



**HAL**  
open science

# Identification of the Primary Factors Determining the Specificity of the human VKORC1 Recognition by Thioredoxin-fold Proteins

Maxim Stolyarchuk, Julie Ledoux, Elodie Maignant, Alain Trouvé, Luba Tchertanov

► **To cite this version:**

Maxim Stolyarchuk, Julie Ledoux, Elodie Maignant, Alain Trouvé, Luba Tchertanov. Identification of the Primary Factors Determining the Specificity of the human VKORC1 Recognition by Thioredoxin-fold Proteins. *International Journal of Molecular Sciences*, 2021, 22 (2), pp.802. 10.3390/ijms22020802 . hal-03042382v2

**HAL Id: hal-03042382**

**<https://hal.science/hal-03042382v2>**

Submitted on 31 Mar 2022

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Article

# Identification of the Primary Factors Determining the Specificity of Human VKORC1 Recognition by Thioredoxin-Fold Proteins

Maxim Stolyarchuk †, Julie Ledoux †, Elodie Maignant, Alain Trouvé and Luba Tchertanov \*

Université Paris-Saclay, ENS Paris-Saclay, CNRS, Centre Borelli, 4 av. des Sciences, F-91190 Gif-sur-Yvette, France; maksim.stoliarchuk@ens-paris-saclay.fr (M.S.); julie.ledoux@ens-paris-saclay.fr (J.L.); elodie.maignant@ens-paris-saclay.fr (E.M.); alain.trouve@ens-paris-saclay.fr (A.T.)

\* Correspondence: Luba.Tchertanov@ens-paris-saclay.fr

† The first two authors contributed equally

**Abstract:** Redox (reduction–oxidation) reactions control many important biological processes in all organisms, both prokaryotes and eukaryotes. This reaction is usually accomplished by canonical disulphide-based pathways involving a donor enzyme that reduces the oxidised cysteine residues of a target protein, resulting in the cleavage of its disulphide bonds. Focusing on human vitamin K epoxide reductase (hVKORC1) as a target and on four redoxins (protein disulphide isomerase (PDI), endoplasmic reticulum oxidoreductase (ERp18), thioredoxin-related transmembrane protein 1 (Tmx1) and thioredoxin-related transmembrane protein 4 (Tmx4)) as the most probable reducers of VKORC1, a comparative in-silico analysis that concentrates on the similarity and divergence of redoxins in their sequence, secondary and tertiary structure, dynamics, intraprotein interactions and composition of the surface exposed to the target is provided. Similarly, hVKORC1 is analysed in its native state, where two pairs of cysteine residues are covalently linked, forming two disulphide bridges, as a target for Trx-fold proteins. Such analysis is used to derive the putative recognition/binding sites on each isolated protein, and PDI is suggested as the most probable hVKORC1 partner. By probing the alternative orientation of PDI with respect to hVKORC1, the functionally related noncovalent complex formed by hVKORC1 and PDI was found, which is proposed to be a first precursor to probe thiol–disulphide exchange reactions between PDI and hVKORC1.

**Keywords:** hVKORC1; Trx-fold proteins; protein folding; dynamics; molecular recognition; thiol–disulphide exchange; protein–protein interactions; PDI–hVKORC1 complex; 3D modelling; molecular dynamics simulation

**Citation:** Stolyarchuk, M.; Ledoux, J.; Maignant, E.; Trouvé, A.; Tchertanov, L. Identification of the Primary Factors Determining the Specificity of the Human VKORC1 Recognition by Thioredoxin-Fold Proteins. *Int. J. Mol. Sci.* **2021**, *22*, 802. <https://doi.org/10.3390/ijms22020802>

Academic Editor: Ron Orbach

Received: 6 December 2020

Accepted: 11 January 2021

Published: 14 January 2021

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



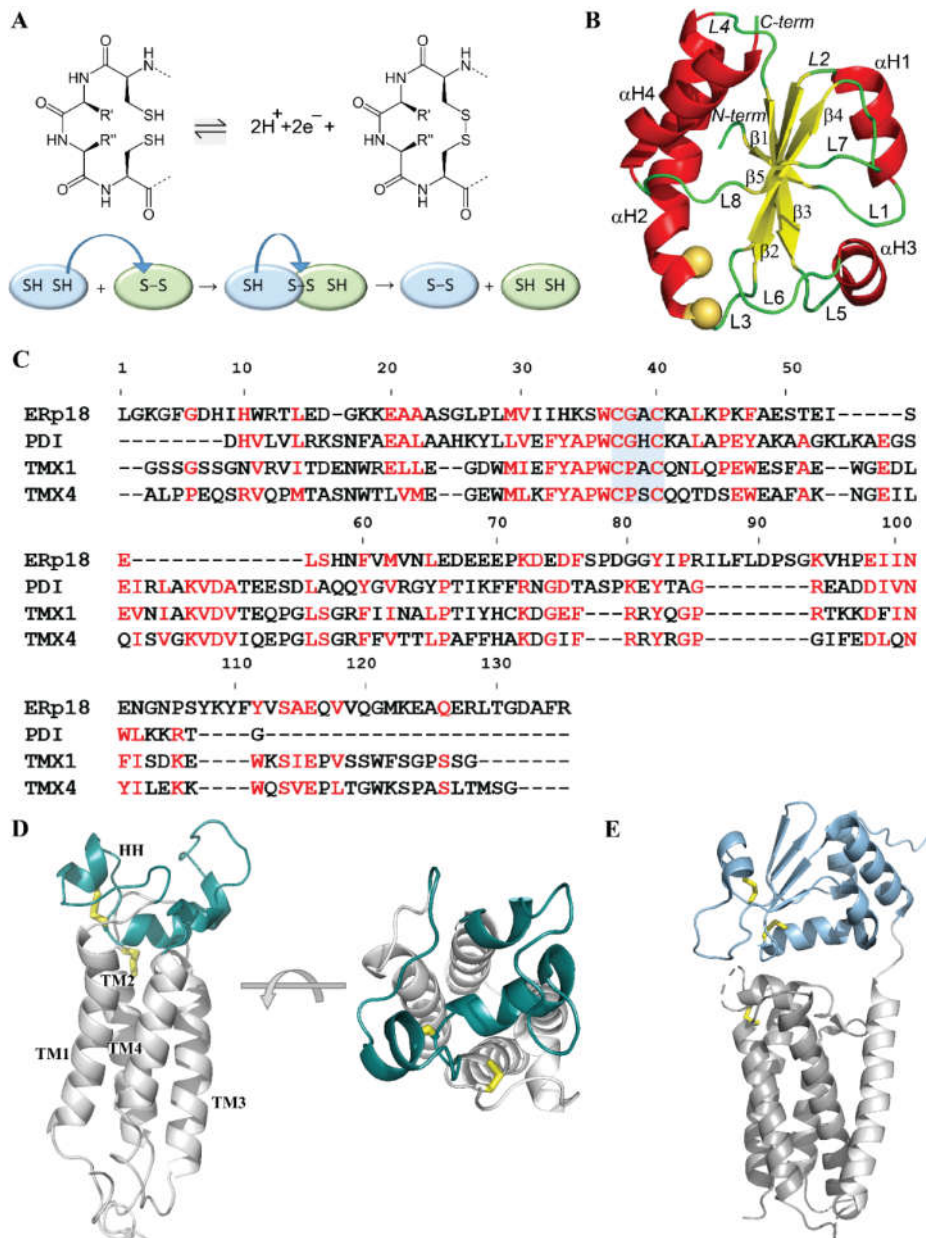
**Copyright:** © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Thioredoxins (Trxs) are disulphide reductases that are responsible for maintaining proteins in their reduced state inside cells. Trxs are involved in a wide variety of fundamental biological functions ([1] and references herein) and, therefore, are vital for all living cells, from archaeobacteria to mammals. The wide variety of Trx reactions is based on their broad substrate specificity and potent capacity to reduce multiple cellular proteins [2]. This broad specificity for thioredoxin and related proteins has made it difficult to distinguish the true physiological partners for the protein from in vitro artefacts.

All membrane-associated Trx proteins possess an active site made up of two vicinal cysteine (C) residues embedded in a conserved CX<sub>1</sub>X<sub>2</sub>C motif. These two cysteines, separated by two residues, play a key role in the transfer of two hydrogen atoms to the oxidised target and the breaking of the Trx–disulphide bond (Figure 1A). This disulphide-

relay pathway is accompanied by an electron transfer in the opposite direction. An intermediate state during the electron transfer is a mixed disulphide bond formed by a pair of cysteine residues from two proteins, which can be resolved by the nucleophilic attack of a thiol group from one of the flanking cysteine residues. Through this mechanism, the disulphide is exchanged within one thiol oxidoreductase or between a disulphide donor and a target protein [3]. Thiol–disulphide exchange reactions occur between redox-sensitive biomolecules if donors and acceptors can interact in the appropriate orientations when attacking and leaving groups [4].



**Figure 1.** Thioredoxin-fold protein as a physiological reductant of human vitamin K epoxide reductase complex 1 (hVKORC1). (A) Oxidation of the two cysteine residues in the CX<sub>1</sub>X<sub>2</sub>C motif of Trx-fold proteins forms a disulphide bond, a process associated with the loss of two hydrogen atoms and, hence, two electrons (top). Mechanism of disulphide exchange between Trx and a target (bottom). The *H*-donor enzyme and a target are coloured in blue and green, respectively. (B) The Trx-fold is illustrated using the X-ray structure of human PDI deposited in PDB [5] (PDB ID: 4ekz). The protein is shown as red ribbons, with two cysteine residues from the CX<sub>1</sub>X<sub>2</sub>C motif as yellow balls. The four  $\alpha$ -helices (in red), five

$\beta$ -strands (in yellow) and eight loops (in green) are numbered. (C) Comparison of the sequences of Trx-fold proteins ERp18, PDI, Tmx1 and Tmx4. Sequences were aligned on ERp18, having the most elongated sequence with ESPript3 (<http://esript.ibcp.fr/>). The solution with the best score is shown. The residues are coloured according to the consensus values: red indicates strict identity or similarity, while nonconserved residues are in black. Blue highlights the CX<sub>1</sub>X<sub>2</sub>C motif. (D) Ribbon diagram of the 3D human VKORC1 model in its inactive state showed in two orthogonal projections. The L-loop is shown in the colour teal, while disulphide bridges formed by cysteine residues C43–C51 and C132–C135 are drawn as yellow sticks. The transmembrane helices (TM) are numbered as in [6]. (E) The structure of VKOR from *Synechococcus sp* (bVKOR; ID PDB: 4nv5) is visualised using ribbons. The structural fragments that have sequences most similar to hVKORC1 and the Trx-like domain are shown in dark grey and light blue, respectively. The disulphide bridges formed by cysteine residues in Trx-like and VKOR-like domains are drawn as yellow sticks.

The thioredoxin fold is the most common structure found in thiol oxidoreductases; it has been carefully described in [7]. It is illustrated with the crystallographic structure of human protein disulphide isomerase (PDI; Figure 1B), which is the best-characterised enzyme that assists in the process of oxidative folding [8,9]. The PDI structure consists of a central five-stranded propeller with four flanking  $\alpha$ -helices, an architecture that contains extra regions compared to the classical thioredoxin fold (a four-stranded  $\beta$ -sheet with three  $\alpha$ -helices formed by about 80 residues).

The dithiol/disulphide group in the CX<sub>1</sub>X<sub>2</sub>C motif, which is located at the head of the  $\alpha$ H2 helix, protrudes from the protein surface and is exposed to a solvent. Such a spatial arrangement of the CX<sub>1</sub>X<sub>2</sub>C motif is probably to ensure the full accessibility of the first cysteine, which is required to react with the cysteine residue of a target to accomplish redox processes. It has been reported that the reactive thiolate of this first cysteine can be stabilised by the positive dipole at the head of the  $\alpha$ H2 helix and by a network of hydrogen (*H*) bonds that are formed between the thiolate and neighbouring residues presented by the helix-turn structure [10].

In the present study, the focus is on the Trx's function as a physiological reductant (*H*-donor) of vitamin K epoxide reductase complex 1 (VKORC1). VKORC1 is an endoplasmic reticulum-resident transmembrane protein that is responsible for the activation of vitamin K-dependent proteins, and it is involved in several vital physiological and homeostasis processes [11]. VKOR is the target of oral anticoagulants like warfarin, which dampens coagulation by limiting the supply of vitamin K. Its functional role is a catalyst in the reduction of vitamin K, requiring cooperation with a redox partner that delivers reducing equivalents. A particularly interesting problem is the enzymatic activation of hVKORC1 by the thiol–disulphide exchange. This process involves “molecular recognition” at the highest level required for proton-transfer reactions.

Recently, 3D models of human VKORC1 (hVKORC1) have been reported along with functionally related enzymatic states [6]. The models that were generated for the metastable states of hVKORC1 and their validation through *in silico* and *in vitro* screening have led to a conceptually plausible mechanism for enzymatic reactions based on a sequence array of hVKORC1-activated states involved in vitamin K transformation. These results suggest several additional questions, the most important being the real enzymatic machinery of hVKORC1 and its activation. Which Trx-fold protein is a specific proton donor of hVKORC1? What are the factors controlling the specificity of hVKORC1 recognition by the Trx protein? What is the exact role of thioredoxin(s) in initiating hVKORC1 reduction?

Since the physiological reductant of hVKORC1 has not yet been identified, initial exploration was made of four human redoxin proteins, namely, protein disulphide isomerase (PDI), endoplasmic reticulum oxidoreductase (ERp18), thioredoxin-related transmembrane protein 1 (Tmx1) and thioredoxin-related transmembrane protein 4 (Tmx4), reported as the most probable *H*-donors of VKOR [12,13]. These proteins have distinct compositions for the active site CX<sub>1</sub>X<sub>2</sub>C-CGHC in PDI, CGAC in ERp18, CPAC in Tmx1 and CPSC in Tmx4—and they show broad but distinct substrate specificity. The nature of this specificity is the main focus of this work. In order to evaluate the one most likely to reduce hVKORC1, a detailed comparison of these redoxins was first provided at different levels

of the protein's organisation—sequence, secondary and tertiary structure, intrinsic dynamics, intraprotein interactions governing structural and conformation properties, and composition of the surface exposed to the targets. Second, hVKORC1 in its native state, in which two pairs of cysteine residues form two disulphide bridges (Figure 1D), was studied as a target of Trx-fold proteins in order to identify the anchor site(s) that enable it to recognise/bind its Trx effector. Finally, modelling of the complex formed by hVKORC1 and PDI, which was suggested as the most probable partner of VKORC1, was carried out using the PDI fragments predicted to be “interacting” as a guide and the VKOR structure from *Synechococcus sp* (bVKOR; PDB ID: 4nv5; Figure 1E; see [14]) as an initial reference. The model of the molecular noncovalent complex formed by PDI and hVKORC1 (PDI–hVKORC1) is proposed as the first useful human precursor for the probing of thiol–disulphide exchange reactions between redoxins as an *H*-donor and hVKORC1 as a substrate.

This study principally leans on molecular dynamics (MD) simulation of the chosen Trx-fold proteins in the reduced state, of human VKORC1 in the inactive (oxidised) state, and on the modelling of the molecular noncovalently bound complex formed by hVKORC1 and PDI. It is suggested that a careful analysis of the simulation data will deliver quantitative and qualitative metrics to shed light on the following questions: (i) Are the 1D, 2D and 3D properties and the dynamic features good indicators for the prediction of the protein fragments participating in hVKORC1 recognition by a Trx? (ii) From the in-silico study of proteins, is it possible to predict which of them is the most likely partner of VKORC1? (iii) How do the predicted results correspond to a model of the complex formed by VKORC1 and its possible partner?

A central goal of this study is to understand, at the atomistic level, the recognition mechanisms between Trx and hVKORC1 (a process preceding the electrons' transfer reaction) and, thereby, identify shared vulnerable sites that can be targeted with anti-hVKORC1 or anti-Trx therapeutics.

## 2. Results

### 2.1. The Trx-Fold Proteins as Possible Partners of VKORC1

#### 2.1.1. Sequences and Structural Data

Structures of PDI (PDB ID: 4ekz; [9]), ERp18 (PDB ID: 1sen; [15]) and Tmx1 (PDB ID: 1x5e; [5]) were used to extract the coordinates of a domain containing the CX<sub>1</sub>X<sub>2</sub>C motif (Table S1, Figure S1). This domain was chosen for the study of all proteins because ERp18, Tmx1 and Tmx4 proteins are only constituted of one Trx-fold domain **a**. The sequences of the four selected Trx proteins show a low identity/similarity (Figure 1C, Table S2) along with the best scores for Tmx1 and Tmx4 (47/68%). The ERp18 sequence differs most from those of PDI, Tmx1 and Tmx4 (23/38%, 15/23% and 15/23%, respectively). A 3D model of Tmx4 was built from Q9H1E5 (<https://www.uniprot.org/uniprot/>), with the Tmx1 structure as a template.

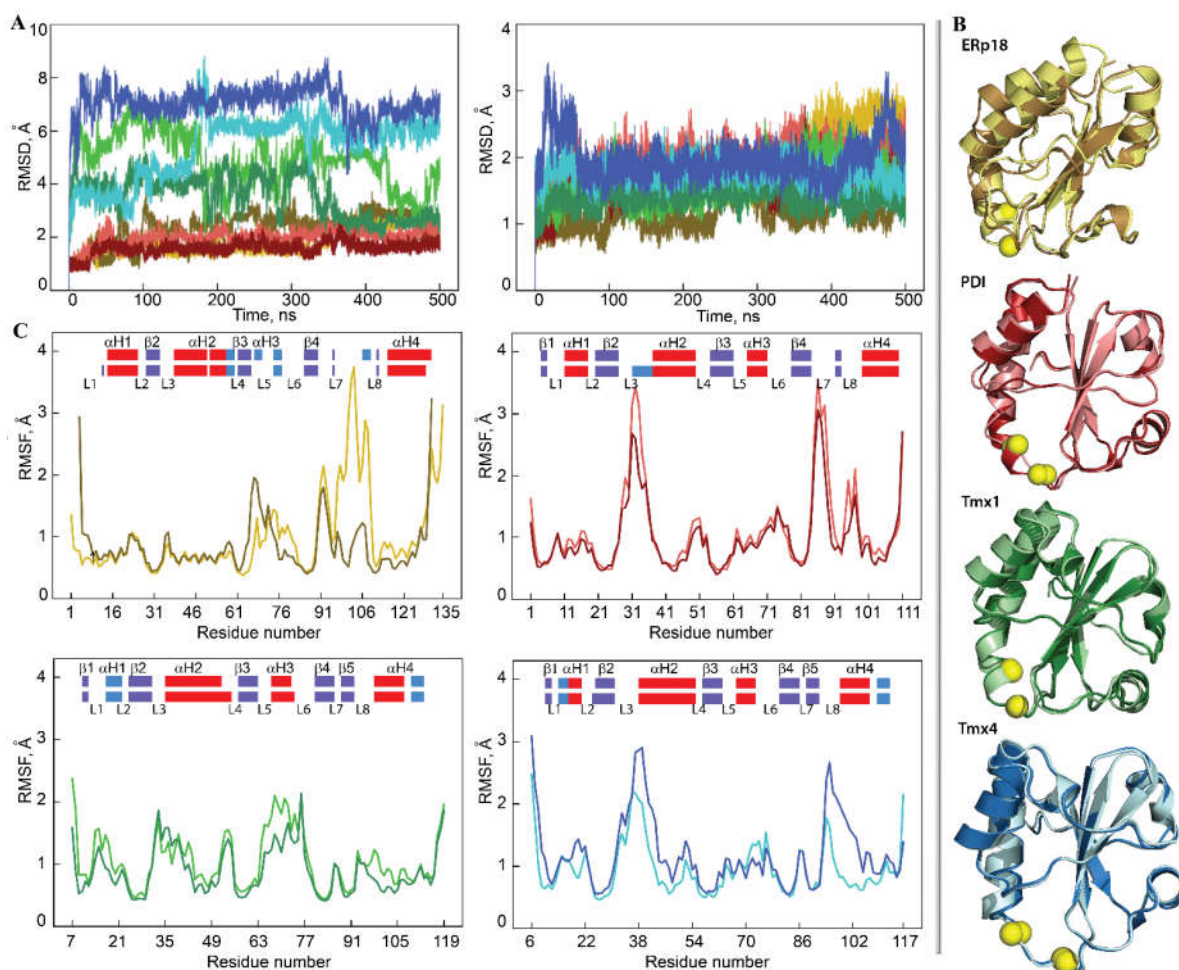
The ERp18, PDI and Tmx1 empirical structures and the Tmx4 homology model were optimised (when necessary) to obtain a CX<sub>1</sub>X<sub>2</sub>C motif in the reduced state. These were then used for the conventional MD simulations (two 500-ns trajectories for each protein), running under strictly identical conditions.

#### 2.1.2. General Characterisation of Trx-Fold Proteins Using MD Simulations

The global stability of each Trx-fold protein over the course of a simulation was estimated using the root mean square deviation (RMSD) that showed (i) similar behaviour for the same protein among both MD replicas and (ii) significant disparity between the different proteins (Figure 2A). Comparable RMSDs for PDI over each replica and between replicas characterise a highly stable protein structure during the simulation. Similar to PDI, the RMSD values for ERp18, Tmx1 and Tmx4 varied within a narrow range after elimination of the largest amplitude N/C-terminal residues. This demonstrates the good

structural stability of each Trx, which is a quality that is typical of well-organised folded regular proteins.

Indeed, in all studied Trx-fold proteins, the properly ordered secondary structures (SS or 2D structure) were shown to be long-lived  $\alpha$ -helices and  $\beta$ -strands. These ordered structures are interconnected by coiled linkers to form a stable globular 3D arrangement that is described as a four- or five-stranded antiparallel  $\beta$ -sheet sandwiched between four  $\alpha$ -helix-bundle structures, which is an archetypical fold of the Trx family of proteins (Figure 2B). Similar to the RMSDs, the root mean square fluctuations (RMSFs) agree well between the pair of replicas for protein (Figures 2C and S2). The most pronounced difference in RMSFs between the two replicas is only observed in ERp18, in which  $\beta$ 5 is partially unfolded and the L7 and L8 loops are joined together, resulting in large fluctuations.



**Figure 2.** Characterisation of the MD simulations for the four Trx-fold proteins ERp18, PDI, Tmx1 and Tmx4. (A) RMSDs from the initial coordinates computed for all  $\text{C}\alpha$ -atoms (right) in each protein after fitting to initial conformation. (B) The superimposed average structures of each protein over replicas 1 and 2. Cysteine residues are shown as yellow balls. RMSD values of 0.5, 0.4, 0.3 and 0.4 Å in ERp18, PDI, Tmx1 and Tmx4, respectively. (C) RMSFs computed for the  $\text{C}\alpha$ -atoms using RMSF amplitude values less than 4 Å for the MD conformation of each protein after fitting to the initial conformation. Highly fluctuating residues (3, 6 and 5 in ERp18, Tmx1 and Tmx4, respectively) were excluded from the RMSD computation. In the insert, the secondary structures— $\alpha$ H- (red),  $3_{10}$ -helices (light blue) and  $\beta$ -strands (dark blue)—were assigned for a mean conformation of every MD trajectory, 1 (top) and 2 (bottom), of each protein and were labelled as in the crystallographic structure of human PDI. (A–C) Proteins are distinguished by colour (first/second replicas): ERp18 (yellow/brown), PDI (light/dark red), Tmx1 (light/dark green) and Tmx4 (light/dark blue). The numbering of the residues in each Trx-fold protein is arbitrary and starts from the first amino acid in the 3D model.

Further characterisation of each protein and a comparison between the proteins is frequently completed by the observations obtained for a randomly chosen single trajectory or concatenated data. This is because the RMSDs and RMSFs in both replicas of each protein display comparable profiles and a similar range of values, and the 2D and 3D structures of each protein are perfectly matched (the RMSD values between the average structures of replicas 1 and 2 are less than 0.5 Å; Figure 2). The exception is PDI, in which the  $\alpha$ H2-helix showed a different length over two replicas that was caused by the distant fold of its *N*-terminal.

How different are the 2D and 3D structures for the four proteins? The organised secondary structures,  $\alpha$ - and  $3_{10}$ -helices and  $\beta$ -strands, involve 55%, 60%, 60% and 56% of the residues in ERp18, PDI, Tmx1 and Tmx4, respectively, where the helical and  $\beta$ -strand fold portions vary from 36% to 42% and from 13% to 22% of total folding, respectively. Although all ordered 2D structures (helices and strands) are generally conserved across the studied proteins, their positions, lengths and qualities (e.g.,  $\alpha$ - or  $3_{10}$ -helix) are slightly different (Figures 2 and S2).

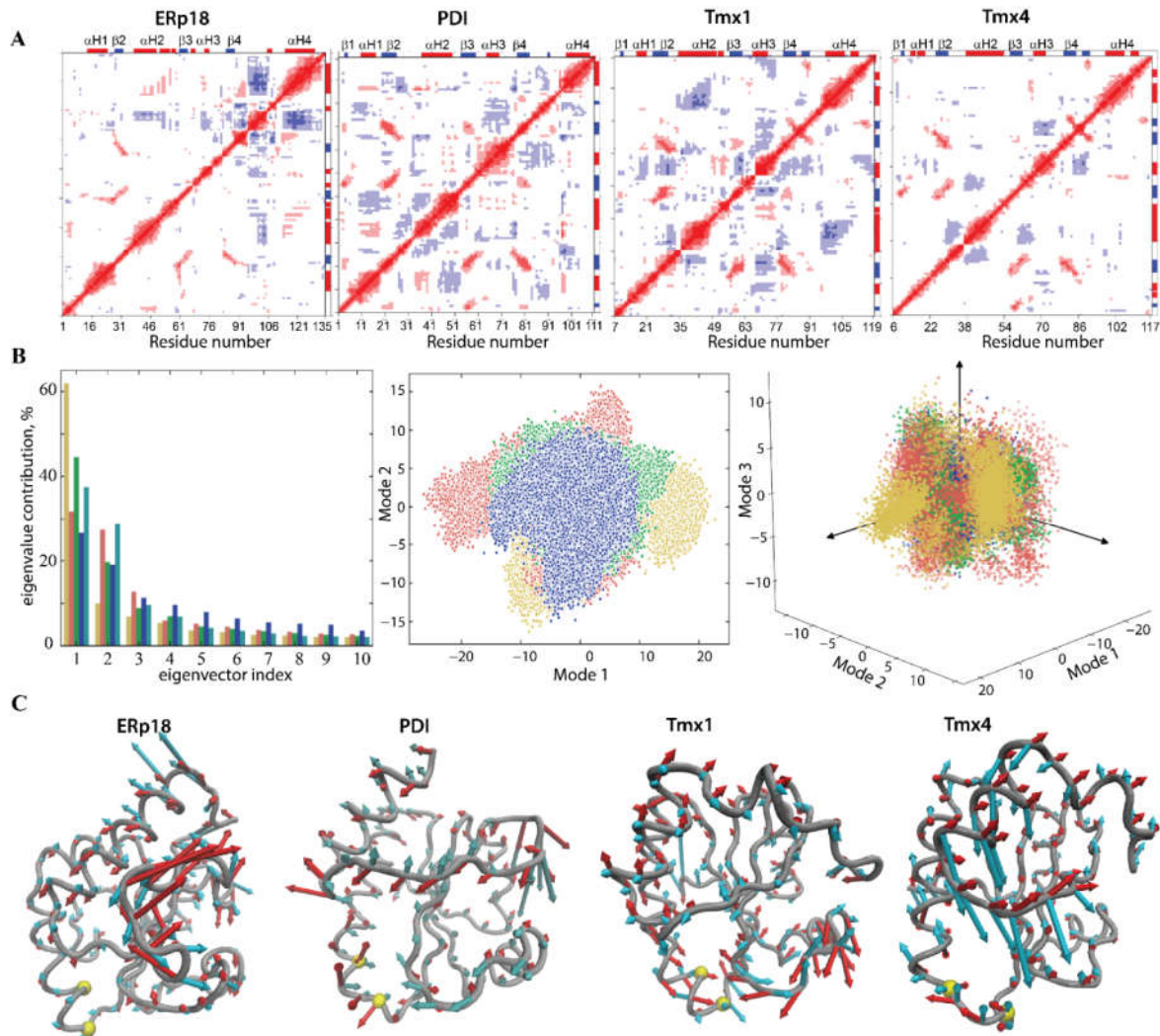
The helical fold of each protein is represented by  $\alpha$ -helices of different length (of 7–18 residues) and by  $3_{10}$ -helices that consist of 3–4 residues. H1, which is a long-lived  $\alpha$ -helix in ERp18 and PDI, is transient and converts between  $\alpha$ - and  $3_{10}$ -helices in Tmx1 and Tmx4. H2, which is the longest  $\alpha$ -helix (14–18 residues) that contains the CX<sub>1</sub>X<sub>2</sub>C motif at its *N*-extremity, is generally conserved in all proteins; however, it may be partially split into two helices (ERp18) or reduced in size (PDI). The folding of the CX<sub>1</sub>X<sub>2</sub>C motif is different in the four proteins, and this represents a part of the regular  $\alpha$ -helix (ERp18 and Tmx1), a transient helix fluctuating between  $\alpha$ - and  $3_{10}$ -helices or/and a turn (PDI) and a coiled structure (Tmx4). In ERp18, H3 consists of a pair of short  $3_{10}$ -helices, while in the other proteins, it is a single and stable  $\alpha$ -helix. H4 is a long and stable  $\alpha$ -helix in ERp18 and PDI, while in Tmx1 and Tmx4, it is folded as a shorter  $\alpha$ -helix and is joined to a  $3_{10}$ -helix.

This analysis illustrates that although the studied proteins share a similar structure, their folding is noticeably different; this reflects their sequence-dependent character.

Additionally, the atomistic RMS fluctuations of the studied proteins show (i) minimal RMSF values for all  $\beta$ -strands forming the antiparallel  $\beta$ -sheet in all proteins, while the helices may have discernable fluctuations (e.g.,  $\alpha$ H2 and  $\alpha$ H3 in Tmx1, and  $\alpha$ H2 in Tmx4), and, as was expected, (ii) strong differences in the fluctuations of the coiled linkers (Figure 2C). These linkers, which interconnect the core  $\beta$ -strands and the surrounding  $\alpha$ -helices, are the most variable elements in the studied proteins in terms of sequence composition, length and conformation. It is also noted that moderate (in the order of 1.5–2.5 Å) but systematically observed fluctuations of fragment L5- $\alpha$ H3-L6 arose in all studied proteins. This fragment is structurally adjacent to the CX<sub>1</sub>X<sub>2</sub>C motif and may play a role in thiol-disulphide exchange reactions.

### 2.1.3. Intrinsic Motion and Its Interdependence on Trx-Folded Proteins

Since a protein's dynamics influence its functional properties, intrinsic motions of Trx-fold proteins were compared. First, a cross-correlation map was computed for all C $\alpha$ -atom pairs of each protein (Figure 3A). The positively correlated motion of  $\beta$ 2-,  $\beta$ 3- and  $\beta$ 4-strands, which was observed in each studied protein, reflects their concerted movement in the  $\beta$ -barrel. To equilibrate structural stability, the other fragments in Trx-fold proteins display a motion that tends to correlate negatively. As such, in ERp18, in addition to the  $\beta$ -barrel coupled motion, the structural moieties with the strongest correlation are L7 and  $\alpha$ H4. In PDI, a regular fractal-like pattern shows the correlated motion of  $\alpha$ H1 with  $\alpha$ H2 and L7 and  $\alpha$ H3 with L7 and L8. In Tmx1, the coupled motion is observed between the  $\alpha$ H2-helix and the  $\alpha$ H4-helix and between the  $\beta$ 2-stand and the  $\alpha$ H3-helix. Tmx4 demonstrates correlated motions between the  $\alpha$ H2-helix and the  $\beta$ 3-strand and between the  $\alpha$ H3-helix and  $\beta$ 4/ $\beta$ 5-strands.



**Figure 3.** Intrinsic motion in the Trx-folded proteins and its interdependence. (A) Inter-residue cross-correlation maps computed for the  $\text{C}\alpha$ -atom pairs of ERp18, PDI, Tmx1 and Tmx4 after the fitting procedure. Secondary structure projected onto the protein sequences ( $\alpha$ -helix/ $\beta$ -strand in red/blue) is shown at the border of matrices. Correlated (positive) and anticorrelated (negative) motions between  $\text{C}\alpha$ -atom pairs are shown as a red–blue gradient. (B) The PCA modes calculated for each protein after least-square fitting of the MD conformations to the average conformation as a reference. The bar chart gives the eigenvalue spectra in descending order for the first 10 modes (left). Projection of ERp18, PDI, Tmx1 and Tmx4 MD conformations with the principal component (PC) in 2D (middle) and 3D subspaces (right). MD conformations were taken every 100 ps (2D) and 10 ps (3D). The protein data is referenced by colour—ERp18 (dark yellow), PDI (brown), Tmx1 (green) and Tmx4 (dark blue and light blue for two replicas). (C) Collective motions characterised by the first two PCA modes. Atomic components in PCA modes 1–2 are drawn as red (1st mode) and cyan (2nd mode) arrows projected on a tube representation of each protein. For clarity, only motion with an amplitude  $\geq 2$  Å is represented. Cysteine residues are shown as yellow balls. All computations were performed on the  $\text{C}\alpha$ -atoms with RMSF fluctuations less than 4 Å for each protein after fitting on the initial conformation.

The collective motion of Trx-fold proteins and its impact on their conformational properties was studied using a principal component analysis (PCA). The principal components (PCs) were determined, and the MD conformations for each protein were projected onto the PC subspace formed by the first two and first three eigenvectors. This indicated that Tmx1 (green) and Tmx4 (blue) conformations were grouped in a unique compact region for each protein, and these regions were perfectly superimposed for both proteins, while the conformations of PDI (red) and ERp18 (yellow) were trapped in two or



three separate regions that were located in a slightly enlarged space (Figure 3B). Randomly selected conformations from the distinct regions in the projection of the first two PCA modes showed that their conformational difference is mainly associated with a motion that leads to a slight skew of the H5-helix and displacement of the H3-helix in ERp18 and a disparity in the H2-helix length in PDI.

From the ten calculated PCA modes describing ~95% of total backbone fluctuations of each Trx-fold protein, the first two most dominant modes were used to illustrate ample collective movements qualitatively (Figure 3C). The PCA modes of the Trx-fold proteins reveal the essential mobility of their fragments, which is either similar in the four proteins or has different features for a given protein. For instance, in ERp18, the greatest mobility is observed for L7 and L8 loops that are joined together due to the unfolding of the  $\beta$ 5-strand. In PDI and Tmx4, the L7 and L8 loops are well separated by the  $\beta$ 5-strand, but each of them shows the coupled motion of a large amplitude. Uniquely, in Tmx1, the  $\alpha$ H3-helix and its joint L5 loop display a high amplitude motion. In PDI, Trx1 and Trx4, the collective motion of the H2-helix and joint L3 loop is comparable in amplitude but differs in directions.

#### 2.1.4. Focus on the Region of Trx-Fold Proteins Potentially Involved in Target Recognition and/or Electron Transfer Reaction

To compare the four Trx-like proteins regarded as probable functional effectors of hVKORC1, the focus was on two fragments that may be involved in target recognition and/or electron transfer reaction. The first fragment, **F1**, comprises L3 and an *N*-extremity of  $\alpha$ H2-helix that includes the CX<sub>1</sub>X<sub>2</sub>C motif and the second, **F2**, which is structurally adjacent to the CX<sub>1</sub>X<sub>2</sub>C motif, is composed of L5- $\alpha$ H3-L6. Both fragments form a frontal region that is exposed to the solvent in each Trx-fold protein, which may interact directly with a target during the electron-exchange process, similar to a bacterial protein containing a Trx-fold domain that is covalently bonded to VKOR (Figure 1E). The delimiting of these two regions is very approximate because the sequences and 2D structures of the studied proteins show significant differences. To have segments of a comparable length in different proteins, the boundaries of fragments were chosen so that their lengths were equal (17 residues; Figure 4).

The **F1** region in PDI, Tmx1 and Tmx4 is similarly initiated by tyrosine, which is the residue reported to be a breaker of secondary structures, while in ERp18, the role of “a breaker” is given to histidine, followed by lysine, which are amino acids that are more likely to be present in disordered regions [16,17]. The following residues of the L3-loop, a pair of hydrophobic residues (APs), are perfectly conserved in PDI, Tmx1 and Tmx4, while in ERp18, these positions are occupied by positively charged and polar residues (KSs). Furthermore, the specific CX<sub>1</sub>X<sub>2</sub>C motif for each studied protein is preceded by tryptophan (W), which is a highly conserved residue in the four proteins. Tryptophan is an amphipathic residue that, similar to tyrosine, is often found at the surface of proteins and is sometimes classified as polar.

It is suggested that the **F1** region of Trx-folded proteins, which contains the CX<sub>1</sub>X<sub>2</sub>C motif, contributes to redox reactions rather than target recognition. Nevertheless, a double action of the **F1** fragment as both redox agent and recognition platform for a target has not been excluded.

The second surface region of Trx-fold proteins, **F2**, which is in the proximity of the CX<sub>1</sub>X<sub>2</sub>C motif, consists of the  $\alpha$ H3-helix and its two adjacent loops, L5 and L6. This fragment shows a negligible or no similarity/identity between the four proteins and, thus, may convey the highest degree of specificity in the discrimination/recognition of a partner. The most critical difference consists of the sequence composition of the L5 loop and the  $\alpha$ H3-helix and the length of the  $\alpha$ H3-helix. In ERp18, a set of five negatively charged amino acids (EDEEEs), which are positioned on the L5 loop and the  $\alpha$ H3-helix, are separated by proline (P) and lysine (K) from the other three negatively charged amino acids (DEDs). This promotes a breakup of the H3-helix into two smaller  $3_{10}$ -helices. In the other proteins,



As the  $\alpha$ H3-helix is moving considerably in Tmx1 and moderately in the other proteins (Figure 3C), we suggest that the  $\alpha$ H3-helix can adapt its orientation to get the best position with respect to the target and, together with its joint loops, L5 and L6, is able to build the recognition (docking) site(s) for target accommodation. The **F2** region is the most dissimilar fragment in the studied proteins, and it has a sequence composed of hydrophobic stretches folded into a polar lipid environment. **F2** also contains polar and charged residues required for stretches of sequence that are exposed to a solvent in cytosolic or extracellular environments [18]. Therefore, **F2**, which is positioned in the proximity of the CX<sub>1</sub>X<sub>2</sub>C motif, is a fragment of a Trx-fold protein that can contribute to VKORC1 recognition.

#### 2.1.5. Geometry of the CX<sub>1</sub>X<sub>2</sub>C Motif

Focusing on the CX<sub>1</sub>X<sub>2</sub>C motif, a key agent in thiol–disulphide exchange reactions, its geometry was characterised in each Trx-fold protein. It was observed that structurally, the CX<sub>1</sub>X<sub>2</sub>C motif constitutes either a part of the  $\alpha$ H2-helix (in ERp18 and Trx1), which is transient in PDI, or an extension of the L3 loop (in Tmx4). Both cysteine residues that are located on a coil are largely exposed to the solvent, whereas only one cysteine is exposed in the folded CX<sub>1</sub>X<sub>2</sub>C, while the other cysteine is buried in the protein chain.

Surprisingly, the folding of the CX<sub>1</sub>X<sub>2</sub>C (CGAC) motif in the calculated conformations (MD simulation) of ERp18 is coherent with those observed in the experimentally determined structure (PDB ID: 1sen), despite the different protein states, namely, reduced (MD simulation), with two protonated thiol groups, and oxidised (X-ray analysis), in which two deprotonated thiol groups form a disulphide bridge. In both protein states studied by the two different methods, the first cysteine from the CGAC motif is the N-cap residue (the last nonhelical residue) of the  $\alpha$ -H2 helix.

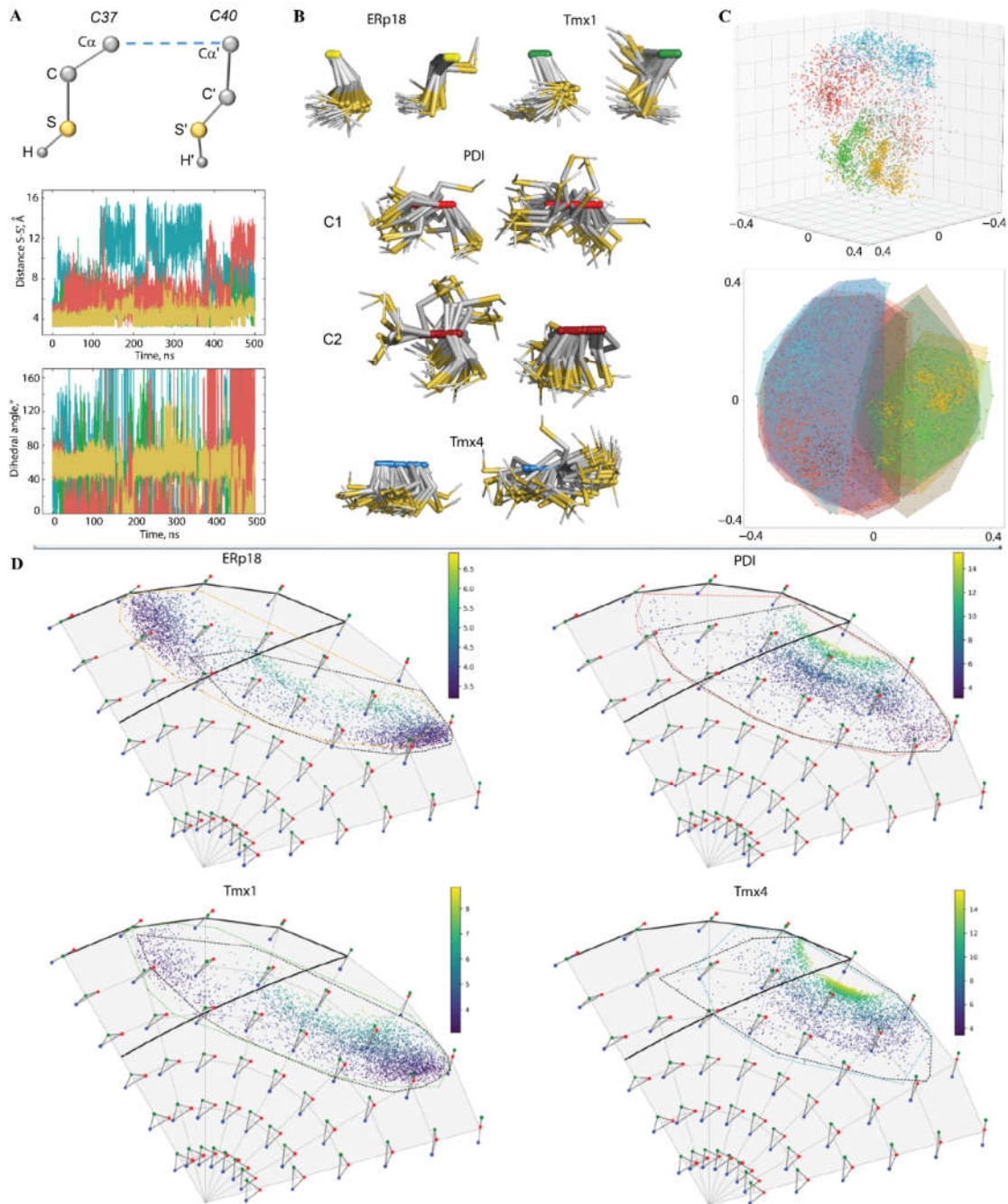
The second unexpected observation is connected to the different folding of the CX<sub>1</sub>X<sub>2</sub>C (CGHC) motif in the calculated (MD simulation) and empirical structures (X-ray analysis) of PDI when studied in the same state (reduced). Indeed, the CX<sub>1</sub>X<sub>2</sub>C motif in the crystallographic structure of PDI was reported as folded, with the C37 positioned at the cap of the  $\alpha$ -H2 helix (PDB ID: 4ekz), while in the MD conformations, the structure of this motif is transient and alternated between the helical fold ( $\alpha$ - or  $3_{10}$ -helices) and the turn/coiled structure, demonstrating high conformational plasticity.

The folding of the CX<sub>1</sub>X<sub>2</sub>C motif in Tmx1 (CPAC) in the MD conformations and the NMR structures (PDB ID: 1x5e; both are in a reduced state) is equivalent, with the first cysteine as an N-cap residue of the downstream  $\alpha$ -H2 helix, similar to ERp18. In Tmx4, a protein with the most similar sequence to Tmx1, the CX<sub>1</sub>X<sub>2</sub>C motif (CPSC) demonstrates a coiled structure. In these two proteins, the conserved proline constitutes the characteristic CPX<sub>2</sub>C motif, and the observed structural differences may be connected either to the X<sub>2</sub> residue or to the long-distance structural effects.

The geometry of the CX<sub>1</sub>X<sub>2</sub>C motif was described by two metrics: a distance S...S' between the protonated sulphur atoms and a dihedral angle S–C $\alpha$ –C $\alpha'$ –S' (Figures 5A and S3). In proteins ERp18 and Tmx1, the mean value (mv) of these parameters (4 Å and 60°, respectively) describe a synclinal configuration (Prelog–Klyne nomenclature) of the sulphur atoms that is well-conserved over the MD simulations. Nevertheless, a rare but not-negligible number of Tmx1 conformations revealed a syn-periplanar or anticlinal orientation of sulphur atoms that promoted a slight increase in the S...S distance. Such restrained geometry of the CX<sub>1</sub>X<sub>2</sub>C motif in ERp18 and Tmx1 is apparently related to its location on the well-folded  $\alpha$ H2-helix. By contrast, the CX<sub>1</sub>X<sub>2</sub>C motif located on a coiled L3 loop in Tmx4 stimulates a highly divergent orientation of sulphur atoms, running from syn-periplanar configuration to an antiperiplanar configuration, as was evidenced by a large variation in the dihedral angle S–C $\alpha$ –C $\alpha'$ –S. The measured metrics, distance S...S and dihedral angle in PDI had values close to those in ERp18 and Tmx1. Nevertheless, a large number of conformations displayed a strongly variant geometry, which is similar to Tmx4. Such richness in PDI conformations corresponds to the transient structure of the N-

terminal of the H2-helix, conversed between the helical fold ( $\alpha$ - and  $3_{10}$ -helices) and turn structure.

To better characterise the dynamical behaviour of the CX<sub>1</sub>X<sub>2</sub>C motif over two trajectories for each protein and to compare the different proteins, 3D skeletal shape trajectories of the motif's atoms were described in Kendall's shape space [19]. For a given integer  $k$ , Kendall's shape space is the manifold of dimension  $3k - 7$  dimension of all possible configurations of  $k$  atoms in  $R^3$  considered up to a rigid transformation (translation, rotation and scaling). It has a Riemannian structure with a computable geodesic distance. The framework allows the use of geometric statistics and dimension reduction methods like multidimensional scaling (MDS) to analyse the shape trajectories [20]. These methods offer various ways of visualising all the data in a common space, summarizing them with a reduced number of variables and comparing them to each other. A tetrahedron, defined for the S- and C $\alpha$ -atoms of two cysteine residues, C37 and C40, was extracted from conformations over MD simulations (Figure 5C). The four proteins can be condensed in two major groups that are weakly overlapping (clearly visible in the 3D view): ERp18 and TMX1 on the one hand, PDI and TMX4 on the other hand, the latter group displaying a larger shape variation.



**Figure 5.** The CX<sub>i</sub>X<sub>2</sub>C motif geometries for ERp18, PDI, Tmx1 and Tmx4. **(A)** Geometry of the CX<sub>i</sub>X<sub>2</sub>C motif (left) is described by distance S...S' (middle) and dihedral angle (right), determined as an absolute value of the pseudo torsion angle S-Cα(C37)-Cα'(C40)-S'. Only one replica 2 is shown. **(B)** Superposition of the thiol groups (Cα-C-S-H) from the CX<sub>i</sub>X<sub>2</sub>C motif of each protein is shown for either only one MD trajectory (ERP18, Tmx1 and Tmx4) or for both (PDI). Samples were taken for each 100-ns frame. **(C)** Multidimensional scaling (MDS) in 2D and 3D on the set of S-C-C-S tetrahedrons. Embedded points have been coloured according to the partner and replica they belong to. **(D)** Evolution of the shape of the triangles S-H...S on Kendall's disk of 3D triangles; each data point is coloured according to the S...S' distance. Representative triangles are regularly sampled on the disk. The thick black line delimits the area of conformations favouring H-bond interaction. The dashed areas are contouring subpopulations according to the S-atom being the H-donor.

This analysis is illustrated by the superposition of the thiol groups (C $\alpha$ -S-H) from the CX<sub>1</sub>X<sub>2</sub>C motif of the MD conformations for each protein (Figure 5B). The orientation of the thiol groups favours H-bond interaction (S-H $\cdots$ S) only in ERp18 and some Tmx1 conformations. In PDI, both the thiol groups are shown to have the most variant orientation within a group and between groups, which reflects their high mobility.

The H-bond between the sulphur atoms of each cysteine is characterised for two cases: (1) the S-atom from C37 is the H-donor to the S-atom of C40, and (2) the S-atom from C40 is the H-donor to the S-atom of C37 (Figures 5D and S4). Monitoring of the geometry of S-H $\cdots$ S (1) showed a very low probability (0.1–0.9%) of such an interaction in all proteins. Contact (2) has a probability of 72% in Erp18 and 27% in Tmx1. Analysis of the contact metrics (distance S $\cdots$ S and angle at H-atom) indicated that a typical S-H $\cdots$ S H-bond is slightly stronger in Tmx1 than ERp18. Such an H-bond was not observed in the other studied proteins.

As expected, the S-H $\cdots$ S H-bond does not influence the folding of the CX<sub>1</sub>X<sub>2</sub>C motif (Figure S5). For instance, this H-bond is observed in conformations from clusters C1, C2 and C4 of ERp18, and it is absent in the others (C3 and C5), although the CX<sub>1</sub>X<sub>2</sub>C motif is well folded in both cases. Interestingly, both thiol groups do not contribute to H-bond interaction in the most prevalent PDI conformation with an unfolded CX<sub>1</sub>X<sub>2</sub>C motif, but K41, which is next to the C40 residue, is H-bound to H39 and L43. In the folded CX<sub>1</sub>X<sub>2</sub>C motif of PDI, C37 is in contact with P35 through the H-bond formed by the main chain atoms. Apparently, this interaction contributes to the stabilisation of the PDI conformation in the folded state, but it is not the unique factor that leads to such a structure. Similar but not-equivalent H-bonds are observed in the well-structured Tmx1 motif and the fully unfolded Tmx4 motif.

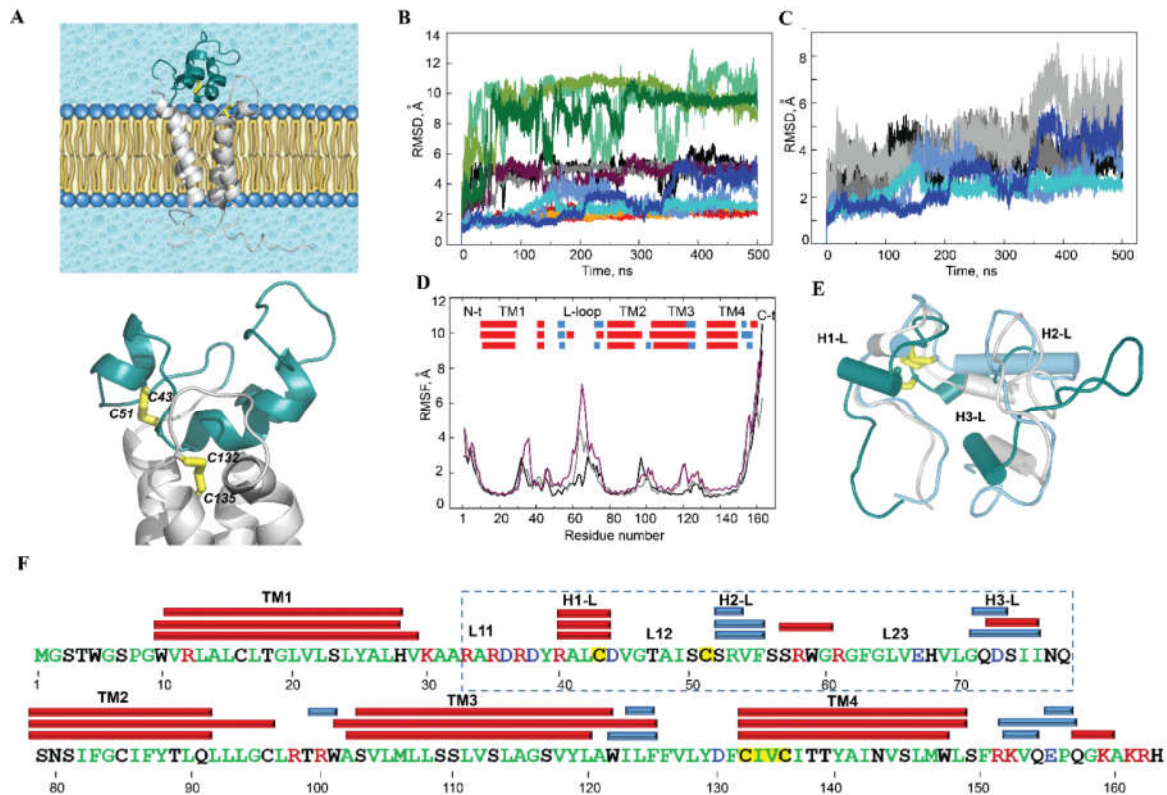
Structure organisation of the CX<sub>1</sub>X<sub>2</sub>C motif strongly influences their reactivity, affecting such properties as their accessibility and protonation state (i.e., pK<sub>a</sub>) [21]. Functional analyses of each cysteine in the consensus CX<sub>1</sub>X<sub>2</sub>C motif demonstrated that N-terminal cysteine is important for the formation of a transient S–S bond with the substrate, whereas C-terminal cysteine is involved in substrate release [22]. In proteins, specific hydrogen-bond donors and an electropositive local environment tend to lower the pK<sub>a</sub> by stabilising thiolate, and a hydrophobic environment or an electronegative local environment tends to raise the pK<sub>a</sub> by destabilising a negatively charged, as opposed to a neutral form, side chain [21,23,24].

## 2.2. Human VKORC1 Viewed as the Target of a Trx-Fold Protein

### 2.2.1. General Characterisation

Human VKORC1 is composed of two domains: the extended luminal loop (L-loop), which contains the cysteine residues that participate in the electron exchange between the redox enzyme and hVKORC1, and the transmembrane domain (TMD), which includes two other cysteine amino acids from the highly conserved CXXC active site that is essential for vitamin K quinone reduction [25,26]. Based on studies of bacterial VKOR homologues, it was proposed that the loop cysteines of hVKORC1 allow protons to be shuttled to the active-site cysteines [12,27].

Earlier, a four-helix transmembrane domain structural model of human VKORC1 in its four functional states was reported [6]. Here, the focus is on the inactive (oxidised) state of hVKORC1, in which two pairs of cysteine residues, C43–C51 and C132–C135, are covalently linked to form disulphide bridges S $\cdots$ S (Figure 1D). hVKORC1 was studied by MD simulations of the model that mimics the protein in its natural environment, i.e., hVKORC1 embedded in the membrane and surrounded by water molecules (Figure 6A). While the extended L-loop (R37–N77 aas) has demonstrated high conformational variability in the protonated forms of hVKORC1 [6], the inactive state of hVKORC1 was studied by repeated 500 ns MD simulations (replicas 1–3) using random initial velocities.



**Figure 6.** hVKORC1 in its inactive state and its conventional MD simulations. (A) 3D model of hVKORC1 in its inactive state; it was inserted into the membrane (top) and zoomed in on the L-loop (bottom). The L-loop is highlighted by the colour teal; disulphide bridges formed by cysteine residues C43–C51 and C132–C135 are drawn as yellow sticks. Transmembrane helices (TMs) are numbered as in [6]. (B,C) RMSDs computed for each MD trajectory (replicas 1–3) from initial coordinates (at  $t = 0$  ns, the same for all replicas) on the  $C\alpha$ -atoms of full-length hVKORC1 (in black, grey and rose brown), of the transmembrane domain (in orange, red and grenadine), of the L-loop (in clear aqua, bleu and navy) and of the N- and C-terminals (in teal, green and deep green) after fitting to the initial conformation of the respective fragment (B); of the L-loop (i) after fitting to its initial conformation (clear aqua, blue and navy blue) and (ii) after fitting of the protein coordinates to the initial conformation of the TMD (black, grey and silver) (C). (D) RMSFs computed for  $C\alpha$ -atoms of the MD conformations (replicas 1–3) after fitting to the initial conformation (at  $t = 0$  ns, the same for all replicas; in black, grey and rose brown). In the insert, the folded secondary structures,  $\alpha$ H- (red) and  $3_{10}$ -helices (blue), were assigned for a mean conformation of each MD trajectory. (E) Superimposition of the L-loop conformations picked from replica 3 at 150 (grey), 250 (light blue) and 375 ns (deep teal). (F) The hVKORC1 sequence (Q9BQB6) and the secondary structure assignment for a mean conformation over each MD trajectory. Residues are coloured according to their properties: positively and negatively charged residues are in red and blue, respectively; hydrophobic residues are in green; polar and amphipathic residues are in black; residues C43, C51 and the CX<sub>1</sub>X<sub>2</sub>C motif are highlighted by a yellow background.  $\alpha$ - and  $3_{10}$ -helices are shown as red and blue batons, respectively. Secondary structure labelling is shown above the VKORC1 sequence. The L-loop sequence is surrounded by dashed lines.

The RMSDs computed for the positions of all  $C\alpha$ -atoms relative to the initial structure ( $t = 0$  ns) showed comparable behaviour over the three MD trajectories, with a mean value (mv) of 5 Å (Figure 6B). The per-domain RMSDs showed that the N- and C-terminals are the fragments that contribute most to large RMSD values (up to 13 Å), while the TMD curves demonstrate a highly stable profile with the smallest RMSDs (2 Å). RMSDs computed for the  $C\alpha$ -atoms of the L-loop, after fitting to its initial conformation, showed alternated values, small or large, that were maintained over a large time scale (50–100 ns). The altered RMSD values, viewed as a set of well-defined slopes, indicate the possible conformational transitions in the L-loop. To check the suggested conformational transitions, MD conformations picked before and after each sudden RMSD change were compared. Three conformations of the L-loop that were chosen from replica 3 at  $t = 150, 250$

and 375 ns showed significant differences in the folding and orientation of the helices and the loops, which revealed structural and conformational transitions (Figure 6E).

The larger RMSD values computed for C $\alpha$ -atoms of the L-loop, after rigid alignment based on the initial conformation of the TMD compared to the RMSDs computed after rigid alignment based on the initial conformation of the L-loop, suggest the displacement of the L-loop from the TMD as a pseudorigid body (Figure 6C). The profile of the RMSF curves is similar in the three MD trajectories, with differences only in the amplitude of the RMS fluctuations of the highly flexible regions of hVKORC1, the N- and C-terminals and the extended L-loop (Figure 6D). The 2D and 3D structures of VKORC1 is generally conserved over the MD trajectories and shows a fully helical fold of the protein, with the four long-living extended (of 15–19 residues) transmembrane  $\alpha$ -helices, TM1–TM4, observed in the reduced forms of hVKORC1 [6] and the three short helices on the L-loop (Figure 6E,F).

### 2.2.2. The Luminal Loop of hVKORC1: Structure and Dynamics

Since the luminal loop (L-loop) is the fragment targeted by a Trx-fold protein, our focus is mainly on its intrinsic structural and dynamical properties and their connection with those of the transmembrane domain of hVKORC1.

L-loop folding, encompassing 30%, 36% and 22% of all residues in replicas 1–3, respectively, is presented by three small (3–4 residues) transient helices, H1-L, H2-L and H3-L, which are partially converted between the  $\alpha$ H- and  $3_{10}$ -helices (Figures 6E and S6). Despite the transient structure of helices, their positions on the sequence are well conserved. The L-loop helices are interconnected by coiled linkers, which, together with the linker joining the L-loop to TM1 from the transmembrane domain of hVKORC1, display RMSF values that suggest the high mobility of these loops (Figure 6D). H1-L, mainly folded as a regular  $\alpha$ -helix, contains C43 at its C-cap, which is linked covalently to C51, an N-cap residue of H2-L helix, forming the S...S bridge between two cysteines. Such covalent bonding significantly restricts the conformational mobility of this fragment. The large coiled linker connecting H2-L and H3-L helices is composed of hydrophobic residues, with the inserted charged and polar amino acids in the proximity of each helix (Figure 6F).

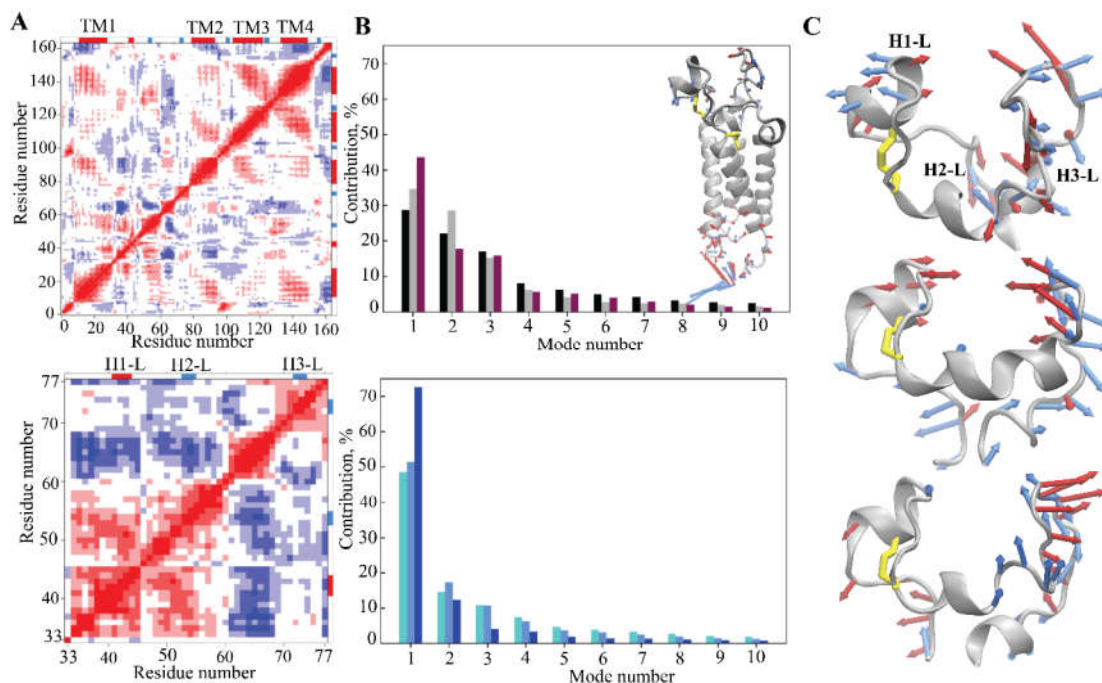
The intrinsic dynamics of hVKORC1 was first analysed with the cross-correlation matrix computed for the C $\alpha$ -atom pairs of the full-length protein and the L-loop. The C $\alpha$ –C $\alpha$  distance pairwise patterns demonstrate the coupled motions within each hVKORC1 domain, the L-loop and the TMD and between two structural domains (Figures 7A and S7). The regular pattern in the TMD reflects the correlated motion of the TM helices that is mainly associated with their collective drift, observed earlier in all metastable states of hVKORC1 [6]. The motion of the L-loop correlates with the movement of the linkers that connect the TM-helices and join the L-loop to the TMD.

The cross-correlations computed on only L-loop atoms display different maps in the three replicas, with either a fine-grained pattern (replicas 1 and 2) or a pattern composed of well-defined blocks of nearly equal size (replica 3), reflecting the highly coupled motion of the L-loop fragments consisting of 10–12 residues from the L-loop helices and their adjacent linkers. The difference in cross-correlation patterns is associated with the disparity of L-loop motion—small or medium in replicas 1 and 2 and broad in replica 3, as evidenced by RMSFs and PCA.

The collective motions of VKORC1, characterised by PCA, showed that ten modes describe ~80–90% of the total fluctuations of both the full-length VKORC1 and the L-loop (Figure 7B). Similar to the RMSF values, the first two PCA modes denote the great mobility of the terminal residues (N- and C-terminus) and the L-loop (Figure 7B insert). PCA analysis performed on only the C $\alpha$ -atoms of the L-loop showed that two first modes characterise most of the L-loop motion that displays the large-amplitude collective movements of the L-loop fragments—helices and adjacent-coiled linkers. The amplitude and direction of motion of the L-loop fragments differ in the three trajectories (Figure 7C), suggesting a



larger conformational space for the L-loop than was observed in each trajectory, probably larger than the total space of all trajectories. The first two modes in replicas 1 and 2 showed a highly coupled motion of the H1-L helix and the L23 linker in a scissors-like manner, while the collective motion in 3 mainly displays a displacement of L23, which is horizontal with respect to the rest of the L-loop and vertical with respect to the TMD (Figures 7C and S7).

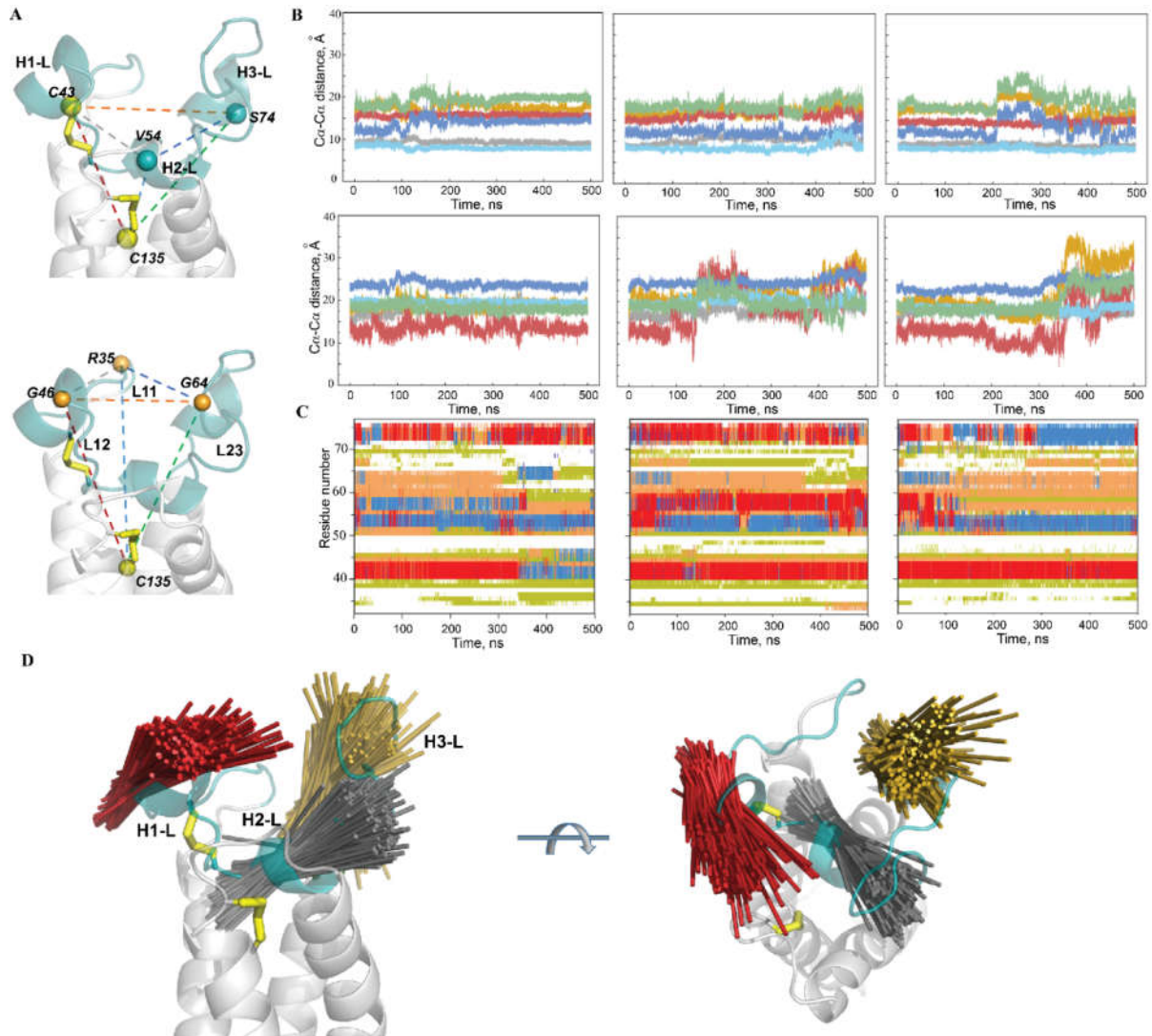


**Figure 7.** Intrinsic motion of hVKORC1 and its L-loop. (A) The inter-residue cross-correlation map computed for the  $\text{C}\alpha$ -atom pairs after fitting to the respective first conformation ( $t = 0$  ns) of the full-length hVKORC1 (top) and the L-loop (bottom) is shown for the three replicas. Correlated (positive) and anticorrelated (negative) motions between the  $\text{C}\alpha$ -atom pairs are shown as a red–blue gradient. (B) The PCA modes of the full-length hVKORC1 (top) and the L-loop (bottom), calculated for each MD trajectory after least-square fitting of the MD conformations to the average conformation of the respective domain as a reference. The bar plot gives the eigenvalue spectra in descending order for the first 10 modes. The data for replicas 1–3 are coloured black, grey and rose brown, respectively, while, for the full-length hVKORC1, the colouring is clear aqua, blue and navy blue for the L-loop. (C) Atomic components in the first PCA modes of the L-loop are drawn as red (1st mode) and blue (2nd mode) arrows projected onto the respective average structure from replicas 1 (top), 2 (middle) and 3 (bottom). Only motion with an amplitude  $\geq 2\text{\AA}$  is shown. The S-S bridge of hVKORC1 is shown using yellow sticks.

To characterise the conformational changes of the L-loop that are associated with a great deal of flexibility and mobility, the most emblematic residues, in view of their fluctuations (RMSFs), were first selected. Two sets of residues—(1) C43, V54 and S74, located on the L-loop helices (the midpoint residues of H1-L, H2-L and H3-L) and showing the minimal values of RMSFs, and (2) R35, G46 and G64, positioned on the L-loop linkers L11, L12 and L13, respectively, and displaying the greatest RMSF values—were chosen (Figure 8A). Each set of residues was completed by residue C135 from the TMD and was then used to define two tetrahedrons, **T1** and **T2**, designed on the  $\text{C}\alpha$ -atoms. It is suggested that light may be shed on the conformational features of the L-loop by analysis of the six straight edges corresponding to the distances between each pair of residues.

Analysis of **T1** geometry showed (i) the great stability of  $\text{C}\alpha$ – $\text{C}\alpha$  distances (d) between C43 (H1-L helix), V54 (H2-L helix) and C135 over nearly all the simulated time and in all the replicas; (ii) high conservation of  $\text{C}\alpha$ – $\text{C}\alpha$  distances between each of the three residues and S74 (H3-L) over a substantial time period (200–300 ns or more), followed by

(iii) a synchronic change of these distances ( $\Delta$  of 6–8 Å), indicating the displacement of the H3-L helix with respect to the other helices, H1-L and H2-L (Figure 8B). As was expected, T2, which is determined using the most fluctuating residues, showed less conserved geometry, displaying synchronic changes in all or at least 3–4 distances ( $\Delta$  of 8–15 Å).



**Figure 8.** Geometry and folding of the L-loop from hVKORC1 in its inactive state. (A) Two tetrahedrons, T1—defined for the C $\alpha$ -atom of C135 and for the midpoint residues of each L-loop helix, and T2—defined for the C $\alpha$ -atom of C135 and for the most fluctuating residues (with the greatest RMSF values), from the L-loop linkers. (B) Distances between each pair of C $\alpha$ -atoms from the tetrahedrons T1 (top) and T2 (bottom) over each MD trajectory. The distance curves and the edges of a tetrahedron are coloured similarly. (C) The time-dependent evolution of the secondary structure of each residue, as assigned by the Define Secondary Structure of Proteins (DSSP) method:  $\alpha$ -helix is in red,  $3_{10}$ -helix is in blue, turn is in orange and bend is in dark yellow. (D) Drift of the L-loop helices observed over the MD simulations (concatenated trajectory, sampled every 100 ps). Superimposed axes of helices from the L-loop are covered on the randomly chosen conformation of hVKORC1 in two orthogonal projections. The axis of each helix is defined as a line connecting the two centroids assigned for the first and the last residues.

Comparison of T1 and T2 metrics showed an absence of coupling between their geometries. Similarly, no evident relation was found between the secondary structure of the L-loop and the T1 or T2 geometries, suggesting that the relative positions of the residues from the L-loop helices and from the linkers connecting these helices are disconnected from the folding–unfolding effects in the L-loop.

This analysis revealed (i) the high stability of the H1-L helix in terms of its secondary structure, as well as in its relative position with respect to TMD, (ii) the quasistable spatial position of the transient H2-L helix relative to H1-L and TMD, (iii) the large displacement of the transient H3-L helix from the anchored structural motif formed by HL-1 and H2-L and of the coiled linkers L11–L13.

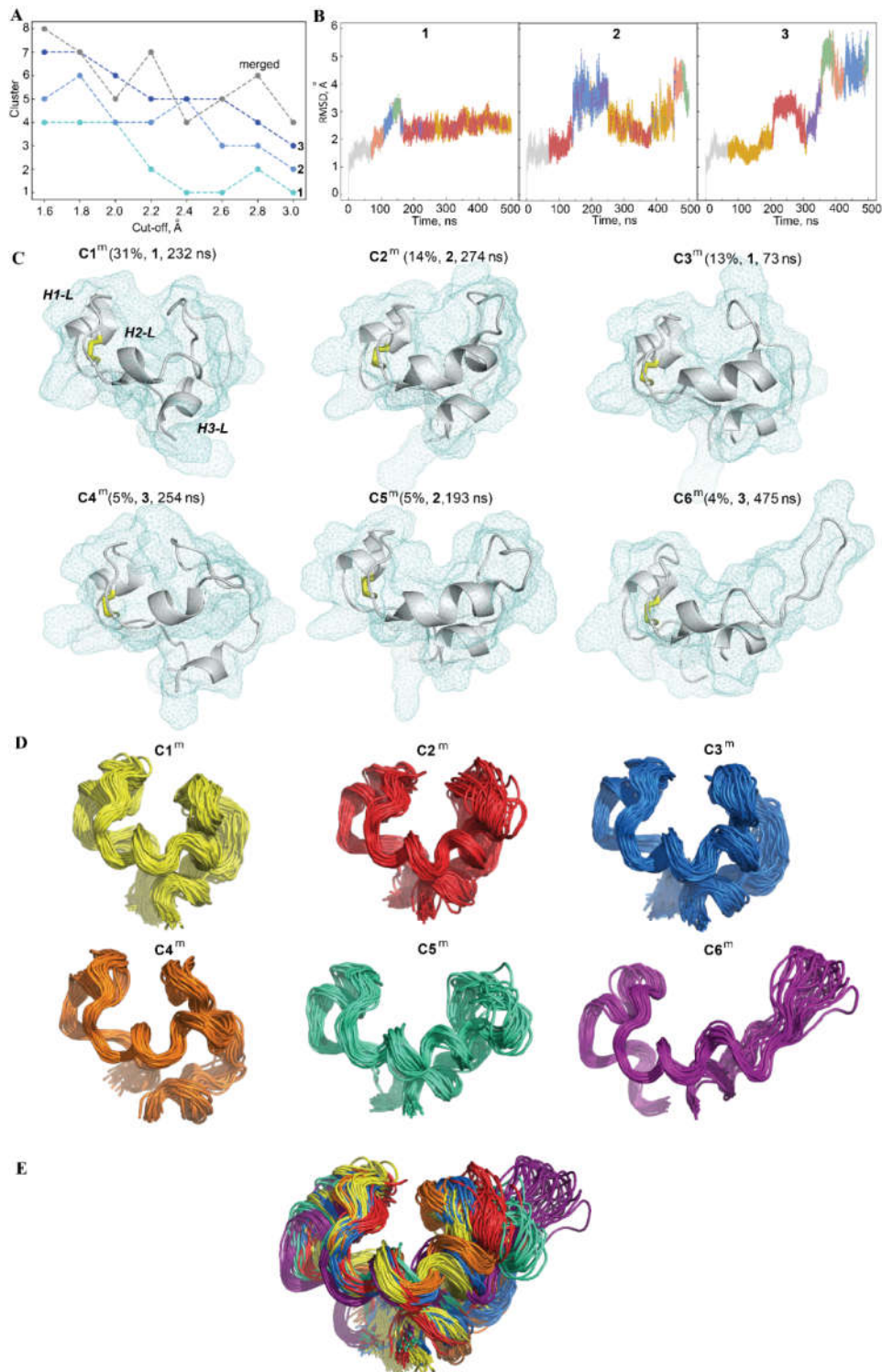
To illustrate the relative orientation of the L-loop helices, their structural drift was analysed. The axis of each helix was defined for the conformations from trajectories 1–3 (sampled every 100 ps, concatenated data), superposed and projected on a randomly chosen conformation of the L-loop (Figure 8D). The superimposed axes (elongated by 50% to better represent their position and direction) form a reep-like distribution for all helices. The axes of the three helices differ in length and their spatial orientation within each reep-like distribution and between the helices.

### 2.2.3. Conformational Variability of the hVKORC1 L-Loop

To characterise the conformational space explored over the MD simulations using the L-loop of hVKORC1 in its inactive state and to distinguish the most probable conformations, the generated conformations were analysed using ensemble-based clustering [28]. Conformations of each MD trajectory were grouped with different RMSD cut-off values that varied from 1.6 to 3.0 Å, with a step of 0.2 Å. Using of cut-off value  $\geq 2.2$  Å results in a poor number of clusters, while more restricted cut-off values of 1.8 and 2 Å were sufficient to regroup the L-loop conformations into clusters that give the best cumulative population (>90%; Figure 9A). Interestingly, clustering with these cut-off values produces an equal number (six) of clusters, with a nonzero population in all replicas (Figure S8). Taking a cut-off of 2.0 Å as the criterion, the population of each cluster obtained for each trajectory was compared. In replica 1, the majority of conformations form the two most-populated clusters, C1 (48%) and C2 (32%); the other conformations are regrouped in clusters with a low population of 0.5–9%. In each of the other replicas (2/3), the MD conformations are regrouped in the three most-populated clusters, with a comparable density between the replicas: C1 (41/33%), C2 (22/23%) and C3 (20/15%). The MD conformations that form the most populated clusters, C1 and C2, are individually regrouped within the narrow time ranges in trajectory 3 only, while in two other simulations, they are observed over a long period for each trajectory as coexisting with the conformations from the other clusters (Figure 9B). The conformations from the lowly populated clusters are usually observed in time ranges where the RMSD varies significantly and may show the transient states of the L-loop.

The representative conformations from different clusters of the same replica are divergent at the folding level (2D) and in 3D-structure organisation (Figure S8). An archetypical example is the considerable disparity between the conformations from clusters C2 and C3 of trajectory 3 that represents the L-loop before and after the transition, which is evidenced by the RMSD curve (Figure 6B). In contrast, some representative conformations of the clusters from different replicas showed a convenient similarity, for instance, C2 and C1 from replicas 1 and 3, respectively.

It is supposed that the L-loop conformational spaces generated by the three independent trajectories are partially overlapped. To verify this hypothesis, a clustering analysis was performed on the merged trajectory composed of the L-loop conformations from three replicas. Accordingly, the number of clusters obtained with the same RMSD cut-off values is significantly lower for the merged data than the sum of clusters obtained individually for each replica (Figure 9A), which confirms the overlapping of the conformational spaces of the L-loop covered over the three replicas of MD simulation of hVKORC1 in its inactive state.



**Figure 9.** Ensemble-based clustering of L-loop MD conformations. **(A)** Number of clusters obtained for each MD trajectory (1, 2 and 3) and the concatenated trajectory. The first 70 ns of every trajectory was omitted from the computation. Clustering was performed on each 10-ps frame of every trajectory using cut-off values that varied from 1.6 to 3.0 Å, with a step of 0.2 Å. **(B)** Location of the MD conformations grouped in clusters, with a cut-off of 2.0 Å for the RMSD curves of trajectories 1–3. Clusters C1–C6 are arbitrarily distinguished by colours in each trajectory: orange (C1), red (C2), blue (C3), rose (C4), green (C5) and violet (C6). **(C)** Representative conformations of the L-loop from clusters (C<sup>m</sup>) with population ≥4%, obtained with a cut-off of 2.0 Å for the merged trajectory. The L-loop is shown as ribbons with a meshed surface, with disulphide bridges C43–C51 drawn as yellow sticks. The L-loop surface is displayed as meshed contours. The population

of each cluster is given in brackets (in %), together with the replica number (in the bold) and the time (in ns) over which the representative conformation was recorded within a replica. **(D)** Conformations of the L-loop (taken every 100 frames) of each cluster ( $C^m$ ) of the merged trajectory, and **(E)** superposed conformations from the  $C1^m$ – $C6^m$  clusters. In **(D,E)**, the L-loop is drawn as a tube.

The first three clusters of the concatenated trajectory (cut-off 2.0 Å) contain 31%, 14% and 12% of all conformations, while the other conformations form the poorly populated clusters. The cumulative population of the clusters, with a density >4% on the merged data, is reduced (72%) with respect to individual trajectories but is still meaningful and statistically rich for the characterization of the most frequent L-loop conformations. Regarding the composition of the clusters, it was found that the dense clusters of the merged trajectory,  $C1^m$  and  $C3^m$ , are composed of conformations from different trajectories ( $C1^m$  and  $C3^m$  are comprised of conformations from replicas **1/2/3**, with proportions of 83/4/12% and 28/12/58%, respectively), while the other clusters are composed of conformations from the unique trajectory—**2** ( $C2^m$  and  $C5^m$ ) and **3** ( $C4^m$  and  $C6^m$ ), respectively (Figure 9C).

The representative conformation of each cluster, generated for the concatenated trajectory, showed that the principal factors leading to the conformational difference of the L-loop consist of (i) a variable length of the H2-L helix; a decrease of that promotes (ii) an elongation of linker L23, which, in turn, encourages (iii) the repositioning of the H3-L helix with respect to the H1-L and H2-L helices (Figure 9 C,D). In contrast to H2-L, the length of the H1-L and H3-L helices is better conserved. The whole shape of the conformations from different clusters well-reflects the “scissor-like” motion of the H1-L helix and the L23 loop that is observed in the PCA modes. The compact shape of the L-loop corresponds to the “closed” position of the H1-L helix and the L23 loop, which is a typical feature of most L-loop conformations (see the highly populated clusters,  $C1^m$ – $C4^m$ ). The conformations grouped in cluster  $C6^m$  show an elongated shape, with an “open” position of the H1-L helix and the L23 loop. Cluster  $C5^m$  is composed of intermediate conformations between the “open” and “closed” forms.

The clustering enabled (i) the splitting of MD conformations of the L-loop into groups composed of similar geometry and shape (within a cut-off), (ii) the assembly of a great majority of conformations into a limited number of clusters, and (iii) a distinction between dense clusters with a statistically reasonable population.

#### 2.2.4. Intra-L-Loop Interactions

To establish the forces that stabilise L-loop conformations, contact maps were computed for each representative conformation from the most populated clusters (>4%) found on the concatenated trajectory. The contact maps show the multiple intra-L-loop interactions between the linkers, between the linkers and helices and between the helices (Figure S9). Nevertheless, the patterns of such contacts differed in clusters  $C1^m$ – $C6^m$ . The most common pattern found in the maps describes the contact of L11 with H2-L and H3-L helices and of L23 with H3-L, which are systematically observed in clusters  $C1^m$ – $C5^m$ .

Analysis of the H-bonds showed that the L-loop conformations are stabilised by mutual H-bonds that form extensive networks (Figure 10, Table S3). Comparing these H-bond networks in “closed” conformations (clusters  $C1^m$ – $C5^m$ ), it is noted that D36, D38, D44, R53, R61 and E67 are the key residues that form the salt bridges. In the “open” conformation (cluster  $C6^m$ ), the set of interacting residues that form the salt bridges is composed of R35, D38, D44 and R58.

The salt bridge that is stabilised by the pairing of charged residues when a combination of two noncovalent interactions is formed, H-bonding and ionic-bonding, is the most commonly observed contribution to the stability of the entropically unfavourable folded conformation of proteins [29]. Indeed, in the highly compact “closed” L-loop conformations from  $C1^m$  and  $C2^m$ , R53 interacts with D36 and D38, forming the R53-based “salt bridge pattern” that stabilises the proximal position of H2-L and the L11 linker. In the



These interactions in C3<sup>m</sup> are completed by the contact of R53 (H2-L) with D36 (L11), causing an overlap of the two “salt bridge patterns”, namely, R61- and R53-based patterns. Additionally, in C3<sup>m</sup>, the other “salt bridge pattern” is formed by R37 contacting with D44, which stabilises H1-L and the L12 loop in a tight spatial position. In C4<sup>m</sup>, the R53- and R61-based “salt bridge patterns” are clearly separated, while each positively charged residue interacts with different subsets of the negatively charged residues, i.e., R61 with D36 and E67 and R53 with D38. These two “salt bridge patterns” gather together two neighbouring helices, H1-L and H2-L, and two distant linkers, L11 and L23. In C4<sup>m</sup>, similar to C3<sup>m</sup>, the “salt bridge pattern” formed with R37 and D44 is clearly separated from the R61- and R53-based “salt bridge patterns”. Such a spatial separation of two “salt bridge patterns” is observed in the “open” conformations of the L-loop from C5<sup>m</sup> and C6<sup>m</sup> clusters, in which two “salt bridge patterns” are formed by R53 (H2-L) interacting with D36 and D38 (C5<sup>m</sup>) or with D38 and D44 (C6<sup>m</sup>), and either by R37 (L11) interacting with R40 and D44 (C5<sup>m</sup>) or by R35 bound to R35 and D38 (C6<sup>m</sup>).

Besides the salt-bridge interactions, the charged residues also contribute to H-bonds by interaction with the different polar and hydrophobic residues, which either act as H-donors or H-acceptors for the atoms in their main or side chains. All these ionic and H-bond interactions between the charged residues and between the charged and polar residues contribute to the tight spatial L-loop arrangement, in which the helices and linkers from the remote sequence segments are localised at close proximity. It is interesting to note that R40, D44, R53 and R61 interact in any conformation of the L-loop, independent of the L-loop’s shape, by forming either the salt bridges or the H-bonds.

Nevertheless, many charged and polar residues that are not involved or are partially involved in intra-L-loop interactions protrude from the L-loop, as illustrated by the “closed” and “open” conformations of the L-loop (Figure 10B). Considering the spatial position of the solvent-exposed residues with respect to L-loop cysteine residues (C43 and C51) that participate in the thiol–disulphide exchange reaction, the residues from sequence S52 to E67 are most likely involved in interactions with a redox protein.

As the L-loop also contains a large number of hydrophobic residues, their contribution to intra- and intermolecular interactions was evaluated. Although hydrophobic forces are known to be relatively weak interactions, such interactions can add up to make an important contribution to the overall stability of a conformer or molecular complex [30].

Multiple contacts between the A41, G46, A48, I49, V54, L70 and L76 hydrophobic residues were observed in “closed” conformations, while in the “open” conformations, such contacts involved V45, F55, F63, L70 and L76 (Figure S10, Table S3). These hydrophobic contacts may reflect the stabilising interactions that complete the H-bond contribution as well as the repulsive forces that equilibrate the strong salt-bridge interactions.

Similar to the charged and polar residues, some hydrophobic side chains are oriented toward the exterior of the L-loop, putting them in positions accessible to the solvent, such that the number of such residues is significantly higher in the “closed” conformations than in the “open” ones (Figure 10C, Table S3). One part of these residues (F55, G56, F63, L65, V66) belongs to the sequence S52–E67, which was postulated to be involved in interactions with a redox protein.

### 2.3 . Modelling of Molecular Complex Formed by hVKORC1 and Its Redox Partners

The molecular complex of hVKORC1 was constructed with PDI to probe our hypothesis on the identification of a hVKORC1 redox partner (see the Discussion section); 3D models of the complex were constructed using the crystallographic structure of the VKOR from bacteria (bVKOR; PDB ID: 4nv5; [14]) as a reference for the initial positioning of PDI relative to hVKORC1.

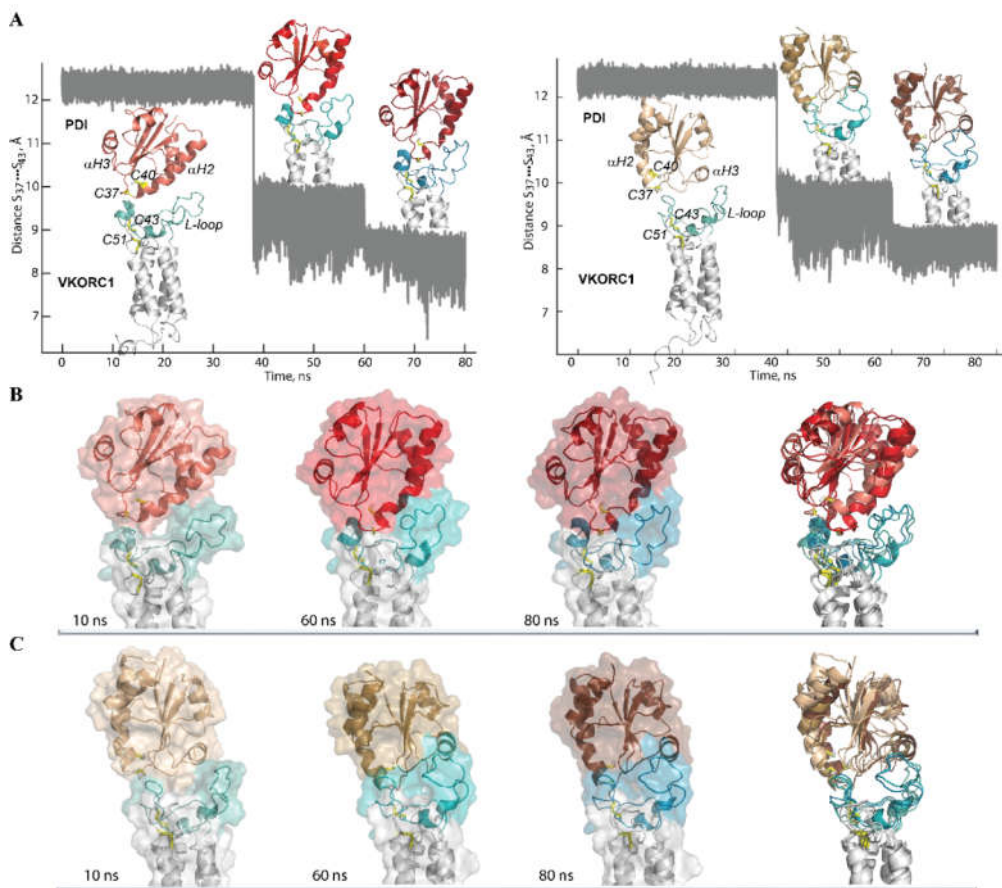
To be most objective in the modelling of the human PDI–VKORC1 complex, the structure of bVKOR was not used as a template because of (i) the suggested alternative VKOR activation mechanisms in bacteria and in eukaryotes, that is, in their respective native environments, which employ significantly different mechanisms for electron transfer [14], (ii) a high structural difference between the Trx- and L-loop domains in bVKOR and human proteins (RMSD values are 4.5 and 4 Å between bVKOR and the “closed” and “open” conformations of hVKORC1, respectively), (iii) very low sequence identity/similarity (15/20%), and (iv) a very large distance between the cysteine residues from the Trx-like and VKORC1-like domains (the minimal S···S distance of 16 Å) in bVKORC1 (Figure S11).

For modelling the human PDI–VKORC1 complex, a conformation of hVKORC1 with the most extended “open” L-loop (the least probable conformation) was chosen as the initial target structure in order to bring the two proteins as close as possible. As for the initial PDI model, the conformation with a well-ordered and long  $\alpha$ H2-helix that is similar to the X-ray structure of PDI [9] was chosen and positioned above hVKORC1 so that (i) the distance between the sulphur atoms from C37 of PDI and C43 of hVKORC1 was as short as possible (12.5 Å) and (ii) each PDI fragment that was suggested to be a fragment able to form the intermolecular interactions with a target, namely, F1 and F2, was alternatively placed above the middle of the L-loop surface. The obtained promodels, **Model 1** and **Model 2**, were explored using MD simulation for conditions (see the Methods section), where restraints apply to the distance S···S between C37 (PDI) and C43 (hVKORC1). The restraints were gradually diminished during a stepped 80-ns MD simulation run (Figure 11).

For both models, structural rearrangement occurred inside each protein and between the proteins, with diminishing S···S distance. In **Model 1**, the extended “open” conformation of the hVKORC1 L-loop was observed at an S···S distance of 12 Å, which then adopted a “closed” conformation at a shortened S···S distance (of 10 and 8 Å), with the  $\alpha$ H1-L helix and the L23 linker located in a proximal position, which is the most probable conformation of the L-loop in isolated VKORC1. The initially well-ordered and long  $\alpha$ H2-helix of PDI is rotated by 30° (at an S···S distance of 10 Å), followed by the bending of the helix, and then (at the S···S distance of 8 Å) by depletion of two helices, a small  $3_{10}$ -helix in the proximity of the CX<sub>1</sub>X<sub>2</sub>C motif and a shortened  $\alpha$ H-helix, which demonstrates a folding–unfolding effect observed in MD simulations of PDI in an isolated state.

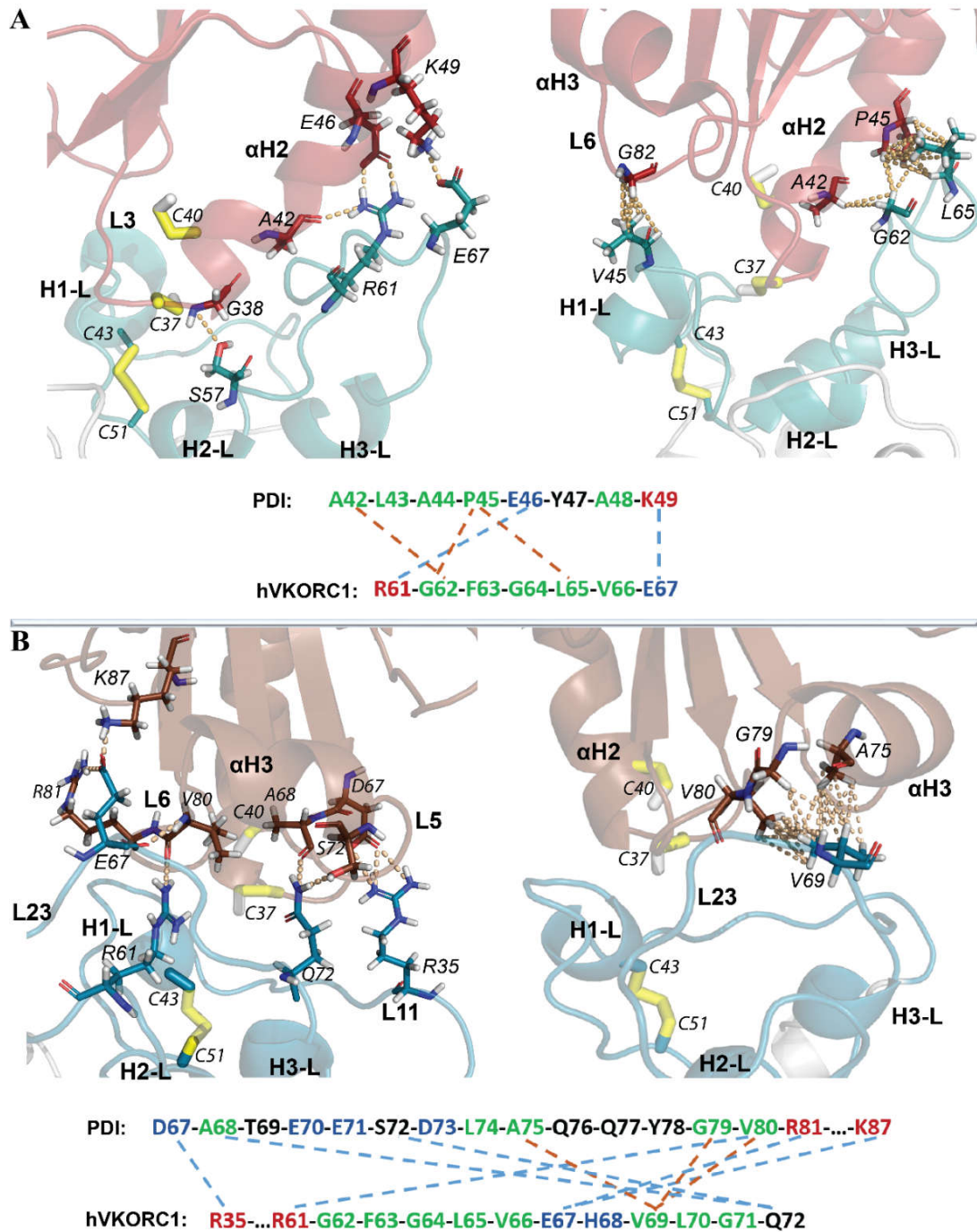
Similarly, in **Model 2**, a gradually diminishing S···S distance from 12 to 8 Å promotes a change in the L-loop conformation from “open” to “closed” in hVKORC1, while in PDI, a departure of the  $\alpha$ H3-helix from its initial position to the location most exposed to the solvent (a 4.5–5.0 Å parallel displacement of the helix) was observed. The conformational changes observed during the simulations of the two PDI–hVKORC1 complex models are reflected in the folding of “interacting” proteins. The extended “open” conformation of the hVKORC1 L-loop, taken as the initial structure for complex modelling, showed increased folding (by 50%) in **Model 1**, with a decrease in the S···S distance from 12 to 8 Å, while in **Model 2**, its helical fold was reduced by 40% (Table S4). As for PDI, the folded content of its initial and final conformations was the same for both models.





**Figure 11.** Modelling of human PDI–VKORC1 complex. (A) MD simulations of 3D model PDI–VKORC1 complexes were performed, with gradually diminished distance (from 12.5 to 8.0 Å) between the sulphur (S) atoms of C37 from PDI and of C43 from the L-loop of hVKORC1. PDI has two orientations with respect to VKORC1, with **F1 (Model 1, left)** and **F2 (Model 2, right)** positioned above the middle of the L-loop surface. Both models of the PDI–VKORC1 complex are shown as snapshots taken at  $t = 10, 60$  and  $80$  ns, with different S...S distances. The reference residues and fragments are labelled. (B,C) Conformations of the PDI–VKORC1 complex, with two different PDI orientations, chosen at  $t = 10, 60$  and  $80$  ns, and their superposition at all three times. In (A–C), the proteins are depicted as ribbons or as ribbons and surfaces and are distinguished by colour: a red palette was used for PDI and a cyan palette for hVKORC1, both nuanced by the tonality from light to dark to distinguish the conformations chosen at  $t = 10, 60$  and  $80$  ns.

Analysis of the intermolecular contacts at the interface between PDI and hVKORC1 (in the conformation taken at  $t = 80$  ns) showed that these two proteins in **Model 1** are linked through two salt bridges formed by R61 (hVKORC1) and E46 (PDI) and by D67 (hVKORC1) and K49 (PDI) (Figure 12A). Hydrophobic contacts were also observed between two pairs of residues: A42 (PDI) and G62 (hVKORC1) and P45 (PDI) and L65 (VKORC1). Moreover, G62 (hVKORC1) interacts with P45 (PDI). The PDI–hVKORC1 interface interactions are completed by an H-bond between the side chain of S57 (hVKORC1) and the main chain of G38 (PDI), the amino acid in the proximity of the CX1X2C motif, and by the hydrophobic interaction between V45 (hVKORC1) and G82 (PDI). All distances between the interacting D...A atoms were ranged from 2.5 to 3.2 Å, which characterise strong interactions.



**Figure 12.** Intermolecular contacts at the interface between PDI and hVKORC1 in two models of the PDI-hVKORC1 complex. The intermolecular H-bonds and hydrophobic contacts between PDI and VKORC1 in Model 1 (A, top) and Model 2 (B, top). (A,B) The proteins are shown as coloured ribbons: PDI in red and brown and VKORC1 in cyan (L-loop), with the interacting residues and thiol groups as sticks. The contacts are indicated by dashed lines: H-bonds in yellow and hydrophobic contacts in salmon. The structural fragments and residues participating in the contacts are labelled. Analysis of the intermolecular contacts was performed on conformations taken at  $t = 80$  ns. (A,B) A pattern of H-bond (in blue) and hydrophobic (in orange) contacts between the PDI and hVKORC1 residues (bottom). Residues are coloured according to their properties: the positively and negatively charged residues are in red and blue, respectively; the hydrophobic residues are in green; the polar and amphipathic residues are in black.

Spatially, two sets of interactions stabilising the PDI–hVKORC1 complex were observed. The first set, which is composed of S57(hVKORC1)···G38(PDI) and V45(hVKORC1)···G82(PDI), is localised in the proximity of the active sites, the CGHC motif of PDI and disulphide bridge C43–C51 of hVKORC1 and probably stabilises their close location, which is induced, in part, by a steric requirement imposed on the sulphur atoms from C37 and C43 to be in the closed position. The second set, which is composed of multiple contacts between the residues from short sequence segments, A42–K49 from PDI and R61–E67 from hVKORC1, forms a very compact regular interaction pattern that describes the highly specific recognition between two molecules that are maintained by two salt bridges and by crosswise hydrophobic interactions. This pattern of interactions stabilises the  $\alpha$ H2-helix of PDI and the L23 linker from hVKORC1 in a close position that is independent of any interaction with Set 1 and, consequently, may present a first step in the PDI–hVKORC1 recognition process.

The residues of hVKORC1 that form salt-bridges and H-bonds are located on the transient H2-L helix and on the L23 linker, which is composed of a segment that was predicted to be the most putative recognition region in an isolated hVKORC1. Similarly, PDI residues participating in hVKORC1 recognition belong to the F1 fragment were regarded as a possible putative recognition site. Surprisingly, a hydrophobic interaction with V45 (hVKORC1) is formed by G82 (PDI), which is a residue from the F2 fragment that is also predicted to be a fragment that contains possible recognition sites.

In **Model 2**, the interaction interface between PDI and hVORC1 is also formed by two salt bridges generated by R35 (hVKORC1) and D67 (PDI) and by E67 (VKORC1) interacting with R81 and K87 from PDI. Other electrostatic interactions are presented by the H-bonds of Q72 (hVKORC1) with A68 and S72 from PDI and of R61 (hVKORC1) with V80 of PDI. Hydrophobic contacts are observed as a three-furcate interaction of the three PDI amino acids (A75, G79 and V80) attached to a unique amino acid (V69) of hVKORC1. Unlike the compact interface contact network in **Model 1**, the interacting residues in both proteins of **Model 2** are distributed over large sequence segments, from D67 to K87 in PDI and from R35 to Q75 in hVKORC1. This highly enlarged interface interaction network seems less probable because of the small probability of a synchronised approach of two space-separated binding sites to the target.

It is interesting that two amino acids, R61 and E67, of hVKORC1 form salt bridges in both models, **Model 1** and **Model 2**, but by selecting different PDI residues. Remarkably, both amino acids belong to a hVKORC1 segment that is predicted to be the putative recognition site by analysis of the isolated protein.

Although very compact and regular, the interface interaction pattern formed by the closely localised residues in both proteins from **Model 1**, together with the increased helical folding of the L-loop by 50%, is a very attractive argument for the choice of this model for being functionally related, though there is still doubt in such a conclusion.

Other characteristics are considered to better justify or challenge our hypothesis. First, from a superimposition of each model on the experimentally defined structure of bVKORC1, the best fit at the level of Trx-like domain orientation with respect to VKOR is observed for **Model 1** (Figure S12), but analysis of the interaction between the Trx-like and VKOR domains in the bacterial protein showed only a single short contact (between Q40 from the  $\alpha$ H2-helix of Trx and L46 of the L-loop), an observation that largely mismatches the interaction patterns observed in both models.

Finally, to check the stability of the interactions between the two proteins in **Model 1** and **Model 2**, the models were simulated at  $t = 80$  ns under more relaxed (“soft”) conditions (see the Methods section), which gives more tolerant restrains on the distance S···S between C37 (PDI) and C43 (hVKORC1).

In the two MD simulations of **Model 1**, which have different “soft” constraints (a time range of 80–100 ns), the distance S···S either varied within an enlarged range (7–11 Å) or, surprisingly, showed a tendency to decrease (6–10 Å) with respect to the simulation

with a more “hard” restriction (a time range of 60–80 ns) (Figure S12). The MD conformations of **Model 1**, generated using different “soft” constraints, showed very similar structures of PDI–hVKORC1 that differed only in the folding of the H2-L helix from the L-loop of hVKORC1 and the  $\alpha$ H2 helix of PDI. Each of these structural effects was observed in isolated proteins. The interface interactions between the residues from the  $\alpha$ H2 helix of PDI and L23 from the L-loop of hVKORC1 were very similar for conformations taken at  $t = 100$  ns and  $t = 80$  ns. With respect to the conformation chosen at  $t = 80$  ns, some novel contacts involving residues from H2-L (hVKORC1) and the L3 loop and of PDI are observed in the conformation taken at  $t = 100$  ns (Figure S13).

These results show that the highly specific recognition between the two molecules is maintained by the strong and stable interactions formed by two salt bridges and by cross-wise hydrophobic interactions preserved in **Model 1**.

The MD conformations of **Model 2**, generated using “soft” and “hard” constraints, showed similar structures of PDI–hVKORC1, which differed only in the position of the  $\alpha$ H3 helix of PDI and the L-loop of hVKORC1. The interactions observed at the interface between the two proteins are nonpreserved, with the exception of a single salt bridge between E67 (hVKORC1) and R81 (PDI) (Figure S14).

### 3. Discussion

Vitamin K epoxide reductase is a membrane protein that reduces vitamin K using a membrane-embedded cysteine-containing redox centre. Such activity requires the cooperation of VKORC1, with a redox partner that delivers reducing equivalents. The physiological redox partner of hVKORC1 remains uncertain; nevertheless, four proteins—PDI, ERp18, Tmx1 and Tmx4—were suggested as the most likely *H*-donors of hVKORC1 [12,13]. Deciphering the molecular origins of VKORC1 recognition by an unknown redox protein is not a trivial task.

We suggested that a careful in-silico study of the isolated proteins would provide useful information. In particular, quantitative metrics and qualitative estimations can shed new light on the target (hVKORC1) features and the peculiarities of redox proteins. Such information may help in predicting (i) the protein fragments participating in VKORC1 recognition by a Trx and (ii) the most probable partner of VKORC1.

What has been learnt from studying VKORC1 and four Trx-fold proteins?

The L-loop is known to bind to and accept reducing equivalents from species-specific partner oxidoreductases essential for VKOR enzymatic function in vivo [31], so this domain was carefully characterised. We found that the L-loop in the inactive (oxidised) state of VKORC1 is noticeably less flexible compared to the reduced states of VKORC1 [6] and more folded, showing three helices connected by coiled linkers. This three-helix fold of the L-loop was generally maintained over the MD simulations, while the length and spatial positions of the helices were highly variable. This variation is reflected in a large number of L-loop conformations, varying from a compact “closed” conformation, which is prevalent, to an extended “open” conformation.

It was established that the H2-L helix is the fundamental actor that controls the conformational features of the L-loop. This transient helix converts between the  $\alpha$ H- and the  $3_{10}$ -folds, adapting in length from short to elongated. The shortened H2-L helix, in which the S56-R61 segment is unfolded, promotes the elongation of the coiled linker L23, connecting the H2-L and H3-L helices. The extended linker L23 shows (i) great mobility with respect to the H1-L helix, which can be described as a “scissor-like” motion, and (ii) a large vertical displacement with respect to the TMD. Moreover, the extended linker L23 delivers increasing mobility to H3-L, evidenced by its displacement with respect to H2-L.

At the sequence level, the L-loop has been reported to be conserved between VKORs from different species [32]. Particularly high conservation was found for the S56–G63 segment, which in hVKORC1 is followed by 5-residue hydrophobic insert GFGLV, which is completed by glutamic acid (E67) and histidine (H68). Sequence conservation, along with observed structural and dynamical properties of the H2-L helix and its adjacent linker

L23, suggests their possible functional role. From the analysis of the H-bonding patterns in the L-loop, regular exposition of the charged (R58, R61, E67 and D73) and polar residues (S56, S57, W59, H68 and N77) to the outer side of the L-loop was observed in positions favourable for contact with a solvent or protein. Therefore, we postulated that the S56–R61 segment, a part of the more extended S53–N77 segment, is a platform for the recognition of a protein partner.

Charged residues have been shown to be instrumental in the definition of binding specificity, while sometimes contributing little binding energy to the interactions themselves [33,34]. In other cases, charged residues were found to promote high-affinity binding [35,36]. They are also the main players in “electrostatic steering”, which is a long-range mechanism in which electrostatic forces can steer a ligand protein to a binding site on the receptor protein; this drastically increases the association rate [37,38]. Often, charged residues that are important for protein–protein interactions are conserved across families of evolutionarily related proteins and protein complexes [39–41].

Moreover, the tryptophan residue (W59) from the S56–R61 segment, following 5-residue hydrophobic insert GFGLV, may act as an anchoring residue that binds the two proteins. Tryptophan residues have been shown to exhibit a strong tendency to remain within the interfacial region [42]. The role of the hydrophobic effect as a driving force in protein folding and assembly is well described [43].

Which of the four studied Trx-fold proteins is the most probable partner for VKORC1?

Regarding the probable redox partners of hVKORC1 (Erp18, PDI, Tmx1 and Tmx4), it was observed that despite their similar architecture, each protein is characterised by its own sequence-dependent structural and dynamical features. In particular, it was observed that the CX<sub>1</sub>X<sub>2</sub>C motif’s different folds are connected to the divergent configuration of the thiol groups—either as part of the well-folded  $\alpha$ H2-helix (Erp18 and Tmx1), with the restrained cis-geometry of sulphur atoms, or as a part of a coiled structure, with the alternating orientation of sulphur atoms that runs from a syn-periplanar configuration to an antiperiplanar configuration (Tmx4), or as part of a transient structure (PDI) reversed between the helical fold ( $\alpha$ - and  $3_{10}$ -helices) and turn-coil structure, leading to a large number of thiol group configurations.

Focusing on the F1 region, suggested to be able to form intermolecular interactions with a target, it is noted that only F1 of PDI and the targeted S56–R61 segment of VKORC1 have similar structural properties, or rather, a structural disorder that describes an intrinsically disordered region (IDR). Indeed, two IDRs, which are the transient N-terminal of the  $\alpha$ -helix H2 in PDI and the transient H2-L helix comprising the S56–R61 segment from the L-loop of VKORC1, show similar structural heterogeneity and plasticity that is consistent with an affinity that is sensitive to changes in local frustration distribution and thermodynamics.

Numerous publications have reported that many protein–protein interactions (PPIs) are mediated by protein regions that are not confined to a single folded conformation prior to binding, namely, IDRs that participate in PPIs (interacting IDRs) [44–46]. IDRs are increasingly recognised for their prevalence and their critical roles in regulatory intermolecular interactions [47]. It has been hypothesised that some traits make IDRs particularly suitable for interactions that involve signalling and regulation, complementing globular domains that more often perform catalytic functions. It has been estimated that IDRs in the human proteome contain ~132,000 binding motifs [48]. Disordered proteins are believed to account for a large fraction of all cellular proteins, playing roles in cell-cycle control, signal transduction, transcriptional and translational regulation, and large macromolecular complexes [49]. Nevertheless, even if fragment F1 is considered the most probable fragment to form intermolecular interactions with a target, the mobility of linker L5 and the  $\alpha$ H3 helix from F2 of PDI means that F2 has strong compatibility with the highly mobile S56–R61 segment of VKORC1. Moreover, F2 shows the most dissimilar sequence in the studied proteins, and it also has a great number of hydrophobic, polar and charged

residues that are exposed to solvents. Consequently, **F2** is also potentially able to contribute to stabilising a supramolecular complex.

These two fragments are very close to the CX<sub>1</sub>X<sub>2</sub>C motif, which is either joined in a sequence (**F1**; sequence vicinity) or adjacent in a 3D structural space (**F2**; spatial vicinity).

This makes clear that we can begin to construct models of the molecular complex formed by hVKORC1 and PDI, where PDI is the most probable redox partner of hVKORC1. Exploring the recognition processes between these two proteins, hVKORC1 and PDI, requires knowledge of the 3D structure of the associated molecular complex.

Direct use of the X-ray structure of VKOR from bacteria (a protein with covalently bound Tmx-like and VKOR-like domains, which has low sequence and structure similarity compared to the human proteins PDI and VKORC1) is not appropriate for the modelling of the human complex but can be a reference for the initial positioning of PDI with respect to hVRORC1. Using conventional MD simulations, two models of the PDI–hVKORC1 complex, with the PDI in two alternative positions, which were either exposed by **F1 (Model 1)** or **F2 (Model 2)** in front of the L-loop of hVKORC1, were studied. In both probed models, proteins bind to each other using a combination of hydrogen bonds, salt bridges and hydrophobic contacts formed by residues from the different protein domains. These domains are small binding clefts and include a few peptides in **Model 1**, while in **Model 2**, the molecular interface represents large areas on each protein and spans widely spaced amino acids in protein sequences.

How do the “interacting” residues predicted by analysis of isolated proteins correspond to the contacts in the complex formed by VKORC1 and PDI?

In **Model 1**, the interface contact network is composed of two salt bridges formed by two pairs of charged residues, R61 and E67, from hVKORC1, which, together with S57, G62 and L65, also contribute to the stabilisation of the two proteins. These residues are amino acids from the L-loop segment that was predicted as a platform for recognition of a protein partner by hVKORC1. In **Model 2**, the interaction interface between PDI and hVROR1 is also completed by two salt bridges formed by R35 and E67 from hVKORC1 interacting with D67, R81 and K87 from PDI and by *H*-bonds formed by Q72 and R61 of hVKORC1 with A68, S72 and V80 of PDI. In both models, two amino acids of hVKORC1, R61 and E67, participate in strong electrostatic interactions, salt bridges or *H*-bonds but with different PDI residues. It is remarkable that both amino acids belong to a hVKORC1 segment that is predicted to be the putative recognition site from analysis of the isolated protein. The contacting PDI residues are mainly predetermined by PDI orientation with respect to the L-loop.

Based on limited data from the stepped finite-time simulations, is it possible to conclude which model is the correct one?

In both models, the optimised (enhanced) orientation of PDI with respect to hVKORC1 is maintained by the multiple interactions between the two molecules.

In **Model 1**, intermolecular contacts are observed between the two short length peptides, R61–E67 from hVKORC1 and A42–K49 from PDI, which form two salt bridges and three crosswise hydrophobic interactions. Such a compact regular interaction pattern may describe highly specific recognition between the two molecules, maintaining the  $\alpha$ H2-helix of PDI and the extended L23 linker of VKORC1 in a close position and, consequently, may present the first step in the PDI–hVKORC1 recognition process. The other set of interactions, S57(hVKORC1)···G38(PDI) and V45(hVKORC1)···G82(PDI), is located in close vicinity to the CGHC motif of PDI and disulphide bridge C43–C51 of hVKORC1. This is induced by a steric requirement imposed on the sulphur atoms from C37 and C43 that holds them in a closed position. Moreover, as these contacts are formed by the main chain atoms, they are rather nonspecific.

In **Model 2**, the interaction interface between PDI and hVROR1 represents a large area for each protein and spans long-spaced amino acids of the protein sequences (D67–K87 in PDI and R61–Q72, completed by R35 in hVKORC1). The two salt bridges, which are formed by R35 (L11 from hVKORC1) and D67 (L5 from PDI) and by E67 (L23 from

VKORC1) interacting with R81 and K87 from L6 of PDI, involve two regions on each protein that are separated by large distances in the sequence and the 3D structure. The other *H*-bonds involve the residues located between the two remote salt bridges. The dense cluster of hydrophobic contacts is realised as a three-furcate interaction of three PDI amino acids (A75, G79, and V80) attached to a single amino acid (V69) of hVKORC1.

In both models of the PDI–hVKORC1 complex, interacting hydrophobic motifs from both proteins form “interacting hydrophobic cores”, which may be the key factors in the recognition process. The total number of noncovalent contacts between PDI and hVKORC1 in **Model 2** is 9, while in **Model 1**, it is only 5. It was reported that the number of connections between each pair of proteins is a strong predictor of how tightly the proteins connect to each other [50].

Nevertheless, despite the large number of *H*-bonds and the dense cluster of hydrophobic contacts, it appears that the enlarged interface interaction network observed in **Model 2** is less likely, due to the low probability of a synchronised approach of the two space-separated binding sites on PDI to the two space-separated binding sites on the target.

Moreover, based on the stepped simulations of **Model 1**, the diminishing distance between the two proteins promoted an increase in the helical folding of the L-loop by 50%, while in **Model 2**, its helical fold was reduced by 40%. While proteins become disordered on their own, their native conformation is stabilised upon binding [51,52]. The folded content of the initial and final PDI conformations is the same in both models; nevertheless, its conformation is adapted in both models by the folding–unfolding of the  $\alpha$ H2-helix in **Model 1** and by the removal of the  $\alpha$ H3 helix in **Model 2**.

The specificity of intermolecular interactions in PDI–hVKORC1 is apparently determined by sequence- and structure-based selectivity, which are the two determining factors in “molecular recognition”. A natural implication of the conformational selection model is the particular range of surface shapes visited by each protein and their collective complementarity, which is adjusted throughout the binding process. It was recognised that cooperativity derives from the hydrophobic effect, the driving force in single-chain protein folding [53]. The hydrophobic folding units that are observed at the interfaces of two-state complexes similarly suggest the cooperative nature of two-chain protein folding, which is also the outcome of the hydrophobic effect [54–56]. Nevertheless, although the hydrophobic effect plays a dominant role in protein–protein binding, it is not as strong as that observed in the interior of protein monomers; its extent is variable. The binding site is not necessarily at the largest patch of the hydrophobic surface. There are high portions of buried charged and polar residues at the interface, suggesting that hydrogen bonds and ion pairs contribute more to the stability of protein-binding than to that of protein-folding. Protein-binding sites have neither the largest total buried surface area nor the most extensive nonpolar buried surface area. They cannot be uniquely distinguished by their electrostatic characteristics, as observed by parameters such as unsatisfied buried charges or the number of hydrogen bonds.

The question is then to test if electrostatic and hydrophobic interactions in the PDI–hVKORC1 complex can be conserved qualitatively. The MD simulations of **Model 1**, performed upon different “soft” constraints that supplied an increased degree of freedom for proteins and allowed them to be removed, proved the stability of the interactions formed by salt bridges and by the crosswise hydrophobic contacts. As **Model 1** of the PDI–hVKORC1 complex showed stable interface interactions under such conditions, it was proposed as the first precursor to probe thiol–disulphide exchange reactions between PDI and hVKORC1.

Returning to the questions stated at the beginning of this work, they seem to have all been answered using a purely in-silico approach. Molecular modelling and molecular dynamics simulations provide powerful tools for the exploration of proteins and their complexes. Such a study is most effective when analysed in close conjunction with experi-

ments on a protein function, which would play an essential role in validating and improving the modelling and simulations. Therefore, we are now waiting for needed experimental validation (currently being undertaken by biologist colleagues) of the predictions given in this article. Experimental validation of the model of the PDI-hVKORC1 complex is essential for the continuation of this research, which will allow a better understanding of the redox chemistry underlying vital cell processes.

## 4. Materials and Methods

### 4.1. 3D Models

**Trx-fold proteins.** Structures of PDI (PDB ID: 4ekz), ERp18 (PDB ID: 1sen) and TMX1 (PDB ID: 1x5e) were retrieved from the PDB database [5] and atomic coordinates of domain **a**, which contains the CX<sub>1</sub>X<sub>2</sub>C motif and is present in all available structures, were extracted. The 3D homology model of hVKORC1 was generated from human sequence Q9H1E5 (<https://www.uniprot.org/uniprot/>) using the Modeller program [57] and the empirical structure of TMX1 (PDB ID: 1x5e) that was used as a template. The 3D model of the h-ERp18 protein was optimised (the cysteine residues were saturated with hydrogen atoms) to obtain a reduced state of the CX<sub>1</sub>X<sub>2</sub>C motif.

**h-VKORC1.** The coordinates of full-length hVKORC1 (sequence M1–H163) in the inactive state was taken from [6].

**Trx-VKORC1 complex.** Each complex of the PDI protein with hVKORC1 (PDI-hVKORC1) was modelled using the structure of bacterial VKOR (bVKOR; PDB ID: 4nv5) as a reference for the initial PDI positioning with respect to hVKORC1. The structures of the human Trx-fold protein and hVKORC1 were carefully superimposed with the respective domains of bVKOR. To eliminate a small intersection between part of the L-loop of hVKORC1 and PDI, the extended conformation of the L-loop was chosen. The PDI protein was placed in two orientations with respect to VKORC1, with (i) L3 and the  $\alpha$ H2-helix (**F1**) and (ii) L5 and the  $\alpha$ H2-helix (**F2**) positioned in front of the predicted “binding fragment” of the L-loop from hVKORC1. The initial distance S...S between the sulphur atoms from C37 of PDI and C43 of hVKORC1 in each built complex was 16 Å.

The stereochemical quality of all 3D models was assessed by Procheck [58], which revealed that more than 95% of nonglycine/nonproline residues have dihedral angles in the most favoured and permitted regions of the Ramachandran plot, as is expected for good models.

### 4.2. Molecular Dynamics Simulation

#### 4.2.1. Preparation of the Systems

For MD simulations, all models of the isolated proteins (PDI, ERp18, Tmx1, Tmx4), hVKORC1, and the two models of the PDI-hVKORC1 complex in two orientations (PDI<sub>F1</sub>-VKORC1 and PDI<sub>F2</sub>-VKORC1) were prepared with the LEAP module of Assisted Model Building with Energy Refinement (AMBER) [59] using the ff14SB all-atom force field parameter set [60]: (i) hydrogen atoms were added; (ii) covalent bond orders were assigned; (iii) protonation states of amino acids were assigned based on their solution for pK values at neutral pH and histidine residues were considered neutral and were protonated for  $\epsilon$ -nitrogen atoms; (vi) the Na<sup>+</sup> counter-ion was added to neutralise the protein charge.

Each membrane protein, hVKORC1 and the two models of complex PDI-VKORC1 (PDI<sub>F1</sub>-VKORC1 and PDI<sub>F2</sub>-VKORC1) were embedded in the equilibrated and hydrated membrane composed of 200 1,2-dilauroyl-sn-glycero-3-phosphocholine (DLPC) lipids using the replacement method available in the CHARMM-GUI membrane builder (<http://www.charmm-gui.org/input/membrane>) [61]. This lipid bilayer was completed with 17293 (hVKORC1), 22047 (PDI<sub>F1</sub>-VKORC1) and 22567 (PDI<sub>F2</sub>-VKORC1) water molecules (TIP3P; [62]), pre-equilibrated during 1.5 ns of MD using the *Lipid14* tool [63] from the AMBER package.



Each protein or protein complex inserted into a membrane was solvated with explicit TIP3P water molecules in a periodic rectangular box with a distance of at least 12 Å between the proteins and the boundary of the water box. Cl<sup>-</sup> ions were randomly placed to neutralise the system.

The total number of atoms in the isolated Trx-fold proteins (protein, water molecules and counter ion) varied from 16,065–26,386. The total number of atoms in the membrane systems (hVKORC1 and its complexes with PDI, including proteins, DLPC lipids, water molecules and counter-ions), was 72683 (hVKORC1), 92570 (PDI<sub>F1</sub>-VKORC1), and 93325 (PDI<sub>F2</sub>-VKORC1). The box size varied in the range of 84 × 84 × 108–141 Å<sup>3</sup>.

#### 4.2.2. Set-Up of the Systems

The set-up of the systems was performed with the Simulated Annealing with NMR-Derived Energy Restraints (SANDER) module [64] of AMBER18. First, each system was minimised successively using the steepest descent and conjugate gradient algorithms, as follows: (i) 10,000 minimisation steps where water molecules have fixed, (ii) 10,000 minimisation steps where the protein backbone is fixed to allow protein side-chains to relax, and (iii) 10,000 minimisation steps without any constraint on the system. The equilibration was performed on the solvent, keeping the solute atoms (except H-atoms) restrained for 100 ps at 310 K and a constant volume (NVT). Protein, membrane and solvent (water and ions) temperatures were separately coupled to the velocity rescale thermostat, which was a modified Berendsen thermostat [65] with a relaxation time of 0.1 ps. Each system was equilibrated for 1 ns (NPT), with all nonhydrogen atoms of the protein and the DLPC membrane harmonically restrained. Semi-isotropic coordinate scaling and Parrinello–Rahman pressure coupling were used to maintain the pressure at 1 bar, with a relaxation time of 5 ps. The Nose–Hoover thermostat [66] was applied to the protein, lipids and solvent (water and ions) separately, with a relaxation time of 0.5 ps to keep the temperature constant at 310 K. Water and ions were allowed to move freely during equilibration.

#### 4.2.3. Production of the MD Trajectories

All trajectories were performed using the AMBER ff14SB force field with the PMEMD module of AMBER 16 and AMBER 18 [59] (GPU-accelerated versions) running on a local hybrid server (Ubuntu, LTS 14.04, 252 GB RAM, 2x CPU Intel Xeon E5-2680 and Nvidia GTX 780ti) and the supercomputer JEAN ZAY at IDRIS.

The 500-ns MD trajectories of each fully relaxed isolated protein were generated (2 replicas for Trx-fold proteins and 3 replicas for hVKORC1) in its natural environment—the water solution for the Trx-fold protein and the solvated bilayer lipid membrane for hVKORC1. Each PDI–hVKORC1 complex that was inserted into the solvated bilayer lipid membrane was simulated for an alternating value of distance S··S from PDI and hVKORC1 (see the next subsection for details). MD simulation of the Trx–VKORC1 complex was first performed for 38 ns, with a constrained S··S distance of 12.8 Å, which was further reduced to 10.2 Å and followed by simulation for 20 ns, and finally to 8.2 Å, followed by the last 20-ns of the simulation.

A time step of 2 fs was used to integrate the equations of motion based on the Leap-Frog algorithm [67]. Coordinate files were recorded every 1 ps. Neighbour searching was performed by the Verlet algorithm [68]. The Particle Mesh Ewald (PME) method [69], with a cutoff of 9.0 Å, was used to treat long-range electrostatic interactions at every time step. The van der Waals interactions were modelled using a 6–12 Lennard–Jones potential. The initial velocities were reassigned according to the Maxwell–Boltzmann distribution.

#### 4.2.4. The Stepped MD Simulations of the PDI–hVKORC1 Complex

In **Model 1** and **Model 2**, to prevent the separation of the PDI protein from hVKORC1 and to bring them together, a restrained harmonic distance was introduced to the S··S atom pair (the sulfur atoms from C37 of PDI and C43 of hVKORC1), which was varied in

a stepwise manner (see Figure 7A). Specifically, the 80-ns simulation was divided into three steps, each with different applied restraints ( $d$ ): from 0 to 38 ns with  $d$  equal to 12.8–11.8 Å (Step A), from 38 to 60 ns with  $d$  equal to 10.2–9.6 Å (Step B), and 60 to 80 ns with  $d$  equal to 9.2–8.2 Å (Step C). To probe the stability of the PDI–hVKORC1 complex, the simulations of **Model 1** and **Model 2** were continued from 80 to 100 ns with two different “soft” restraints applied to distance S··S (see Figure S13). While the lower limit value remained at 8.2 Å, as in the previous simulation steps (A–C), the upper limit in Step D was increased to 10.2 Å (as in the 60–80 ns step) and 12.8 Å (as in the 0–38 ns step).

#### 4.3. Data Analysis

##### 4.3.1. Conventional Analysis of the MD Trajectories

Unless otherwise stated, all recorded MD trajectories were analysed (RMSFs, RMSDs, DSSP, clustering) with the standard routines of the CPPTRAJ 4.15.0 program [70] of AMBER 18 Suite.

- (1) The RMSD and RMSF values were calculated for the  $C\alpha$ -atoms using the initial model (at  $t = 0$  ns) as a reference. All analysis was performed on the MD conformations (every 10 ps) by considering either all simulations or the production part of the simulation, which was generated after the removal of non-well-equilibrated conformations (0–70 ns), as was shown by the RMSDs, or on residues with a fluctuation of less than 4 Å, as shown by the RMSFs. For hVKORC1, the RMSDs were individually calculated for each domain after least-square fittings of the MD conformations to the initial conformation of a domain, thus removing rigid-body motion from the analysis.
- (2) Secondary structural propensities for all residues were calculated using the Define Secondary Structure of Proteins (DSSP) method [71]. The secondary structure types were assigned for residues based on backbone -NH and -CO atom positions. Secondary structures were assigned every 10 and 20 ps for the individual and concatenated trajectories, respectively.
- (3) The dynamic cross-correlation (DCC) between all atoms within a molecule quantifies the correlation coefficients of motions between atoms, i.e., the degree to which the atoms move together [72]. Calculations were performed on backbone  $C\alpha$ -atoms on the productive simulation time of each MD trajectory using an ensemble-based approach [28]. The correlation values vary between  $-1$  and  $1$ , where  $1$  illustrates a complete correlation,  $-1$  a complete anti-correlation and  $0$  no correlation.
- (4) The collective motions of proteins were investigated by principal component analysis (PCA). For an  $N$ -atom system, a trajectory matrix contains, in each column, Cartesian coordinates for a given atom at each time step ( $x(t)$ ). Fitting the coordinate data to a reference structure results in the proper trajectory matrix ( $X$ ). The trajectory data is then used to generate a covariance matrix ( $C$ ), elements of which are defined as in Equation (1)

$$c_{ij} = \langle (x_i - \langle x_i \rangle)(x_j - \langle x_j \rangle) \rangle \quad (1)$$

where  $\langle (x_i - \langle x_i \rangle)(x_j - \langle x_j \rangle) \rangle$  denotes an average performed over all the time steps of the trajectory.

The principal components (PCs) are obtained by a diagonalisation of the covariance matrix  $C$  (see Equation (2)).

$$C = VAV^T \quad (2)$$

This results in a diagonal matrix  $A$  containing eigenvalues as diagonal entries and a matrix  $V$  containing the corresponding eigenvectors. If the eigenvectors are sorted such that their eigenvalues are in decreasing order, the eigenvector with the largest eigenvalues (i.e., the first PCs) accounts for the highest proportion of variance within

the data. The second component is orthogonal to the first one and accounts for the second-highest proportion of variance, and so on.

- (5) The extent to which the fluctuations of a system are correlated depends on the magnitude of the cross-correlation coefficient ( $CC_{ij}$ ). The  $CC_{ij}$  of the atomic fluctuations obtained from the MD simulations ( $CC^{PCA}$ ) were computed using Equation (3):

$$CC_{ij}^{PCA} = \frac{\langle \Delta r_i^T \Delta r_j \rangle}{\langle \Delta r_i^T \Delta r_i \rangle^{1/2} \langle \Delta r_j^T \Delta r_j \rangle^{1/2}} \quad (3)$$

where  $i$  and  $j$  are two atoms of  $C\alpha$ ;  $\Delta r_i$  and  $\Delta r_j$  are displacement vectors of  $i$  and  $j$ ;  $\Delta r^T$  denotes the transpose of a column vector. If  $CC(ij) = 1$ , the fluctuations of  $i$  and  $j$  are completely correlated.

If  $CC(ij) = -1$ , the fluctuations of  $i$  and  $j$  are completely anticorrelated. If  $CC(ij) = 0$ , the fluctuations of  $i$  and  $j$  are not correlated. All snapshots were fitted using the transmembrane domain  $C\alpha$  as a reference before performing cross-correlation analysis.

The Normal Mode Wizard (NMWiz) plugin [73] of the Visual Molecular Dynamics (VMD) 1.9.3 program [74] was used to visualise the motions along with the principal components.

- (6) Clustering analysis was performed on the productive simulation time of each MD trajectory using an ensemble-based approach [28]. The first 70 ns were omitted from the analysis of Trx-fold proteins. The analysis was performed every 100 ps.

The algorithm extracts representative MD conformations from a trajectory by clustering the recorded snapshots according to their Ca-atom RMSDs. The procedure for each trajectory can be described as follows: (i) a reference structure is randomly chosen in the MD conformational ensemble, and all conformations within an arbitrary cutoff  $r$  are removed from the ensemble; this step is repeated until no conformation remains in the ensemble, providing a set of reference structures at a distance of at least  $r$ ; (ii) the MD conformations are grouped into  $n$  reference clusters based on their RMSDs from each reference structure. The cut-off was set to 2 Å for both clustered proteins or domains (Trx-fold and L-loop) to allow the comparison.

- (7) Drift analysis of helices was performed on the L-loop from h-VKORC1 using the centroids ( $C_i$ ) defined for the main-chain atoms for amino acids (aas) at the top and bottom of each helix. Positions of these centroids were monitored over the MD simulations, and their coordinates were projected on the  $x$ - $z$  and  $y$ - $z$  planes. The geometry of the  $CX_1X_2C$  motif from the Trx-fold proteins was described by the distance  $S \cdots S$  between two sulphur atoms from cysteine residues C37 and C40 and the dihedral angle determined as an absolute value of pseudotorsion angle  $S-C\alpha_{37}-C40\alpha-S$ .

- (8) H-bonds between heavy atoms ( $N$ ,  $O$ , and  $S$ ) as potential donors/acceptors were calculated with the following geometric criteria: donor/acceptor distance cut-off was set to 3.6 Å, and the bond angle cut-off was set to 120°. Hydrophobic contacts were considered for all hydrophobic residues with side chains within a distance of 4 Å of each other.

Visual inspection of the conformations and figure preparation was made with PyMOL (<https://pymol.org/2/>). The VMD 1.9.3 program [74] was used to prepare the protein MD animations. To visualise the motions along the principal components, the Normal Mode Wizard (NMWiz) plugin [73], which is distributed with the VMD program, was utilised.

#### 4.3.2. Advanced Methods of Analysis

- (1) Metric multidimensional scaling (MDS) is an algorithm for dimension reduction and visualization; it computes an embedding of a set of points (a shape trajectory in our case) in a lower dimension space with respect to the pairwise distances (Kendall's ones in our case) in the original set [19].

The algorithm consists of a minimisation of the cost (see Equation (4)):

$$\sum_{i \neq j} (d_{ij} - \|x_i - x_j\|)^2 \quad (4)$$

where  $D = (d_{ij})$  is the pairwise distance matrix, and  $\{x_i\}_i$  are the embedded points. It can be implemented using the manifold. MDS class in Python's scikit learn library.

- (2) The Fréchet mean of a set is a point minimising the sum of squared distances to each point of the set. As an example, the Fréchet mean  $\bar{T}$  of one set  $\{T_i\}_i$  of tetrahedrons is defined as in Equation (5):

$$\bar{T} \in \underset{T}{\operatorname{argmin}} \sum_i d(T, T_i)^2 \quad (5)$$

When the distance is the Euclidean distance, the Fréchet mean is no other than the classical mean we know.

- (3) Kendall's shape space of 3D triangles is isometric to the northern hemisphere of a 3D sphere of radius  $\frac{1}{2}$ , where the equilateral triangle is at the north pole [20]. We use a planar representation of the half-sphere as a disk with the equilateral triangle at the centre by the transformation  $(\varphi, \theta) \rightarrow (r = \sin(\theta), \varphi)$  from the spherical coordinates to the polar coordinates. Each 3D triangle, up to translation rotation and scaling, is represented by a unique point of the disk.

**Supplementary Materials:** Supplementary materials can be found at [www.mdpi.com/1422-0067/22/2/802/s1](http://www.mdpi.com/1422-0067/22/2/802/s1). Supplementary Material (S) Content: Table S1. Structures of the thioredoxin-fold proteins (from human) deposited in Protein Database PDB. Table S2. Sequence identity / similarity in ERp18, PDI, Tmx1 and Tmx4. Table S3. L-loop of hVKORC1: Residues involved in non-covalent intramolecular interactions and those available for intermolecular interactions. Table S4. Helical fold (top) of L-loop in isolated hVKORC1. Figure S1. Crystallographic structure of PDI. Figure S2. MD simulations of ERp18, PDI, Tmx1 and Tmx4. Figure S3. Geometry of CX<sub>1</sub>X<sub>2</sub>C motif for ERp18, PDI, Tmx1 and Tmx4. Figure S4. Hydrogen bond in CX<sub>1</sub>X<sub>2</sub>C fragment from ERp18. Figure S5. The intramolecular interactions in the CX<sub>1</sub>X<sub>2</sub>C motif with neighbour residues was characterised for the conformations regrouped into clusters (cut-off of 2.0 Å) from the concatenated MD trajectory. Figure S6. Folding of hVKORC1 over the MD simulations. Figure S7. Intrinsic motion of hVKORC1 and its L-loop. Figure S8. Clustering analysis of the L-loop conformations by using the ensemble-based clustering. Figure S9. The contact maps of pairwise distances C $\alpha$ -C $\alpha$  (< 10 Å) computed for each conformation from the mostly populated clusters (> 4%) found on the concatenated trajectory. Figure S10. The hydrophobic contacts (yellow dashed lines) in L-loop of hVKORC1. Figure S11. Comparison of the Trx-domain and L-loop from bVKOR with the PDI and L-loop from hVKORC1. Figure S12. Superposition of Model 1 (left) and of Model 2 (right) of the PDI-hVKORC1 complex into the structure of bVKOR (PDB ID: 4NV5). Figure S13. Modelling of the human PDI-hVKORC1 complex and their MD simulations. Figure S14. Intermolecular contacts at the interface between PDI and hVKORC1 in the Model 1 of the PDI-hVKORC1 complex. Figure S15. Intermolecular contacts at the interface between PDI and hVKORC1 in the Model 2 of the PDI-hVKORC1 complex.

**Author Contributions:** Conceptualization, L.T.; Methodology, L.T. and A.T.; Software, E.M. and A.T.; Formal analysis, M.S. and J.L.; Investigation, M.S., J.L., E.M., A.T., and L.T.; Sesources, M.S. and J.L.; Data Curation, M.S., J.L., and E.M.; Writing—Original Draft Preparation, A.T. and L.T.; Writing—Review and Editing, L.T.; Visualization, M.S., J.L., E.M., and L.T.; Supervision, L.T.; Project Administration, L.T.; Funding Acquisition, L.T. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by Agence Nationale de la Recherche (ANR), France (<https://anr.fr/>), grant number 18-CE20-0025-01.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** The numerical model simulations upon which this study is based are too large to archive or to transfer. Instead, we provide all the information needed to replicate the simulations. The models coordinates are available from L. Tchertanov at ENS Paris-Saclay.

**Acknowledgments:** This research was supported by Centre National de la Recherche Scientifique (CNRS), Institut Farman and Ecole Normale Supérieure (ENS) Paris-Saclay. The authors were granted access to high-performance computing (HPC) resources at the French National Computing Centre CINES (DARI A0070710973) by GENCI (Grand Equipement National de Calcul Intensif). Calculations were performed on the Jean Zay cluster at IDRIS (101081).

**Conflicts of Interest:** The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

### Abbreviations

bVKOR	Vitamin K epoxide reductase from bacteria
ERp18	Endoplasmic reticulum oxidoreductase
hVKORC1	Human vitamin K epoxide reductase
ID	Identification code
L-loop	Luminal loop
MD	Molecular dynamics
PCA	Principal component analysis
PDB	Protein Database
PDI	Protein disulphide isomerase
RMSD	Root mean square deviation
RMSF	Root mean square fluctuation
Tmx1	Thioredoxin-related transmembrane protein 1
Tmx4	Thioredoxin-related transmembrane protein 4
Trx	Thioredoxin

### References

- Hatahet, F.; Ruddock, L.W. Protein Disulfide Isomerase: A Critical Evaluation of Its Function in Disulfide Bond Formation. *Antioxidants Redox Signal.* **2009**, *11*, 2807–2850, doi:10.1089/ars.2009.2466.
- Lee, S.; Kim, S.M.; Lee, R.T. Thioredoxin and Thioredoxin Target Proteins: From Molecular Mechanisms to Functional Significance. *Antioxidants Redox Signal.* **2013**, *18*, 1165–1207, doi:10.1089/ars.2011.4322.
- Hudson, D.A.; Gannon, S.A.; Thorpe, C. Oxidative protein folding: From thiol–disulfide exchange reactions to the redox poise of the endoplasmic reticulum. *Free Radic. Biol. Med.* **2015**, *80*, 171–182, doi:10.1016/j.freeradbiomed.2014.07.037.
- Winther, J.R.; Thorpe, C. Quantification of thiols and disulfides. *Biochim. Biophys. Acta* **2014**, *1840*, 838–846, doi:10.1016/j.bbagen.2013.03.031.
- 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, doi:10.1093/nar/28.1.235.
- Chatron, N.; Chalmond, B.; Trouvé, A.; Benoît, E.; Caruel, H.; Lattard, V.; Tchertanov, L. Identification of the functional states of human vitamin K epoxide reductase from molecular dynamics simulations. *RSC Adv.* **2017**, *7*, 52071–52090, doi:10.1039/c7ra07463h.
- Martin, J.L. Thioredoxin—A fold for all reasons. *Structure* **1995**, *3*, 245–250, doi:10.1016/s0969-2126(01)00154-x.
- Dobson, C.M.; Karplus, M. The fundamentals of protein folding: Bringing together theory and experiment. *Curr. Opin. Struct. Biol.* **1999**, *9*, 92–101, doi:10.1016/s0959-440x(99)80012-8.
- Wang, C.L.; Li, W.; Ren, J.; Fang, J.; Ke, H.; Gong, W.; Feng, W.; Wang, C.-C. Structural Insights into the Redox-Regulated Dynamic Conformations of Human Protein Disulfide Isomerase. *Antioxidants Redox Signal.* **2013**, *19*, 36–45, doi:10.1089/ars.2012.4630.
- Guddat, L.W.; Bardwell, J.C.; Martin, J.L. Crystal structures of reduced and oxidized DsbA: Investigation of domain motion and thiolate stabilization. *Structure* **1998**, *6*, 757–767, doi:10.1016/s0969-2126(98)00077-x.
- Goodstadt, L.; Ponting, C.P. Vitamin K epoxide reductase: Homology, active site and catalytic mechanism. *Trends Biochem. Sci.* **2004**, *29*, 289–292, doi:10.1016/j.tibs.2004.04.004.
- Schulman, S.; Wang, B.; Li, W.; Rapoport, T.A. Vitamin K epoxide reductase prefers ER membrane-anchored thioredoxin-like redox partners. *Proc. Natl. Acad. Sci. USA* **2010**, *107*, 15027–15032.
- Wajih, N.; Hutson, S.M.; Wallin, R. Disulfide-dependent Protein Folding Is Linked to Operation of the Vitamin K Cycle in the Endoplasmic Reticulum. *J. Biol. Chem.* **2007**, *282*, 2626–2635, doi:10.1074/jbc.m608954200.

14. Tie, J.-K.; Jin, D.-Y.; Stafford, D.W. Mycobacterium tuberculosis Vitamin K Epoxide Reductase Homologue Supports Vitamin K-Dependent Carboxylation in Mammalian Cells. *Antioxid. Redox Signal.* **2012**, *16*, 329–338, doi:10.1089/ars.2011.4043.
15. Rowe, M.L.; Ruddock, L.W.; Kelly, G.; Schmidt, J.M.; Williamson, R.A.; Howard, M.J. Solution Structure and Dynamics of ERp18, a Small Endoplasmic Reticulum Resident Oxidoreductase. *Biochemistry* **2009**, *48*, 4596–4606, doi:10.1021/bi9003342.
16. Dunker, A.K.; Lawson, J.; Brown, C.J.; Williams, R.M.; Romero, P.; Oh, J.S.; Oldfield, C.J.; Campen, A.M.; Ratliff, C.M.; Hipps, K.W.; et al. Intrinsically disordered protein. *J. Mol. Graph. Model.* **2001**, *19*, 26–59, doi:10.1016/s1093-3263(00)00138-8.
17. Imai, K.; Mitaku, S. Mechanisms of secondary structure breakers in soluble proteins. *Biophysics* **2005**, *1*, 55–65, doi:10.2142/biophysics.1.55.
18. Schwartz, R.; King, J. Frequencies of hydrophobic and hydrophilic runs and alternations in proteins of known structure. *Protein Sci.* **2006**, *15*, 102–112, doi:10.1110/ps.051741806.
19. Kendall, D.G. Shape Manifolds, Procrustean Metrics, and Complex Projective Spaces. *Bull. Lond. Math. Soc.* **1984**, *16*, 81–121, doi:10.1112/blms/16.2.81.
20. Dryden, I.L.; Mardia, K.V. *Statistical Shape Analysis: With Applications in R*, 2nd ed.; John Wiley & Sons: Hoboken, NJ, USA, 2016; 496p.
21. Kortemmea, T.; Creighton, T.E. Ionisation of Cysteine Residues at the Termini of Model  $\alpha$ -Helical Peptides. Relevance to Unusual Thiol pKa Values in Proteins of the Thioredoxin Family. *J. Mol. Biol.* **1995**, *253*, 799–812, doi:10.1006/jmbi.1995.0592.
22. Xu, S.; Sankar, S.; Neamati, N. Protein disulfide isomerase: A promising target for cancer therapy. *Drug Discov. Today* **2014**, *19*, 222–240, doi:10.1016/j.drudis.2013.10.017.
23. Dyson, H.J.; Jeng, M.-F.; Tennant, L.L.; Slaby, I.; Lindell, M.; Cui, D.-S.; Kuprin, A.S.; Holmgren, A. Effects of Buried Charged Groups on Cysteine Thiol Ionization and Reactivity in *Escherichia coli* Thioredoxin: Structural and Functional Characterization of Mutants of Asp 26 and Lys 57. *Biochemistry* **1997**, *36*, 2622–2636, doi:10.1021/bi961801a.
24. Pinitglang, S.; Noble, M.; Verma, C.; Thomas, E.W.; Brocklehurst, K. Studies on the enhancement of the reactivity of the (Cys-25)-S-(His159)-Im+H ion-pair of papain by deprotonation across pKa 4. *Biochem. Soc. Trans.* **1996**, *24*, 468s, doi:10.1042/bst024468s.
25. Tie, J.-K.; Stafford, D.W. Structural and functional insights into enzymes of the vitamin K cycle. *J. Thromb. Haemost.* **2016**, *14*, 236–247, doi:10.1111/jth.13217.
26. Tie, J.-K.; Stafford, D.W. Functional Study of the Vitamin K Cycle Enzymes in Live Cells. *Methods Enzymol.* **2017**, *584*, 349–394, doi:10.1016/bs.mie.2016.10.015.
27. Hatahet, F.; Boyd, D.; Beckwith, J. Disulfide bond formation in prokaryotes: History, diversity and design. *Biochim. Biophys. Acta* **2014**, *1844*, 1402–1414, doi:10.1016/j.bbapap.2014.02.014.
28. Lyman, E.; Zuckerman, D.M. Ensemble-Based Convergence Analysis of Biomolecular Trajectories. *Biophys. J.* **2006**, *91*, 164–172, doi:10.1529/biophysj.106.082941.
29. Pylaeva, S.; Brehm, M.; Sebastiani, D. Salt Bridge in Aqueous Solution: Strong Structural Motifs but Weak Enthalpic Effect. *Sci. Rep.* **2018**, *8*, 1–7, doi:10.1038/s41598-018-31935-z.
30. Pace, C.N.; Fu, H.; Fryar, K.L.; Landua, J.; Trevino, S.R.; Shirley, B.A.; Hendricks, M.M.; Iimura, S.; Gajiwala, K.; Scholtz, J.M.; et al. Contribution of Hydrophobic Interactions to Protein Stability. *J. Mol. Biol.* **2011**, *408*, 514–528, doi:10.1016/j.jmb.2011.02.053.
31. Rishavy, M.A.; Usualieva, A.; Hallgren, K.W.; Berkner, K.L. Novel Insight into the Mechanism of the Vitamin K Oxidoreductase (VKOR). *J. Biol. Chem.* **2011**, *286*, 7267–7278, doi:10.1074/jbc.m110.172213.
32. Bevans, C.G.; Krettler, C.; Reinhart, C.; Watzka, M.; Oldenburg, J. Phylogeny of the Vitamin K 2,3-Epoxy Reductase (VKOR) Family and Evolutionary Relationship to the Disulfide Bond Formation Protein B (DsbB) Family. *Nutrients* **2015**, *7*, 6224–6249, doi:10.3390/nu7085281.
33. Davis, S.J.; Davies, E.A.; Tucknott, M.G.; Jones, E.Y.; Van Der Merwe, P.A. The role of charged residues mediating low affinity protein-protein recognition at the cell surface by CD2. *Proc. Natl. Acad. Sci. USA* **1998**, *95*, 5490–5494.
34. Slagle, S.P.; Kozack, R.E.; Subramaniam, S. Role of Electrostatics in Antibody-Antigen Association: Anti-Hen Egg Lysozyme/Lysozyme Complex (HyHEL-5/HEL). *J. Biomol. Struct. Dyn.* **1994**, *12*, 439–456, doi:10.1080/07391102.1994.10508750.
35. Nelson, C.A.; Viner, N.J.; Young, S.P.; Petzold, S.J.; Unanue, E.R. A negatively charged anchor residue promotes high affinity binding to the MHC class II molecule I-Ak. *J. Immunol.* **1996**, *157*, 755–762.
36. Stenlund, P.; Lindberg, M.J.; Tibell, L.A.E. Structural Requirements for High-Affinity Heparin Binding: Alanine Scanning Analysis of Charged Residues in the C-Terminal Domain of Human Extracellular Superoxide Dismutase†. *Biochemistry* **2002**, *41*, 3168–3175, doi:10.1021/bi011454r.
37. Schreiber, G. Kinetic studies of protein-protein interactions. *Curr. Opin. Struct. Biol.* **2002**, *12*, 41–47, doi:10.1016/s0959-440x(02)00287-7.
38. Wade, R.C.; Gabdouliline, R.R.; Lüdemann, S.K.; Lounnas, V. Electrostatic steering and ionic tethering in enzyme-ligand binding: Insights from simulations. *Proc. Natl. Acad. Sci. USA* **1998**, *95*, 5942–5949.
39. Haberland, J.; Gerke, V. Conserved charged residues in the leucine-rich repeat domain of the Ran GTPase activating protein are required for Ran binding and GTPase activation. *Biochem. J.* **1999**, *343*, 653–662.
40. Unkles, S.E.; Rouch, D.A.; Wang, Y.; Siddiqi, M.Y.; Glass, A.D.M.; Kinghorn, J.R. Two perfectly conserved arginine residues are required for substrate binding in a high-affinity nitrate transporter. *Proc. Natl. Acad. Sci. USA* **2004**, *101*, 17549–17554.
41. Zhao, N.; Pang, B.; Shyu, C.-R.; Korkin, D. Charged residues at protein interaction interfaces: Unexpected conservation and orchestrated divergence. *Protein Sci.* **2011**, *20*, 1275–1284, doi:10.1002/pro.655.

42. De Planque, M.R.R.; Bonev, B.B.; Demmers, J.A.A.; Greathouse, D.V.; Ii, R.E.K.; Separovic, F.; Watts, S.A.A.; Killian, J.A. Interfacial Anchor Properties of Tryptophan Residues in Transmembrane Peptides Can Dominate over Hydrophobic Matching Effects in Peptide–Lipid Interactions†. *Biochemistry* **2003**, *42*, 5341–5348, doi:10.1021/bi027000r.
43. Chandler, D. Interfaces and the driving force of hydrophobic assembly. *Nat. Cell Biol.* **2005**, *437*, 640–647, doi:10.1038/nature04162.
44. Mohan, A.; Oldfield, C.J.; Radivojac, P.; Vacic, V.; Cortese, M.S.; Dunker, A.K.; Uversky, V.N. Analysis of Molecular Recognition Features (MoRFs). *J. Mol. Biol.* **2006**, *362*, 1043–1059, doi:10.1016/j.jmb.2006.07.087.
45. Tompa, P.; Davey, N.E.; Gibson, T.J.; Babu, M.M. A Million Peptide Motifs for the Molecular Biologist. *Mol. Cell* **2014**, *55*, 161–169, doi:10.1016/j.molcel.2014.05.032.
46. Van Der Lee, R.; Buljan, M.; Lang, B.; Weatheritt, R.J.; Daughdrill, G.W.; Dunker, A.K.; Fuxreiter, M.; Gough, J.; Gsponer, J.; Jones, D.T.W.; et al. Classification of Intrinsically Disordered Regions and Proteins. *Chem. Rev.* **2014**, *114*, 6589–6631, doi:10.1021/cr400525m.
47. Wright, P.E.; Dyson, H.J. Intrinsically disordered proteins in cellular signalling and regulation. *Nat. Rev. Mol. Cell Biol.* **2015**, *16*, 18–29, doi:10.1038/nrm3920.
48. Wong, E.T.C.; So, V.; Guron, M.; Kuechler, E.R.; Malhis, N.; Bui, J.M.; Gsponer, J. Protein–Protein Interactions Mediated by Intrinsically Disordered Protein Regions Are Enriched in Missense Mutations. *Biomolecules* **2020**, *10*, 1097, doi:10.3390/biom10081097.
49. Dosztányi, Z.; Chen, J.; Dunker, A.K.; Simon, I.; Tompa, P. Disorder and Sequence Repeats in Hub Proteins and Their Implications for Network Evolution. *J. Proteome Res.* **2006**, *5*, 2985–2995, doi:10.1021/pr060171o.
50. Vangone, A.; Bonvin, A.M.J.J. Contacts-based prediction of binding affinity in protein–protein complexes. *eLife* **2015**, *4*, e07454, doi:10.7554/elife.07454.
51. Keskin, O.; Gursoy, A.; Ma, B.; Nussinov, R. Principles of Protein–Protein Interactions: What are the Preferred Ways for Proteins To Interact? *Chem. Rev.* **2008**, *108*, 1225–1244, doi:10.1021/cr040409x.
52. Keskin, O.; Ma, B.; Nussinov, R. Hot Regions in Protein–Protein Interactions: The Organization and Contribution of Structurally Conserved Hot Spot Residues. *J. Mol. Biol.* **2005**, *345*, 1281–1294, doi:10.1016/j.jmb.2004.10.077.
53. Dill, K.A.; Bromberg, S.; Yue, K.; Chan, H.S.; Ftebig, K.M.; Yee, D.P.; Thomas, P.D. Principles of protein folding—A perspective from simple exact models. *Protein Sci.* **2008**, *4*, 561–602, doi:10.1002/pro.5560040401.
54. Tsai, C.-J.; Lin, S.L.; Wolfson, H.J.; Nussinov, R. Studies of protein–protein interfaces: A statistical analysis of the hydrophobic effect. *Protein Sci.* **1997**, *6*, 53–64, doi:10.1002/pro.5560060106.
55. Tsai, C.-J.; Ma, B.; Sham, Y.Y.; Kumar, S.; Nussinov, R. Structured disorder and conformational selection. *Proteins Struct. Funct. Bioinform.* **2001**, *44*, 418–427, doi:10.1002/prot.1107.
56. Tsai, C.-J.; Nussinov, R. Hydrophobic folding units at protein–protein interfaces: Implications to protein folding and to protein–protein association. *Protein Sci.* **1997**, *6*, 1426–1437, doi:10.1002/pro.5560060707.
57. Webb, B.; Sali, A. Comparative Protein Structure Modeling Using MODELLER. *Curr. Protoc. Bioinform.* **2016**, *54*, 5.6.1–5.6.37, doi:10.1002/cpbi.3.
58. Laskowski, R.A. PDBsum: Summaries and analyses of PDB structures. *Nucleic Acids Res.* **2001**, *29*, 221–222, doi:10.1093/nar/29.1.221.
59. Case, D.A.; Iii, T.E.C.; Darden, T.; Gohlke, H.; Luo, R.; Merz, K.M., Jr.; Onufriev, A.; Simmerling, C.; Wang, B.; Woods, R.J. The Amber biomolecular simulation programs. *J. Comput. Chem.* **2005**, *26*, 1668–1688, doi:10.1002/jcc.20290.
60. Maier, J.A.; Martinez, C.; Kasavajhala, K.; Wickstrom, L.; Hauser, K.E.; Simmerling, C. ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. *J. Chem. Theory Comput.* **2015**, *11*, 3696–3713, doi:10.1021/acs.jctc.5b00255.
61. Jo, S.; Kim, T.; Iyer, V.G.; Im, W. CHARMM-GUI: A web-based graphical user interface for CHARMM. *J. Comput. Chem.* **2008**, *29*, 1859–1865, doi:10.1002/jcc.20945.
62. Jorgensen, W.L.; Chandrasekhar, J.; Madura, J.D.; Impey, R.W.; Klein, M.L. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **1983**, *79*, 926–935, doi:10.1063/1.445869.
63. Dickson, C.J.; Madej, B.D.; Skjevik, Å.A.; Betz, R.M.; Teigen, K.; Gould, I.R.; Walker, R.C. Lipid14: The Amber Lipid Force Field. *J. Chem. Theory Comput.* **2014**, *10*, 865–879, doi:10.1021/ct4010307.
64. Salomon-Ferrer, R.; Case, D.A.; Walker, R.C. An overview of the Amber biomolecular simulation package. *WIREs Comput. Mol. Sci.* **2013**, *3*, 198–210, doi:10.1002/wcms.1121.
65. Peters, E.; Goga, N.; Berendsen, H.J.C. Stochastic Dynamics with Correct Sampling for Constrained Systems. *J. Chem. Theory Comput.* **2014**, *10*, 4208–4220, doi:10.1021/ct500380x.
66. Evans, D.J.; Holian, B.L. The Nose–Hoover thermostat. *J. Chem. Phys.* **1985**, *83*, 4069–4074, doi:10.1063/1.449071.
67. Van Gunsteren, W.F.; Berendsen, H.J.C. A Leap-frog Algorithm for Stochastic Dynamics. *Mol. Simul.* **1988**, *1*, 173–185, doi:10.1080/08927028808080941.
68. Grubmüller, H.; Heller, H.; Windemuth, A.; Schulten, K. Generalized Verlet Algorithm for Efficient Molecular Dynamics Simulations with Long-range Interactions. *Mol. Simul.* **1991**, *6*, 121–142, doi:10.1080/08927029108022142.
69. Hockney, R.W.; Eastwood, J.W. *Computer Simulation Using Particles*; McGraw-Hill: New York, NY, USA, 1981.
70. Roe, D.R.; Cheatham, I.T.E. PTRAJ and CPPTRAJ: Software for Processing and Analysis of Molecular Dynamics Trajectory Data. *J. Chem. Theory Comput.* **2013**, *9*, 3084–3095, doi:10.1021/ct400341p.

71. Kabsch, W.; Sander, C. Dictionary of protein secondary structure: Pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* **1983**, *22*, 2577–2637, doi:10.1002/bip.360221211.
72. Hünenberger, P.; Mark, A.; Van Gunsteren, W. Fluctuation and Cross-correlation Analysis of Protein Motions Observed in Nanosecond Molecular Dynamics Simulations. *J. Mol. Biol.* **1995**, *252*, 492–503, doi:10.1006/jmbi.1995.0514.
73. Bakan, A.; Meireles, L.M.; Bahar, I. ProDy: Protein Dynamics Inferred from Theory and Experiments. *Bioinformatics* **2011**, *27*, 1575–1577, doi:10.1093/bioinformatics/btr168.
74. Humphrey, W.; Dalke, A.; Schulten, K. VMD: Visual molecular dynamics. *J. Mol. Graph.* **1996**, *14*, 33–38, doi:10.1016/0263-7855(96)00018-5.