a reasonable approximation of all local protein 3D structures [72]. PBs are very efficient in tasks such as protein superimpositions [73] and MD analyses [74]. They are labeled from *a* to *p*: PBs *m* and *d* can be roughly described as prototypes for an α-helix and central β-strand, respectively; PBs *a* to *c* primarily represent β-strand N-caps; PBs *e* and *f* represent β-strand C-caps; and PBs *a* to *j* are specific to coils, PBs *k* and *l* to α-helix N-caps, while PBs *n* to *p* to α-helix C-caps. Each PB assignment was carried out using our PBxplore tool (available at GitHub) [69].

PB assignments were undertaken for each residue of the C-domain and over every snapshot extracted from the MD simulations. The equivalent number of PBs (N_{eq}) is a statistical measurement similar to entropy that represents the average number of PBs for a residue at a given position. N_{eq} is calculated as follows [71]:

$$N_{eq} = exp\left(-\sum_{x}^{16} f_x \, ln \, f_x \right)$$

where f_x is the probability of PB x. An N_{eq} value of 1 indicates that only one type of PB is observed, while a value of 16 is equivalent to a random distribution. To underline the main differences between one system and another one for each position, the absolute difference ΔN_{eq} between corresponding N_{eqs} values was computed.

However, since the same ΔN_{eq} value can be obtained with different types of blocks in similar proportions, we defined a complementary measure, ΔPB, that evaluates a change in the PB profile by calculating the absolute sum of the differences for each PB between the probabilities of a PB x to be present in the first and the second forms (x goes from PB *a* to PB *p*). ΔPB is calculated as follows [67]:

$$\Delta PB = \sum_{x=1}^{16} \left| f_x^{S1} - f_x^{S2} \right| \tag{1}$$

where f^{S1}_x and f^{S2}_x are the percentages of occurrence of a PB x in the first (S1) and the second system (S2), respectively. A value of 0 indicates a perfect PB identity between the 1st and 2nd systems, while a score of 2 indicates a total difference.

It should be noted that all the protocols (modeling, molecular dynamics, and analyses) were similar for all the systems, for both monomers and dimers.

3. Results

3.1. Analysis of CALRwt Monomer Dynamics

CALRwt is a calcium-binding chaperone residing in the endoplasmic reticulum (ER), thanks to the KDEL motif localized at the end of its C-terminus, which functions as an endoplasmic reticulum retention signal [75]. As the CALR C-domain is incomplete in the PDB structure (PDB ID 6ENY chain E structure [35] (see Figure 2a)), it was previously completed with comparative modeling, with MODELLER [58] and PSIPRED [76] predictions used as constraints (see Materials and Methods section and Figure 2b). The electrostatics shows highly negative values in the C-terminus region.

This first part consists of using the wild-type reference form of CALR to assess whether it is possible via MD to reproduce the physiological behavior of the C-domain of CALR, i.e., to capture Ca^{2+} ions. We therefore simulated the dynamics of CALRwt in monomeric form over 500 nsec. Thirteen Ca^{2+} ions were added to neutralize the negatively charged system due to numerous aspartic acid and glutamic acid residues.

The RMSD values reached a plateau around the correct value of 1.25 Å from 20 nsec to the end of the simulation (these same results were found for other replicates, but the data are not shown), i.e., the simulation stabilized quickly and could therefore be used. The MD simulations were analyzed in light of the RMSFs and Protein Blocks (PBs). For the map of the PB highlights regions with high flexibility and rigid regions, see Figure 3a,c.

In particular, the beginning (residues 352 to 360) and end (residues 382 to 386) of the helix show a high variation in PBs with N_{eq} values of four and seven, respectively (see Figure 3c), which indicates that the start of the helix tends to be a deformable region, whereas the end of the helix appears to be disordered (as defined in [77]). The helical region is characterized by PB *m* (see Figure 3a). The second part of the structure is constituted of unstructured regions (residues 387 to 417), but some residues show very high intensities in terms of PB. For instance, residue E405 fixes calcium ions with another residue (D408), which stabilizes its local conformation (see Figure 3a).
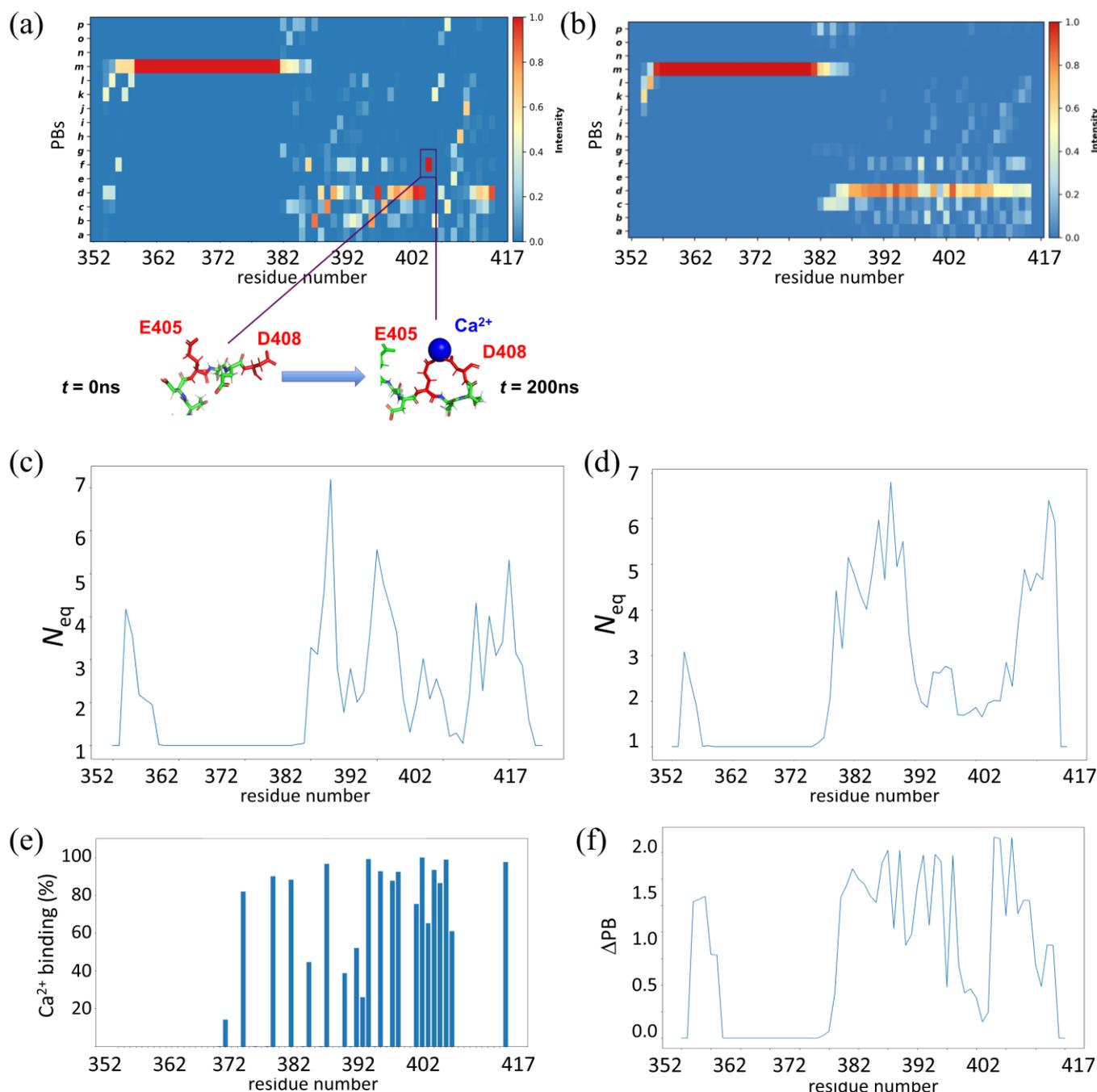
Figure 3. MD of CALRwt. (**a**) PB analysis of CALRwt with Ca^{2+}, with a focus on the Ca^{2+} bound by residues E405 and D408; (**b**) PB analysis of CALRwt with Na^+; Neq of CALRwt (**c**) with Ca^{2+} and (**d**) with Na^+; (**e**) site occupation of Ca^{2+} (for CALRwt with Ca^{2+}); and (**f**) ΔPB between MD with Ca^{2+} and with Na^+.

Initially, CALRwt was without any Ca^{2+} ions fixed onto it. However, after 50 nsec, almost all Ca^{2+} ions were bonded to the CALRwt structure (see Videos S1 and S2). This produces local structural constraints from calcium ions bonded between two residues (see Figure 3e). Hence, the region between residue 392 and residue 405 has a high density of calcium fixation within a radius of 3 Å from each residue. This density is also visible in the PB map with a high intensity of blocks (in red-orange) (see Figure 3a) as well as in the RMSF and N_{eq} plot with a decreasing value in this region (see Figure 3c). The distribution of the binding sites for Ca^{2+} ions does not seem to be deterministic. Indeed, the binding sites between the simulations of 400 nsec and 100 nsec (see Figure S1a,b) are different, which led to different PB maps; it also led to different local conformations (see Figures 3a,c and S2a,b).

To try and check the impact of calcium ions on the structure, we simulated the same system in similar conditions but with 26 Na^+ ions instead of 13 Ca^{2+} ions. Na^+ ions do not bind to the structure of CALRwt, which leaves the CALRwt structure without any structural constraints. As with the simulation with Ca^{2+} ions, the end of the helix was also highly flexible with an N_{eq} value between six and seven (see Figure 3d). The second part of the structure could not be stabilized with Ca^{2+} ion binding; however, the PB profile for the unstructured regions remained around PB *d* (see Figure 3b) and related conformations, with some high N_{eq} values often higher than four. With only Na^+, replicates of the simulation provided similar profiles (see Figures 3b,d and S2c,d and Video S3).

Local structural differences between the two systems (with Ca^{2+} or with Na^+) can be compared using Protein Blocks. Ca^{2+} binding induces local conformations that are not present in the simulation of CALRwt with Na^+ ions. Indeed, the region around residue 407 has a ΔPB superior to one (see Figure 3f), indicating that these simulations have different PB distributions for this region. In fact, a Ca^{2+} ion bound to residue E403 and L417 and provokes a specific fold that is strongly different from the CALRwt Na^+ simulation (see Figure S3). Some regions are still found to be similar between the two systems. The region around residue 401 has a low ΔPB, i.e., this region is less constrained due to calcium ion binding. An additional half-microsecond simulation was conducted with Ca^{2+}, and, in addition to Na^+ and Cl^-, the CALRwt behaviors followed those of the simulations with only Ca^{2+} (see Figure S4) as these ions bind very quickly, as seen previously.

The results of this first part show that the CALRwt models allow a good fixation of Ca^{2+} ions. Therefore, these simulations properly mimic the natural role of CALRwt in binding Ca^{2+} ions. These MD simulations also underline the strong dynamic differences between a calcium medium and a sodium medium. For the first, the interactions strongly stabilize local conformations, while the second makes it possible to leave the disordered tendencies of the C-domain C-terminus. These data suggest that this approach can be applied to different CALRs.

3.2. Frameshifts Induce Distinct Dynamics for CALRm Classes A, B, and C

The CALR variants were classified into five classes from A to E (see Figure 1). Classes A, B, and C are characterized by frameshifts and loss of the endoplasmic reticulum retention signal. They are characterized by a strong change in their charges (see Figure 2c–e), and they are not supposed to be able to bind Ca^{2+} ions anymore, having become very positive. They also carry new cysteines to allow their dimerization [52]. The last two classes do not share these cysteines. The class E sequence is similar to CALRwt, with one or more punctual mutations (while keeping the ER retention signal), and class D is roughly half of CALRwt (without the ER retention signal).

In this second part, we focus on the first three CALRm classes as monomers. They correspond perfectly to what is expected from CALRms associated with ET. The models were built from the available CALRwt structure and completed as before, with secondary structural constraints taken from PSIPRED. The selected models, therefore, have a perfect match with the predicted helical residue and the predicted loops. The simulations were carried out in a similar way as with CALRwt. The only difference was the charges being

different to neutralize them, with Cl- ions having been added. The example of CALRm class A is the classic canonical type I [19,20]. CALRm class A (see Figure 4a–c,j and Video S3) is characterized by a unique main α-helix ranging from residue 366 to 388 (see Figure 4c). The beginning of the fragment is quite rigid with a succession of small turns characterized by PBs *k*, *b*, and *f* and low values of less than 2.5 (see Figure 4b). The helix is highly rigid with an N_{eq} value equal to one, beginning with the N-cap PBs *fkl* (positions 362 to 364) to position 382, the C-cap (positions 383 to 388 are less constrained, with N_{eq} values of two). Positions 395 and 402–404 are the only positions with some local constraints (N_{eq} values of two or less), while the rest of the C-termini region is highly flexible, with some N_{eq} values of higher than four (see Figure 4b). The RMSF values are well correlated with this analysis. The rigid helix is associated with the lowest RMSF values (see Figure 4a), while the C-termini has the highest RMSF values. This behavior is well described in Figure 4j, which shows that since the initial model (n = 0 nsec), the adjacent parts move, especially the C-terminal part (n = 200 and n = 400 nsec). The superposition of a large number of conformations shows the particularly flexible aspect of these regions around the helix of class A well.

Thus, CALRm class A is characterized by a single rigid helix of about 20 residues, the end of the mutated C-domain being particularly flexible with few local constraints.

The example of CALRm class B corresponds to the classic canonical type II [19,20]. Very surprisingly, the dynamics of the CALRm class B monomer is strongly different from that of class A. CALRm class B has three helices (see Figure 4d–f,k and Video S5); the first helix spans from residue 354 to 374, the second and larger helix from residue 379 to 414, and the last and smaller one from residue 417 to 425 (see Figure 4d). The N-cap of CALRm class B is disordered (with an N_{eq} value of eight at position 354), the N-cap of the first helix (positions before 358) is unstable, while the rest of this helix is rigid (with N_{eq} values of one), and even the transition of the second helix is highly constrained with PB series *nopadcdfblkl* (positions 372 to 382 with N_{eq} values of less than two). The second long helix is rigid till position 401. A disordered region is then found between positions 406 and 417. The last helix is well-defined with an N_{eq} of one. Due to the disordered regions, the RMSF values are higher and could provide a false impression of instability. This is well explained in Figure 4k, which shows the separated movements of the different helices that are maintained as coherent.

Thus, CALRm class B is characterized by 3 rigid helices of about 14, 22, and 6 residues, respectively, and with one main disordered region of 10 residues.

The example of CALRm class C came from the former other type [19,20]. It has the length of the CALRm class B example but looks like CALRm class A, with only one long helix ranging from residues 384 to 406 (see Video S6). The original second smaller one (positions 418 to 422) disappeared at the beginning of the simulations. The class C dynamics is highly particular with highly flexible/disordered N- and C-termini with N_{eq} values of higher than six, associated with the high sampling of different local conformations. Before the α-helix, beginning at position 369, local constraints become important with the series of PBs *dddfbdcfbfbdfbf*. This series is rarely observed in ordered proteins, but here, is found with a low N_{eq} (of less than two). As for the CALRm class B example, the RMSF values can be confusing as disordered regions can provide the wrong impression of very high values. This is, again, well presented in Figure 4k. The three different times of the simulation show the flexibility of the regions outside the helix (and the constraints observed before it) and the superimposition of the complexity of these different zones well.

Thus, CALRm class C is characterized by a single rigid helix of about 19 residues, with 10 residues before it having constrained local conformations and the end of the mutated C-domain being highly flexible.
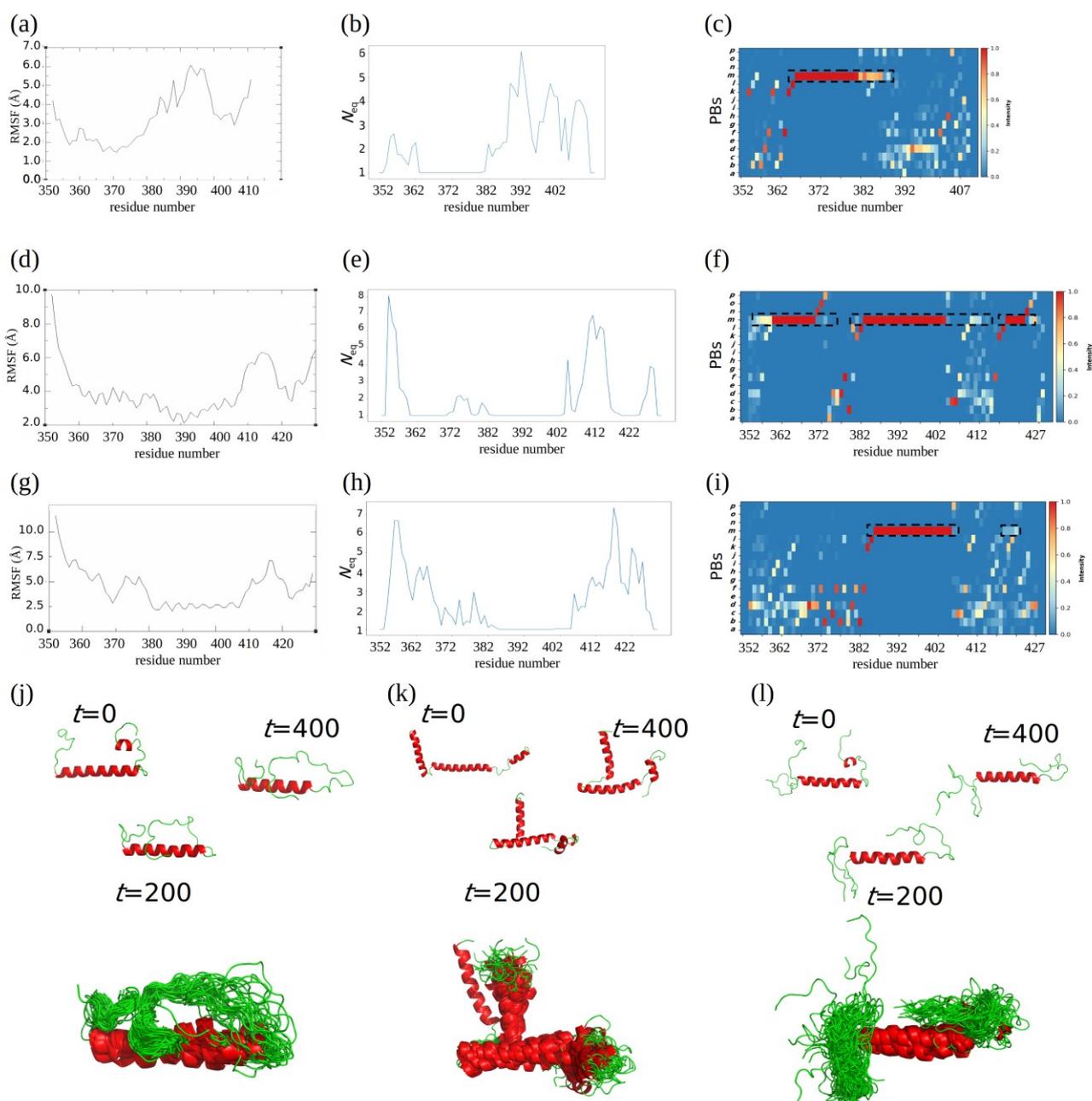
Figure 4. MD of CALRm classes A, B, and C. (**a**,**d**,**g**) Show RMSF values; (**b**,**e**,**h**) show N_{eq} values; (**c**,**f**,**i**) PB analysis of (**a**–**c**) class A, (**d**–**f**) class B, and (**g**–**i**) class C; (**j**–**l**) show visualizations for 3 times (0 nsec, 200 nsec, and 400 nsec) of MD simulations (up) and superimposition of multiple observed conformations (down) of (**j**) class A, (**k**) class B, and (**l**) class C. Please note that the choice of snapshots after that of the initial conformation is arbitrary but allows us to note the movements of the systems well (see also the Supplementary Videos).

The results of this second part show that the CALRm models of classes A, B, and C exhibit unfolding at the ends of the α-helices, with the presence of extremely flexible zones (see Videos S4–S6). The reductions in the α-helix size are identical to those observed for CALRwt dynamics with Na$^+$ ions (this is not observed with Ca^{2+} ions (see Video S1)). The second point of interest is the difference in behavior between the three classes. They are clearly distinct in terms of dynamics. Class A (herein, represented by canonical type I) has an extremity that is highly flexible ($N_{eq} > 4$–6) with some more constrained residues (with N_{eq} values of two). Class B (herein, canonical type II) has three helices that decrease in size

but remain present. Only one true region is disordered between the second and the last helix. We, therefore, have a set of three rigid bodies quite distinct from class A. Even more surprisingly, class C is composed, from the N- to C-termini, of (i) a disordered zone, (ii) a succession of local constraints that correspond to β-turns, followed by (iii) a central helix (corresponding to the second helix of class B), and then (iv) a very disordered C-terminus region ($N_{eq} > 4$–6). These three different behaviors were not too expected but are clear. It should be noted that Ca^{2+}, Na^+, and Cl^- mixtures were made without this changing the results of the MD.

3.3. Unexpected Similarities between CALRm Classes D and E and CALRwt Dynamics

This third part focuses on classes D and E. The former corresponds to a short form of the C-terminal of CALRwt, while the latter has only a point mutation with CALRwt. It was carried out using SCWRL software [78]. They are, therefore, very close to CALRwt, i.e., they are cysteine-free and therefore cannot be dimerized by the covalent bond of cystine. Ca^{2+} ions were, therefore, added, as with CALRwt, and all the parameters were identical to the previous simulations.

CALRm class D is extremely short (see Figure 5 and Video S7). It is mainly composed of one long α-helix ranging from residue 353 to 386 (see Figure 5c). This helix remains stable during 400 nsec (with N_{eq} values of one), but its extremities are flexible with an N_{eq} value of around five for the end of the helix (see Figure 5b). The RMSF values reflect the same observation (see Figure 5a). Interestingly, some replicates show more flexibility in the N-cap than the C-cap region. Figure 5g shows how Ca^{2+} interacts with the C-cap region, and the superimposition underlines the great stability of CALRm class D.

The CALRm class E sequence is the same as that of CALRwt, with one residue substituted at position 367 (L367I). CALRm class E has one main α-helix from residue 352 to residue 386 (see Figure 5f and Video S8). As with CALRwt, it binds very quickly to Ca^{2+} ions (see Figure S5a,b). Surprisingly, the beginning of the α-helix is highly flexible and tends to be a disordered region with an N_{eq} value of seven (see Figure 5e). On the contrary, the end of the helix is quite stable with an N_{eq} value of two. This is found in all replicates. These observations do not correspond to the CALRwt dynamics with Ca^{2+} ions, wherein the N-cap was stable but not the C-cap of the α-helix (see ΔPB in Figure S5c). This unstructured conformation can be induced by the position of the mutation L367I (symbolized by a bar in the PB map in Figure 5f), having a potential allosteric effect on the stability of the helix. Ca^{2+} ions also bind to the structure of the class E variant, as with CALRwt, stabilizing the regions in the second part of the structure (indicated by the red blocks in the PB maps). However, in every case, the core of the helix is always highly rigid. We also notice that the Ca^{2+} site, as seen previously, often fluctuates, providing different local conformations (see Figure S5c). Figure 5h shows the complexity of the N-termini region conformations, even with Ca^{2+} ions.

Both classes are different from CALRm classes A, B, and C since they do not share any cysteines to form dimers through disulfide bonds. Moreover, the dynamics of class D does not look like the dynamics of the main classes. Indeed, the destructuring of the helix observed in class D is not on the same level as that for CALRm class A, B, or C but is closer to CALRwt behaviors.

CALRm class E has some differences from CALRwt, particularly in the stability of the helix and the beginning of the helix, which tends to be disordered. Nevertheless, class E shares some characteristics with CALRwt, such as calcium ions fixation, which stabilizes the structure in some regions, or the absence of cysteines in the sequence. Finally, CALRm class E is the only one that still possesses the KDEL motif. Hence, variants from class E would probably be maintained in the ER, as with CALRwt, without the capacity to induce ET by binding onto the MPL of megakaryocytes. We will look at this question in the following sections.
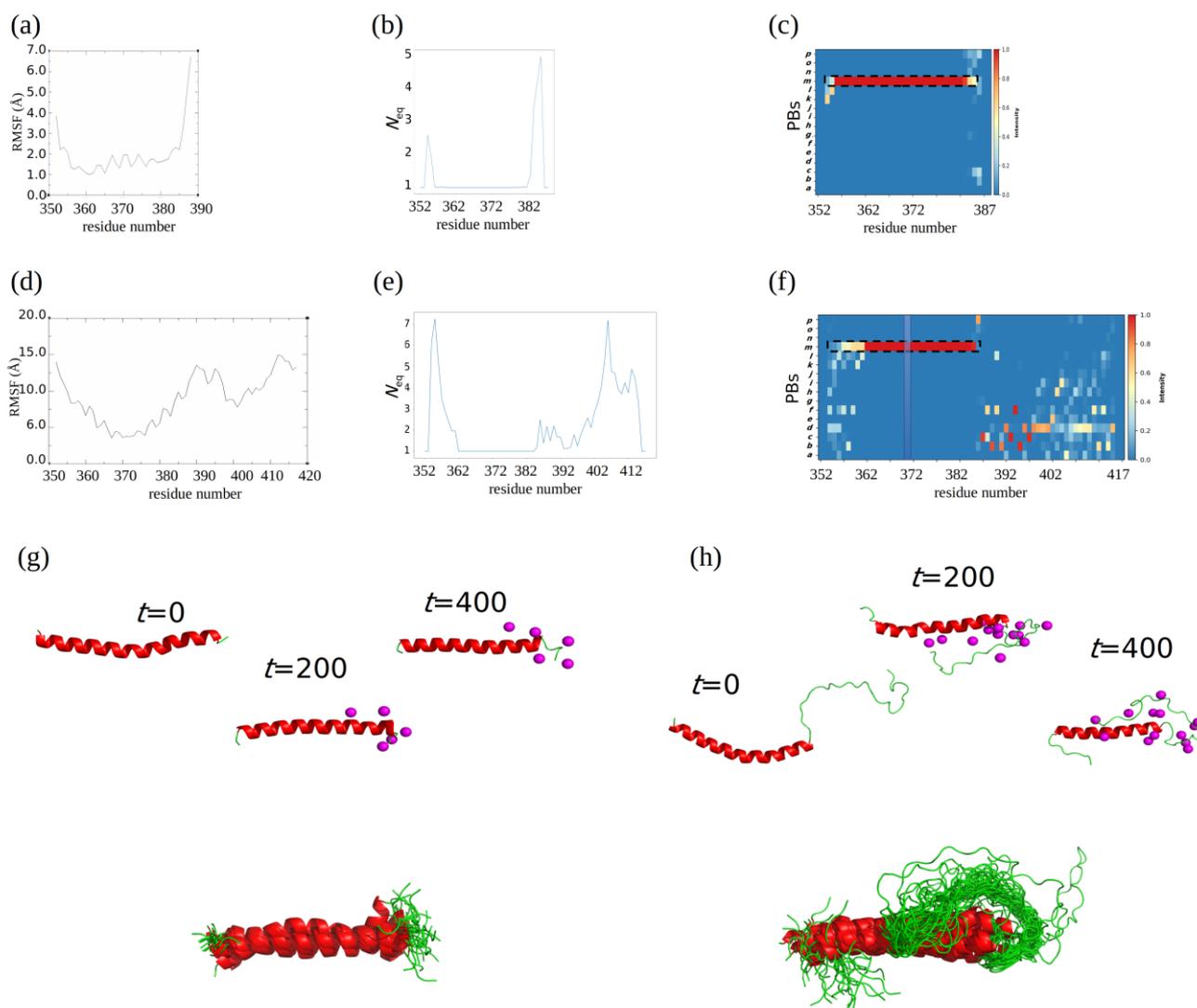
Figure 5. MD of CALRm classes D and E. (**a**,**d**) show RMSF values; (**b**,**e**) show N_{eq} values; (**c**,**f**) PB analysis of (**a**–**c**) class D and (**d**–**f**) class E; (**g**,**h**) show visualizations of 3 times (0 nsec, 200 nsec, and 400 nsec) of MD simulations (up) and superimposition of multiple observed conformations (down) of (**g**) class D and (**h**) class E.

The results of this third part show that the CALRm models of classes D and E are (i) quite distinct from the dynamics of classes A, B, and C, and (ii) close to the dynamics of CALRwt. Like the latter, they interact with Ca^{2+} ions and therefore have similar local constraints.

3.4. CALRwt Dimer Does Not Stay as a Dimer

In this fourth part, we focus on CALRwt as a dimer. As noted in the Materials and Method section, different options could have been considered to build the CALR dimers, such as docking. We preferred a constrained approach that is perfectly appropriate for the case of CALRms with disulfide bridges, i.e., CALRms A, B, and C (see Figure S6). We, therefore, used MODELLER for creating the disulfide bridges. The selected 3D models that were used for molecular dynamics were used as structural templates for both chains. We applied the same approach, with a distance criterion in the area corresponding to the cysteines, for CALRwt, and also for CALRm classes D and E, to start from comparable conformations (see Section 3.6). Figure 6 shows the results of the approach. It slightly modifies some of the 3D structural models, such as CALRm A (see Figures 2c and 6b), but

not significantly. The visualization of the electrostatics already shows that the charges are important and quite often opposed and at a short distance between the two chains.
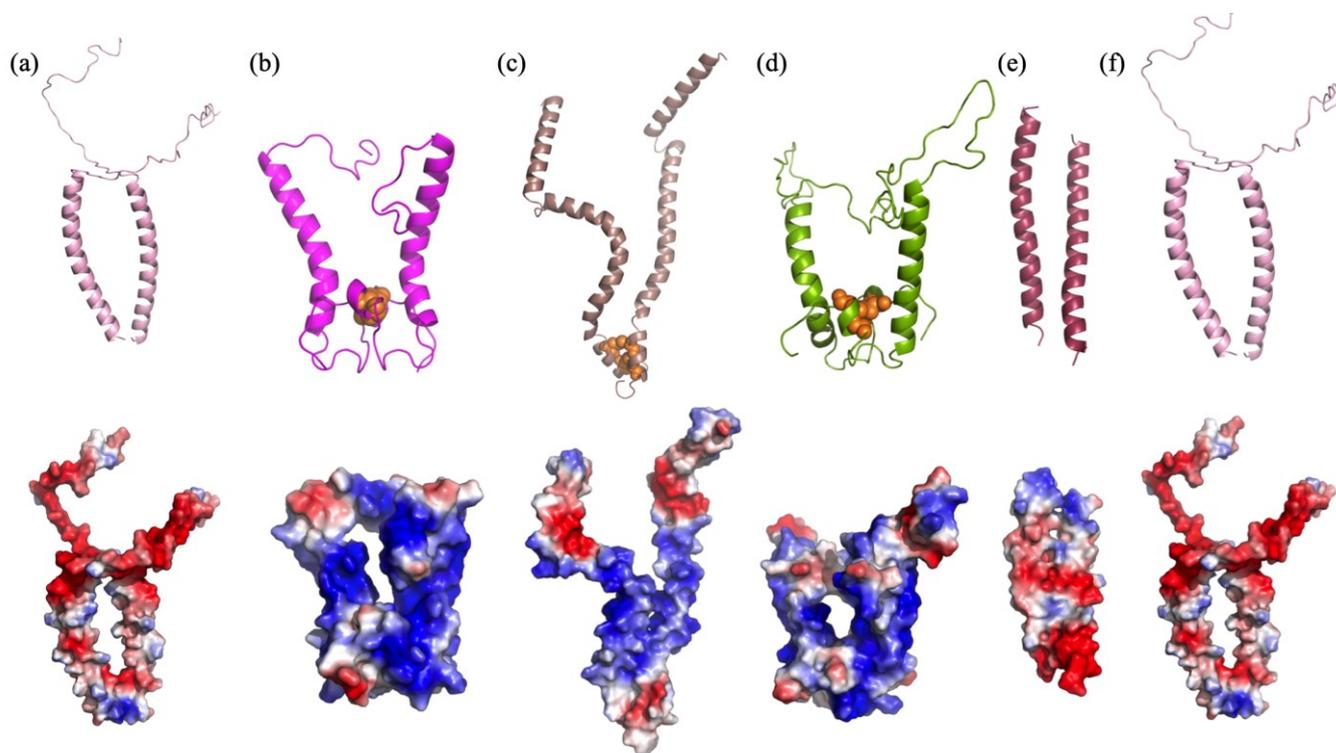
Figure 6. All the initial dimer systems. (**a**) CALRwt, (**b**) CALRm A, (**c**) CALRm B, (**d**) CALRm C, (**e**) CALRm D, and (**f**) CALRm E; (up) secondary structure visualizations and (down) electrostatics. (**b**–**d**) Cysteines implicated in disulfide bridges are indicated by orange balls.

The dynamics of the dimeric form of CALRwt was simulated for 500 nsec (with 2 replicates of 250 nsec (see Figure 7)), using Ca^{2+} ions to neutralize the system. Each chain has stable dynamics thanks to Ca^{2+} ions binding to the structures. Indeed, the helix remains mainly well structured. The Ca^{2+} ions binding to the structure allows interaction between chains through Ca^{2+} ions, similar to the interaction seen between residues on the monomeric form of CALRwt. A simulation of the CALRwt dimer was computed using Na^{+} ions instead of Ca^{2+} ions over 250 nsec. The dynamics of each chain is quite close to the dynamics of the CALRwt monomer with Na^{+} ions (see Figures S2c and 7a,d). Indeed, the ΔPB values are below one for almost the whole structure, meaning that they share similar PBs between the chains and the monomeric form of CALRwt with Na^{+} ions (see Figure S7a–d). This also indicates that the instability of the C-cap helix of both chains is at the same level as that observed with the monomer (see Figures 3b,e and 7b,e).

As in the previous systems, both chains have different dynamics (see Figure 7c,f) sampling different local conformations and flexibility levels. The distance computation shows a clear separation process for the dimer of CALRwt after 100 nsec (see Figure 7g). The two chains are not able to stay together, as seen in the evolution of the two chains (see Figure 7h–j). These results clearly show that the C-domain of CALRwt cannot simply dimerize and that its natural form must be monomeric.
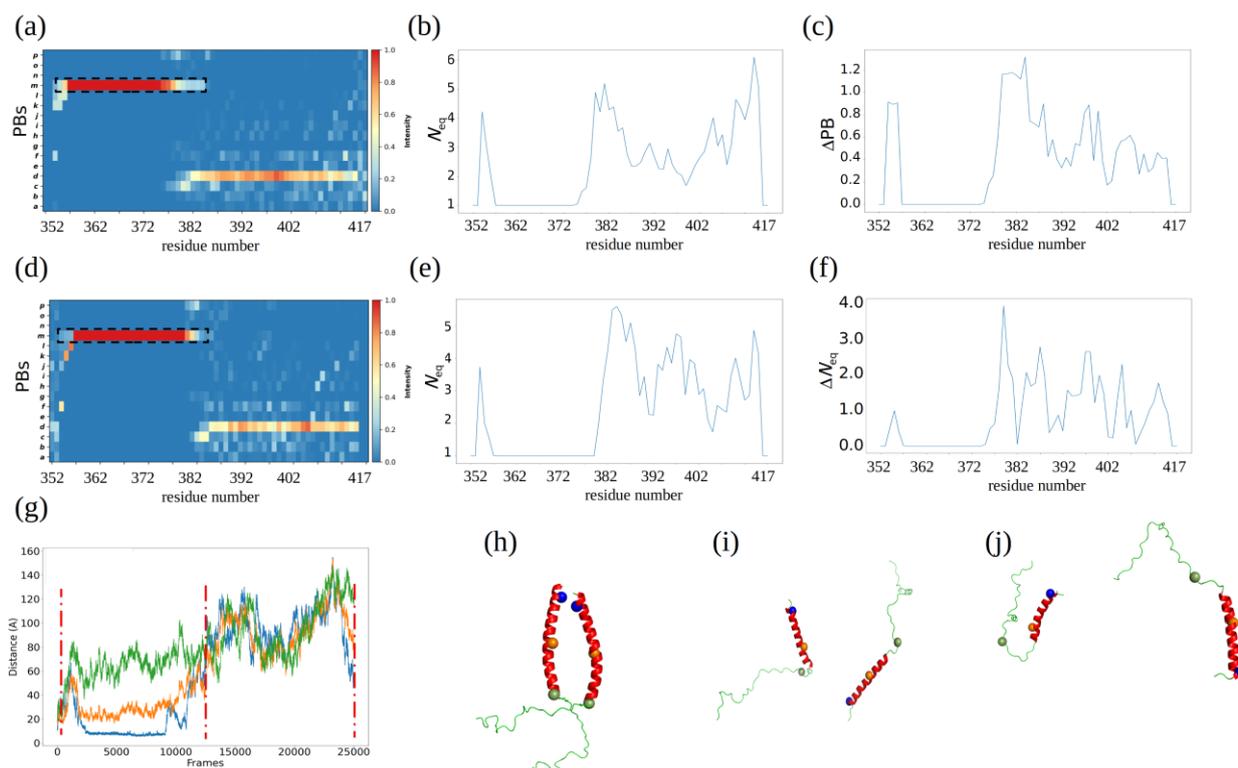
Figure 7. MD of CALRwt dimer. PB analysis of (**a**) chain 1 and (**d**) chain 2; (**b**) and (**e**) show corresponding N_{eq}; (**c**) ΔPB and (**f**) ΔN_{eq} between the two chains; (**g**) distances between different positions of the two chains (in blue, residue 356, in orange, 371, and in green, 386); and (**h**–**j**) 3D conformations observed during the simulations (positions in the simulation shown with red bars in (**g**)).

3.5. Dimerization of CALRm Classes A, B, and C Mediated by Disulfide Bonds

In this fifth part, we focus on the first three CALRm classes as dimers. The dimers were created with MODELLER (i) using previous modeled classes and (ii) adding specific disulfide constraints between cysteines. These systems correspond to what is expected in CALRms associated with ET [52,53]. The simulations were carried out in a similar way as with CALRm classes A, B, and C with an appropriate number of Cl⁻ ions.

The dynamics of the CALRm class A dimer was simulated over 500 nsec to analyze the structural impact of having disulfide bonds. The dynamics of each chain is distinct with some local similarities (see Video S9).

The main helix of chain 1 (see Figure 8a) appears to be more unstructured than that of chain 2 (see Figure 8d), especially at the beginning (residues 365 to 370) and the end (residues 382 to 387) of the helix, with N_{eq} values of one (see Figure 8b). The second helix remains mainly structured for chain 1, with N_{eq} values of one. On the contrary, the main helix of chain 2 is well-structured, whereas the second helix has a more variable PB profile and tends to be flexible, with an N_{eq} value of around four. The ΔPB values in these regions are superior to one, which means that they have different PB distributions (see Figure 8c). Despite these structural differences between the chains, they remain steadily linked through disulfide bonds, which maintain the dimeric form of CALRm class A. This linkage is probably responsible for the stabilization of the second helix of each chain since they remain mainly structured, unlike the dynamics of class A in monomeric form. Indeed, a comparison with the previous monomer of CALRm class A showed that the flexibility is distinct, with ΔN_{eq} values often between two and six, underlining a difference in terms of flexibility behaviors with monomeric forms (see Figure S8b,d). In a similar way, outside the helical regions, ΔPB often reaches values higher than 1.5, underlining that distinct conformations are explored due to the dimerization of the chains.
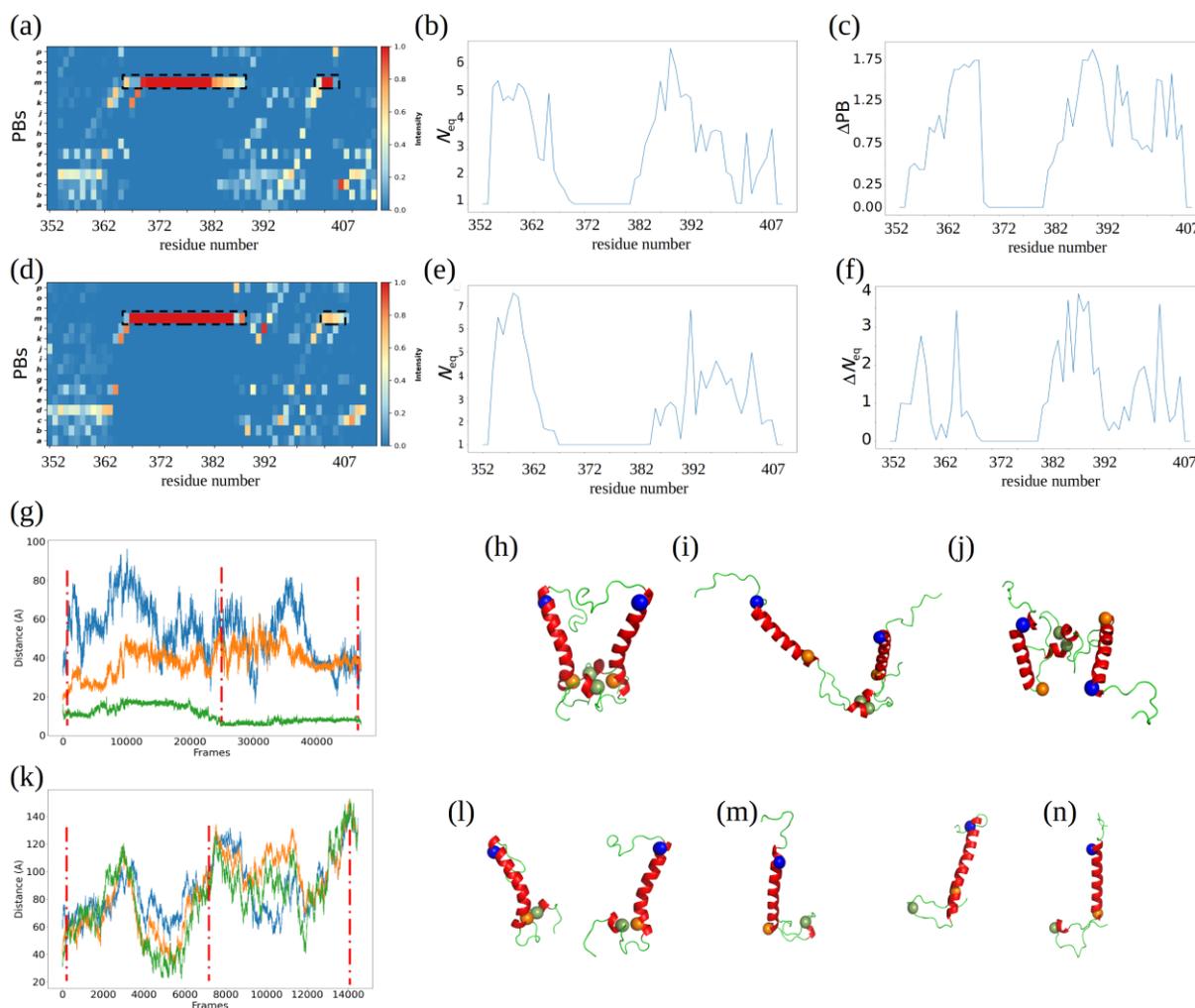
Figure 8. MD of CALRm dimer class A. (**a**–**j**) CALRm dimer class A with disulfide bridge, and (**k**) to (**n**) without disulfide bridge. PB analysis of (**a**) chain 1 and (**d**) chain 2; (**b**,**e**) show corresponding N_{eq}; (**c**) ΔPB and (**f**) ΔN_{eq} between the two chains; (**g**,**k**) distance between different positions of the two chains (in blue, residue 366, in orange, 385, and in green, 400); and (**h**–**j**) and (**l**–**n**) show 3D conformations observed during the simulations (positions in the simulation shown with red bars in (**g**,**k**), respectively).

Indeed, this dimer is highly dynamic as it does not seem to like to stay as a dimer. A distance calculation between the two chains was performed to observe changes in the distance between the two chains during the simulation. Three positions were selected—residue 366 on chain 1 (the blue color sphere), residue 385 (the orange color sphere), and residue 400 (the green color sphere)—which correspond to the cysteines implicated in the disulfide bridge. The distance between the N-terminal of each chain varies between 25 and 85 Å (blue sphere). The distance in the second position varies between 20 and 50 Å (orange sphere), while the distance in the disulfide bond position remains always very high for a cystine, i.e., around 7 Å. It would have been graded as a bad-grade cystine [79,80]. In fact, the variation in the distance diminishes as we go from the N-terminal to the C-terminal sequence where the disulfide bonds are located. Indeed, positive electrostatic patches from the middle of both chains (see Figure 6b) induce a repulsive force, which can explain the relatively high distance between them before reaching the disulfide bonds. These observations suggest that disulfide bonds are required/essential to maintain the dimer form; otherwise, the electrostatic forces between chains would have made it impossible, and both chains would have been separated.

To ensure this result, a new system was made by reducing the disulfide bridge to have free cysteines. A short molecular dynamics of 140 nsec was performed and gave very clear results. Again, the helix retains its coherence quite well, and the rest strongly samples the conformational space quite significantly (see Figure S9 and Video S10) but corresponding to what was seen with the monomer. On the other hand, the dimer does not hold at all. From the first few seconds, the two chains repel each other, as suggested in the previous paragraphs. The distance variations range easily from 20 A to over 100 A, as seen in Figure 8k. The visualizations (see Figure 8l–n) show that they have become fully independent again and that the disulfide bridge is essential to maintain the dimerization of the C-domain of CALRm class A.

The dimeric form of CALRm class B was modeled in the same way as the dimeric form of class A, its dynamics simulated over 500 nsec (2 times 250 nsec, as seen in Videos S11 and S12). Both chains have some differences in the PB distribution, especially on the main helix (see Figure 9a,d). Indeed, chain 1 has the main helix C-cap unstructured (from residue 406 to 410). The two other helices remain mainly structured. The helices in chain 2 are highly stable, N_{eq} plots underline mainly the loop regions (see Figure 9b,e), The ΔN_{eq} is more limited than that for the previous system as the helix limits local fluctuations (see Figures 8f and 9f). The highest ΔPB value remains on the main helix C-cap and the third helix (see Figure 9c). As with the class A dimer, the disulfide bonds are likely stabilizing the helices around them, which allows these helices to be well-structured, unlike the dynamics of class B in monomeric form. Indeed, a comparison with the monomer shows that the intermediate regions have different flexibilities (see Figure S10b,d) and different sampled local conformations (see Figure S10a,c).
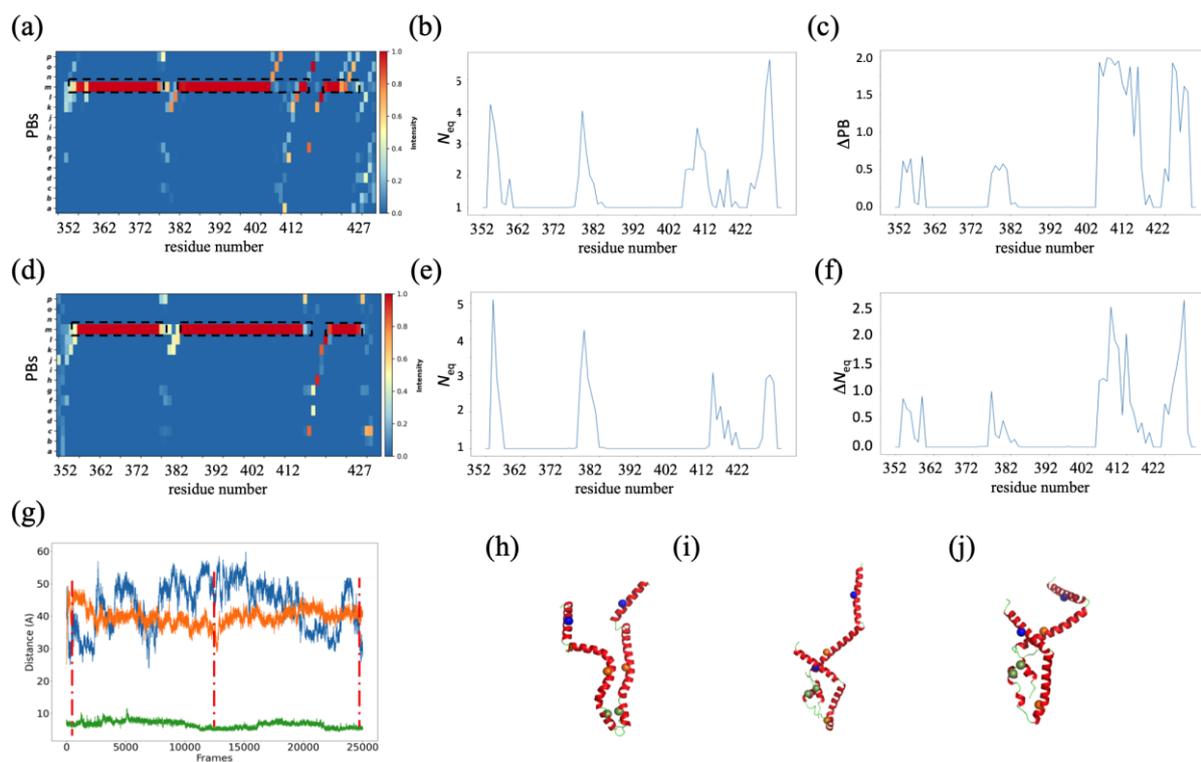
Figure 9. MD of CALRm dimer class B. PB analysis of (**a**) chain 1 and (**d**) chain 2; (**b**) and (**e**) show corresponding N_{eq}; (**c**) ΔPB and (**f**) ΔN_{eq} between the two chains; (**g**) distance between different positions of the two chains (in blue, residue 366, in orange, 400, and in green, 419, the last being the cysteine implicated in the disulfide bridge); and (**h–j**) 3D conformations observed during the simulations (positions in the simulation shown with red bars in (**g**)).

Similar to the class A dimer, distances were computed on three different anchors. The dimer quickly shows great movements that are translated into the subsequent separation of the two chains. The central position (orange ball) also moves quickly around 45 Å. It continues to oscillate during all the dynamics. The N-termini (blue ball) also starts to move away quickly at great distances (see Figure 9g–j) and oscillates between 40 and 60 Å. The main difference from the previous system is that the helices are numerous and move like small rigid bodies. After moving away, they do not move back together as their charges are identical and therefore unfavorable for interactions. Without the disulfide bridge, the two chains could not interact. A specificity of class B is the persistence of negative charges associated with the first helix that interact weakly with the positive charges of the second helix; this behavior is not observed in other cases.

Interestingly, the second MD replica shows exactly the same conformation between chains. However, the main and the third helices of chain 2 merge together to form only one helix that spans from residue 461 to 505. This surprising observation could perhaps highlight the long-range impact of disulfide bonds on stabilizing the structure.

The dynamics of the dimeric form of CALRm class C shows a high level of de-structuring in the first helix at the extremities and a large destruction of the second helix for both chains (see Video S13). The last observation follows the dynamics of CALRm class C in monomeric form. As with the other system, the dimeric form is not symmetric. Chain 2 shows more stable residues in helices in regard to the monomeric simulation. The distance between chains is quite low in the position close to the disulfide bonds and high in the N-terminal part of each chain.

The results of this fifth part underline that due to their charges, the CALRm classes A, B, and C cannot stay as dimers without the disulfide bridges. When the disulfide bridge is broken (see Video S10), the two chains separate instantly. Interestingly, this dimerization also influenced the dynamics of the protein chains. As is often seen in dimers, both chains are not equivalent (see Figures 6c, 8c, and 9c, with a high ΔPB between chains). On average, we can consider that there is even a slight stiffening of the systems (the N_{eq} remains quite high but slightly less than that for the monomer forms).

These results clearly show that the C-domain of these CALRms must dimerize simply with disulfide bridges and that otherwise, they must have monomeric behavior. Their sequences make them clearly distinct from CALRwt.

3.6. Incapacity to Dimerize of CALRm Classes D and E

This last part focuses on the CALRs that cannot dimerize through disulfide bonds, since they do not have any cysteines in their sequences. This includes CALRm classes D and E. As previously, the models of the monomers were used for the dimers. However, due to the absence of cysteines, these dimers were modeled using distance constraints between the residues corresponding to cysteine. Similarly, these three CALRs bind Ca^{2+} ions, stabilizing the structure, since the whole system is negative, contrary to the main CALRm classes A, B, and C (see Figure 6a,e,f). MDs were performed with similar parameters to before.

The dynamics of class E in the dimeric form was simulated over 250 nsec using Ca^{2+} ions to neutralize the system. As with the monomer, Ca^{2+} binding participates to stabilize each chain separately with some PB with high intensity (see Figure 10a,d). The helical regions remain well-structured and are more structured than the monomeric form of class E (see Figures 5e,f and 10b,e). Indeed, the ΔPB between the chains and the monomer shows a value superior to one for the N-terminal part of the helix and, more specifically, for chain 1 (see Figure S11a), and it is slightly less for chain 2 (see Figure S11b). As with the other systems, the two chains have different flexibility properties (see Figures 10f and S11b–d). As with the previous dimer, the distances between the chains were computed according to three anchor points (see Figure 10g). The chains do not remain in interaction and are quickly separated and reach a maximum distance of approximately 250 Å in 160 nsec. This is particularly striking when observing the different conformations observed in the simulation (see Figure 10h–j).
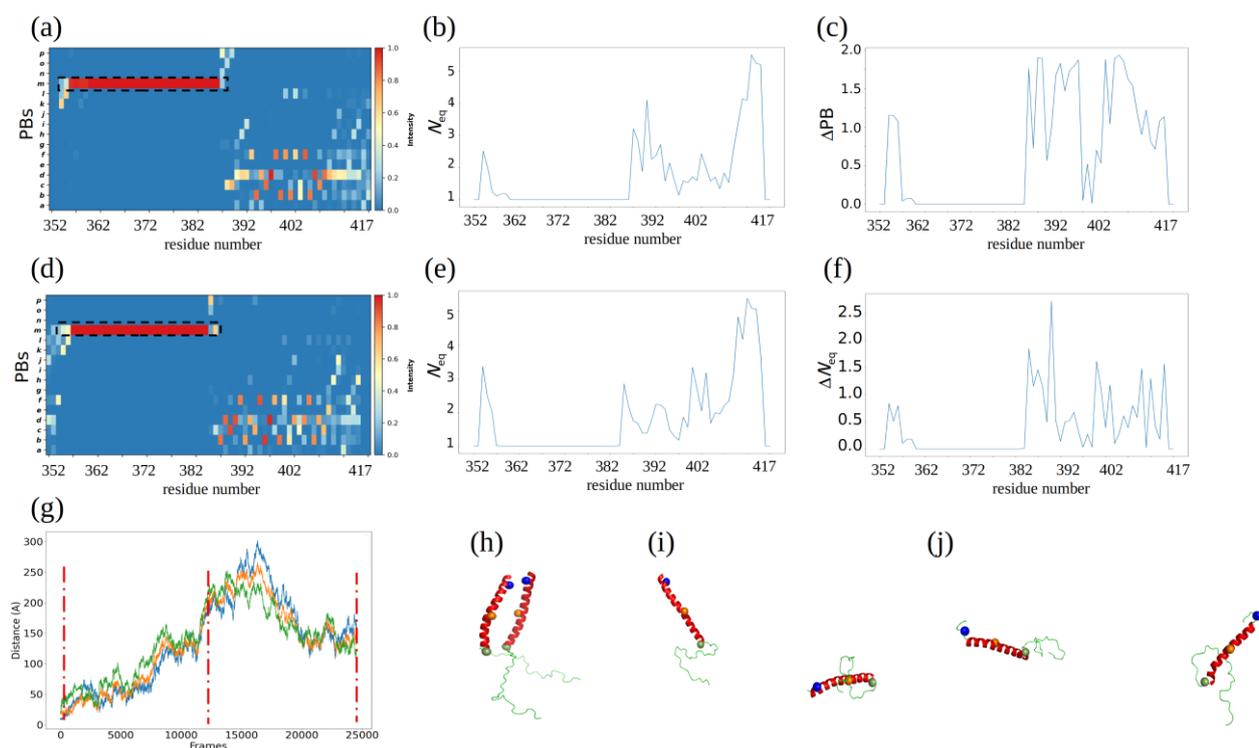
Figure 10. MD of CALRm dimer class E. PB analysis of (**a**) chain 1 and (**d**) chain 2; (**b**) and (**e**) show corresponding N_{eq}; (**c**) ΔPB and (**f**) ΔN_{eq} between the two chains; (**g**) distance between different positions of the two chains (in blue, residue 356, in orange, 371, and in green, 386); and (**h–j**) 3D conformations observed during the simulations (positions in the simulation shown with red bars in (**g**)).

Another simulation was conducted using Na^+ ions instead of Ca^{2+} ions (for a total duration of 110 nsec (see Video S14 and also Video S15 for comparison with CALRwt)). The helical regions remain stable with a slight unfolding of the helix C-cap. The second part of the sequence for each chain has a dynamic similar to the dynamics of CALRwt with Na^+ ions. Indeed, the coil regions are free of binding since there is no Ca^{2+}; therefore, they adopt a stable conformation, indicating that they are not moving a lot. For these regions, the ΔPB values between the chains and the CALRwt with Na^+ ions are below one, which means that they share more similar dynamics in this particular region than seen in other systems (but remain highly dynamic).

The MD simulations of CALRm class D were conducted over 245 nsec using Ca^{2+} ions and 200 nsec using Na^+ ions. The first simulation with Ca^{2+} ions shows a similar behavior of the ions as with the dimer of CALRwt. Ca^{2+} bindings stabilize the structure. Each chain has some variation in PB at the extremities of the helix, as with the monomeric form of class D with a ΔPB slightly above one. Some transient interactions could be provided by Ca^{2+} ions, punctually stabilizing the dimeric form, but these cases are not realistic due to the need for a high concentration of Ca^{2+} ions.

The simulation with Na^+ ions shows a different dynamic (see Video S16). The helices are stable for both chains (see Figure 11a,d). However, the extremities of the helices are much more unstructured. The C-cap helix of chain 1 has an N_{eq} value of 6, which means that these regions are close to being considered as disordered regions, and a value of 3.5 for chain 2, i.e., still highly flexible (see Figure 11b,d). The ΔPB values range from 1.20 to 1.75, respectively (see Figure 11c), which is comparable with the values achieved with the monomeric class D (see Figure S12a–d). Hence, its extremities are flexible and sometimes close to being disordered regions (see Figure 11e). The distance calculation shows a clear

separation between the chains, reaching an impressive distance of approximately 200 Å (see Figure 11g) and underlying the independent dynamics of both chains (see Figure 11i–k).
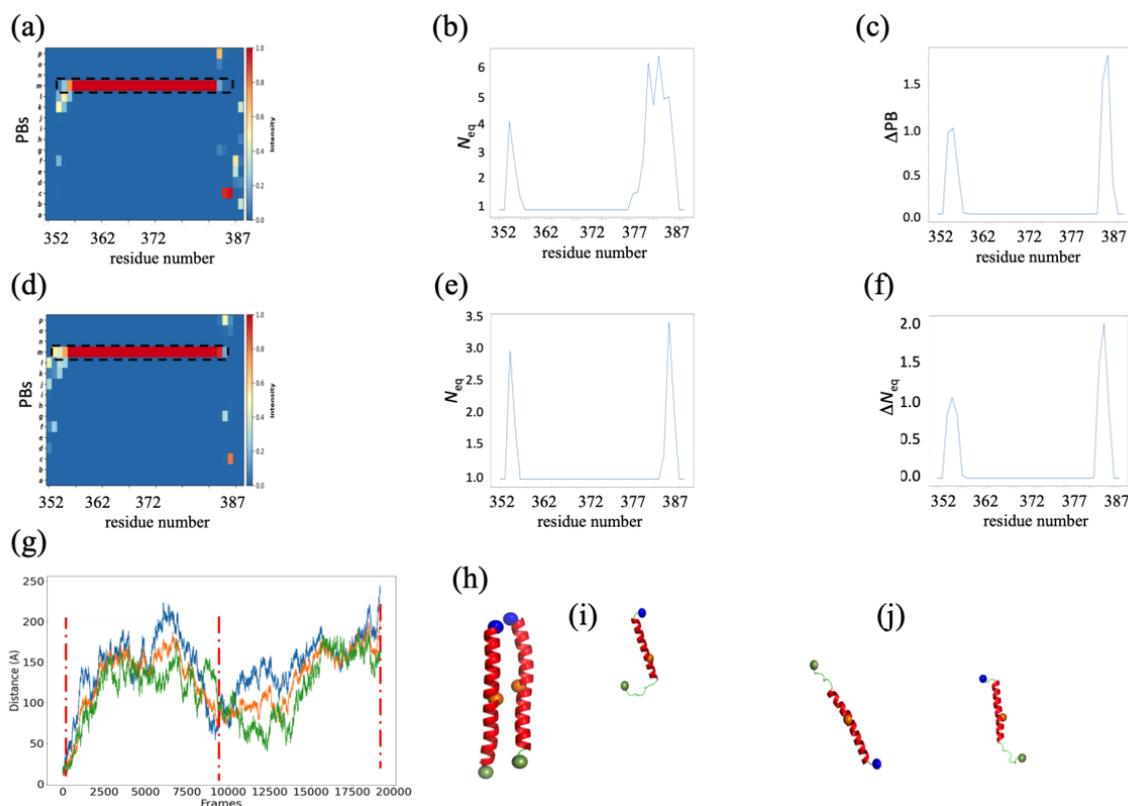
Figure 11. MD of CALRm dimer class D. PB analysis of (**a**) chain 1 and (**d**) chain 2; (**b**) and (**e**) show corresponding N_{eq}; (**c**) ΔPB and (**f**) ΔN_{eq} between the two chains; (**g**) distance between different positions of the two chains (in blue, residue 352, in orange, 370, and in green, 388); and (**h–j**) 3D conformations observed during the simulations (positions in the simulation shown with red bars in (**g**)).

Thus, the results of this last part show that neither the CALRwt nor class E and class D are able to naturally form a dimer. They all keep the rigid/flexible dichotomy between the helical and non-helical regions, as seen with the different monomers (see Section 3.2). Hence, the absence of cysteines prevents a strong interaction between chains in order to remain stable. The different visualizations of the chains (see Figures 7i,j, 10i,j and 11i,j) and corresponding distances (see Figures 7g, 10g and 11g) show without any doubt that the chains move independently. Hence, these results clearly show that the C-domains of these two types of CALRm are strongly equivalent or even similar to those of CALRwt.

4. Discussion

Essential thrombocythemia is a blood cancer characterized by an overproduction of platelets mediated via a pathway in megakaryocytes involving variants of JAK2, MPL, and CALR [36,39,81]. Variants of CALR were classified into 3 groups: type 1, type 2, and other [82]. We thought that this classification could be improved to clearly cluster the properties of all CALRms. We previously proposed a novel classification based on five classes named A to E; this classification is different from the previous classification that underlines three types mainly centered on type I and II, which represent 80% of patients. The aim of our novel classification was to provide a better understanding of the role of CALRm in essential thrombocythemia cancer [44]. CALRm dimerization can lead to the activation of MPL via binding, thus inducing ET [36]. They are able to dimerize due to novel cysteines in their C-terminal sequences that come from frameshift mutations [52,83].

Our goal was to elucidate the dynamic properties of each class of CALRm and to investigate the viability of the dimerization of CALRms through disulfide bonds. Additionally, we wanted to verify our hypothesis that class E is only a polymorphism of CALRwt and is not associated with ET. To achieve this, we used molecular dynamics simulations to reproduce the theoretical dynamics of each CALRm class in monomeric and dimeric forms, and we analyzed them with Protein Blocks [71,84].

We, therefore, proposed a comparative modeling protocol to generate the 169 structural models of CALRms present to date in CALR-ETdb [47]. For this purpose, we used the C-domain from the PDB, which was incomplete and needed to be completed (please notice that the one obtained via the cryo-EM structure in [35] is the same as the one obtained previously via X-ray structures in [57,85], probably because of the molecular replacement). Predictions of secondary structures were made with PSI-PRED [76,86]; their confidence indices are excellent. In the first step, CALRwt and CALRm types I and II were built using MODELLER [58]. The other CALRms were derived from these, sometimes with other constraints to avoid having too-long helices. We mainly focused on CALRwt, CALRm class A (we took the classical type I as the reference for this), and CALRm class B (we took the classical type II as the reference for this class). CALRm class D corresponds well to the X-ray structure because it is very short. CALRm class E is simply CALRwt with a point mutation obtained using SCWRL [78]. For CALRm class C, we used class B as a support (see the alignment in Figure 1).

With the increase in Protein Data Bank structural data [87] and new Deep Learning approaches, we wanted to test (or re-test) these approaches. We made a wide choice with threading approaches: Phyre2 [64], de novo, i.e., I-TASSER [61], and Robetta [62], and Deep Learning AlphaFold2 [63], locally, and also with CollabFold (on a reduced basis, as in [88]). The confidence indices remain very low for the disordered part of CALRwt (similar to class E) but also for classes A, B, and C. We have to remember that the new sequences obtained via the frameshifts do not exist in nature, and are not, therefore, in the databases, i.e., the multiple sequence alignments are completely empty and cannot provide any information. Therefore, we kept our models (see Data S1).

From a dynamic point of view, it is interesting to see that in the ER, the Ca^{2+} ions allow CALRwt and the two CALR classes D and E to strongly rigidify them. The number of possible Ca^{2+} binding sites is impressive and made such that depending on the simulation, they could be different. These results are in good agreement with the previous experimental and predicted results [37,38,52,83,89]. Indeed, these results show that disulfide points are of major importance for the dimerization of the C-domain of CALRm. Venkatesan and co-workers effectively highlighted this with type I, our class A. We reproduced these results by mutating the cysteines, reaching the same conclusion [52,53,83]. The analysis of the dynamics is also in good agreement with analyses performed using predictive approaches for disorders by Uversky and collaborators [37,89]. They highlighted the difficulty of seeing ordered areas in C-termini [38]. Herein, we presented, in addition, the localization of rigid fragments surrounded by very deformable regions ($N_{eq} > 6$). This leads to the localization of mobile zones [74] usable for therapeutic research approaches.

In more detail, we also highlighted the possibility of having strongly constrained areas in disordered regions (see Figure 3). The three classes with frameshifts and losses of the retention signal also have strongly disordered regions (see Figure 4) but show very different behaviors. The cases of CALRm class A (canonical type I) and CALRm class B (canonical type II) are striking, such that the first one has a particularly flexible C-terminal region, while for the second one, three helices are observed with an intermediate disordered region (thus, in agreement with the previous analyses [37,38,89] but more precise in terms of the structural characteristics).

The analysis of dimers is more complex because the flexible parts are more difficult to analyze. However, CALRm classes A, B, and C necessarily need the creation of the disulfide bridges to remain associated. CALRwt and CALRm classes D and E show no affinity for remaining in dimer form. The examples of broken disulfide bonds in CALRm class A

highlight the necessity of having disulfide bonds in these variants and, more importantly, indicate that these cysteines are probably determinants of ET. An interesting and very reassuring point is that it is not necessary to carry out very long dynamics to arrive at a sufficient conformation sampling. Indeed, with less than 200 nsec of simulations, all the conformations were observed, and the values of RMSF and N_{eq} no longer evolved. We have carried out tests over much longer times with several replicas (the data are not shown).

To conclude, Table 1 summarizes most of our findings. Our results suggest that class E is effectively very similar to CALRwt in terms of its dynamics and behavior in monomeric and dimeric forms. Hence, our hypothesis that class E is not associated with ET could be correct, and it may only correspond to human polymorphism. Class D is roughly half of CALRwt and has different dynamics compared with CALRwt but also cannot dimerize due to the absence of cysteines in its sequence, which is similar to CALRwt and class E. It may also not be a driving force of ET. We hope that these results will lead to new reflections on the sequencing of ETs as well as on therapeutic possibilities.

Table 1. Provided for each system are the length of CALR fragment, the presence (or not) of ER retention signal, the complete charge of the fragment, the % of residues associated with helical content, the presence (or not) of cysteine, and whether the dimer is stable.

System	Length	KDEL	Charge	Helix (%)	Cysteine	Stable Dimer
CALRwt	66	Yes	26 −	43	No	No
CALRm A	60	No	15 +	38	Yes	Yes
CALRm B	79	No	9 +	55	Yes	Yes
CALRm C	78	No	15 +	28	Yes	Yes
CALRm D	37	No	9 −	89	No	No
CALRm E	66	Yes	26 −	43	No	No

These results are important as ET is associated with a large number of complications [9], and it remains difficult to associate the clinical significance with the prognoses of the different CALR types [90,91]. Herein, we underlined that these types must also be carefully taken into consideration. These results are also critical in regard to the search for new drugs and antibodies focusing on CALRms [40,92].

References

1. Arber, D.A.; Orazi, A.; Hasserjian, R.; Thiele, J.; Borowitz, M.J.; Le Beau, M.M.; Bloomfield, C.D.; Cazzola, M.; Vardiman, J.W. The 2016 revision to the World Health Organization classification of myeloid neoplasms and acute leukemia. *Blood* **2016**, *127*, 2391–2405. [CrossRef] [PubMed]
2. Chuzi, S.; Stein, B.L. Essential thrombocythemia: A review of the clinical features, diagnostic challenges, and treatment modalities in the era of molecular discovery. *Leuk. Lymphoma* **2017**, *58*, 2786–2798. [CrossRef] [PubMed]
3. Szuber, N.; Mudireddy, M.; Nicolosi, M.; Penna, D.; Vallapureddy, R.R.; Lasho, T.L.; Finke, C.; Begna, K.H.; Elliott, M.A.; Hook, C.C.; et al. 3023 Mayo Clinic Patients with Myeloproliferative Neoplasms: Risk-Stratified Comparison of Survival and Outcomes Data Among Disease Subgroups. *Mayo Clin. Proc.* **2019**, *94*, 599–610. [CrossRef] [PubMed]
4. Kupfer, H.G.; Ebbels, B.J.; Miller, J.N.; Thoma, G.W.; Russi, S. Essential thrombocythemia. *Ann. Intern. Med.* **1958**, *48*, 685–697.
5. Dameshek, W. Some Speculations on the Myeloproliferative Syndromes. *Blood* **1951**, *6*, 372–375. [CrossRef]
6. Fenaux, P.; Simon, M.; Caulier, M.T.; Lai, J.L.; Goudemand, J.; Bauters, F. Clinical course of essential thrombocythemia in 147 cases. *Cancer* **1990**, *66*, 549–556. [CrossRef]
7. Araki, M.; Komatsu, N. The role of calreticulin mutations in myeloproliferative neoplasms. *Int. J. Hematol.* **2020**, *111*, 200–205. [CrossRef]
8. Maleknia, M.; Shahrabi, S.; Ghanavat, M.; Vosoughi, T.; Saki, N. Essential thrombocythemia: A hemostatic view of thrombogenic risk factors and prognosis. *Mol. Biol. Rep.* **2020**, *47*, 4767–4778. [CrossRef]
9. Babakhanlou, R.; Masarova, L.; Verstovsek, S. A review of essential thrombocythemia and its complications. *Clin. Adv. Hematol. Oncol.* **2023**, *21*, 76–84.
10. Mascarenhas, J.; Mesa, R.; Prchal, J.; Hoffman, R. Optimal therapy for polycythemia vera and essential thrombocythemia can only be determined by the completion of randomized clinical trials. *Haematologica* **2014**, *99*, 945–949. [CrossRef]
11. Mazza, G.L.; Mead-Harvey, C.; Mascarenhas, J.; Yacoub, A.; Kosiorek, H.E.; Hoffman, R.; Dueck, A.C.; Mesa, R.A. Symptom burden and quality of life in patients with high-risk essential thrombocythaemia and polycythaemia vera receiving hydroxyurea or pegylated interferon alfa-2a: A post-hoc analysis of the MPN-RC 111 and 112 trials. *Lancet Haematol.* **2022**, *9*, e38–e48. [CrossRef]
12. James, C.; Ugo, V.; Le Couédic, J.-P.; Staerk, J.; Delhommeau, F.; Lacout, C.; Garçon, L.; Raslova, H.; Berger, R.; Bennaceur-Griscelli, A.; et al. A unique clonal *JAK2* mutation leading to constitutive signalling causes polycythaemia vera. *Nature* **2005**, *434*, 1144–1148. [CrossRef]
13. Lundberg, P.; Takizawa, H.; Kubovcakova, L.; Guo, G.; Hao-Shen, H.; Dirnhofer, S.; Orkin, S.H.; Manz, M.G.; Skoda, R.C. Myeloproliferative neoplasms can be initiated from a single hematopoietic stem cell expressing *JAK2-V617F*. *J. Exp. Med.* **2014**, *211*, 2213–2230. [CrossRef]
14. Rumi, E.; Pietra, D.; Pascutto, C.; Guglielmelli, P.; Martínez-Trillos, A.; Casetti, I.; Colomer, D.; Pieri, L.; Pratcorona, M.; Rotunno, G.; et al. Clinical effect of driver mutations of *JAK2*, *CALR*, or *MPL* in primary myelofibrosis. *Blood* **2014**, *124*, 1062–1069. [CrossRef]

15. Tefferi, A.; Barbui, T. Polycythemia vera and essential thrombocythemia: 2021 update on diagnosis, risk-stratification and management. *Am. J. Hematol.* **2020**, *95*, 1599–1613. [CrossRef]

16. Pikman, Y.; Lee, B.H.; Mercher, T.; McDowell, E.; Ebert, B.L.; Gozo, M.; Cuker, A.; Wernig, G.; Moore, S.; Galinsky, I.; et al. MPLW515L Is a Novel Somatic Activating Mutation in Myelofibrosis with Myeloid Metaplasia. *PLoS Med.* **2006**, *3*, e270. [CrossRef]

17. Szuber, N.; Hanson, C.A.; Lasho, T.L.; Finke, C.; Ketterling, R.P.; Pardanani, A.; Gangat, N.; Tefferi, A. MPL-mutated essential thrombocythemia: A morphologic reappraisal. *Blood Cancer J.* **2018**, *8*, 121. [CrossRef]

18. Guglielmelli, P.; Calabresi, L. The mpl mutation. *Int. Rev. Cell Mol. Biol.* **2021**, *365*, 163–178.

19. Klampfl, T.; Gisslinger, H.; Harutyunyan, A.S.; Nivarthi, H.; Rumi, E.; Milosevic, J.D.; Them, N.C.C.; Berg, T.; Gisslinger, B.; Pietra, D.; et al. Somatic Mutations of Calreticulin in Myeloproliferative Neoplasms. *N. Engl. J. Med.* **2013**, *369*, 2379–2390. [CrossRef]

20. Nangalia, J.; Massie, C.E.; Baxter, E.J.; Nice, F.L.; Gundem, G.; Wedge, D.C.; Avezov, E.; Li, J.; Kollmann, K.; Kent, D.G.; et al. Somatic *CALR* mutations in myeloproliferative neoplasms with nonmutated *JAK2*. *N. Engl. J. Med.* **2013**, *369*, 2391–2405. [CrossRef]

21. Michalak, M.; Corbett, E.F.; Mesaeli, N.; Nakamura, K.; Opas, M. Calreticulin: One protein, one gene, many functions. *Biochem. J.* **1999**, *344 Pt 2*, 281–292. [CrossRef] [PubMed]

22. Khanna, N.C.; Tokuda, M.; Waisman, D.M. Conformational changes induced by binding of divalent cations to calregulin. *J. Biol. Chem.* **1986**, *261*, 8883–8887. [CrossRef] [PubMed]

23. Khanna, N.C.; Tokuda, M.; Waisman, D.M. Comparison of calregulins from vertebrate livers. *Biochem. J.* **1987**, *242*, 245–251. [CrossRef]

24. Khanna, N.C.; Tokuda, M.; Waisman, D.M. Calregulin: Purification, cellular localization, and tissue distribution. *Methods Enzymol.* **1987**, *139*, 36–50.

25. Michalak, M.; Groenendyk, J.; Szabo, E.; Gold, L.I.; Opas, M. Calreticulin, a multi-process calcium-buffering chaperone of the endoplasmic reticulum. *Biochem. J.* **2009**, *417*, 651–666. [CrossRef] [PubMed]

26. Raghavan, M.; Wijeyesakere, S.J.; Peters, L.R.; Del Cid, N. Calreticulin in the immune system: Ins and outs. *Trends Immunol.* **2013**, *34*, 13–21. [CrossRef]

27. Rutkevich, L.A.; Williams, D.B. Participation of lectin chaperones and thiol oxidoreductases in protein folding within the endoplasmic reticulum. *Curr. Opin. Cell Biol.* **2011**, *23*, 157–166. [CrossRef]

28. Carpio, M.A.; Decca, M.B.; Lopez Sambrooks, C.; Durand, E.S.; Montich, G.G.; Hallak, M.E. Calreticulin-dimerization induced by post-translational arginylation is critical for stress granules scaffolding. *Int. J. Biochem. Cell Biol.* **2013**, *45*, 1223–1235. [CrossRef]

29. Nauseef, W.M.; McCormick, S.J.; Clark, R.A. Calreticulin Functions as a Molecular Chaperone in the Biosynthesis of Myeloperoxidase. *J. Biol. Chem.* **1995**, *270*, 4741–4747. [CrossRef]

30. Wada, I.; Imai, S.-I.; Kai, M.; Sakane, F.; Kanoh, H. Chaperone Function of Calreticulin When Expressed in the Endoplasmic Reticulum as the Membrane-anchored and Soluble Forms. *J. Biol. Chem.* **1995**, *270*, 20298–20304. [CrossRef]

31. Baksh, S.; Michalak, M. Expression of calreticulin in Escherichia coli and identification of its Ca2+ binding domains. *J. Biol. Chem.* **1991**, *266*, 21458–21465. [CrossRef]

32. Nakamura, K.; Zuppini, A.; Arnaudeau, S.; Lynch, J.; Ahsan, I.; Miyazaki-Krause, R.; Papp, S.; De Smedt, H.; Parys, J.B.; Müller-Esterl, W.; et al. Functional specialization of calreticulin domains. *J. Cell Biol.* **2001**, *154*, 961–972. [CrossRef]

33. Wijeyesakere, S.J.; Gafni, A.A.; Raghavan, M. Calreticulin Is a Thermostable Protein with Distinct Structural Responses to Different Divalent Cation Environments. *J. Biol. Chem.* **2011**, *286*, 8771–8785. [CrossRef]

34. Ellgaard, L.; Riek, R.; Herrmann, T.; Güntert, P.; Braun, D.; Helenius, A.; Wüthrich, K. Nmr structure of the calreticulin p-domain. *Proc. Natl. Acad. Sci. USA* **2001**, *98*, 3133–3138. [CrossRef]

35. Blees, A.; Januliene, D.; Hofmann, T.; Koller, N.; Schmidt, C.; Trowitzsch, S.; Moeller, A.; Tampé, R. Structure of the human MHC-I peptide-loading complex. *Nature* **2017**, *551*, 525–528. [CrossRef]

36. Shide, K. Calreticulin mutations in myeloproliferative neoplasms. *Int. Rev. Cell Mol. Biol.* **2021**, *365*, 179–226.

37. Falchi, M.; Varricchio, L.; Martelli, F.; Marra, M.; Picconi, O.; Tafuri, A.; Girelli, G.; Uversky, V.N.; Migliaccio, A.R. The Calreticulin control of human stress erythropoiesis is impaired by *JAK2V617F* in polycythemia vera. *Exp. Hematol.* **2017**, *50*, 53–76. [CrossRef]

38. Migliaccio, A.R.; Uversky, V.N. Dissecting physical structure of calreticulin, an intrinsically disordered ca^{2+}-buffering chaperone from endoplasmic reticulum. *J. Biomol. Struct. Dyn.* **2018**, *36*, 1617–1636. [CrossRef]

39. Fucikova, J.; Spisek, R.; Kroemer, G.; Galluzzi, L. Calreticulin and cancer. *Cell Res.* **2021**, *31*, 5–16. [CrossRef]

40. How, J.; Hobbs, G.S.; Mullally, A. Mutant calreticulin in myeloproliferative neoplasms. *Blood* **2019**, *134*, 2242–2248. [CrossRef]

41. Cabagnols, X.; Defour, J.-P.; Ugo, V.; Ianotto, J.C.; Mossuz, P.; Mondet, J.; Girodon, F.; Alexandre, J.H.; Mansier, O.; Viallard, J.F.; et al. Differential association of calreticulin type 1 and type 2 mutations with myelofibrosis and essential thrombocytemia: Relevance for disease evolution. *Leukemia* **2015**, *29*, 249–252. [CrossRef] [PubMed]

42. Pietra, D.; Rumi, E.; Ferretti, V.V.; Di Buduo, C.A.; Milanesi, C.; Cavalloni, C.; Sant'Antonio, E.; Abbonante, V.; Moccia, F.; Casetti, I.C.; et al. Differential clinical effects of different mutation subtypes in *CALR*-mutant myeloproliferative neoplasms. *Leukemia* **2016**, *30*, 431–438. [CrossRef] [PubMed]

43. Pérez Encinas, M.M.; Sobas, M.; Gómez-Casares, M.T.; Abuin Blanco, A.; Noya Pereira, M.S.; Raya, J.M.; Andrade-Campos, M.M.; Álvarez Larrán, A.; Lewandowski, K.; Łukasz, S.; et al. The risk of thrombosis in essential thrombocythemia is associated with the type of *CALR* mutation: A multicentre collaborative study. *Eur. J. Haematol.* **2021**, *106*, 371–379. [CrossRef] [PubMed]

44. Cottin, L.; Riou, J.; Orvain, C.; Ianotto, J.C.; Boyer, F.; Renard, M.; Truchan-Graczyk, M.; Murati, A.; Jouanneau-Courville, R.; Allangba, O.; et al. Sequential mutational evaluation of *CALR*-mutated myeloproliferative neoplasms with thrombocytosis reveals an association between *CALR* allele burden evolution and disease progression. *Br. J. Haematol.* **2020**, *188*, 935–944. [CrossRef] [PubMed]

45. Chen, R.; Shi, X.; Wang, L.; Wang, X.; Wei, J.; Kang, X.; Du, F.; Gao, S.; Yang, F.; Jiang, W. Essential thrombocythemia with *CALR* mutation and recurrent stroke: Two case reports and literature review. *Ther. Adv. Neurol. Disord.* **2022**, *15*, 17562864221092093. [CrossRef]

46. Imashuku, S.; Kotani, S.; Hinami, J.; Nishimura, K. Splenic mass in a case of *CALR* -mutated essential thrombocythemia. *Clin. Case Rep.* **2018**, *6*, 2291–2292. [CrossRef]

47. El Jahrani, N.; Cretin, G.; de Brevern, A.G. CALR-ETdb, the database of calreticulin variants diversity in essential thrombocythemia. *Platelets* **2022**, *33*, 157–167. [CrossRef]

48. Tate, J.G.; Bamford, S.; Jubb, H.C.; Sondka, Z.; Beare, D.M.; Bindal, N.; Boutselakis, H.; Cole, C.G.; Creatore, C.; Dawson, E.; et al. COSMIC: The Catalogue of Somatic Mutations in Cancer. *Nucleic Acids Res.* **2019**, *47*, D941–D947. [CrossRef]

49. Araki, M. Development of myeloproliferative neoplasms by mutant calreticulin: Underlying mechanisms. *Rinsho Ketsueki* **2018**, *59*, 1072–1077.

50. Araki, M.; Yang, Y.; Imai, M.; Mizukami, Y.; Kihara, Y.; Sunami, Y.; Masubuchi, N.; Edahiro, Y.; Hironaka, Y.; Osaga, S.; et al. Homomultimerization of mutant calreticulin is a prerequisite for MPL binding and activation. *Leukemia* **2019**, *33*, 122–131. [CrossRef]

51. Charonis, A.S.; Michalak, M.; Groenendyk, J.; Agellon, L.B. Endoplasmic reticulum in health and disease: The 12th International Calreticulin Workshop, Delphi, Greece. *J. Cell. Mol. Med.* **2017**, *21*, 3141–3149. [CrossRef]

52. Venkatesan, A.; Geng, J.; Kandarpa, M.; Wijeyesakere, S.J.; Bhide, A.; Talpaz, M.; Pogozheva, I.D.; Raghavan, M. Mechanism of mutant calreticulin-mediated activation of the thrombopoietin receptor in cancers. *J. Cell Biol.* **2021**, *220*, e202009179. [CrossRef]

53. Venkatesan, A.; Geng, J.; Kandarpa, M.; Wijeyesakere, S.; Bhide, A.; Talpaz, M.; Pogozheva, I.; Raghavan, M. MPN-274: Mpl is Activated by Dimers of MPN-Linked Calreticulin Mutants Stabilized by Disulfide Bonds and Ionic Interactions. *Clin. Lymphoma Myeloma Leuk.* **2020**, *20*, S335. [CrossRef]

54. Consortium, U. Uniprot: A worldwide hub of protein knowledge. *Nucleic Acids Res.* **2019**, *47*, D506–D515. [CrossRef]

55. Altschul, S.F.; Madden, T.L.; Schäffer, A.A.; Zhang, J.; Zhang, Z.; Miller, W.; Lipman, D.J. Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucleic Acids Res.* **1997**, *25*, 3389–3402. [CrossRef]

56. Berman, H.M.; Battistuz, T.; Bhat, T.N.; Bluhm, W.F.; Bourne, P.E.; Burkhardt, K.; Feng, Z.; Gilliland, G.L.; Iype, L.; Jain, S.; et al. The protein data bank. *Acta Crystallogr. D Biol. Crystallogr.* **2002**, *58*, 899–907. [CrossRef]

57. Moreau, C.J.; Cioci, G.; Iannello, M.; Laffly, E.; Chouquet, A.; Ferreira, A.; Thielens, N.M.; Gaboriaud, C. Structures of parasite calreticulins provide insights into their flexibility and dual carbohydrate/peptide-binding properties. *IUCrJ* **2016**, *3*, 408–419. [CrossRef]

58. Šali, A.; Blundell, T.L. Comparative Protein Modelling by Satisfaction of Spatial Restraints. *J. Mol. Biol.* **1993**, *234*, 779–815. [CrossRef]

59. Shen, M.-Y.; Sali, A. Statistical potential for assessment and prediction of protein structures. *Protein Sci.* **2006**, *15*, 2507–2524. [CrossRef]

60. DeLano, W.L. *PyMOL*; Delano Scientific: San Carlos, CA, USA, 2002; p. 700.

61. Yang, J.; Yan, R.; Roy, A.; Xu, D.; Poisson, J.; Zhang, Y. The I-TASSER Suite: Protein structure and function prediction. *Nat. Methods* **2015**, *12*, 7–8. [CrossRef]

62. Song, Y.; DiMaio, F.; Wang, R.Y.-R.; Kim, D.; Miles, C.; Brunette, T.; Thompson, J.; Baker, D. High-Resolution Comparative Modeling with RosettaCM. *Structure* **2013**, *21*, 1735–1742. [CrossRef] [PubMed]

63. Jumper, J.; Evans, R.; Pritzel, A.; Green, T.; Figurnov, M.; Ronneberger, O.; Tunyasuvunakool, K.; Bates, R.; Žídek, A.; Potapenko, A.; et al. Highly accurate protein structure prediction with AlphaFold. *Nature* **2021**, *596*, 583–589. [CrossRef] [PubMed]

64. Kelley, L.A.; Mezulis, S.; Yates, C.M.; Wass, M.N.; Sternberg, M.J.E. The Phyre2 web portal for protein modeling, prediction and analysis. *Nat. Protoc.* **2015**, *10*, 845–858. [CrossRef]

65. Van Der Spoel, D.; Lindahl, E.; Hess, B.; Groenhof, G.; Mark, A.E.; Berendsen, H.J. Gromacs: Fast, flexible, and free. *J. Comput. Chem.* **2005**, *26*, 1701–1718. [CrossRef] [PubMed]

66. Huang, J.; MacKerell, A.D., Jr. Charmm36 all-atom additive protein force field: Validation based on comparison to NMR data. *J. Comput. Chem.* **2013**, *34*, 2135–2145. [CrossRef]

67. Goguet, M.; Narwani, T.J.; Petermann, R.; Jallu, V.; de Brevern, A.G. In silico analysis of glanzmann variants of *Calf* -1 domain of α(iib)β(3) integrin revealed dynamic allosteric effect. *Sci. Rep.* **2017**, *7*, 8001. [CrossRef]

68. Narwani, T.J.; Craveur, P.; Shinada, N.K.; Floch, A.; Santuz, H.; Vattekatte, A.M.; Srinivasan, N.; Rebehmed, J.; Gelly, J.C.; Etchebest, C.; et al. Discrete analyses of protein dynamics. *J. Biomol. Struct. Dyn.* **2020**, *38*, 2988–3002. [CrossRef]

69. Barnoud, J.; Santuz, H.; Craveur, P.; Joseph, A.P.; Jallu, V.; de Brevern, A.G.; Poulain, P. PBxplore: A tool to analyze local protein structure and deformability with Protein Blocks. *PeerJ* **2017**, *5*, e4013. [CrossRef]
70. Michaud-Agrawal, N.; Denning, E.J.; Woolf, T.B.; Beckstein, O. MDAnalysis: A toolkit for the analysis of molecular dynamics simulations. *J. Comput. Chem.* **2011**, *32*, 2319–2327. [CrossRef]
71. De Brevern, A.G.; Etchebest, C.; Hazout, S. Bayesian probabilistic approach for predicting backbone structures in terms of protein blocks. *Proteins Struct. Funct. Bioinform.* **2000**, *41*, 271–287. [CrossRef]
72. Joseph, A.P.; Agarwal, G.; Mahajan, S.; Gelly, J.-C.; Swapna, L.S.; Offmann, B.; Cadet, F.; Bornot, A.; Tyagi, M.; Valadié, H.; et al. A short survey on protein blocks. *Biophys. Rev.* **2010**, *2*, 137–145. [CrossRef]
73. Joseph, A.P.; Srinivasan, N.; de Brevern, A.G. Progressive structure-based alignment of homologous proteins: Adopting sequence comparison strategies. *Biochimie* **2012**, *94*, 2025–2034. [CrossRef]
74. Craveur, P.; Joseph, A.P.; Esque, J.; Narwani, T.J.; Noël, F.; Shinada, N.; Goguet, M.; Leonard, S.; Poulain, P.; Bertrand, O.; et al. Protein flexibility in the light of structural alphabets. *Front. Mol. Biosci.* **2015**, *2*, 20. [CrossRef]
75. Fliegel, L.; Burns, K.; MacLennan, D.H.; Reithmeier, R.A.; Michalak, M. Molecular cloning of the high affinity calcium-binding protein (calreticulin) of skeletal muscle sarcoplasmic reticulum. *J. Biol. Chem.* **1989**, *264*, 21522–21528. [CrossRef]
76. Jones, D.T. Protein secondary structure prediction based on position-specific scoring matrices. *J. Mol. Biol.* **1999**, *292*, 195–202. [CrossRef]
77. Akhila, M.V.; Narwani, T.J.; Floch, A.; Maljković, M.; Bisoo, S.; Shinada, N.K.; Kranjc, A.; Gelly, J.-C.; Srinivasan, N.; Mitić, N.; et al. A structural entropy index to analyse local conformations in intrinsically disordered proteins. *J. Struct. Biol.* **2020**, *210*, 107464. [CrossRef]
78. Krivov, G.G.; Shapovalov, M.V.; Dunbrack, R.L., Jr. Improved prediction of protein side-chain conformations with SCWRL4. *Proteins* **2009**, *77*, 778–795. [CrossRef]
79. Kalmankar, N.V.; Pavalam, M.; Indrakumar, S.; Srinivasan, N.; Sowdhamini, R. DSDBASE 2.0: Updated version of Disulphide dataBASE, a database on disulphide bonds in proteins. *Database J. Biol. Databases Curation* **2022**, *2022*, baac005. [CrossRef]
80. Sowdhamini, R.; Srinivasan, N.; Shoichet, B.; Santi, D.V.; Ramakrishnan, C.; Balaram, P. Stereochemical modeling of disulfide bridges. Criteria for introduction into proteins by site-directed mutagenesis. *Protein Eng. Des. Sel.* **1989**, *3*, 95–103. [CrossRef]
81. Soliman, E.A.; El-Ghlban, S.; El-Aziz, S.A.; Abdelaleem, A.; Shamaa, S.; Abdel-Ghaffar, H. *JAK2*, *CALR*, and *MPL* Mutations in Egyptian Patients with Classic Philadelphia-negative Myeloproliferative Neoplasms. *Clin. Lymphoma Myeloma Leuk.* **2020**, *20*, e645–e651. [CrossRef]
82. Prins, D.; González Arias, C.; Klampfl, T.; Grinfeld, J.; Green, A.R. Mutant Calreticulin in the Myeloproliferative Neoplasms. *Hemasphere* **2020**, *4*, e333. [CrossRef] [PubMed]
83. Venkatesan, A.; Satin, L.S.; Raghavan, M. Roles of calreticulin in protein folding, immunity, calcium signaling and cell transformation. *Prog. Mol. Subcell. Biol.* **2021**, *59*, 145–162. [PubMed]
84. De Brevern, A.G. Analysis of protein disorder predictions in the light of a protein structural alphabet. *Biomolecules* **2020**, *10*, 1080. [CrossRef] [PubMed]
85. Chouquet, A.; Païdassi, H.; Ling, W.L.; Frachet, P.; Houen, G.; Arlaud, G.J.; Gaboriaud, C. X-ray structure of the human calreticulin globular domain reveals a peptide-binding area and suggests a multi-molecular mechanism. *PLoS ONE* **2011**, *16*, e17886. [CrossRef] [PubMed]
86. Buchan, D.W.A.; Jones, D.T. The PSIPRED Protein Analysis Workbench: 20 years on. *Nucleic Acids Res.* **2019**, *47*, W402–W407. [CrossRef]
87. Berman, H.M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T.N.; Weissig, H.; Shindyalov, I.N.; Bourne, P.E. The protein data bank. *Nucleic Acids Res.* **2000**, *28*, 235–242. [CrossRef]
88. Mirdita, M.; Schütze, K.A.; Moriwaki, Y.A.; Heo, L.; Ovchinnikov, S.; Steinegger, M. ColabFold: Making protein folding accessible to all. *Nat. Methods* **2022**, *19*, 679–682. [CrossRef]
89. Varricchio, L.; Falchi, M.; Dall'Ora, M.; De Benedittis, C.; Ruggeri, A.; Uversky, V.N.; Migliaccio, A.R. Calreticulin: Challenges Posed by the Intrinsically Disordered Nature of Calreticulin to the Study of Its Function. *Front. Cell Dev. Biol.* **2017**, *5*, 96. [CrossRef]
90. Afana, M.; Abu-Tineh, M.; Ellahie, A.; Ismail, O.; Sideeg, D.; Albattah, A.; Yassin, M.A. Myocardial Infarction as an Initial Presentation of Essential Thrombocythemia with Calreticulin (CALR) Mutation (None Type 1, None Type 2). *Cureus* **2023**, *15*, e33612. [CrossRef]
91. Guglielmelli, P.; Maccari, C.; Sordi, B.; Balliu, M.; Atanasio, A.; Mannarelli, C.; Capecchi, G.; Sestini, I.; Coltro, G.; Loscocco, G.G.; et al. Phenotypic correlations of *CALR* mutation variant allele frequency in patients with myelofibrosis. *Blood Cancer J.* **2023**, *13*, 21. [CrossRef]
92. Tvorogov, D.; Thompson-Peach, C.A.L.; Foßelteder, J.; Dottore, M.; Stomski, F.; Onnesha, S.A.; Lim, K.; Moretti, P.A.B.; Pitson, S.M.; Ross, D.M.; et al. Targeting human CALR-mutated MPN progenitors with a neoepitope-directed monoclonal antibody. *EMBO Rep.* **2022**, *23*, e52904. [CrossRef]