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

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

Article

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 and Luba Tchertanov[*]

Université Paris-Saclay, ENS Paris-Saclay, CNRS, Centre Borelli, 4 av. des Sciences, F-91190 Gif-sur-Yvette, France

# The first two authors contributed equally
* Correspondence: Luba.Tchertanov@ens-paris-saclay.fr

Received: 30 November 2020; Accepted: date; Published: date

Abstract: Redox (reduction–oxidation) reactions control many important biological processes in all organisms, both prokaryotes and eukaryotes. This reaction is usually accomplished by canonical disulfide-based pathways involving a donor enzyme that reduces the oxidized cysteine residues of a target protein resulting in the cleavage of its disulfide bonds. Focusing on the human vitamin K epoxide reductase (hVKORC1) as a target and on the four redoxins (PDI, ERp18, Tmx1 and Tmx4) as the most probable reducers of VKORC1, a comparative *in silico* analysis is provided that concentrates on the similarity and divergence of redoxins in their sequence, secondary and tertiary structure, dynamics, intra-protein interactions and composition of the surface exposed to the target. Similarly, the hVKORC1 is analysed in the 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 the Protein Disulfide Isomerase (PDI) is suggested as the most probable hVKORC1 partner. By probing the alternative orientation of PDI with respect to hVKORC1, the functionally related non-covalent complex formed by hVKORC1 and PDI is found, which is proposed as a first precursor to probe the thiol-disulfide 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

1. Introduction

Thioredoxins (Trxs) are disulfide 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 archaebacteria 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_1X_2C motif. These two cysteines separated by two residues play a key role in the transfer of the two hydrogen atoms to the oxidised target, and the breaking of the Trx disulfide bond (Figure 1A). This disulfide-relay pathway is accompanied by an electron transfer in the opposite direction. An intermediate state during the electron transfer is a mixed disulfide 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 disulfide is exchanged within one thiol oxidoreductase or between a disulfide donor and a target protein [3]. Thiol-disulfide exchange reactions occur between redox-sensitive biomolecules if donors and acceptors can interact in the appropriate orientations when attacking and leaving groups [4].

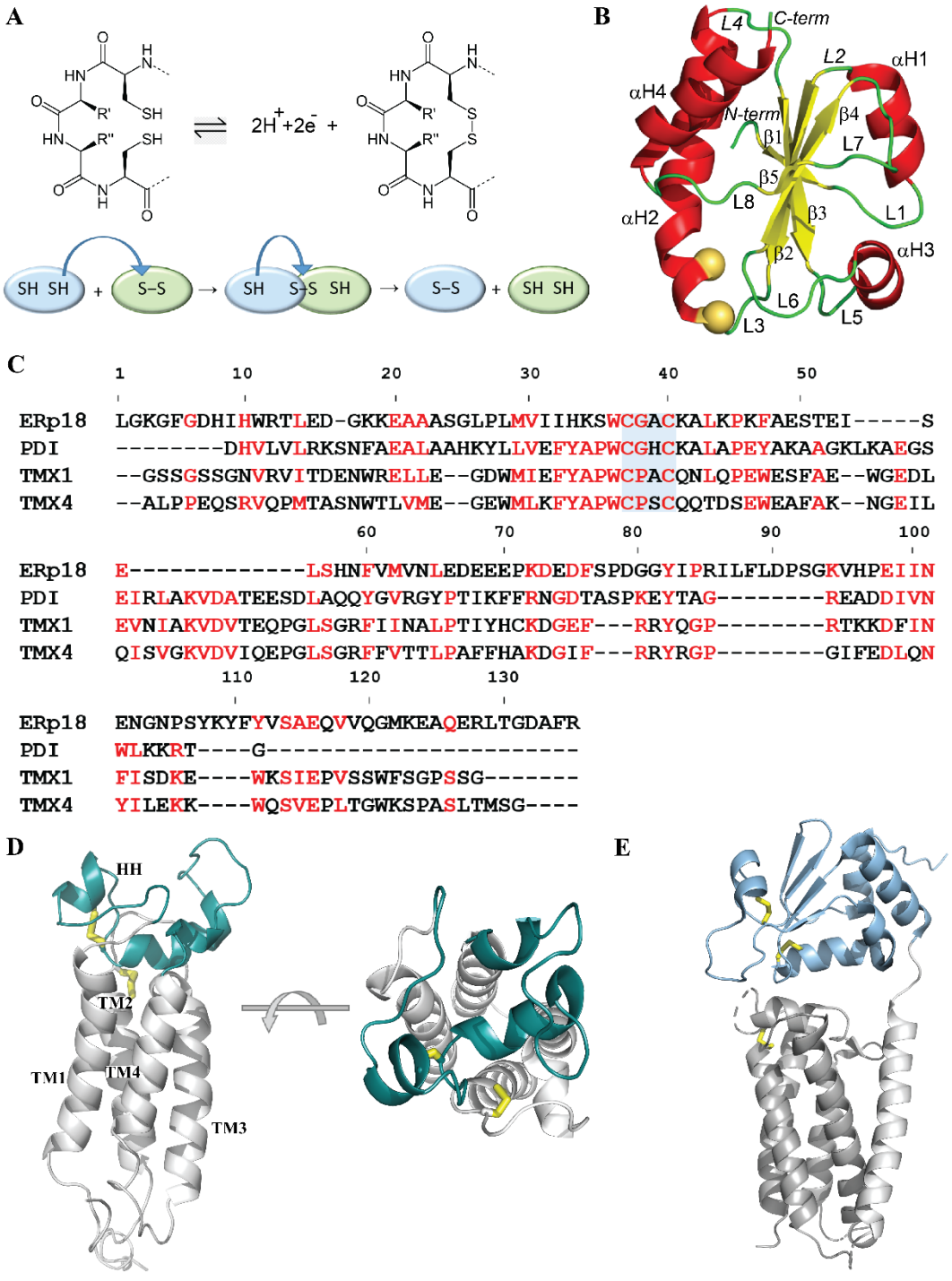

Figure 1. Thioredoxin-fold protein as a physiological reductant of the human vitamin K epoxide reductase complex 1 (hVKORC1). (**A**) Oxidation of the two cysteine residues in the CX_1X_2C motif of Trx-fold protein forms for a disulfide bond, a process associated with the loss of two hydrogen atoms and hence two electrons (top). Mechanism of disulfide exchange between Trx and a target (bottom). The H-donor enzyme and a target are respectively coloured in blue and green. (**B**) The Trx fold is illustrated using the X-ray structure of the human PDI deposed in PDB [5] (PDB ID: 4ekz). The protein is shown as red ribbons, with two cysteine residues from the CX_1X_2C motif as yellow balls. The four α-helices (in red), five β-

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://espript.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 non conserved residues are in black. Blue highlights the CX_1X_2C 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 disulfide 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 fragment that has the sequence most similar to hVKORC1 and the Trx-like domain are respectively shown in dark grey and light blue. The disulfide 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 and is carefully described in [7]. It is illustrated with the crystallographic structure of the human Protein Disulfide Isomerase (PDI) (Figure 1B**)**, which is the best characterized enzyme that assists in the process of oxidative folding [8, 9]. The PDI structure consists of a central five-stranded propeller with four flanking α-helices, an architecture that contains extra regions compared to the classical thioredoxin fold (a four-stranded β-sheet and three α-helices formed by about 80 residues).

The dithiol/disulphide group in the CX_1X_2C motif, which is located at the head of the αH2 helix, protrudes from the protein surface and is exposed to a solvent. Such a spatial arrangement of the CX_1X_2C motif is probably to ensure the full accessibility of the first cysteine, which is required to react with cysteine residue of a target to accomplish redox processes. It was reported that the reactive thiolate of this first cysteine can be stabilized by the positive dipole at the head of the αH2 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 the 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 is involved in several vital physiological and homeostasis processes [11].

Recently, 3D models of the human VKORC1 (hVKORC1) have been reported along with functionally-related enzymatic states [6]. The models that were generated of the metastable states of hVKORC1 and their validation through *in silico* and *in vitro* screening has led to the concept of a plausible mechanism for enzymatic reactions based on a sequential array of the 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 factors controlling the specificity of hVKORC1 recognition by the Trx protein? What is the exact role of thioredoxin(s) in initiating the hVKORC1 reduction?

Since the physiological reductant of hVKORC1 has not yet been identified, initial exploration was made of four human redoxin proteins – namely the protein disulfide isomerase (PDI), the endoplasmic reticulum oxidoreductase (ERp18), the thioredoxin-related transmembrane protein 1 (Tmx1) and the thioredoxin-related transmembrane protein 4 (Tmx4) – reported to be the most probable H-donors of VKOR [12, 13]. These proteins have distinct compositions for the active site CX_1X_2C − CGHC in the PDI, CGAC in ERp18, CPAC in Tmx1 and CPSC in Tmx4 − and they show a broad, but distinct substrate specificity. The nature of this specificity is the main focus of this work. In order to evaluate that one is the most likely to reduce hVKORC1, a detailed comparison of these redoxins is first provided at different levels of the protein's organisation − sequence, secondary and tertiary structure, intrinsical dynamics, intra-protein interactions governing structural and conformation properties, and composition of the surface exposed to the targets. Second, the hVKORC1 in the native state, in which two pairs of cysteine residues form two disulphide bridges (Figure 1D), was studied as a target of Trx-

fold proteins, with the aim of identifying the anchor site(s) that enable to recognize / bind its Trx effector. Finally, modelling of the complex formed by hVKORC1 and PDI, that was suggested as the most probable partner of VKORC1, was carried out using the PDI fragments predicted as 'interacting' as a guide and the VKOR structure from *Synechococcus sp* (bVKOR) (ID PDB: 4NV5) [14] (Figure 1E) as an initial reference. The model of the molecular non-covalent complex formed by PDI and hVKORC1 (PDI-hVKORC1) is proposed as a first human precursor useful for the probing of the thiol-disulfide exchange reactions between redoxin 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 the human VKORC1 in the inactive (oxidised) state, and on the modelling of the molecular non-covalently 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 followed questions: (i) Are the 1D, 2D and 3D properties and the dynamic features good indicators for prediction of the protein fragments participating in the hVKORC1 recognition by a Trx ? (ii) From *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 atomistic level, the recognition mechanisms between Trx and hVKORC1, which is 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 the 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) (PDB, [5] were used to extract the coordinates of a domain containing the CX_1X_2C motif (Table A1; Figure A1). This domain was chosen for 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 A2) 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 the Q9H1E5 (https://www.uniprot.org/uniprot/) along with the Tmx1 structure as a template.

The ERp18, PDI and Tmx1 empirical structures and the Tmx4 homology model were optimized (when necessary) to obtain a CX_1X_2C 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 during the two MD replicas and (ii) significant disparity between the different proteins (Figure 2A). The comparable RMSDs for PDI over each replica and between the replicas characterise a highly stable protein structure during the simulation. Similar to PDI, the RMSD values for ERp18, Tmx1 and Tmx4 vary within a narrow range after elimination of the largest amplitude N/C-terminal residues. This demonstrates a 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 α-helices and β-strands. These ordered structures are interconnected by coiled linkers to form a stable globular 3D arrangement which is described as a four or five-stranded antiparallel β-sheet sandwiched between four α-helix-bundle structures, which is an

archetypical fold of the Trx family of proteins **(**Figure 2 B). Similar to the RMSDs, the root mean square fluctuations (RMSFs) agree well between the two trajectories of each protein (Figure 2C; Figure A2). The most pronounced difference in RMSF between the two replicas is only observed in ERp18, in which β5 is partially unfolded and the L7 and L8 loops join together, resulting in large fluctuations.

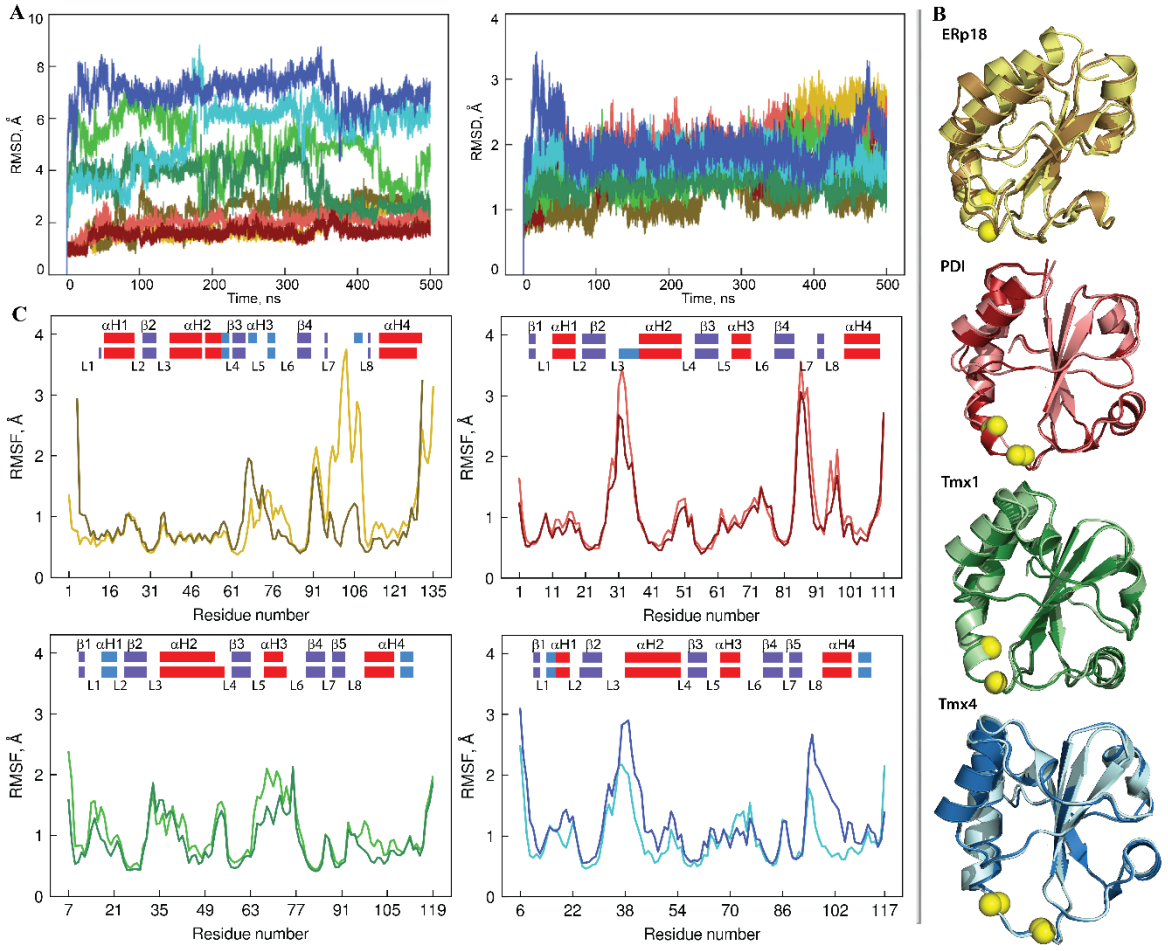

Figure 2. Characterisation of the MD simulations for the four Trx-fold proteins − ERp18, PDI, Tmx1 and Tmx4. (**A**) RMSDs from the initial coordinates computed for all Cα-atoms (right) in each protein after fitting to initial conformation. (**B**) The superimposed average structures of each protein over replicas **1** and **2**. 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 Cα atoms using the RMSF amplitude values less than 4 Å for the MD conformation of each protein after fitting to the initial conformation. The highly fluctuating residues (3, 6 and 5 aas in ERP18, Tmx1 and Tmx4, respectively) were excluded from the RMSD computation. In the insert, the secondary structures, αH- (red), 3_{10}-helices (light blue) and β-strand (dark blue), were assigned for a mean conformation of every MD trajectory, **1** (top) and **2** (bottom), of each protein and they were labelled as in the crystallographic structure of the human PDI. (**A-C**) Proteins are distinguished by colour (1^{st}/2^{nd} replicas) − ERp18 (yellow/brown), PDI (light/dark red), Tmx1 (light/dark green) and Tmx4 (light/dark blue). Numbering of the residues in each Trx-fold protein is arbitrary and started from the first amino acid in the 3D model.

Further characterization of each protein and comparison between the proteins will be frequently completed by the observations obtained for a randomly chosen single trajectory or for a concatenated data. This is because the RMSDs and RMSFs in the two 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 αH2-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, α-, 3_{10}-helices and β-strands, involve 55, 60, 60 and 56% of the residues in ERp18, PDI, Tmx1 and Tmx4, respectively, where the helical and β-strand fold portions vary respectively from 36 to 42% and from 13 to 22% of the total folding. Though all ordered 2D structures, helices and strands, are generally conserved across the studied proteins, their positions, lengths and qualities (e.g., α- or 3_{10}-helix) are slightly different (Figure 2; Figure A2).

The helical fold of each protein is represented by α-helices of different length (of 7-18 residues) and by the 3_{10}-helices that consist of 3-4 residues. H1, which is a long-lived α-helix in ERp18 and PDI, is transient and it converts between α- and the 3_{10}-helix in Tmx1 and Tmx4. H2, which is the longest α-helix (14-18 residues) that contains the CX_1X_2C 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_1X_2C motif is different in the four proteins and this represents a part of the regular α-helix (ERp18 and Tmx1), a transient helix conversed between the α- 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 α-helix. H4 is the long and stable α-helix in ERP18 and PDI, while in Tmx1 and Tmx4, it is folded as the shorter α-helix and is joint to a 3_{10}-helix.

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

Also, the atomistic RMS fluctuations of the studied proteins show (i) the minimal RMSF values of all β-strands forming the antiparallel β-sheet in all proteins, while the helices may have discernable fluctuations (e.g., αH2 and αH3 in Tmx1, and αH2 in Tmx4), and as was expected, (ii) the strong differences in the fluctuations of the coiled linkers (Figure 2C). These linkers, which interconnected the core β-stands and surround α-helices, are the most variable elements in the studied proteins in terms of sequence composition, length and conformation. It is also noted that moderate (order of 1.5-2.5 Å), but systematically observed, fluctuations of fragment L5-αH3-L6 arose in all studied proteins. This fragment is structurally adjacent to the CX_1X_2C motif and may play a role in the thiol-disulfide exchange reactions.

2.1.3. Intrinsic motion and its interdependence in Trx-folded proteins

Since a protein's dynamics influence its functional properties, intrinsic motions of Trx-fold proteins were compared. First, the cross-correlation map was computed for all Cα-atom pairs of each protein (Figure 3A). The positively correlated motion of β2-, β3- and β4-strands, which was observed in each studied protein, reflects their concerted movement in the β-barrel. To equilibrate the structural stability, the other fragments in Trx-fold proteins displayed a motion that tends to correlate negatively. As such, in ERp18, in addition to the β-barrel coupled motion, the structural moieties with the strongest correlation are L7 and αH4. In PDI, a regular fractal-like pattern showed a correlated motion of αH1 with αH2 and L7, and of αH3 with L7 and L8. In Tmx1 the coupled motion is observed between αH2-helix and αH4-helix, and between β2-stand and αH3-helix. The Tmx4 demonstrates correlated motions between αH2-helix and β3-strand, and between αH3-helix and β4/β5-strands.

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 three eigenvectors. This indicated that the Tmx1 (green) and Tmx4 (blue) conformations were grouped in a unique compact region for each protein and these regions were perfectly superimposed for the 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). The 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 to

displacement of the H3-helix in ERp18, and with a disparity in the H2-helix length in PDI (data not shown).

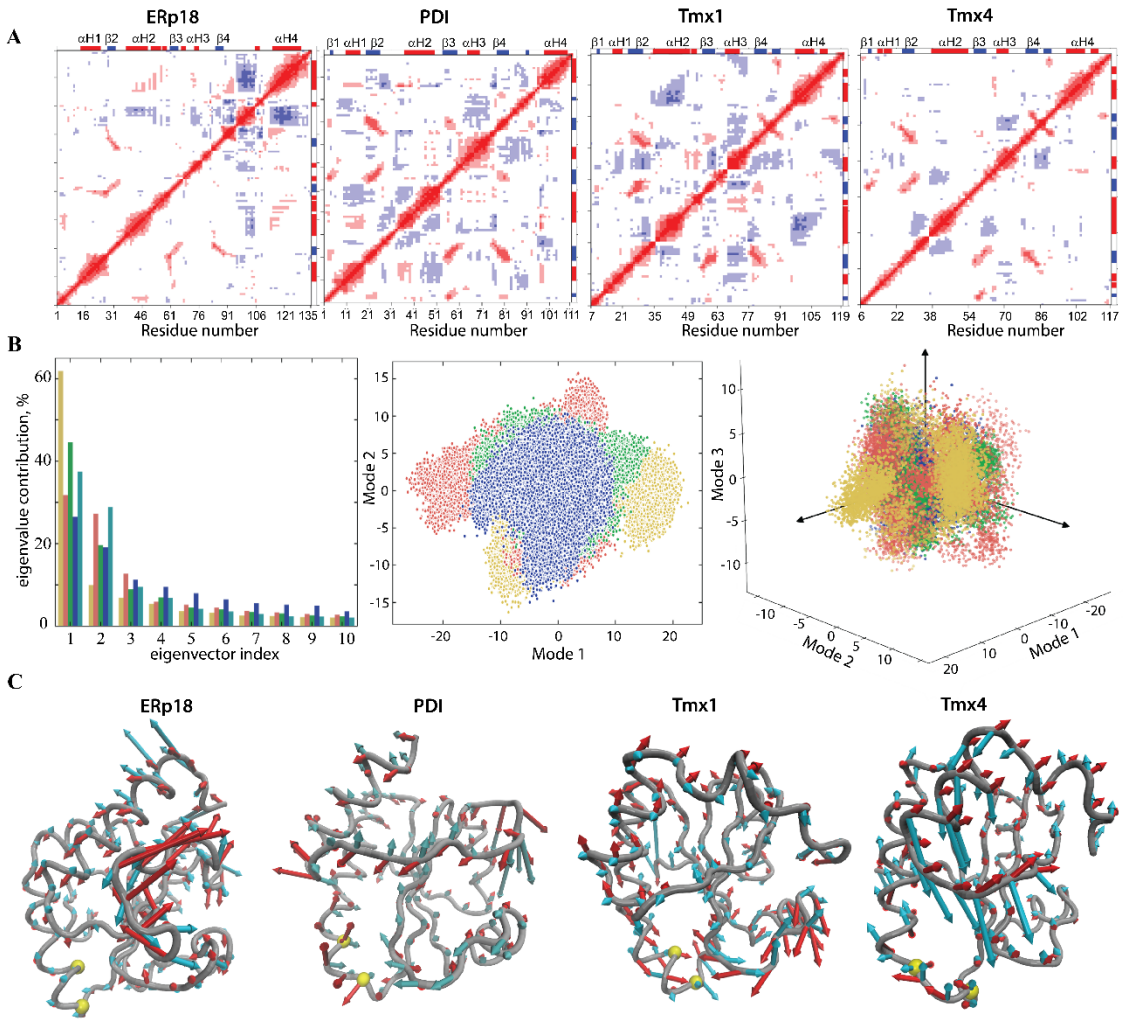

Figure 3. **Intrinsic motion in the Trx-folded proteins and its interdependence.** (**A**) Inter-residue cross-correlation maps computed for the Cα-atom pairs of ERp18, PDI, Tmx1 and Tmx4 after the fitting procedure. Secondary structure projected onto the protein sequences (α-helix/β-strand in red/blue) is shown at the border of matrices. Correlated (positive) and anti-correlated (negative) motions between Cα-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 the ERp18, PDI, Tmx1 and Tmx4 MD conformations with the principal component (PC) in 2D (middle) and in 3D subspace (right). The 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 and cyan arrows projected on a tube representation of each protein. For clarity, only motion with an amplitude ≥ of 2 Å was represented. Cysteine residues are shown as yellow balls. All computations were performed on the Cα-atoms with the RMSF fluctuations less than 4 Å for each protein after fitting on the initial conformation.

From the ten calculated PCA modes describing ~95% of the total backbone fluctuations of each Trx-fold protein, the first two most dominant modes were used to illustrate qualitatively the ample collective

movements (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 join together due to unfolding of the β5 strand. In PDI and in Tmx4, the L7 and L8 loops are well separated by the β5 strand, but each of them shows the coupled motion of a large amplitude. Uniquely, in Tmx1, the αH3 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 the target recognition and/or in the electron transfer reaction

To compare the four Trx-like proteins regarded as the probable functional effectors of hVKORC1, the focus was on two fragments that may be involved in the target recognition and/or in the electron transfer reaction. The first fragment, ***F1***, comprises L3 and a N-extremity of αH2-helix that includes the CX_1X_2C motif and the second, ***F2***, which is structurally adjacent to the CX_1X_2C motif, is composed of L5-αH3-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, similarly to a bacterial protein containing a Trx-fold domain 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 the segments of a comparable length in different proteins, the boundaries of fragments were chosen so that their length was equal (17 residues) (Figure 4).

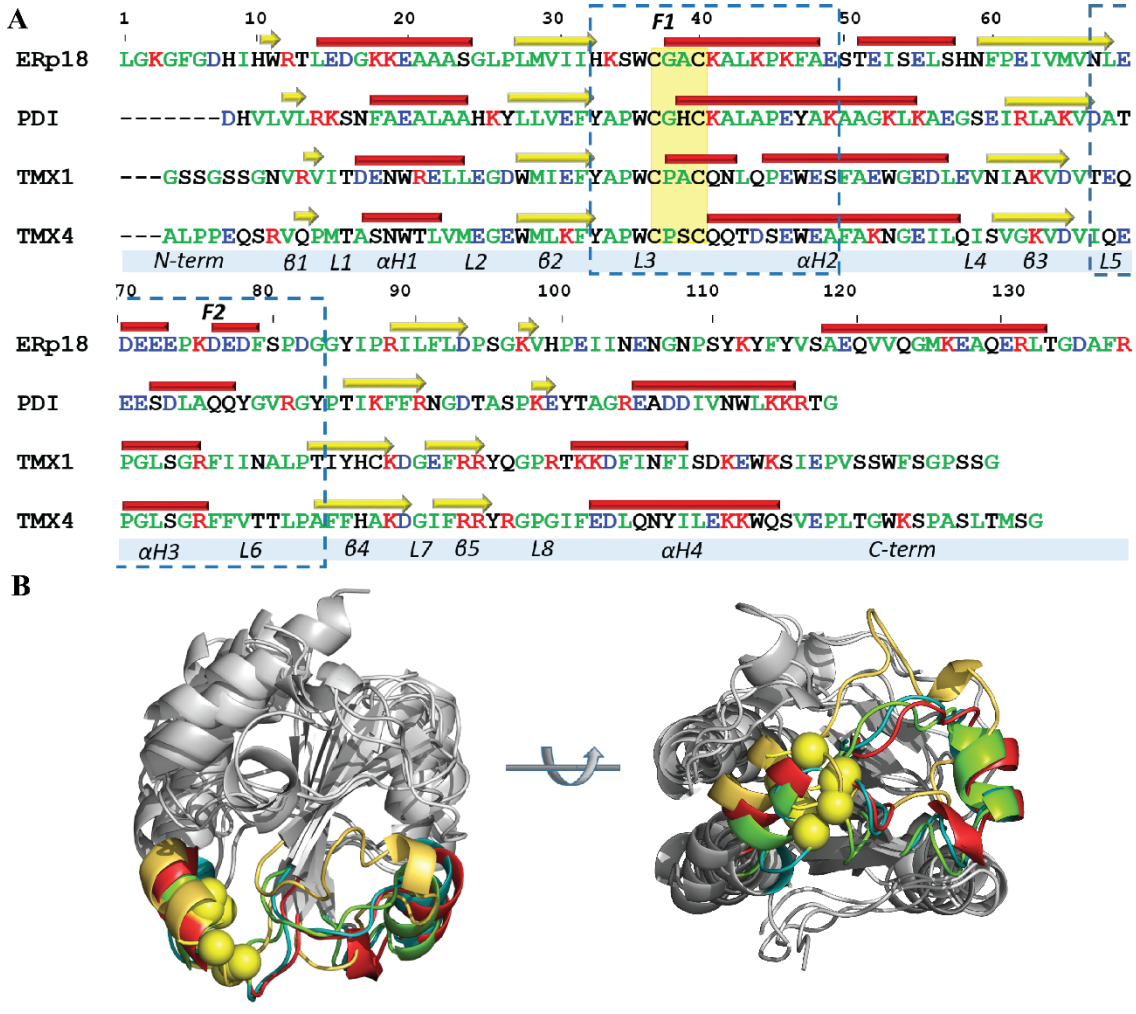

Figure 4. The sequence and folding of the Trx-like proteins. (**A**) Alignment of the sequences and the secondary structure assigned to a mean conformation of the concatenated trajectory of each studied protein. Residues are colored according to their properties – the positively and negatively charged residues are respectively in red and blue; the hydrophobic residues are in green; the polar and amphipathic residues are in black; the CX_1X_2C motif is highlighted by a yellow background. The α-helices and β-strands are shown as red batons and yellow arrows respectively. The secondary structure labelling is shown below the Tmx4 sequence. (**B**) The superimposed 3D structures of the Trx-fold proteins are shown in two orthogonal projections. The proteins are drawn as ribbons with the cysteine residue as yellow balls. The ***F1*** and ***F2*** regions that are potentially involved in the target recognition and/or in the electron transfer reaction are outlined by dashed lines in (**A**) and differentiated by color in (**B**) to distinguish between the proteins − ERp18 (dark yellow), PDI (red), Tmx1 (green) and Tmx4 (dark blue).

The ***F1*** region, in PDI, Tmx1 and Tmx4, is similarly initiated by tyrosine, which is the residue reported to be a breaker of the 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 (AP), are perfect conserved in PDI, Tmx1 and Tmx4, while in ERp18 these positions are occupied by positively charged and polar residues (KS). Furthermore, the specific CX_1X_2C motif for each studied protein is preceded by tryptophan (W), which is the highly conserved residue in the four proteins. Tryptophan is an amphipathic residue that, similar to the tyrosine, is often found at the surface of proteins and is sometimes also classified as polar.

It is suggested that the ***F1*** region of Trx-folded proteins that contains the CX_1X_2C motif contributes to the redox reactions rather than to the target recognition. Nevertheless, a double action of the ***F1*** fragment as both the redox agent and the recognition platform for a target is not excluded.

The second surface region of Trx-fold proteins, ***F2***, which is in the proximity of the CX_1X_2C motif, consists of the α-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 and αH3-helix, and of the length of the αH3-helix. In ERp18 a set of the five negatively charged amino acids (EDEEE), which are positioned on L5 and αH3-helix, are separated by proline (P) and lysine (K) from the other three negatively charged amino acids (DED). This promotes a breakup of the H3-helix into two small 3_{10}-helices. In the other proteins, the number of the negatively charged residues in this region is diminished to four in PDI and to one in Tmx1 and Tmx4. The two last proteins, Tmx1 and Tmx4, have the same αH3-helix content and differ only in the combination of the amino acids in L5. Despite a great difference in the αH3-helix composition of PDI compared to that of Tmx1 and Tmx2, the length of helix in the three proteins is equivalent (6 aas). In all studied proteins, the short loop L5 contains at least one negatively charged residue and one polar residue, while the extended L6 loop is mainly composed of hydrophobic residues enriched by one or two polar residues with an inserted charged amino acid – the negative in ERp18 and the positive in PDI.

As the αH3-helix is moving, considerately in Tmx1 and moderately in the other proteins (Figure 3C), it is suggested that the αH3-helix can adapt its orientation to give 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 the 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 the apolar 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_1X_2C motif, is a fragment of a Trx-fold protein that can contribute to VKORC1 recognition.

2.1.5. Geometry of the CX_1X_2C motif

Focusing on the CX_1X_2C motif, a key agent in the thiol-disulfide exchange reactions, its geometry was characterised in each Trx-fold protein. It was observed that structurally, the CX_1X_2C motif constitutes either a part of the αH2-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_1X_2C, while the other cysteine is buried into the protein chain.

Surprisingly, the folding of the CX_1X_2C (CGAC) motif in the calculated conformations (MD simulation) of ERp18 is coherent with those observed in the experimentally determined structure (PDB ID: 1sen), despite of the different protein states – reduced (MD simulation) with two protonated thiol groups, and oxidized (X-Ray analysis) in which two deprotonated thiol groups form a disulfide bridge. In both protein states studied using two different methods, the first cysteine from the CGAC motif is the N-cap residue (the last non-helical residue) of the α-H2 helix.

The second unexpected observation is connected to the different folding of the CX_1X_2C (CGHC) motif in the calculated (MD simulation) and empirical structures (X-Ray analysis) of PDI studied in the same state (reduced). Indeed, the CX_1X_2C motif in the crystallographic structure of PDI was reported as folded with the C37 positioned at the cap of the α-H2 helix (PDB ID: 4ekz), while in the MD conformations the structure of this motif is transient and alternated between the helical fold (α- or 3_{10}-helices) and the turn/coiled structure, demonstrating a high conformational plasticity.

The folding of the CX_1X_2C motif in Tmx1 (CPAC) in the MD conformations and in the NMR structures (PDP ID: 1x5e) (both are in a reduced state) is equivalent, with the first cysteine as a N-cap residue of the downstream α-H2 helix, similarly to ERp18. In Tmx4, a protein with the most similar sequence to Tmx1, the CX_1X_2C motif (CPSC), demonstrates a coiled structure. In these two proteins the conserved proline constitutes the characteristic CPX_2C motif, and the observed structural differences may be connected either to the X_2 residue or to the long-distance structural effects.

The geometry of the CX_1X_2C motif was described by two metrics, a distance S···S′ between the protonated sulphur atoms and a dihedral angle S−Cα−Cα′−S′ (Figure 5A; Figure A3). 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 the sulphur atoms that promoted a slight increase in the S···S distance. Such a restrained geometry of the CX_1X_2C motif in Erp18 and Tmx1 is apparently related to its location on the well-folded αH2-helix. By contrast, the CX_1X_2C motif located on a coiled L3 in Tmx4 stimulates a highly divergent orientation of the sulphur atoms, running from syn-periplanar to anti-periplanar configuration, as was evidenced by the large variation in the dihedral angle S−Cα−Cα−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 the PDI conformations corresponds to the transient structure of the N-terminal of H2-helix, conversed between the helical fold (α- and 3_{10}-helices) and turn structure.

To better characterise the dynamical behaviour of the CX_1X_2C motif over two trajectories for each protein and compare between 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 $3k-7$of all possible configurations of k atoms in $\mathbb{R}^3$ 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 Multi-Dimensional Scaling (MDS), to analyse the shape trajectories [20]. These methods offer various ways of visualizing the data all together in a common space, summarizing them with a reduced number of variables, and comparing them with each other. A tetrahedron defined for the S- and Cα-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 weakly overlapping (clearly

visible 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.

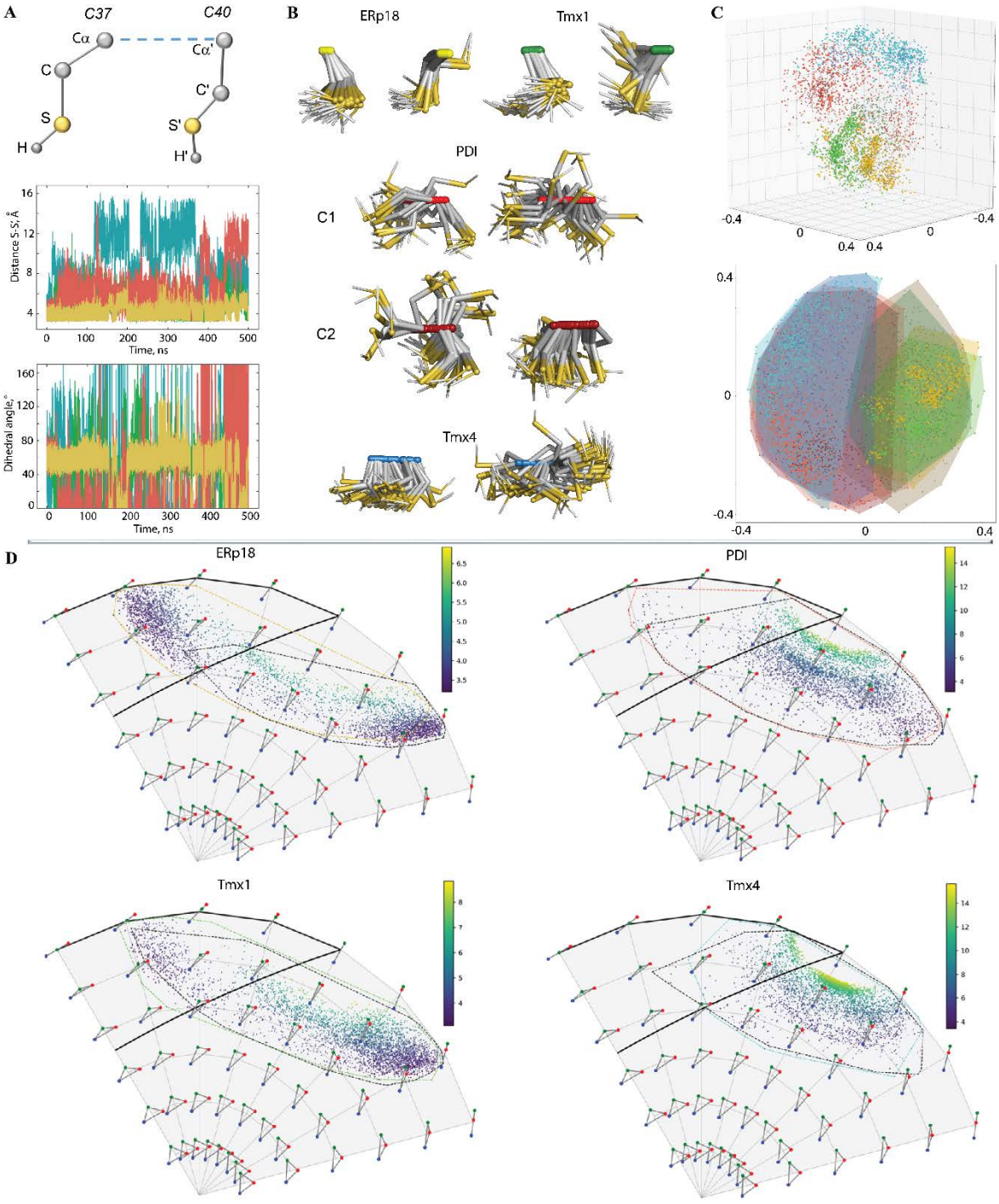

Figure 5. The CX_1X_2C motif geometries for ERp18, PDI, Tmx1 and Tmx4. (**A**) Geometry of CX_1X_2C motif (left) is described by distance S···S′ (middle) and dihedral angle (right) determined as an absolute value of the pseudo torsion angle S−Cα(C37)−Cα′(C40)−S′. Only one replica **2** was shown. (**B**) Superposition of the thiol groups (Cα-C-S-H) from the CX_1X_2C motif of each protein showed either for only one MD trajectory (ERP18, Tmx1 and Tmx4) or for both (PDI). Samples were taken for each 100 ns frame. (**C**) MDS in 2D and 3D on the set of S-C-C-S tetrahedrons. Embedded points have been colored 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 the S···S distance. Representative triangles are regularly sampled on the disk. The thick

blackline delimits the area of conformations favouring the H-bond interaction. The dashed areas are contouring the sub-populations according the S-atom being the H-donor.

This analysis was illustrated by superposition of the thiol groups (Cα-S-H) from the CX_1X_2C motif of the MD conformations for each protein (Figure 5B). The orientation of the thiol groups favours the H-bond interaction (S-H···S) only in ERp18 and in some Tmx1 conformations. In PDI, both the thiol groups are shown to have the most variant orientation within a group, as well as between the groups, which reflects their high mobility.

The H-bond between the sulphur atoms of each cysteine was characterized 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 (Figure 5D; Figure A4). Monitoring of a geometry of S-H···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···S and angle at H-atom) indicated that a typical S-H···S H-bond is slightly stronger in Tmx1 with respect to ERp18. Such a H-bond was not observed in the other studied proteins.

As expected, the S-H···S H-bond does not influence the folding of the CX_1X_2C motif (Figure A5). 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_1X_2C motif is well folded in both cases. It is interesting that both thiol groups do not contribute to the H-bond interaction in the most prevalent PDI conformation with an unfolded CX_1X_2C motif, but K41, which is next to the C40 residue, is H-bound to H39 and to L43. In the folded CX_1X_2C 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 apparently 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 in the fully unfolded Tmx4 motif.

Structure organisation of CX_1X_2C motif influences strongly their reactivity, affecting such properties as their accessibility and protonation state (i.e., pK_a) [21]. Functional analyses of each cysteine in the consensus CX_1X_2C motif demonstrated that the N-terminal cysteine is important for the formation of a transient S-S bond with the substrate whereas the C-terminal cysteine is involved in the substrate release [22]. In proteins, specific hydrogen-bond donors and an electropositive local environment tend to lower the pK_a by stabilizing the thiolate, and a hydrophobic environment or an electronegative local environment tends to raise the pK_a by destabilizing a negatively-charged as opposed to neutral form of the side chain [21, 23, 24].

2.2. The human VKORC1 viewed as the target of a Trx-fold protein

2.2.1. General characterisation

The hVKORC1 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 disulfide bridges S···S (Figure 1D). The hVKORC1 was studied by MD simulations of the model that mimics the protein in its natural environment, i.e. hVKORC1 is 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.

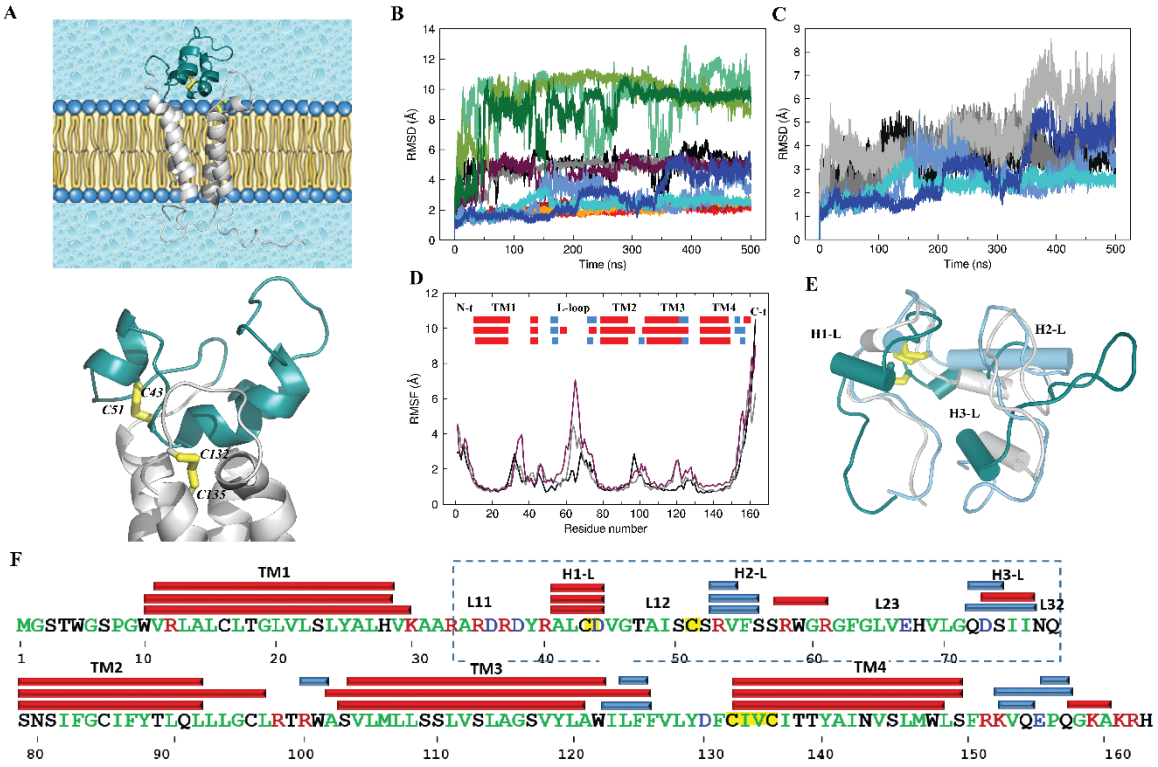

Figure 6. The hVKORC1 in the inactive state and its conventional MD simulations. (**A**) 3D model of hVKORC1 in its inactive state as it was inserted into the membrane (top) and zoomed in on the L-loop (bottom). The L-loop is highlighted by the colour teal, disulfide bridges formed by cysteine residues C43-C51 and C132-C135 are drawn as yellow sticks. The transmembrane helices (TM) are numbered as in [6]. (**B-C**) RMSDs computed for each MD trajectory (**1**-**3** replica) from initial coordinates (at t=0 ns, the same for all replicas) on the Cα-atoms of the 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**); and 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 the Cα-atoms of the MD conformations (**1**-**3** replicas) 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, αH- (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 – 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, the residues C43, C51 and the CX_1X_2C motif are highlighted by a yellow background. The α- and 3_{10}-helices are shown as red and blue batons, respectively. The 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 Cα-atoms relative to the initial structure (t=0 ns) showed comparable behaviours 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 the large RMSD values (up to 13 Å), while the TMD curves demonstrate a highly stable profile with

the smallest RMSDs (2 Å). The RMSDs computed for the Cα-atoms of the L-loop after fitting to its initial conformation, showed alternated values, small or large, that was maintained during 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. The three conformations of the L-loop that were chosen from replica **3** at t = 150, 250 and 375 ns showed significant differences in folding and in orientation of the helices and of the loops, which revealed the structural and conformational transitions (Figure 6E).

The enlarged RMSD values computed for the Cα-atoms of L-loop, after the coordinate fitting for the initial conformation of the TMD with respect to the RMSDs computed after the coordinate fitting for the initial conformation of L-loop, suggest the displacement of L-loop from the TMD as a pseudo-rigid body (Figure 6C). The profile of the RMSF curves is similar in 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 structure of VKORC1 is generally conserved over the MD trajectories and showed a fully helical fold of the protein with the four long-living extended (of 15-19 residues) transmembrane α-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 L-loop is the fragment targeted by a Trx-fold protein, the focus is mainly on its intrinsic structural and dynamical properties and their connection with those of the transmembrane domain of hVKORC1.

The L-loop folding, encompassing 30, 36 and 22% of all residues in **1**-**3** replicas respectively, is presented by three small (3-4 residues) transient helices, H1-L, H2-L and H3-L, which are partially converted between the αH- and 3_{10}-helices (Figure 6E; Figure A6). Despite the transient structure of helices, their positions on the sequence are well conserved. The L-loop helices are interconnected by the coiled linkers, which together with the linker joining L-loop to TM1 from the transmembrane domain of hVKORC1, display the RMSF values that suggests a high mobility of these loops (Figure 6D). The H1-L, mainly folded as a regular α-helix, contains at its C-cap the C43 that is linked covalently to C51, a N-cap residue of H2-L helix, forming the S···S bridge between two cysteines. Such a covalent bonding restricts significantly the conformational mobility of this fragment. The large coiled linker connecting H2-L and H3-L helices is composed of the 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 Cα-atom pairs of the full-length protein and of the L-loop. The Cα-Cα distance pairwise patterns demonstrate the coupled motions within each hVKORC1 domain, L-loop and TMD, and between two structural domains (Figure 7A; Figure A7). The regular pattern in TMD reflects the correlated motion of the TM helices that is mainly associated with their collective drift, early observed in all metastable states of hVKORC1 [6]. The motion of L-loop correlates with the movement of the linkers connecting the TM-helices and joining the L-loop to TMD.

The cross-correlations computed on only the L-loop atoms display different maps in the three replicas, with either a poor pattern (replicas **1** and **2**) or a pattern composed of well-defined blocks of a 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 the cross-correlation patterns is associated to the disparity of the L-loop motion, a small or medium in replicas **1** and **2**, and a broad in **3**, as evidenced by RMSFs and PCA.

The collective motions of VKORC1, characterized by the PCA, showed that ten modes describe ~80-90% of the total fluctuations of both the full-length VKORC1 and 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). The PCA analysis performed on only the Cα-atoms of 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 differs in the three trajectories (Figure 7C), suggesting a larger conformational space for the L-loop than was observed in each trajectory, and probably also larger than the total space of all trajectories. The first two modes in replicas 1 and **2** showed a highly coupled motion of H1-L helix and 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 rest of the L-loop and vertical with respect to the TMD (Figure 7C, Figure A7).

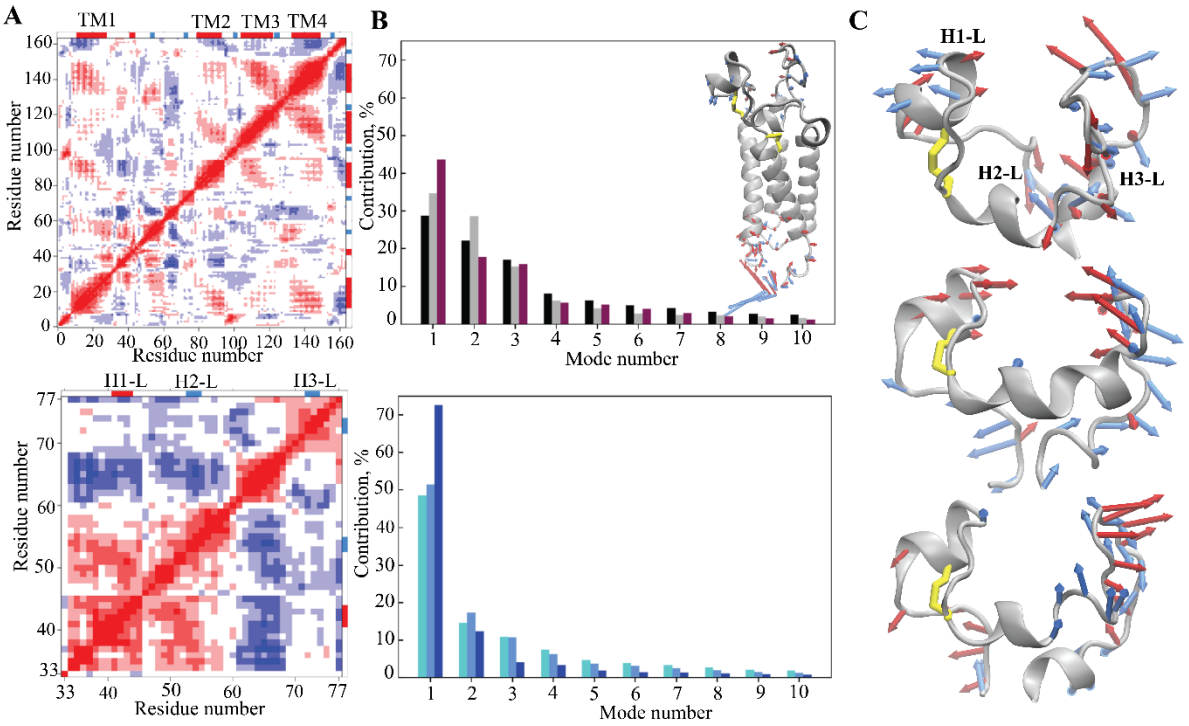

Figure 7. **Intrinsic motion of hVKORC1 and its L-loop.** (**A**) The inter-residue cross-correlation map computed for the Cα-atom pairs after fitting to the respective first conformation (t=0ns) of the full-length hVKORC1 (top) and of the L-loop (bottom) is shown for the **3** replicas. Correlated (positive) and anti-correlated (negative) motions between the Cα-atom pairs are shown as a red-blue gradient. (**B**) The PCA modes of the full-length hVKORC1 (top) and of 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 respectively colour black, grey and rose brown, 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 ≥ 2 Å was shown. The S-S bridge of hVKORC1 is shown using yellow sticks.

To characterise the conformational changes of the L-loop 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 TMD and was then used to define two tetrahedrons, **T1** and **T2**, designed on the Cα-atoms. It is suggested that light may be shed

on the conformational features of L-loop by analysis of the six straight edges corresponding to the distances between each pair of residues.

Analysis of the **T1** geometry showed (i) a great stability of the Cα-Cα 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) a high conservation of the Cα-Cα distances between each of three residues and S74 (H3-L) during a substantial time period (200-300 ns or more), followed by (iii) a synchronic change of these distances (Δ 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, shows a less conserved geometry that displays the synchronic changes of all or at least 3-4 distances (Δ of 8-15 Å).

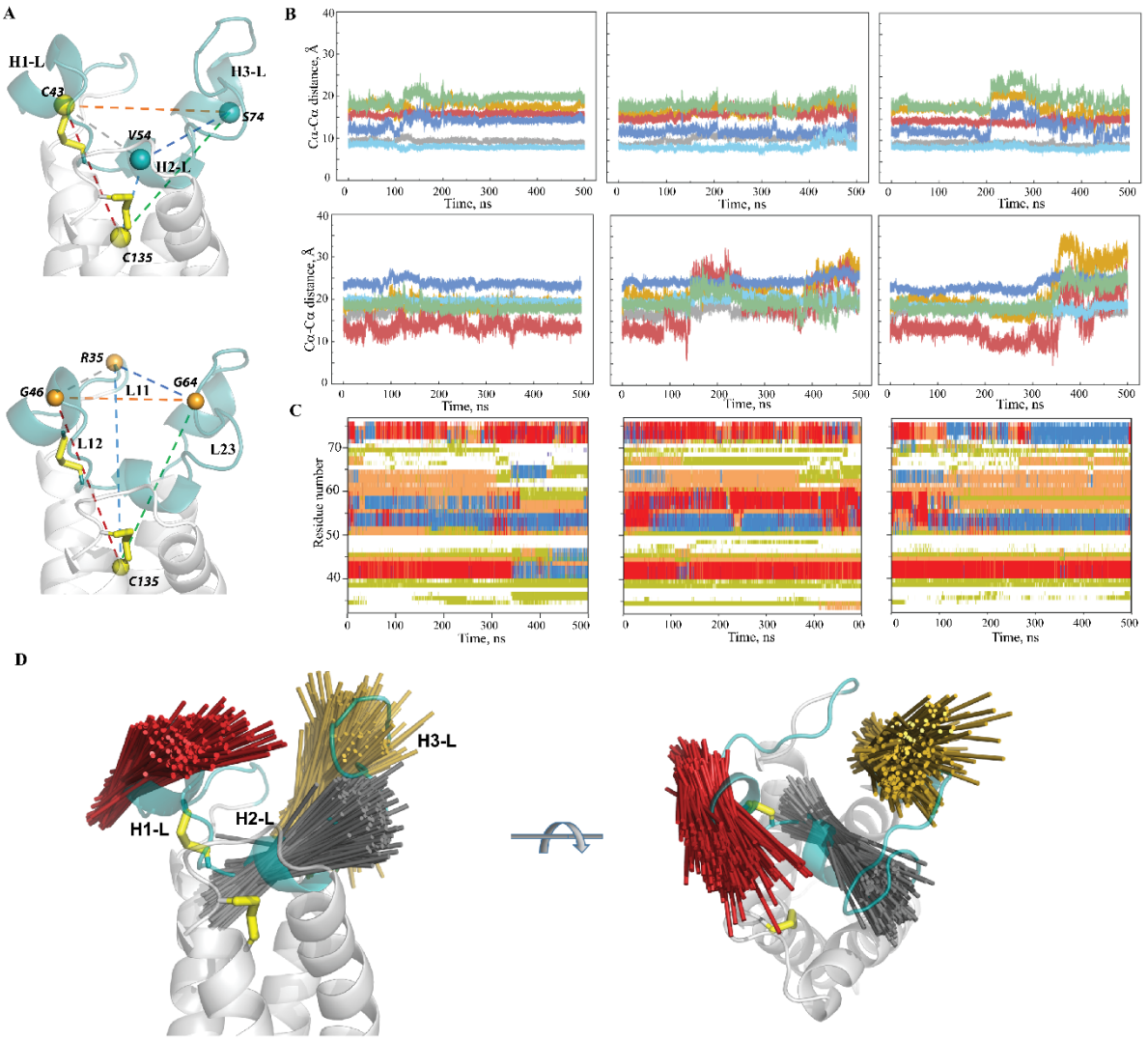

Figure 8. Geometry and folding of the L-loop from hVKORC1 in the inactive state. (**A**) Two tetrahedrons, **T1** – defined for the Cα-atom of C135 and for the midpoint residues of each L-loop helix, and **T2** – defined for the Cα-atom of C135 and for the most fluctuating residues (with the greatest RMSF values) from the L-loop likers. (**B**) Distances between each pair of Cα-atoms from the tetrahedron **T1** (top) and **T2** (bottom) over each MD trajectory. The distance curves and the edges of a tetrahedron are coloured similarly. (**C**) The time-depended evolution of the secondary structure of each residue as assigned by DSSP: α-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 L-loop are covered on the randomly chosen conformation of hVKORC1 in two orthogonal projections.

The axis of each helix was 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) a high stability of the H1-L helix in terms of the secondary structure, as well as in its relative position with respect to TMD, (ii) the quasi-stable spatial position of the transient H2-L helix relative to H1-L and to TMD, (iii) the large displacement of 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 L-loop (Figure 8D). The superimposed axes (elongated by 50% to better represent their position and direction) form a reap-like distribution for all helices. The axes of three helices differ in length and in their spatial orientation within each reap-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 the inactive state and to distinguish the most probable conformations, the generated conformations were analysed using the 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 ≥ 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 in 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 non-zero population in all replicas (Figure A8). 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 two mostly populated clusters, C1 (48%) and C2 (32%); the other conformations are regrouped in clusters with a low population (of 9-0.5%). 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 only in trajectory **3**, 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 folding level (2D) and in the 3D structure organisation (Figure A8). 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 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. Systematically, 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), that confirms the overlapping of the conformational spaces of the L-loop covered over the three replicas of MD simulation of the hVKORC1 in the inactive state.

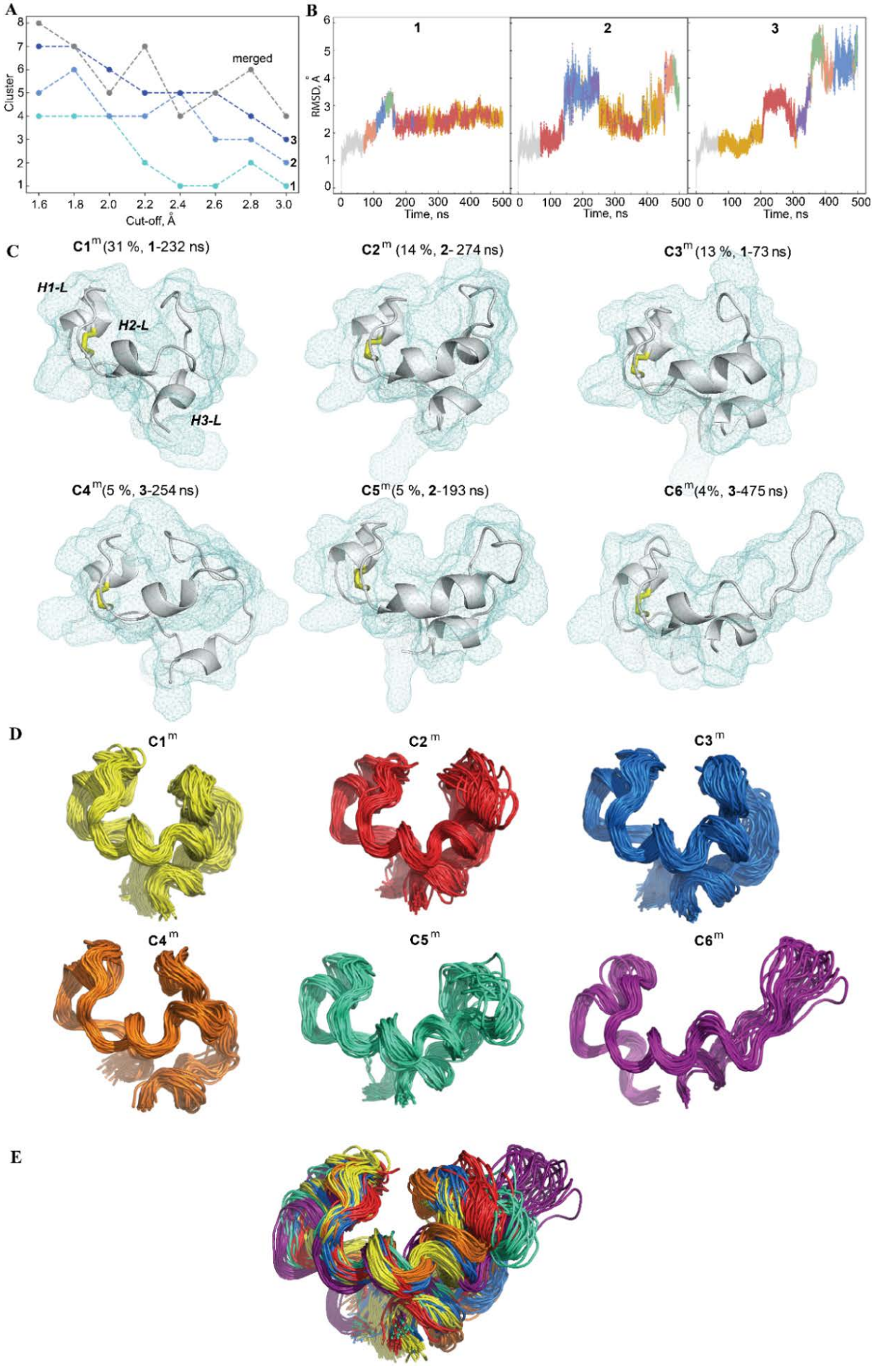
A
Cluster
Cut-off, Å
merged
3
2
1
B
RMSD, Å
Time, ns
1
2
3
C
C1m (31 %, 1-232 ns)
C2m (14 %, 2- 274 ns)
C3m (13 %, 1-73 ns)
C4m (5 %, 3-254 ns)
C5m (5 %, 2-193 ns)
C6m (4%, 3-475 ns)
H1-L
H2-L
H3-L
D
C1m
C2m
C3m
C4m
C5m
C6m
E

Figure 9. E**nsemble-based clustering of the L-loop MD conformations**. (**A**) Number of clusters obtained for each MD trajectory (**1**, **2** and **3**) and for 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 **1**-**3** trajectories. 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**) The *representative conformations* of the L-loop from the clusters (C^m) 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 and with the disulfide bridge 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 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 the 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 of 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 conformations* of each clusters generated for the concatenated trajectory showed that the principle factors leading to the conformational difference of L-loop consist of (i) a variable length of H2-L helix, and a decrease of that promotes (ii) an elongation of linker L23, which in turn, encourages (iii) 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 reflects well 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 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 the elongated shape with the '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 the MD conformations of the L-loop into groups composed of similar geometry and shape (within a cut-off), (ii) the assembly of the great majority of conformations in a limited number of clusters and (iii) the distinction between the dense clusters with a statistically reasonable population.

2.2.4. The intra-L-loop interactions

To establish the forces that stabilise the L-loop conformations, the 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 A**9**). Nevertheless, the patterns of such contacts are differed in clusters $C1^m$-$C6^m$. The most common pattern found in the maps describes the contacts 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 the mutual H-bonds that form the extensive networks (Figure 10, Table A3). 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 which form 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 stabilized by pairing of the charged residues when a combination of two non-covalent 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 the H2-L and L11 linker. In conformations from the cluster C3^{m}, the 'salt bridge pattern' is formed by R61 interacting with D36 and E67, which stabilises the tight location of two distant linkers, L11 and L23.

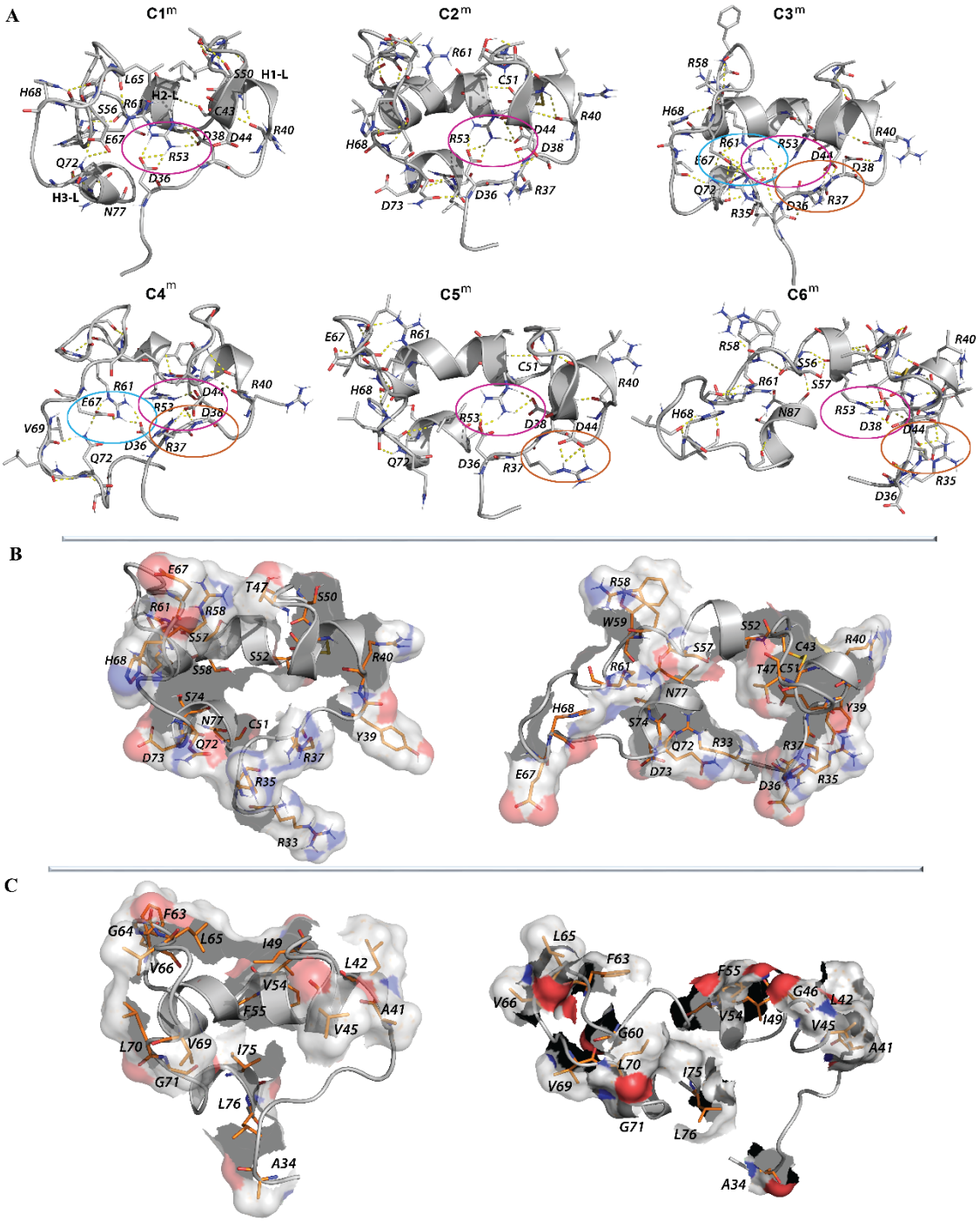

Figure 10. Interacting residues in the L-loop conformations. **(A)** The intra-loop H-bond interactions in the L-loop conformations from clusters $C1^m$ –$C6^m$. The H-bonds D-H···A (D···A < 3.6 Å, ∠DHA ≥ 120°), where D and A are H-donor and H-acceptor (O/N) atoms, were analysed in a *representative conformation* from each cluster of the merged trajectories. Interactions that stabilised the helices were not considered. The L-loop is shown as ribbons, with the interacting residues as sticks and H-bond traces as dashed lines. The common H-bonding motifs are encircled by magenta (at R53), blue (at R61) and orange (at R37). The most characteristic donor and acceptor groups are labelled. The N, O and C atoms are in blue, red and grey respectively. (**B**) The charged and polar residues protruded outside of the L-loop. (**C**) The hydrophobic residues protruded outside of the L-loop. The L-loop is shown as ribbons, with the residues exposed to the solvent displayed as sticks with a space filling encountering. In (**B**-**C**), the N, O and C atoms are in blue, red and orange respectively.

These interactions in $C3^m$ 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. Additionally, in $C3^m$ the other 'salt bridge pattern' is formed by R37 contacting with D44, which stabilises the H1-L and L12 loop in a tight spatial position. In $C4^m$, 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 $C4^m$, similar to $C3^m$, 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 L-loop from $C5^m$ and $C6^m$ clusters in which two 'salt bridge patterns' are formed by R53 (H2-L) interacting with D36 and D38 ($C5^m$), or with D38 and D44 ($C6^m$), and either by R37 (L11) interacting with R40 and D44 ($C5^m$), or by R35 bound to R35 and D38 ($C6^m$).

Besides the salt-bridge interactions, the charged residues are also contributing to H-bonds by interaction with the different polar and hydrophobic residues, which either act as the 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 are interacting in any conformation of L-loop, independent of the L-loop's shape, by forming either the salt bridges or the H-bonds.

Nevertheless, the many charged and polar residues that are not involved or are partially involved in the intra-L-loop interactions, are protrude outside of 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 the L-loop cysteine residues (C43 and C51) that are participating in the thiol-disulfide exchange reaction, the residues from sequence S52 to E67 are most likely involved in interactions with a redox protein.

As the L-loop is also comprised of a large number of hydrophobic residues, their contribution to intra- and inter-molecular 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 involve V45, F55, F63, L70 and L76 (Figure A10, Table A3). These hydrophobic contacts may reflect the stabilising interactions which complete the H-bond contribution, as well as the repulsive forces that eqilibriate 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, at that the number of such residues is significantly higher in the 'closed' conformations than in the 'open' (Figure 10C, Table A3**)**.

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.

3.1. 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 Discussion). 3D models of the complex were constructed using the crystallographic structure of the VKOR from bacteria (bVKOR) (PDB code: 4NV5) [14] as a reference for the initial positioning of PDI relative to hVKORC1.

In order to be the most objective in the modelling of the human PDI-VKORC1 complex, the structure of bVKOR was not used as a template because (i) of 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) of a high structural difference between the Trx- and L-loop domains in bVKOR and in human proteins (RMSD values are 4.5 and 4 Å between the bVROR and the 'closed' and 'open' conformations of hVKORC1 respectively), (iii) of a very low sequence identity/similarity (15/20%), and (iv) of 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 A11).

For modelling of 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 to bring the two proteins as close as possible. As for the initial PDI model, the conformation with a well-ordered and long αH2-helix that is similar to the X-ray structure of PDI [9] was chosen and positioned above the hVKORC1 so that (i) the distance between the sulphur atoms from C37 of PDI and from C43 of hVKOR1 was as short as possible (12.5 Å) and (ii) each PDI fragment, ***F1*** and ***F2***, which was suggested to be a fragment able to form the intermolecular interactions with a target, was alternatively placed above the middle of the L-loop surface. The obtained pro-models, **Model 1** and **Model 2**, were explored using MD simulation for conditions (see Methods) where restrains 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 observed at an S···S distance of 12 Å, then adopts the 'closed' conformation at a shortened S···S distance (of 10 and 8 Å) with the αH1-L helix and 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 αH2-helix of PDI rotated by 30° (at an S···S distance of 10 Å) followed by 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_1X_2C motif and a shortened αH-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 αH3-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 A4). As for PDI, the folded content of its initial and final conformations is the same for both models.

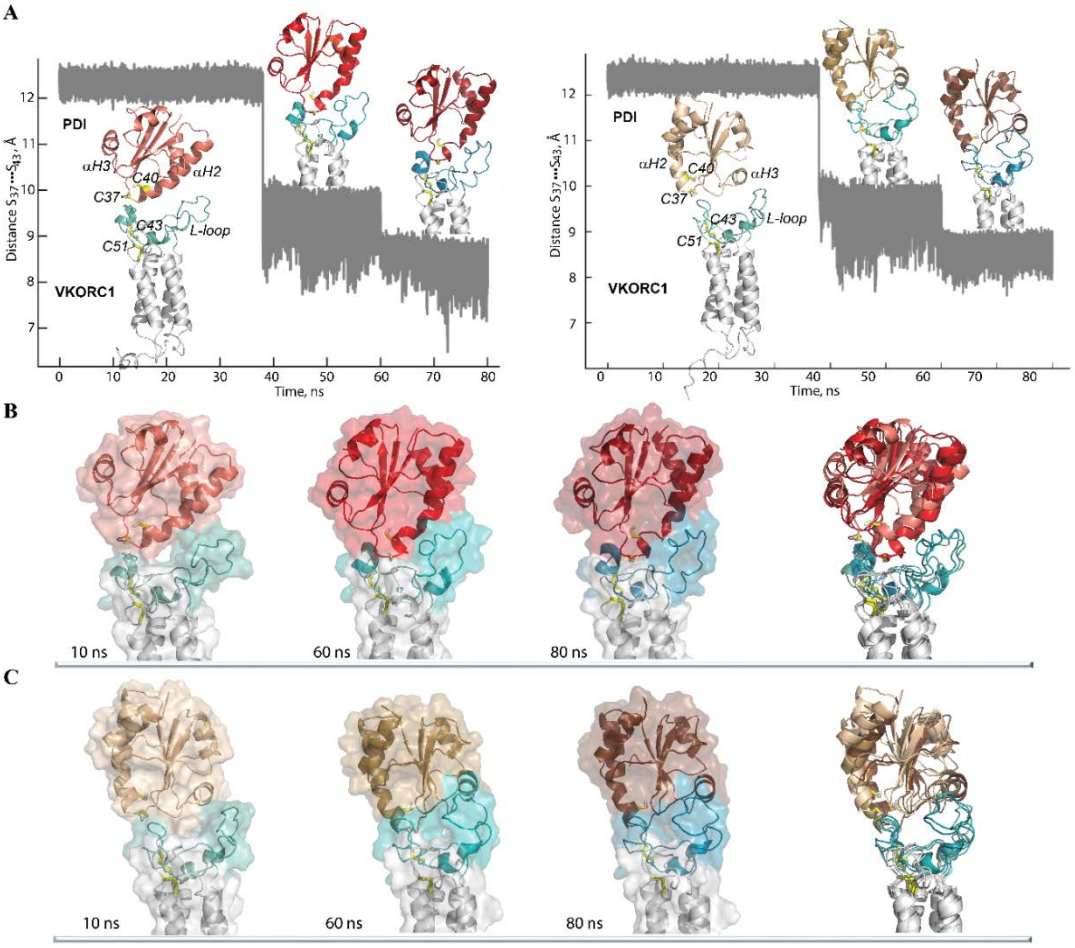

Figure 11. **Modelling of the human PDI-VKORC1 complex**. (**A**) The MD simulations of 3D models PDI-VKORC1 complexes were performed with a gradually diminished distance (from 12.5 to 8.0 Å) between the sulphur (S) atoms of C37 from PDI and of C43 from L-loop of hVKORC1. The PDI has two orientations 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 are 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 – red palette was used for PDI and 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 the 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 a 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 are ranged from 2.5 to 3.2 Å that characterise strong interactions.

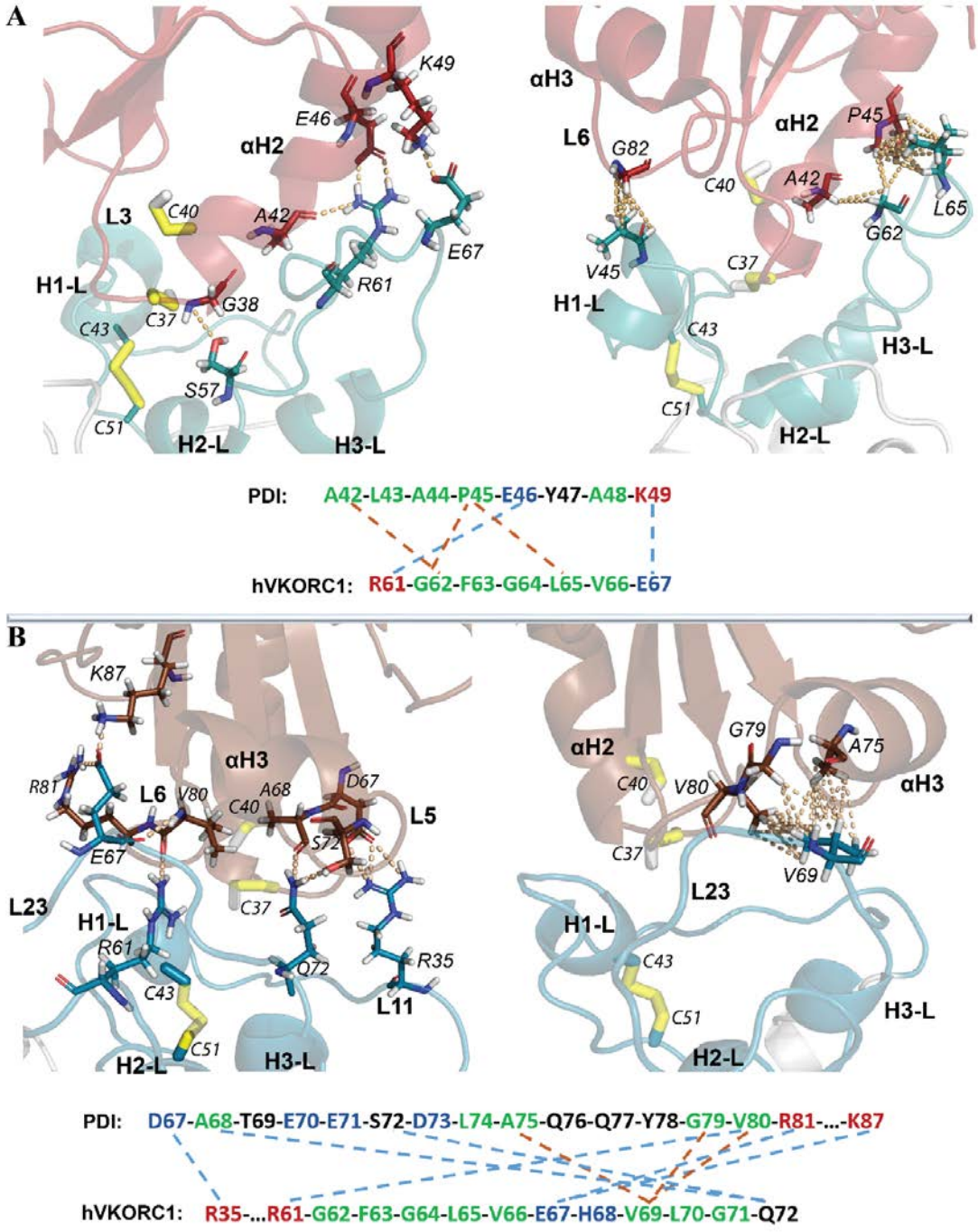

Figure 12. Intermolecular contacts at the interface between PDI and hVKORC1 in two models of the PDI-hVKORC1 complex. The intermolecular H-bonds and hydrophobic contacts between PDI and VKORC1 in the **Model 1** (**A**, top) and in **Model 2 (B**, top**)**. (**A**-**B**) The proteins a shown as coloured 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 in salmon. The structural fragments and residues participating in the contacts are labelled. Analysis of 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 disulfide 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 the short sequence segments, A42-K49 from PDI and R61-E67 from hVKORC1, forms a very compact regular interaction pattern which describes the highly specific recognition between two molecules that are maintained by two salt bridges and by cross vise hydrophobic interactions. This pattern of interactions stabilises the αH2-helix of PDI and the L23 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 L23 linker, which is composed of a segment that was predicted to be the mostly putative recognition region in an isolated hVKORC1. Similarly, the PDI residues participating in hVKORC1 recognition belong to the ***F1*** fragment that was regarded as a possible putative recognition site. Surprisingly, the hydrophobic interaction with V45 (hVKORC1) is formed by G82 (PDI), which is a residue from the ***F2*** fragment, and is also predicted to be a fragment that contains possible recognition sites.

In **Model 2**, the interaction interface between PDI and hVROR1 is also formed by the 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. The 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 the both models, **Model 1** and **Model 2**, but by selecting different PDI residues. It is remarkable that 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 the both proteins from **Model 1**, together with the increased helical folding of L-loop by 50%, are very attractive arguments for the choice of this model as functionally related, though there is still doubt in such a conclusion.

The other characteristics will be searched which will better justify or disprove our hypothesis. First, from a superimposition of each model with the experimentally defined structure of bVKORC1, the best fit at level of the Trx-like domain orientation with respect to VKOR is observed for **Model 1** (Figure A12), 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 αH2-helix of Trx and L46 of 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 taken at t=80 ns under more relaxed ('soft') conditions (see Methods) were simulated, 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' constrains (a time range of 80-100 ns), the distance S···S either varied within an enlarged range (7-11 Å) or, surprisingly, shows 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 A12). The MD conformations of **Model 1** generated using different 'soft' constrains 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 of the αH2 helix of PDI. Each of these structural effects were observed in isolated proteins. The interface interactions between the residues from the αH2 helix of PDI and L23 from L-loop of hVKORC1 are very similar for conformations taken at t=100 ns and at t=80 ns. With respect to the conformation chosen at t=80 ns, some novel contacts involving residues from H2-L (hVKORC1) and from L3 loop and of PDI are observed in the conformation taken at t= 100 ns (Figure A13).

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

The MD conformations of **Model 2** generated using 'soft' and 'hard' constrains showed similar structures of PDI-hVKORC1 that differed only in the position of the αH3 helix of PDI and of the L-loop of hVKORC1. The interactions observed at the interface between two proteins are non-preserved, with the exception of a single salt-bridge between E67 (hVKORC1) and R81 (PDI) (Figure A14).

3. Discussion

The vitamin K epoxide reductase (VKORC1) is a crucial enzyme in blood coagulation and the target of the most commonly used oral anticoagulant warfarin. Its functional role is a catalyst in the reduction of vitamin K, requiring cooperation with a redox partner that delivers reducing equivalents. The particularly interesting problem is the enzymatic activation of hVKORC1 by the thiol-disulfide exchange. This process involves 'molecular recognition' at the highest level required for the proton-transfer reactions.

The physiological redox partner of hVKORC1 remains uncertain, nevertheless the four proteins − PDI, ERp18, Tmx1 and Tmx4 – were suggested as the most likely H-donors of hVKORC1 [12, 13]. Consequently, an understanding of the molecular origins of the VKORC1 recognition by an unknown redox protein is not a trivial task.

To attack the problem, it is first suggested that a careful *in silico* study of the isolated proteins will provide useful information. In particular, the quantitative metrics and qualitative estimations that can shed light on the target (hVKORC1) features and on peculiarity of redox proteins. Such information may help in predicting (i) the protein fragments participating in the VKORC1 recognition by a Trx, and (ii) the most probable partner of VKORC1.

What has been learned from studying VKORC1 and the 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 has been carefully characterized. It was found that the L-loop in the inactive (oxidized) 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 L-loop is generally maintained over the MD simulations, while the length and spatial positions of the helices are highly variable. This variation is reflected in a large number of L-loop conformations, varying from the compact 'closed' conformation, which is prevalent, to the extended 'open' conformation.

It was established that the H2-L helix is the fundamental actor that controls the conformational features of L-loop. This transient helix converts between the αH- and 3_{10}-fold, adapts in length from short to elongated. The shortened H2-L helix, in which the S56-R61 segment is unfolded, promotes an elongation of the coiled linker L23, connecting the H2-L and H3-L helices. The extended linker L23 shows (i) a 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 that is evidenced by its displacement with respect to H2-L.

At sequence level, the L-loop has been reported to be conserved between VKORs from different species [32]. A particularly high conservation was found for the S56-G63 segment, which in hVKORC1 is followed by the 5-residue hydrophobic insert GFGLV that is completed by the glutamic acid (E67) and histidine (H68). The sequence conservation along with the observed structural and dynamical properties of the H2-L helix and its adjacent linker L23 suggest their possible functional role. From analysis of the H-bonding patterns in L-loop, regular exposition of the charged (R58, R61, E67 and D73) and polar residues (S56, S57, W59, H68 and N77) to the outside of L-loop was observed in positions favourable to a contact with the solvent or a protein. Therefore, it was postulated that the S56-R61 segment, a part of the more extended S53-N77 segment, is a platform for recognition of a protein partner.

The charged residues are 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 a 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 into a binding site on the receptor protein, and this drastically increases the association rate [37, 38]. Often, the charged residues that are important for protein–protein interactions, are conserved across families of evolutionary related proteins and protein complexes [39-41].

Moreover, the tryptophan residue (W59) from the S56-R61 segment and the following 5 -residue hydrophobic insert GFGLV may act as anchoring residues that bound together 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_1X_2C motif's different fold is connected to the divergent configuration of the thiol groups – either as a part of the well-folded αH2-helix (Erp18 and Tmx1) with the restrained cis-geometry of the sulphur atoms, or as a part of a coiled structure with alternating orientation of the sulphur atoms that runs from syn-periplanar to anti-periplanar configuration (Tmx4), or as a part of a transient structure (PDI) reversed between the helical fold (α- and 3_{10}-helices) and turn-coil structure leading to a large number of thiol group configurations.

Focusing on the ***F1*** region, that is suggested to be a fragment able to form the intermolecular interactions with a target, it is noted that the only ***F1*** of PDI and the targeted S56-R61 segment of VKORC1 have a similar structural propriety, or rather, a structural disorder that describes an intrinsically disordered region (IDR). Indeed, two IDRs, which are the transient N-terminal of α-helix H2 in PDI and the transient H2-L helix comprising the S56-R61 segment from the L-loop of VKORC1, show the best structural compatibility from the point of view of their concerted structural reorganization during the arrangement of a transient supramolecular complex.

A large number of 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 recognized for their prevalence and their critical roles in regulatory intermolecular interactions [47]. It has been hypothesized 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 and to play roles in cell-cycle

control, signal transduction, transcriptional and translational regulation, and large macromolecular complexes [49].

Nevertheless, even if fragment ***F1*** is considered as the most probable fragment to form intermolecular interactions with a target, the mobility of linker L5 and of αH3 helix from ***F2*** of PDI means that ***F2*** has a 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 which are exposed to solvent. Consequently, ***F2*** is also potentially able to contribute in stabilising of a supramolecular complex.

These two fragments are very close to the CX_1X_2C motif which is either joined in a sequence (***F1***) (sequence vicinity) or adjacent in a 3D structural space (***F2***) (spatial vicinity).

It has been made 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.

As well a direct use of the X-ray structure of VKOR from bacteria (a protein with covalently bound Tmx-like and VKOR-like domains, which has a poor sequence and structure similarity compared to the human proteins PDI and VKORC1) is not appropriate for modelling of the human complex, it was used as 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 by ***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 the two salt-bridges formed by the two pairs of charged residues, R61 and E67, from hVKORC1, which, together with S57, G62 and L65, also contribute to the stabilisation of two proteins. These residues are the amino acids from the L-loop segment 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 the two salt bridges formed by R35 and E67 from hVKORC1 interacting with D67, R81 and K87 from PDI, and by the H-bonds formed by Q72 and R61 of hVKORC1 with A68, S72 and V80 of PDI. In both models, the two amino acids of hVKORC1, R61 and E67 participate in the strong electrostatic interactions, the 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 pre-determined by the PDI orientation with respect to L-loop.

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

In the 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 cross vice hydrophobic interactions. Such a compact regular interaction pattern may describe the highly specific recognition between two molecules, which maintains the αH2-helix of PDI and the extended L23 of VKORC1 in a

close positon 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 the disulfide 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 non-specific.

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 the 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, the interacting hydrophobic motifs from both proteins form the 'interacting hydrophobic cores', which may the key factors in the recognition process. The total number of non-covalent contacts between PDI and hVKORC1 in **Model 2** is 9, while in **Model 1** this 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 an enlarged interface interaction network observed in **Model 2** is less likely, due to the low probability of a synchronized 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 two proteins promotes an increase in the helical folding of L-loop by 50%, while in **Model 2**, its helical fold was reduced by 40%. While proteins became disordered on their own, their native conformation is stabilized 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 folding-unfolding of the αH2-helix in **Model 1**, and by removal of the αH3 helix in **Model 2**.

Apparently, specificity of intermolecular interactions in PDI-hVKORC1 is 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 variety 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 the folding of a single-chain protein a 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 the 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, and its extent is variable. The binding site is not necessarily at the largest patch of the hydrophobic surface. There are high proportions 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 non polar 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, could electrostatic and hydrophobic interactions in the PDI-hVKORC1 complex be conserved qualitatively? The MD simulations of **Model 1** different 'soft' constrains that supplied an increased degree of freedom for proteins and allowed them to be removed, which proved the stability of the interactions formed by salt bridges and by the crosswise hydrophobic contacts. As the **Model 1** of the PDI-hVKORC1 complex showed stable interface interactions under such conditions,

it was proposed as a first precursor to probe the thiol-disulfide exchange reactions between PDI and hVKORC1.

Returning to the questions proposed 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 experiments on a protein function, which would play an essential role in validating and improving the modelling and simulations. Therefore, the authors are now waiting for the first results needed for 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 that will allow a better understand 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 the atomic coordinates of domain **a**, that contains the CX_1X_2C motif and is present in all available structures, were extracted. The 3D homology model of h-TMX4 was generated from the human sequence Q9H1E5 (https://www.uniprot.org/uniprot/) using the Modeller program [57] and the empirical structure of the TMX1 (PDP ID: 1x5e) that was used as a template. The 3D model of the h-ERp18 protein was optimized (the cysteine residues were saturated with hydrogen atoms) to obtain a reduced state of the CX_1X_2C motif.

h-VKORC1. The coordinates of the 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 the 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 the PDI protein, the extended conformation of L-loop was chosen. The PDI protein was placed in two orientations with respect to VKORC1, with (i) the L3 and αH2-helix (***F1***), and (ii) the L5 and αH2-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 the non-glycine/non-proline residues have dihedral angles in the most favoured and the 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-VKORC1 complex in two orientations (PDI_{F1}-VKORC1 and PDI_{F2}-VKORC1) − were prepared with the LEAP module of AMBER 16 [59] using *ff14SB* all-atom force filed 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, histidine residues were considered neutral and were protonated for ε-nitrogen atoms, (vi) Na^+ counter-ion was added to

neutralize the protein charge.

Each membrane protein, hVKORC1 and the two models of complex PDI-VKORC1 (PDI_{F1}-VKORC1 and PDI_{F2}-VKORC1), was embedded in the equilibrated and hydrated membrane composed of 200 1,2-dilauroyl-sn-glycero-3-phosphocho-line (DLPC) lipids using the replacement method available in the CHARMM-GUI Membrane Builder (http://www.charmm-gui.org/input/membrane) [61]. This lipid bilayer had been completed with 17293 (hVKORC1), 22047 (PDI_{F1}-VKORC1) 22567 (PDI_{F2}-VKORC1) water molecules (TIP3P [62] and 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^- ions were randomly placed to neutralize the system.

The total number of atoms in the isolated Trx-fold proteins (protein, water molecules and counter ion) varies 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_{F1}-VKORC1), and 93325 (PDI_{F2}-VKORC1). The box size was varied in the range of 84 × 84 × 108–141 $Å^3$.

4.2.2. Set up of the systems

The set-up of the systems was performed with the SANDER module [64] of AMBER18. First, each system was minimized successively using the steepest descent and conjugate gradient algorithms as follows: (i) 10,000 minimization steps where the water molecules have fixed protein atoms, (ii) 10,000 minimization steps where the protein backbone is fixed to allow protein side chains to relax, and (iii) 10,000 minimization 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). The 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 has been equilibrated during 1 ns (NPT) with all non-hydrogen atoms of the protein and the DLPC membrane harmonically restrained. A semi-isotropic coordinate scaling and the 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 the 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 on 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 hVRORC1) in its natural environment − the water solution for the Trx-fold protein and the solvated bilayer lipid membrane for h-VKORC1. 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 next subsection for the details). The MD simulation of the Trx-VKORC1 complex was first performed during 38 ns with a constrained S···S distance of 12.8 Å, that was further reduced to 10.2 Å, followed by simulation during 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 modeled 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 the two models, **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 from C43 of hVKORC1), which was varied in a step-wise manner (see Figure 7A). Specifically, the 80-ns simulation was divided into three steps each with a 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 D). To probe the stability of the PDI-hVKORC1 complex, the simulations of **Model 1** and of **Model 2**) were continued from 80 to 100 ns with two different 'soft' restraints applied to distance S···S (see Figure A13). 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 analyzed (RMSFs, RMSDs, DSSP, clustering) with the standard routines CPPTRAJ 4.15.0 program [70] of AMBER 18 Suite.

The RSMD and RMSF values were calculated for the Cα atoms using the initial model (at t = 0 ns) as a reference. All analysis was performed on the MD conformations (every 10 ps) considering either all simulation or the production part of the simulation, which was generated after 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 RMSF. For hVKORC1, the RMSDs were individually calculated for each domain after least-squares fitting of the MD conformations to the initial conformation of a domain, thus removing rigid-body motion from the analysis.

Secondary structure. The 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.

Dynamic cross-correlations. 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 the backbone Cα-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.

Principal Component Analysis. The collective motions of proteins were investigated by principal components analysis (PCA). For an N-atoms system, a trajectory matrix contains in each column a Cartesian coordinates for a given atom at each time step $\big(x(t)\big)$. Fitting the coordinate data to a reference structure results the proper trajectory matrix (X). The trajectory data is then used to generate a covariance matrix (C), elements of which are defined as (1):

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

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

The principle components (PCs) are obtained by a diagonalization of the covariance matrix C (2).

$$C = V\Lambda V^T \quad (2)$$

This results in a diagonal matrix Λ containing the 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.

Cross-Correlations Analysis. 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 (3) (63):

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

where i and j are two atoms Cα; Δr_i and Δr_j are displacement vectors of i and j; and Δr^T denotes the transpose of a column vector.

If $CC(ij) = 1$ the fluctuations of i and j are completely correlated (same phase and period), if $CC(ij) = -1$ the fluctuations of i and j are completely anticorrelated, and if $CC(ij) = 0$ the fluctuations of i and j are not correlated. All snapshots were fitted using the transmembrane domain Cα as a reference before performing cross-correlations analysis.

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

Conformation clustering. 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 for the Trx-fold proteins. Analysis was performed each 100 ps.

The algorithm extracts representative MD conformations from a trajectory by clustering the recorded snapshots according to their Cα-atoms 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 comparison.

The other structural measurements: The drift analysis of the helices was performed on 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···S between two sulphur atoms from C37 and C40, and a dihedral angle determined as an absolute value of pseudo-torsion angle S− Cα37− C40α−S.

Non-covalent distance monitoring. The H-bond between heavy atoms (N, O, and S) as potential donors/acceptors were calculated with the 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.

Graphics. 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

visualize the motions along the principal components, the Normal Mode Wizard (NMWiz) plugin [73] that is distributed with VMD was utilized.

4.3.2. Advanced methods of analysis

Multi-dimensional scaling. Metric Multi-Dimensional Scaling (MDS) is an algorithm for dimension reduction and visualisation: 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 minimization of the cost

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

where $D = (d_{ij})$ is the pairwise distance matrix, and the $\{x_i\}_i$ are the embedded points.

It can be implemented using the $manifold.MDS$ class in the python's $scikit\ learn$ library.

Fréchet mean. The Fréchet mean of a set is a point minimizing the sum of squared distances to each point of the set. As an example, the Fréchet mean $\overline{T}$ of one set $\{T_i\}_i$ of tetrahedrons is defined as

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

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

Kendall's disk of 3D triangles. Kendall's shape space of 3D triangles is isometric to north hemisphere of a 3D sphere of radius ½ where the equilateral triangle is at the north pole [20]. We use planar representation of the half sphere as a disk with the equilateral triangle at the center, by the transformation $(\varphi, \theta) \rightarrow (r = \sin(\theta), \varphi)$of the spherical coordinates to the polar coordinates. Each 3D triangle up the translation rotation and scaling is representing by a unique point of the disk.

Supplementary Materials: Supplementary materials can be found at www.mdpi.com/xxx/s1. **Appendix A**

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.; resources, 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), grant number 18-CE20-0025-01.

Acknowledgments: The authors thank Isaure Chauvot de Beauchene for the useful discussions, and John Redford for English Language editing. This research was supported by Centre National de la Recherche Scientifique (CNRS), Institut Farma 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 the GENCI (Grand Equipement National de Calcul Intensif). Calculations were performed on Jean Zay cluster at the 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

hVKORC1	the human vitamin K epoxide reductase Multidisciplinary Digital Publishing Institute
bVKOR	the vitamin K epoxide reductase from bacteria
L-loop	luminal loop
Trx	thioredoxin
PDI	Protein Disulfide Isomerase
ERp18	the endoplasmic reticulum oxidoreductase
Tmx1	the thioredoxin-related transmembrane protein 1
Tmx4	the thioredoxin-related transmembrane protein 4
MD	molecular dynamics
PDB	Protein Database
ID	identification code
RMSD	the root mean square deviation
RMSF	the root mean square fluctuation
PCA	the principal component analysis

References

1. Hatahet F, Ruddock LW. Protein disulfide isomerase: a critical evaluation of its function in disulfide bond formation. Antioxidants & redox signaling. 2009;11(11):2807-50. Epub 2009/05/30. doi: 10.1089/ars.2009.2466. PubMed PMID: 19476414.
2. Lee S, Kim SM, Lee RT. Thioredoxin and thioredoxin target proteins: from molecular mechanisms to functional significance. Antioxidants & redox signaling. 2013;18(10):1165-207. Epub 2012/05/23. doi: 10.1089/ars.2011.4322. PubMed PMID: 22607099; PubMed Central PMCID: PMCPMC3579385.
3. Hudson DA, Gannon SA, Thorpe C. Oxidative protein folding: from thiol-disulfide exchange reactions to the redox poise of the endoplasmic reticulum. Free radical biology & medicine. 2015;80:171-82. Epub 2014/08/06. doi: 10.1016/j.freeradbiomed.2014.07.037. PubMed PMID: 25091901; PubMed Central PMCID: PMCPMC4312752.
4. Winther JR, Thorpe C. Quantification of thiols and disulfides. Biochimica et biophysica acta. 2014;1840(2):838-46. Epub 2013/04/10. doi: 10.1016/j.bbagen.2013.03.031. PubMed PMID: 23567800; PubMed Central PMCID: PMCPMC3766385.
5. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, et al. The Protein Data Bank. Nucleic acids research. 2000;28(1):235-42. Epub 1999/12/11. doi: 10.1093/nar/28.1.235. PubMed PMID: 10592235; PubMed Central PMCID: PMCPMC102472.
6. Chatron N, Chalmond B, Trouvé A, Benoît E, Caruel H, Lattard V, et al. Identification of the functional states of human vitamin K epoxide reductase from molecular dynamics simulations. RSC Advances. 2017;7(82):52071-90. doi: 10.1039/C7RA07463H.
7. Martin JL. Thioredoxin--a fold for all reasons. Structure (London, England : 1993). 1995;3(3):245-50. Epub 1995/03/15. doi: 10.1016/s0969-2126(01)00154-x. PubMed PMID: 7788290.
8. Dobson CM, Karplus M. The fundamentals of protein folding: bringing together theory and experiment. Current opinion in structural biology. 1999;9(1):92-101. Epub 1999/02/27. doi: 10.1016/s0959-440x(99)80012-8. PubMed PMID: 10047588.
9. Wang CL, W.; Ren, J.; Fang, J.; Ke, H.; Gong, W.; Feng, W.;Wang, C. . Structural Insights into the Redox-Regulated Dynamic Conformations of Human Protein Disulfide Isomerase. Antioxidants & redox signaling. 2013;19(1):36-45. doi: 10.1089/ars.2012.4630. PubMed PMID: 22870953.
10. Guddat LW, Bardwell JC, Martin JL. Crystal structures of reduced and oxidized DsbA: investigation of domain motion and thiolate stabilization. Structure (London, England : 1993). 1998;6(6):757-67. Epub 1998/07/10. doi: 10.1016/s0969-2126(98)00077-x. PubMed PMID: 9655827.
11. Goodstadt L, Ponting CP. Vitamin K epoxide reductase: homology, active site and catalytic mechanism. Trends in biochemical sciences. 2004;29(6):289-92. Epub 2004/07/28. doi: 10.1016/j.tibs.2004.04.004. PubMed PMID: 15276181.
12. Schulman S, Wang B, Li W, Rapoport TA. Vitamin K epoxide reductase prefers ER membrane-anchored thioredoxin-like redox partners. Proceedings of the National Academy of Sciences of the United States of

America. 2010;107(34):15027-32. Epub 2010/08/11. doi: 10.1073/pnas.1009972107. PubMed PMID: 20696932; PubMed Central PMCID: PMCPMC2930587.
13. Wajih N, Hutson SM, Wallin R. Disulfide-dependent protein folding is linked to operation of the vitamin K cycle in the endoplasmic reticulum. A protein disulfide isomerase-VKORC1 redox enzyme complex appears to be responsible for vitamin K1 2,3-epoxide reduction. The Journal of biological chemistry. 2007;282(4):2626-35. Epub 2006/11/25. doi: 10.1074/jbc.M608954200. PubMed PMID: 17124179.
14. Tie JK, Jin DY, Stafford DW. Mycobacterium tuberculosis vitamin K epoxide reductase homologue supports vitamin K-dependent carboxylation in mammalian cells. Antioxidants & redox signaling. 2012;16(4):329-38. Epub 2011/09/24. doi: 10.1089/ars.2011.4043. PubMed PMID: 21939388; PubMed Central PMCID: PMCPMC3246416.
15. Rowe ML, Ruddock LW, Kelly G, Schmidt JM, Williamson RA, Howard MJ. Solution Structure and Dynamics of ERp18, a Small Endoplasmic Reticulum Resident Oxidoreductase. Biochemistry. 2009;48(21):4596-606. doi: 10.1021/bi9003342.
16. Dunker AK, Lawson JD, Brown CJ, Williams RM, Romero P, Oh JS, et al. Intrinsically disordered protein. Journal of molecular graphics & modelling. 2001;19(1):26-59. Epub 2001/05/31. doi: 10.1016/s1093-3263(00)00138-8. PubMed PMID: 11381529.
17. Imai K, Mitaku S. Mechanisms of secondary structure breakers in soluble proteins. Biophysics (Nagoya-shi, Japan). 2005;1:55-65. Epub 2005/10/19. doi: 10.2142/biophysics.1.55. PubMed PMID: 27857553; PubMed Central PMCID: PMCPMC5036629.
18. Schwartz R, King J. Frequencies of hydrophobic and hydrophilic runs and alternations in proteins of known structure. Protein science : a publication of the Protein Society. 2006;15(1):102-12. Epub 2005/12/24. doi: 10.1110/ps.051741806. PubMed PMID: 16373477; PubMed Central PMCID: PMCPMC2242367.
19. Kendall DG. Shape Manifolds, Procrustean Metrics, and Complex Projective Spaces. Bulletin of the London Mathematical Society. 1984;16(2):81-121. doi: https://doi.org/10.1112/blms/16.2.81.
20. Dryden ILM, K. V. Statistical Shape Analysis: With Applications in R, 2nd Edition2016. 496 p.
21. Kortemme T, Creighton TE. Ionisation of cysteine residues at the termini of model alpha-helical peptides. Relevance to unusual thiol pKa values in proteins of the thioredoxin family. Journal of molecular biology. 1995;253(5):799-812. Epub 1995/11/10. doi: 10.1006/jmbi.1995.0592. PubMed PMID: 7473753.
22. Xu S, Sankar S, Neamati N. Protein disulfide isomerase: a promising target for cancer therapy. Drug discovery today. 2014;19(3):222-40. Epub 2013/11/05. doi: 10.1016/j.drudis.2013.10.017. PubMed PMID: 24184531.
23. Dyson HJ, Jeng MF, Tennant LL, Slaby I, Lindell M, Cui DS, et al. 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(9):2622-36. Epub 1997/03/04. doi: 10.1021/bi961801a. PubMed PMID: 9054569.
24. Pinitglang S, Noble M, Verma C, Thomas EW, 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. Biochemical Society transactions. 1996;24(3):468s. Epub 1996/08/01. doi: 10.1042/bst024468s. PubMed PMID: 8879012.
25. Tie JK, Stafford DW. Structural and functional insights into enzymes of the vitamin K cycle. Journal of thrombosis and haemostasis : JTH. 2016;14(2):236-47. Epub 2015/12/15. doi: 10.1111/jth.13217. PubMed PMID: 26663892; PubMed Central PMCID: PMCPMC5073812.
26. Tie JK, Stafford DW. Functional Study of the Vitamin K Cycle Enzymes in Live Cells. Methods in enzymology. 2017;584:349-94. Epub 2017/01/10. doi: 10.1016/bs.mie.2016.10.015. PubMed PMID: 28065270; PubMed Central PMCID: PMCPMC5812275.
27. Hatahet F, Boyd D, Beckwith J. Disulfide bond formation in prokaryotes: history, diversity and design. Biochimica et biophysica acta. 2014;1844(8):1402-14. Epub 2014/03/01. doi: 10.1016/j.bbapap.2014.02.014. PubMed PMID: 24576574; PubMed Central PMCID: PMCPMC4048783.
28. Lyman E, Zuckerman DM. Ensemble-based convergence analysis of biomolecular trajectories. Biophysical journal. 2006;91(1):164-72. Epub 2006/04/18. doi: 10.1529/biophysj.106.082941. PubMed PMID: 16617086; PubMed Central PMCID: PMCPMC1479051.
29. Pylaeva S, Brehm M, Sebastiani D. Salt Bridge in Aqueous Solution: Strong Structural Motifs but Weak Enthalpic Effect. Scientific reports. 2018;8(1):13626. Epub 2018/09/13. doi: 10.1038/s41598-018-31935-z. PubMed PMID: 30206276; PubMed Central PMCID: PMCPMC6133928.
30. Pace CN, Fu H, Fryar KL, Landua J, Trevino SR, Shirley BA, et al. Contribution of hydrophobic interactions to protein stability. Journal of molecular biology. 2011;408(3):514-28. Epub 2011/03/08. doi: 10.1016/j.jmb.2011.02.053. PubMed PMID: 21377472; PubMed Central PMCID: PMCPMC3086625.
31. Rishavy MA, Usubalieva A, Hallgren KW, Berkner KL. Novel insight into the mechanism of the vitamin K oxidoreductase (VKOR): electron relay through Cys43 and Cys51 reduces VKOR to allow vitamin K reduction

and facilitation of vitamin K-dependent protein carboxylation. The Journal of biological chemistry. 2011;286(9):7267-78. Epub 2010/10/28. doi: 10.1074/jbc.M110.172213. PubMed PMID: 20978134; PubMed Central PMCID: PMCPMC3044983.

32. Bevans CG, Krettler C, Reinhart C, Watzka M, Oldenburg J. Phylogeny of the Vitamin K 2,3-Epoxide Reductase (VKOR) Family and Evolutionary Relationship to the Disulfide Bond Formation Protein B (DsbB) Family. Nutrients. 2015;7(8):6224-49. Epub 2015/08/01. doi: 10.3390/nu7085281. PubMed PMID: 26230708; PubMed Central PMCID: PMCPMC4555120.
33. Davis SJ, Davies EA, Tucknott MG, Jones EY, van der Merwe PA. The role of charged residues mediating low affinity protein-protein recognition at the cell surface by CD2. Proceedings of the National Academy of Sciences of the United States of America. 1998;95(10):5490-4. Epub 1998/05/20. doi: 10.1073/pnas.95.10.5490. PubMed PMID: 9576909; PubMed Central PMCID: PMCPMC20404.
34. Slagle SP, Kozack RE, Subramaniam S. Role of electrostatics in antibody-antigen association: anti-hen egg lysozyme/lysozyme complex (HyHEL-5/HEL). Journal of biomolecular structure & dynamics. 1994;12(2):439-56. Epub 1994/10/01. doi: 10.1080/07391102.1994.10508750. PubMed PMID: 7702779.
35. Nelson CA, Viner NJ, Young SP, Petzold SJ, Unanue ER. A negatively charged anchor residue promotes high affinity binding to the MHC class II molecule I-Ak. Journal of immunology (Baltimore, Md : 1950). 1996;157(2):755-62. Epub 1996/07/15. PubMed PMID: 8752926.
36. Stenlund P, Lindberg MJ, Tibell LA. 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(9):3168-75. Epub 2002/02/28. doi: 10.1021/bi011454r. PubMed PMID: 11863456.
37. Schreiber G. Kinetic studies of protein-protein interactions. Current opinion in structural biology. 2002;12(1):41-7. Epub 2002/02/13. doi: 10.1016/s0959-440x(02)00287-7. PubMed PMID: 11839488.
38. Wade RC, Gabdoulline RR, Lüdemann SK, Lounnas V. Electrostatic steering and ionic tethering in enzyme-ligand binding: insights from simulations. Proceedings of the National Academy of Sciences of the United States of America. 1998;95(11):5942-9. Epub 1998/05/30. doi: 10.1073/pnas.95.11.5942. PubMed PMID: 9600896; PubMed Central PMCID: PMCPMC34177.
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. The Biochemical journal. 1999;343 Pt 3(Pt 3):653-62. Epub 1999/10/21. PubMed PMID: 10527945; PubMed Central PMCID: PMCPMC1220598.
40. Unkles SE, Rouch DA, Wang Y, Siddiqi MY, Glass AD, Kinghorn JR. Two perfectly conserved arginine residues are required for substrate binding in a high-affinity nitrate transporter. Proceedings of the National Academy of Sciences of the United States of America. 2004;101(50):17549-54. Epub 2004/12/04. doi: 10.1073/pnas.0405054101. PubMed PMID: 15576512; PubMed Central PMCID: PMCPMC536016.
41. Zhao N, Pang B, Shyu CR, Korkin D. Charged residues at protein interaction interfaces: unexpected conservation and orchestrated divergence. Protein science : a publication of the Protein Society. 2011;20(7):1275-84. Epub 2011/05/13. doi: 10.1002/pro.655. PubMed PMID: 21563227; PubMed Central PMCID: PMCPMC3149200.
42. de Planque MR, Bonev BB, Demmers JA, Greathouse DV, Koeppe RE, 2nd, Separovic F, et al. Interfacial anchor properties of tryptophan residues in transmembrane peptides can dominate over hydrophobic matching effects in peptide-lipid interactions. Biochemistry. 2003;42(18):5341-8. Epub 2003/05/07. doi: 10.1021/bi027000r. PubMed PMID: 12731875.
43. Chandler D. Interfaces and the driving force of hydrophobic assembly. Nature. 2005;437(7059):640-7. Epub 2005/09/30. doi: 10.1038/nature04162. PubMed PMID: 16193038.
44. Mohan A, Oldfield CJ, Radivojac P, Vacic V, Cortese MS, Dunker AK, et al. Analysis of molecular recognition features (MoRFs). Journal of molecular biology. 2006;362(5):1043-59. Epub 2006/08/29. doi: 10.1016/j.jmb.2006.07.087. PubMed PMID: 16935303.
45. Tompa P, Davey NE, Gibson TJ, Babu MM. A million peptide motifs for the molecular biologist. Molecular cell. 2014;55(2):161-9. Epub 2014/07/20. doi: 10.1016/j.molcel.2014.05.032. PubMed PMID: 25038412.
46. van der Lee R, Buljan M, Lang B, Weatheritt RJ, Daughdrill GW, Dunker AK, et al. Classification of intrinsically disordered regions and proteins. Chemical reviews. 2014;114(13):6589-631. Epub 2014/04/30. doi: 10.1021/cr400525m. PubMed PMID: 24773235; PubMed Central PMCID: PMCPMC4095912.
47. Wright PE, Dyson HJ. Intrinsically disordered proteins in cellular signalling and regulation. Nature reviews Molecular cell biology. 2015;16(1):18-29. Epub 2014/12/23. doi: 10.1038/nrm3920. PubMed PMID: 25531225; PubMed Central PMCID: PMCPMC4405151.
48. Wong ETC, So V, Guron M, Kuechler ER, Malhis N, Bui JM, et al. Protein-Protein Interactions Mediated by Intrinsically Disordered Protein Regions Are Enriched in Missense Mutations. Biomolecules. 2020;10(8). Epub 2020/07/30. doi: 10.3390/biom10081097. PubMed PMID: 32722039; PubMed Central PMCID: PMCPMC7463635.

49. Dosztányi Z, Chen J, Dunker AK, Simon I, Tompa P. Disorder and sequence repeats in hub proteins and their implications for network evolution. Journal of proteome research. 2006;5(11):2985-95. Epub 2006/11/04. doi: 10.1021/pr060171o. PubMed PMID: 17081050.
50. Vangone A, Bonvin AM. Contacts-based prediction of binding affinity in protein-protein complexes. eLife. 2015;4:e07454. Epub 2015/07/21. doi: 10.7554/eLife.07454. PubMed PMID: 26193119; PubMed Central PMCID: PMCPMC4523921.
51. Keskin O, Gursoy A, Ma B, Nussinov R. Principles of protein-protein interactions: what are the preferred ways for proteins to interact? Chemical reviews. 2008;108(4):1225-44. Epub 2008/03/22. doi: 10.1021/cr040409x. PubMed PMID: 18355092.
52. Keskin O, Ma B, Nussinov R. Hot regions in protein--protein interactions: the organization and contribution of structurally conserved hot spot residues. Journal of molecular biology. 2005;345(5):1281-94. Epub 2005/01/13. doi: 10.1016/j.jmb.2004.10.077. PubMed PMID: 15644221.
53. Dill KA, Bromberg S, Yue K, Fiebig KM, Yee DP, Thomas PD, et al. Principles of protein folding--a perspective from simple exact models. Protein science : a publication of the Protein Society. 1995;4(4):561-602. Epub 1995/04/01. doi: 10.1002/pro.5560040401. PubMed PMID: 7613459; PubMed Central PMCID: PMCPMC2143098.
54. Tsai CJ, Lin SL, Wolfson HJ, Nussinov R. Studies of protein-protein interfaces: a statistical analysis of the hydrophobic effect. Protein science : a publication of the Protein Society. 1997;6(1):53-64. Epub 1997/01/01. doi: 10.1002/pro.5560060106. PubMed PMID: 9007976; PubMed Central PMCID: PMCPMC2143524.
55. Tsai CJ, Ma B, Sham YY, Kumar S, Nussinov R. Structured disorder and conformational selection. Proteins. 2001;44(4):418-27. Epub 2001/08/03. doi: 10.1002/prot.1107. PubMed PMID: 11484219.
56. Tsai CJ, Nussinov R. Hydrophobic folding units at protein-protein interfaces: implications to protein folding and to protein-protein association. Protein science : a publication of the Protein Society. 1997;6(7):1426-37. Epub 1997/07/01. doi: 10.1002/pro.5560060707. PubMed PMID: 9232644; PubMed Central PMCID: PMCPMC2143752.
57. Webb B, Sali A. Comparative Protein Structure Modeling Using MODELLER. Current protocols in bioinformatics. 2016;54:5.6.1-5.6.37. Epub 2016/06/21. doi: 10.1002/cpbi.3. PubMed PMID: 27322406; PubMed Central PMCID: PMCPMC5031415.
58. Laskowski RA. PDBsum: summaries and analyses of PDB structures. Nucleic acids research. 2001;29(1):221-2. Epub 2000/01/11. doi: 10.1093/nar/29.1.221. PubMed PMID: 11125097; PubMed Central PMCID: PMCPMC29784.
59. Case DA, Cheatham TE, 3rd, Darden T, Gohlke H, Luo R, Merz KM, Jr., et al. The Amber biomolecular simulation programs. Journal of computational chemistry. 2005;26(16):1668-88. Epub 2005/10/04. doi: 10.1002/jcc.20290. PubMed PMID: 16200636; PubMed Central PMCID: PMCPMC1989667.
60. Maier JA, Martinez C, Kasavajhala K, Wickstrom L, Hauser KE, Simmerling C. ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. Journal of chemical theory and computation. 2015;11(8):3696-713. Epub 2015/11/18. doi: 10.1021/acs.jctc.5b00255. PubMed PMID: 26574453; PubMed Central PMCID: PMCPMC4821407.
61. Jo S, Kim T, Iyer VG, Im W. CHARMM-GUI: a web-based graphical user interface for CHARMM. Journal of computational chemistry. 2008;29(11):1859-65. Epub 2008/03/21. doi: 10.1002/jcc.20945. PubMed PMID: 18351591.
62. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML. Comparison of simple potential functions for simulating liquid water. The Journal of Chemical Physics. 1983;79(2):926-35. doi: 10.1063/1.445869.
63. Dickson CJ, Madej BD, Skjevik ÅA, Betz RM, Teigen K, Gould IR, et al. Lipid14: The Amber Lipid Force Field. Journal of chemical theory and computation. 2014;10(2):865-79. doi: 10.1021/ct4010307.
64. Salomon-Ferrer R, Case DA, Walker RC. An overview of the Amber biomolecular simulation package. WIREs Computational Molecular Science. 2013;3(2):198-210. doi: https://doi.org/10.1002/wcms.1121.
65. Peters EA, Goga N, Berendsen HJ. Stochastic Dynamics with Correct Sampling for Constrained Systems. Journal of chemical theory and computation. 2014;10(10):4208-20. Epub 2015/11/21. doi: 10.1021/ct500380x. PubMed PMID: 26588119.
66. Evans DJ, Holian BL. The Nose–Hoover thermostat. The Journal of Chemical Physics. 1985;83(8):4069-74. doi: 10.1063/1.449071.
67. Van Gunsteren WF, Berendsen HJC. A Leap-frog Algorithm for Stochastic Dynamics. Molecular Simulation. 1988;1(3):173-85. 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. Molecular Simulation. 1991;6(1-3):121-42. doi: 10.1080/08927029108022142.
69. Hockney RWE, J. W. Computer Simulation Using Particles New York: McGraw-Hill; 1981.

70. Roe DR, Cheatham TE. PTRAJ and CPPTRAJ: Software for Processing and Analysis of Molecular Dynamics Trajectory Data. Journal of chemical theory and computation. 2013;9(7):3084-95. 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(12):2577-637. Epub 1983/12/01. doi: 10.1002/bip.360221211. PubMed PMID: 6667333.
72. Hünenberger PH, Mark AE, van Gunsteren WF. Fluctuation and cross-correlation analysis of protein motions observed in nanosecond molecular dynamics simulations. Journal of molecular biology. 1995;252(4):492-503. Epub 1995/09/29. doi: 10.1006/jmbi.1995.0514. PubMed PMID: 7563068.
73. Bakan A, Meireles LM, Bahar I. ProDy: protein dynamics inferred from theory and experiments. Bioinformatics (Oxford, England). 2011;27(11):1575-7. Epub 2011/04/08. doi: 10.1093/bioinformatics/btr168. PubMed PMID: 21471012; PubMed Central PMCID: PMCPMC3102222.
74. Humphrey W, Dalke A, Schulten K. VMD: visual molecular dynamics. Journal of molecular graphics. 1996;14(1):33-8, 27-8. Epub 1996/02/01. doi: 10.1016/0263-7855(96)00018-5. PubMed PMID: 8744570.