Journal Name

Electronic Supplementary Information, ESI

Identification of the Functional States of Human Vitamin K Epoxide Reductase from Molecular Dynamics Simulations

Content
Received 00th July 20xx,
Accepted 00th January 20xx

DOI: 10.1039/x0xx00000x

www.rsc.org/

N. Chatron,[a,b,c], B. Chalmond[a], A. Trouvé[a], E. Benoît[b], H. Caruel[c], V. Lattard[b], L. Tchertanov[a]

A. Figures S1- S11

B. Table S1

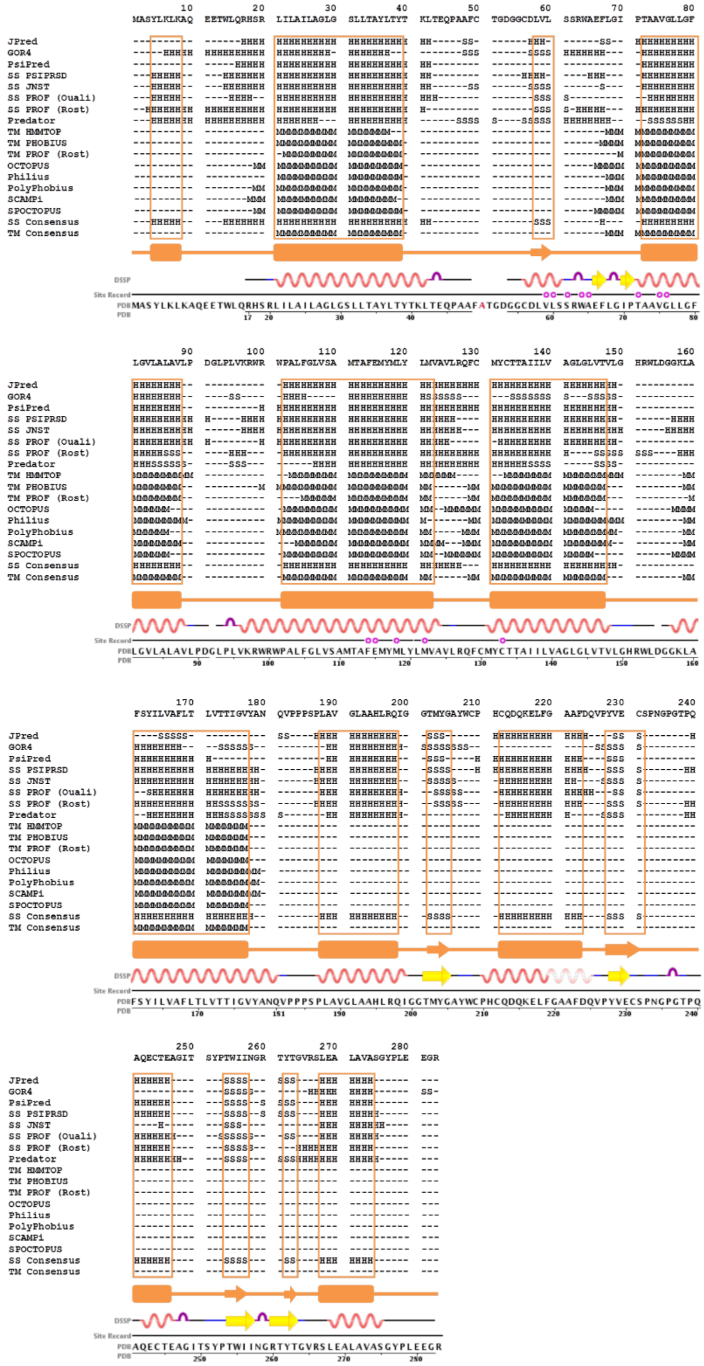
C. Modelling Methods

D. PDB files contain the atomic coordinates of the models

[a.] *Centre de Mathématiques et de Leurs Applications (CMLA), ENS Cachan, CNRS, Université Paris-Saclay, 61 avenue du Président Wilson, Cachan, 94235 France*
[b.] *USC 1233 INRA-Vetagro Sup, Veterinary School of Lyon, Marcy l'Etoile, France*
[c.] *Liphatech, Bonnel, 47480 Pont du Casse, France*
† Functional States of VKOR/N. Chatron *at al*.
Electronic Supplementary Information (ESI) available: [11 Figures, 1 Table and Methods details]. See DOI: 10.1039/x0xx00000x

Electronic Supplementary Information, ESI

Fig. S1 Validation of the methods used for prediction of secondary structure. The bVKOR topology and its membrane content were predicted by 8 different methods for each kind of forecast. Two types of consensus– for the secondary structure prediction and for the membrane content prediction–were considered. The secondary structure interpretation (DSSP) of the crystallographic structure 1NV5.



Please do not adjust margins

Fig. S2 The homology modeling of hVKORC1. (**A**) Alignment of the sequence's segment (13-181 residues) of bVKOR with the sequence of hVKORC1. The identical residues are distinguished by grey background; the similar residues are shown in blue. The secondary structure interpretation of the template sequence (top) and the predicted secondary structure composition (consensus) of the target sequence (below) are shown. (**B**) Structure 4NV5 of the bVKOR is presented as cartoon. The structural fragment showing the best sequence similarity with hVKORC1 is denoted in orange. (**C**) The hVKORC1 topology and its membrane content were predicted by 8 different methods for each kind of forecast. Two types of consensus – for the secondary structure prediction and for the membrane content prediction – were schematized (in deep teal).

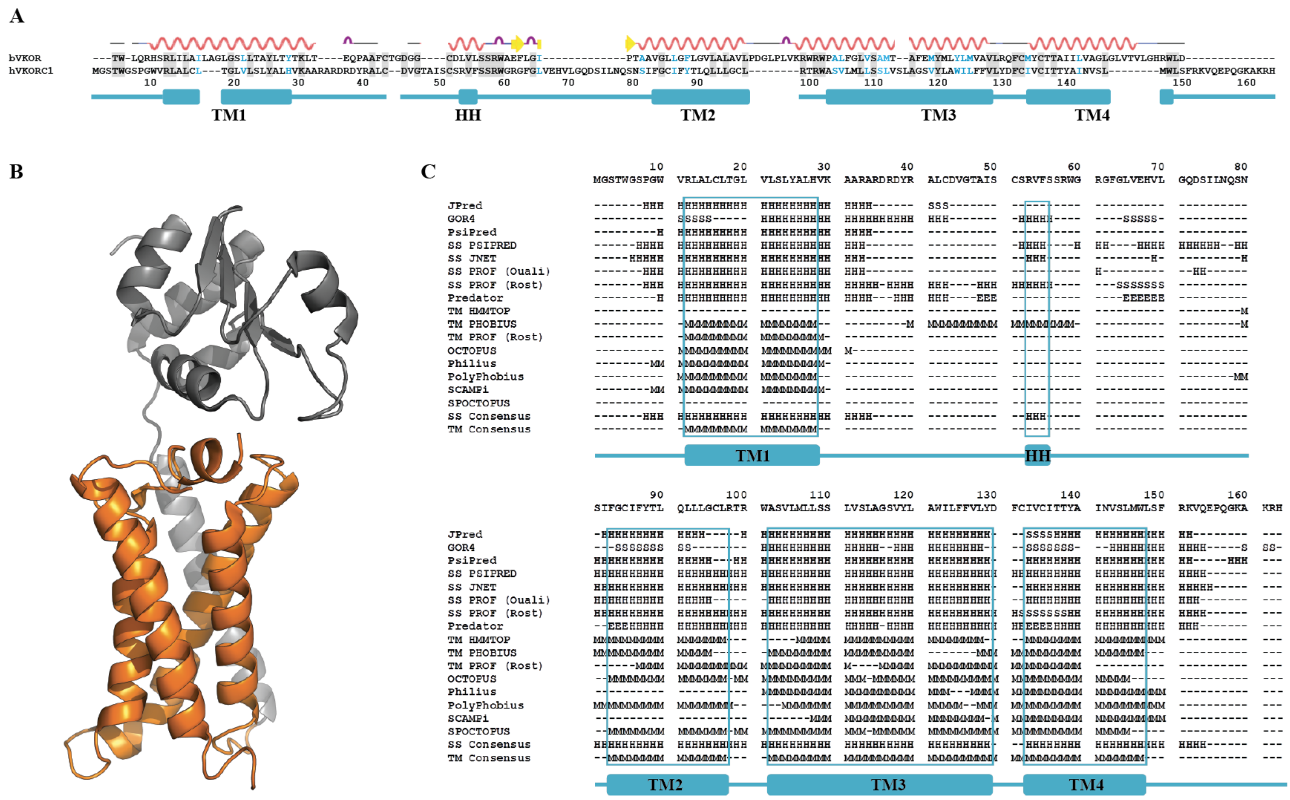


Fig. S3 Hydrophobic interactions stabilising the hVKORC1 structure of. Protein is presented in two orthogonal views as cartoon with the hydrophobic residues showed as space-filling spheres.

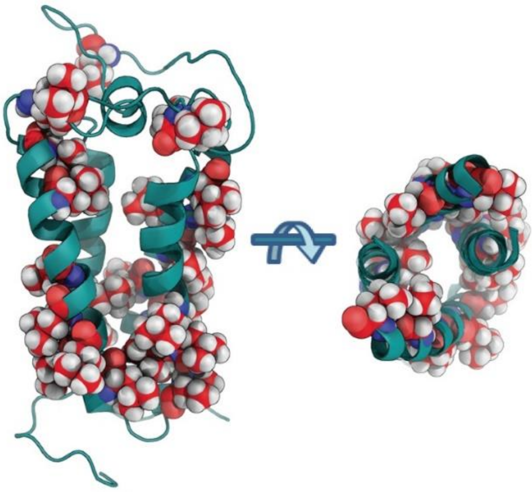


Fig. S4 Structural features of hVKORC1. (**A**) The secondary structure assignment for protein was done for each 1-µs trajectory, **1''** (top) and **3''** (bottom). The α-helix, 3^{10} helix, β-bridge, turn and loop are shown in red, orange, light-green, blue and cyan, respectively, and referred to the predicted helices. (**B**) Drift of helices was monitored over the extended simulations **1''** (left) and **3''** (right). Two centroids, assigned on the last four residues at the top and at the bottom of each helix, were defined. A sole centroid for HH was assign on seven residues. . Coordinates of each centroid were computed for each MD simulation time step and presented as the lines connecting the two TM centroids, in 3D space (top) and as the points projected on the *x*-*y* plane (bottom) coloured from blue (t=0) to cyan (t=1 µs) for the top of TM helices; from black (t=0) to green (t=1 µs) for the bottom; and from red (t=0) to yellow (t=1 µs) for HH.

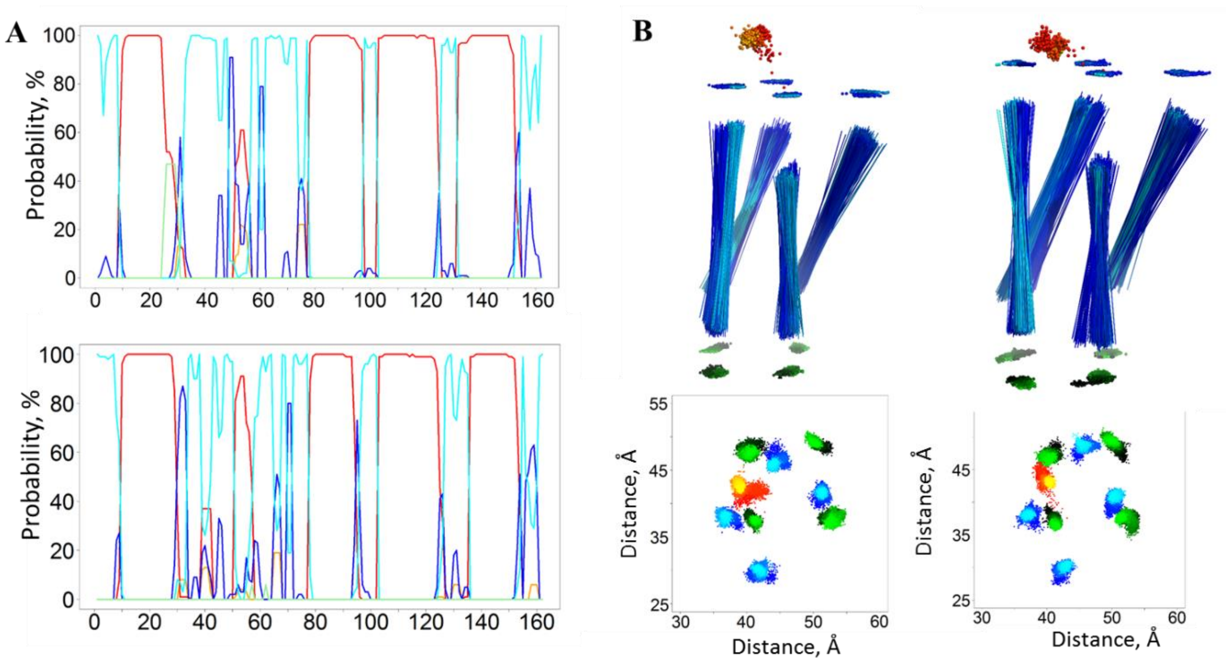


Fig. S5 Molecular dynamics simulation of hVKORC1. The RMSFs profiles of the short (100 ns **1**, **2** and **3**) and extended (1-µs) simulations **1''** and **3''**. The RMSFs were computed on the Cα atoms.

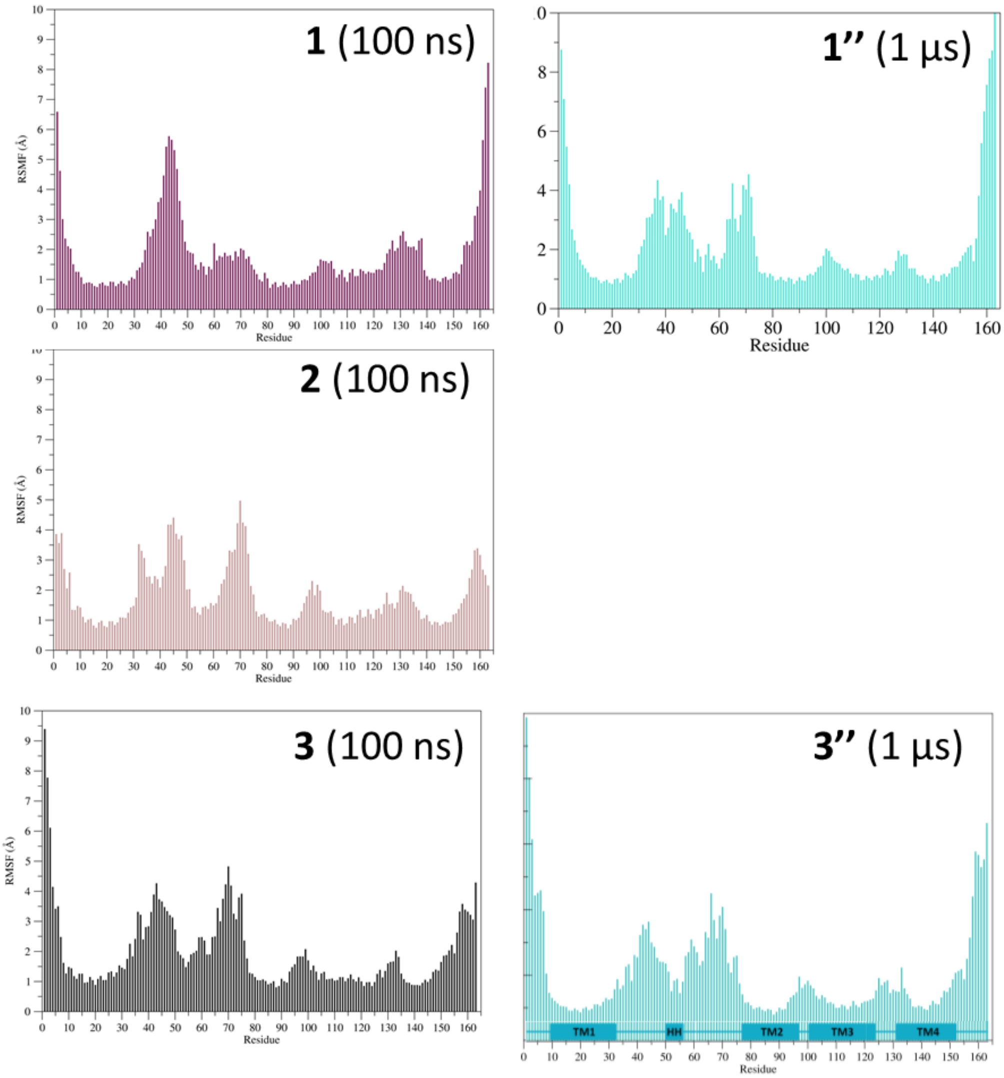


Fig. S6 Global motions and their correlations. The inter-residue cross-correlations maps resulting from PCA of the short (100 ns) (left) and extended to 500 ns (middle) and to 1 µs (right) trajectories, **1″** (the upper half-matrix) and **3″** (the lower half-matrix). Correlated (positive) and anti-correlated (negative) motions between atom pairs are presented as a gradient between red and color colors.

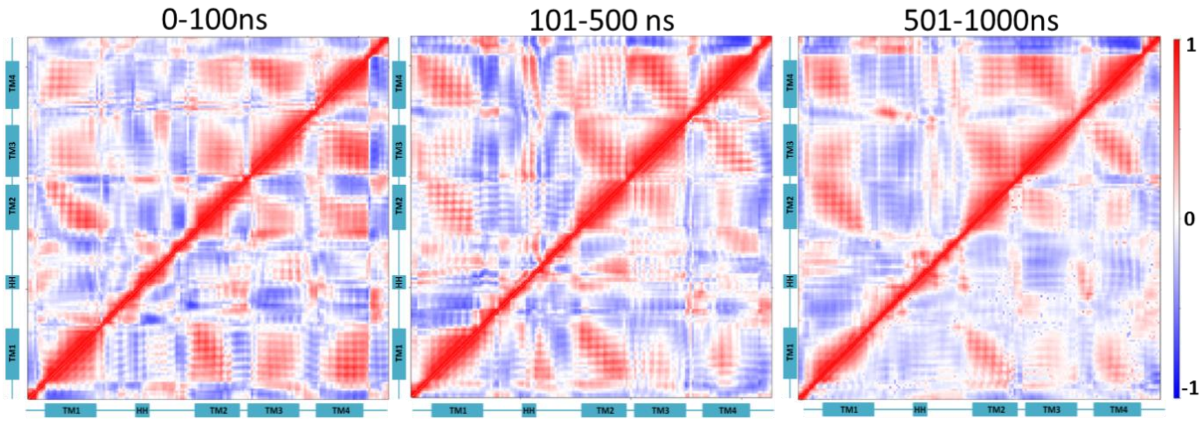


Fig. S7 Spectral analysis of trajectory 3" after PCA. (**A**) Similarity matrix W; (**B**) The largest eigenvalues; (**C**) First slow variable; (**D**) Second slow variable. By searching the changes of sign of this slow variable, it appears that the main basin is approximately located between the frame12500 (t=250 ns) and the frame 30000 (t=600 ns), and the second basin between the frame 37500 (t=750 ns) and the last frame 50000 (t=1000 ns).

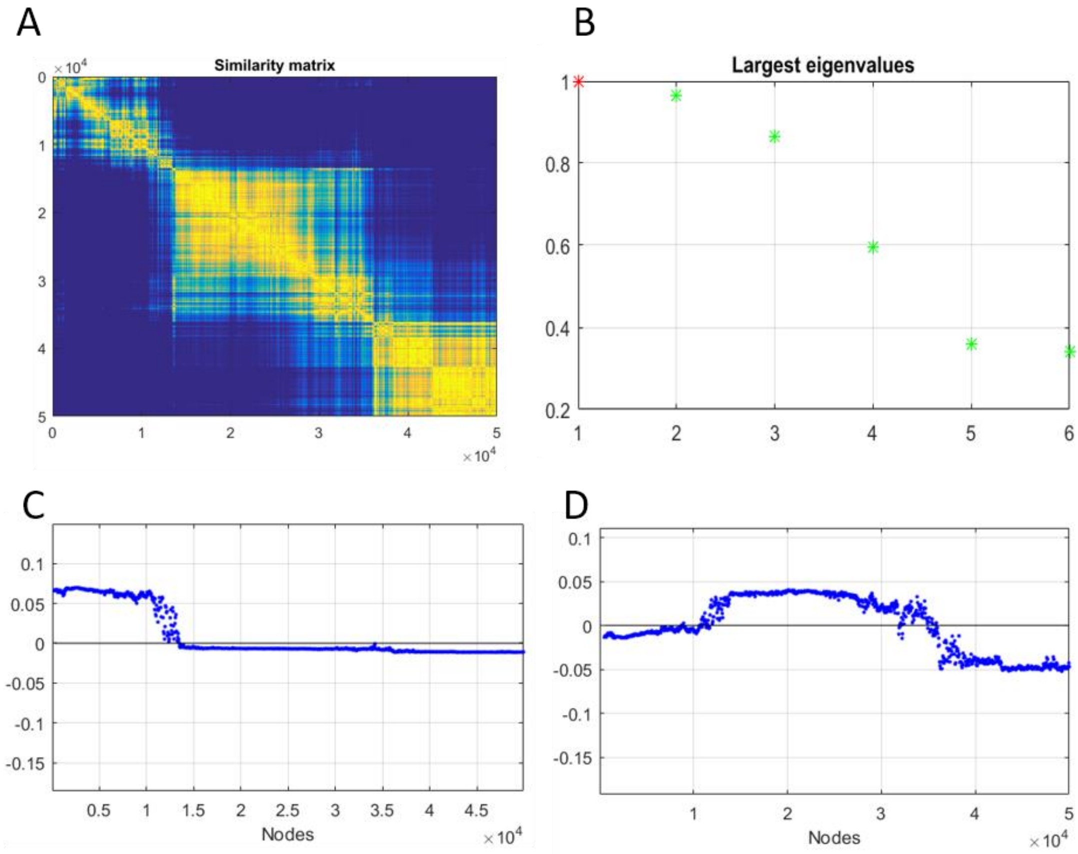


Fig. S8 Spectral analysis of trajectory 3" without PCA reduction. (**A**) Similarity matrix W; (**B**) The largest eigenvalues; (**C**) First slow variable; (**D**) Second slow variable.

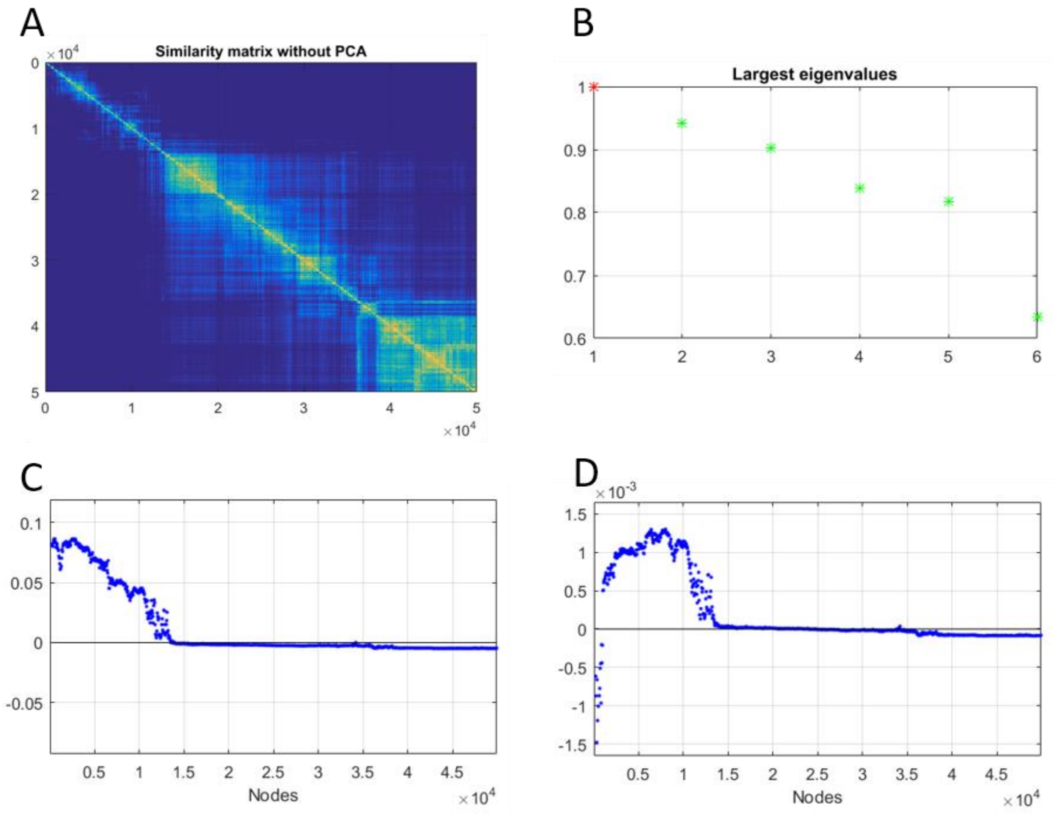


Fig. S9. Structural features of models M[I]-M[IV]. (**A**) The secondary structure evolution over MD simulations (all trajectories were merged). The α-helix, 3^{10} helix, β-bridge, turn and loop are shown in red, orange, light-green, blue and cyan, respectively, and referred to the predicted helices (top). For each model, the cysteine residues forming S•••S bridge is noted at top.

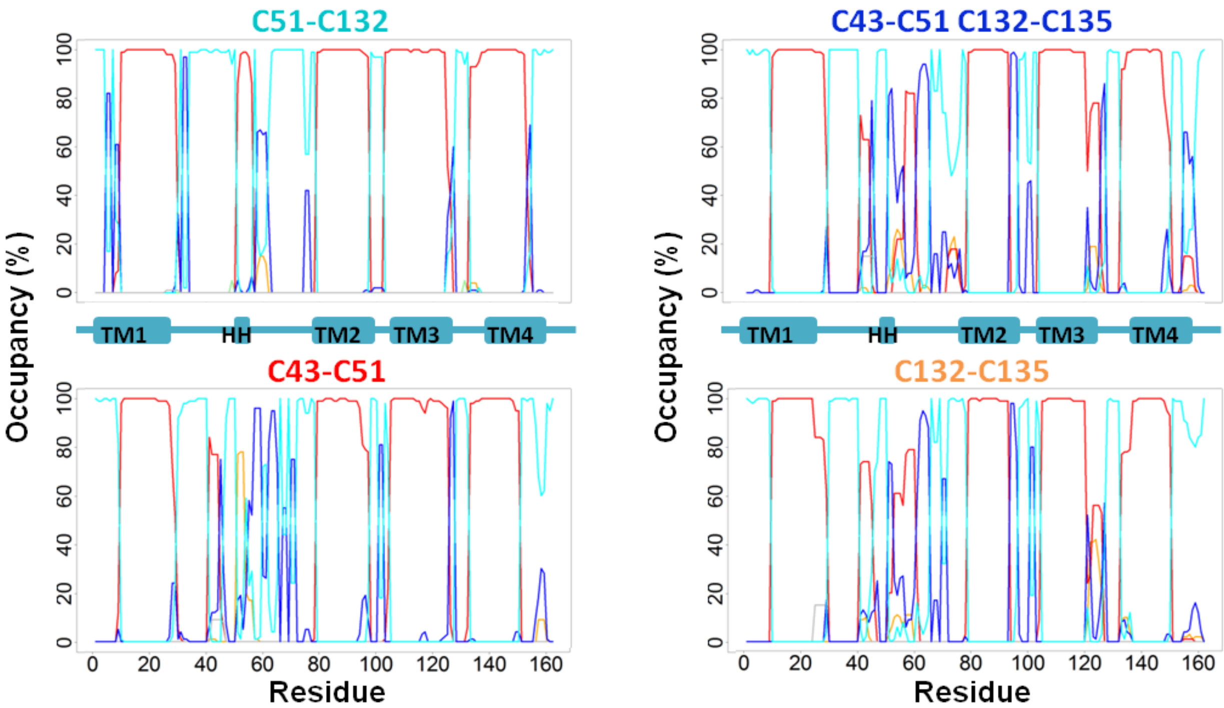


Fig. S10. The clusters of conformations defined with cutoff of 2.0 Å and their population (in %) calculated for MD simulation (2x100 ns) of models **M[I]-M[IV]**. For each model, the cysteine residues forming S•••S bridge is noted at top.

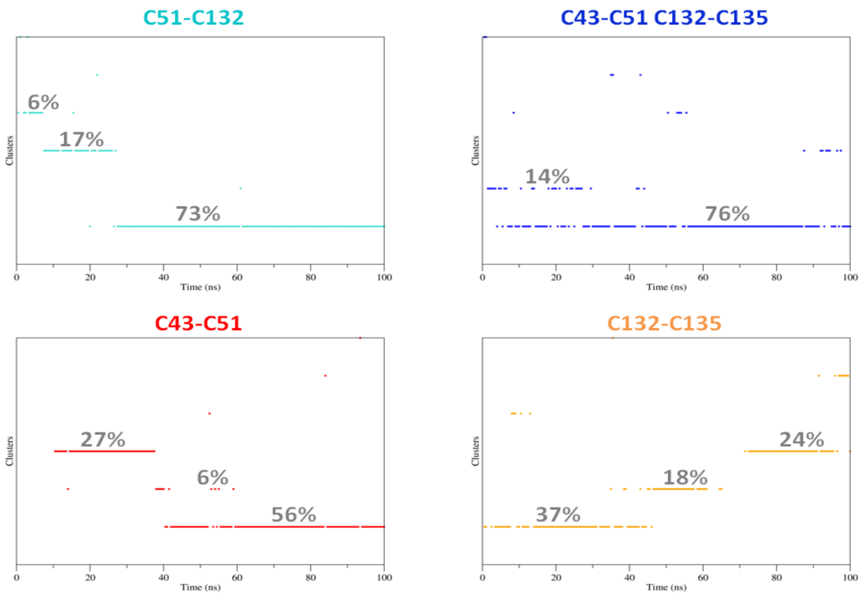
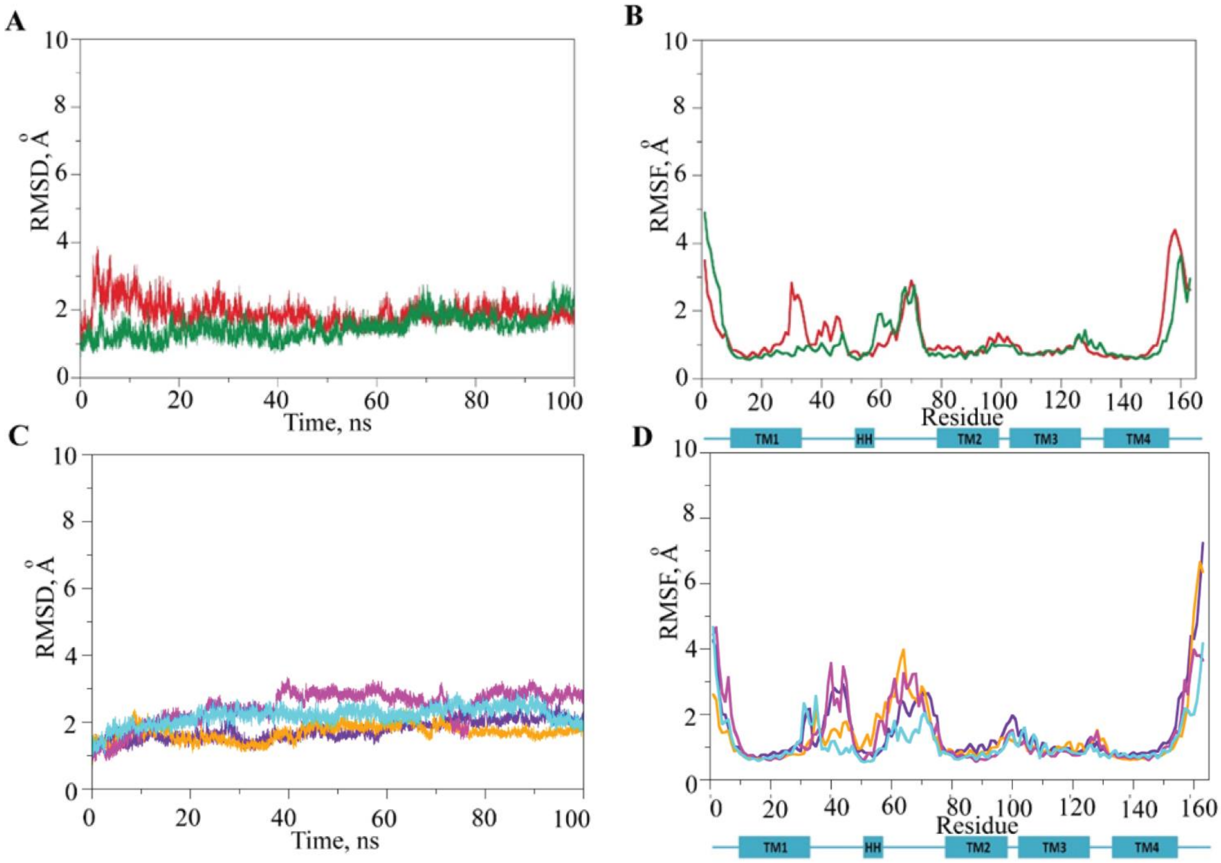


Fig. S11. Molecular dynamics simulations of the hVKORC1 complexes. (A) The RMSDs from the initial coordinates (t=0 ns) are computed on the Cα atoms of 100-ns MD trajectories of vitK^{EP}•**T**^{III} (green) and vitK$^{2\text{-}OH}$•**T**^{IV} (red). (**B**) The RMSFs are computed on the Cα atoms over of 100-ns MD trajectories of vitK^{EP}•**T**^{III} (green) and vitK$^{2\text{-}OH}$•**T**^{IV} (red). (**C**) The RMSDs from the initial coordinates (t=0 ns) are computed on the Cα atoms of 100-ns MD trajectories of A•**T**^{III} (violet), D•**T**^{III} (orange) and P•**T**^{III} (purple) ad W•**T**^{III} (cyan). (**D**) The RMSFs are computed on the Cα atoms over of 100-ns MD trajectories of A•**T**^{III} (violet), D•**T**^{III} (orange) and P•**T**^{III} (purple) ad W•**T**^{III} (cyan).



C. Table S1

Table S1. **Crystallographic structures of VKOR reported in Protein Data Base (PDB)** [1]. The PDB identification code (ID), resolution (Res, in Å), the sequence length, missing residues, mutated residues, cysteine residues, distances (d, in Å) between the sulfur atoms from cysteine residues and reference are denoted for each structure.

PDB ID	Res, (Å)	Sequence length	Missing residues	Cysteine**	S···S, d (Å)	Ref
3KP9	3.6	16 - 279	53-55, 91-92	C50(43), C56S(51) C130(132), C133(135)	C50···S56, 13 S56···C130, 8 C130···C133, 2	[2]
4NV2	3.6	17 - 282	92-93	C50A(43), C56S(51), C130(132), C133(135)	A50···C56, 4 C56···C130, 2 C130···C133, 4	[3]
4NV5 A*	2.8	17 - 282	49-53, 92,155	C56(51), C130(132), C133(135)	C56···C130, 2 C130···C133, 4	[3]
4NV5 B*	2.8	13 - 279	49-53	C56(51), C130(132), C133(135)	C56···C130, 2 C130···C133, 4	[3]
4NV6	4.2	17 - 282	92, 155	C50(43), C56(51) C130(132), C133(135)	C50···C56, 14 C56···C130, 2 C130···C133, 4	[3]

*Two chains, A and B.

**The cysteine residue or its mutant, as numbered in a structure sequence and its corresponding number in hVKORC1 sequence denoted in brackets.

C. Modelling Methods

1. Secondary structure prediction

PREDATOR [4] algorithm is based on potentially hydrogen-bonded residues recognition in the target sequence using a structural database information. First, the propensity of residues to form α-helix or β-sheet type hydrogen bonds is calculated for each single residue. Then, for each residue, the influence of the nearest neighbors is considered. Finally, the secondary structure consensus is established for each residue, by combination of the two predictions.

GOR IV [5] uses information theory and Bayesian statistics to calculate the propensity of each residue from the target sequence to form specific secondary structure, considering the nearest neighbors of the local segment.

PSIPRED [6] is based on position-specific scoring matrices (PPSM) combined to neural network application. The target sequence is submitted to the PSI-BLAST algorithm and, after three iterations, the PPSM is picked. This matrix is then split in fifteen residues length windows, each one used as input for the neural network. A second neural network filters successive outputs from the first one. Ten percent of the data is kept aside as an evaluation set. The neural network is trained on the remaining data for weights optimization, according to an on-line back-propagation procedure. Training is stopped when the accuracy of the network on the evaluation set starts to decrease.

PROF [7] is based on neural network, but requires five steps. This algorithm starts with GOR method, complemented with evolutionary information. The predicted by GOR results, combined with neural networks, are trained either in an unbalanced or in a balanced way using different profiles, producing the classifiers. The results are analyzed by linear discrimination or neural networks, yielding new classifiers which are then used again for a training. Finally, the two resulting classifiers are averaged to give a unique prediction for each residue.

JPRED [8], a consensus secondary structure predictor, uses evolutionary information and is based on six methods – NNSSP (nearest neighbors prediction), PHD (a jury decision neural networks), DSC (linear discrimination), MULPRED (a consensus single sequence method combination), ZPRED (conservation number weighted prediction) and PREDATOR (hydrogen bonding propensities).

2. Topology prediction

HMMTOP [9] predicts the localization of helical transmembrane segments and the topology of transmembrane proteins. It bases on the concept, that the transmembrane proteins topology is determined by the maximum divergence of amino acid composition of sequence segments. The method localizes certain sequence segments in areas used as structural parts (inside, outside, inside helix tail, outside helix tail and membrane helix). The method accuracy is enhanced by hidden Markov model, which controls the different segments length.

OCTOPUS [10] uses a combination of hidden Markov models and artificial neural networks. It first performs a homology search using BLAST to create a sequence profile used as the input to a set of neural networks which predict the location preference for each residue – transmembrane, interface, globular, loop, inside or outside. These predictions are used as input to a two-track hidden Markov model, which uses them to calculate the most likely topology. Results of these two sets are finally combined to obtain each residue preference.

SPOCTOPUS [11] uses the same algorithm as OCTOPUS, enforced with a predicting the signal peptide in the target sequence. Location of residues from this signal peptide is determined by a hidden Markov model prior the topology prediction by OCTOPUS, providing more accurate topology prediction.

PHOBIUS and POLYPHOBIUS [12] use a hidden Markov model, decoding algorithm that combines probabilities for sequence features of homologs by considering the average of the posterior label probability of each position in a global sequence alignment.

PHILIUS [13] (www.yeastrc.org/philius) is based on a hidden Markov model similarly to PHOBIUS and POLYPHOBIUS, expanded to using the more powerful class of dynamic Bayesian networks.

SCAMPI [14], a simple generic topology model, uses a contribution of position-specific amino acids to the free energy of membrane insertion that performs on a par with the current best statistics-based topology predictors. It is similar to a hidden Markov model in the sense that states and state transitions are used to define an underlying grammar.

References

1. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE (2000) The Protein Data Bank. Nucleic Acids Res 28: 235-242. gkd090 [pii].
2. Li W, Schulman S, Dutton RJ, Boyd D, Beckwith J, Rapoport TA (2010) Structure of a bacterial homologue of vitamin K epoxide reductase. Nature 463: 507-512. nature08720 [pii];10.1038/nature08720 [doi].
3. Liu S, Cheng W, Fowle GR, Shen G, Li W (2014) Structures of an intramembrane vitamin K epoxide reductase homolog reveal control mechanisms for electron transfer. Nat Commun 5: 3110. ncomms4110 [pii];10.1038/ncomms4110 [doi].
4. Frishman D, Argos P (1996) Incorporation of non-local interactions in protein secondary structure prediction from the amino acid sequence. Protein Eng 9: 133-142.
5. Sen TZ, Jernigan RL, Garnier J, Kloczkowski A (2005) GOR V server for protein secondary structure prediction. Bioinformatics 21: 2787-2788. bti408 [pii];10.1093/bioinformatics/bti408 [doi].
6. Jones DT (1999) Protein secondary structure prediction based on position-specific scoring matrices. J Mol Biol 292: 195-202. 10.1006/jmbi.1999.3091 [doi];S0022-2836(99)93091-7 [pii].
7. Ouali M, King RD (2000) Cascaded multiple classifiers for secondary structure prediction. Protein Sci 9: 1162-1176. 10.1110/ps.9.6.1162 [doi].
8. Cuff JA, Clamp ME, Siddiqui AS, Finlay M, Barton GJ (1998) JPred: a consensus secondary structure prediction server. Bioinformatics 14: 892-893. btb130 [pii].
9. Tusnady GE, Simon I (1998) Principles governing amino acid composition of integral membrane proteins: application to topology prediction. J Mol Biol 283: 489-506. S0022-2836(98)92107-6 [pii];10.1006/jmbi.1998.2107 [doi].
10. Viklund H, Elofsson A (2008) OCTOPUS: improving topology prediction by two-track ANN-based preference scores and an extended topological grammar. Bioinformatics 24: 1662-1668. btn221 [pii];10.1093/bioinformatics/btn221 [doi].
11. Viklund H, Bernsel A, Skwark M, Elofsson A (2008) SPOCTOPUS: a combined predictor of signal peptides and membrane protein topology. Bioinformatics 24: 2928-2929. btn550 [pii];10.1093/bioinformatics/btn550 [doi].
12. Kall L, Krogh A, Sonnhammer EL (2005) An HMM posterior decoder for sequence feature prediction that includes homology information. Bioinformatics 21 Suppl 1: i251-i257. 21/suppl_1/i251 [pii];10.1093/bioinformatics/bti1014 [doi].
13. Reynolds SM, Kall L, Riffle ME, Bilmes JA, Noble WS (2008) Transmembrane topology and signal peptide prediction using dynamic bayesian networks. PLoS Comput Biol 4: e1000213. 10.1371/journal.pcbi.1000213 [doi]
14. Bernsel A, Viklund H, Falk J, Lindahl E, von HG, Elofsson A (2008) Prediction of membrane-protein topology from first principles. Proc Natl Acad Sci U S A 105: 7177-7181. 0711151105 [pii];10.1073/pnas.0711151105 [doi].

D. PDB files contain the atomic coordinates of the models

S1. The atomic coordinates of the vitK^{EP}-hVKORC1 complex at t=0
S1-1 and S1-2. The atomic coordinates of the vitK^{EP}-hVKORC1 complex at t=100 ns of each replica (1 and 2)

S2. The atomic coordinates of the W-hVKORC1 complex at t=0

S2-1 and S2-2. The atomic coordinates of the W-hVKORC1 complex at t=100 ns of each replica (1 and 2)

Please do not adjust margins