Check for updates

# Identification of the functional states of human vitamin K epoxide reductase from molecular dynamics simulations†‡

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

In mammalians, the enzymatic activity of vitamin K epoxide reductase (VKORC1) requires a protein conformational reorganisation that includes several transient enzymatic states involving a dynamic electron transfer. Regarding the structurally non-characterised human enzyme (hVKORC1), this process remains poorly explained and the different redox states of the enzyme generated by its biochemical transformation are unknown. Here, we report a 3D model of the fully reduced hVKORC1 at the atomistic level. By exploring this model through molecular dynamics (MD) simulations, we established the most probable intermediate states of the enzyme which were used for generation of the putative functionally-related enzymatic states. Enzymatic functionality of each state was assigned by probing their recognition properties with respect to vitamin K in its quinone and hydroxyquinone forms. Two states were identified as contributing to the two-step vitamin K transformation. The state highly selective for native vitamin K was further validated through analyses of its free energy of binding with vitamin K agonists (VKAs) that showed a high correlation with the experimental inhibiting constants.

## Introduction

The mammalian vitamin K epoxide reductase (VKOR) VKORC1 is an endoplasmic reticulum (ER)-resident transmembrane protein involved in γ-carboxylation of glutamate residues of vitamin K-dependent proteins (PVKD). These proteins include blood clotting factors, matrix Gla protein and osteocalcin,[1,2] and their enzymatic modification by VKORC1 is mandatory for essential physiological processes, such as coagulation,[3] calcium homeostasis,[4] energy metabolism,[5] signal transduction and cell development.[6,7] Specifically, VKORC1 converts the inactive form of vitamin K (vitamin K 2,3-epoxide) into vitamin K quinone and its reduced active form, vitamin K hydroquinone, the cofactor used for PVKD activation.[1] The active site of VKORC1 contains a highly conserved CXXC motif[8] that is essential for vitamin K quinone reduction.[9] Two other cysteine amino acids, which participate in the electron exchange between the active site and a thioredoxin-like domain in bacterial species, are also conserved among

species.[10–12] The function of the latter cysteine residues in human VKORC1 remain controversial[1,13,14] as the regulation of human enzyme activity differs from that of the bacterial enzyme, *i.e.* in the absence of a thioredoxin domain, hVKORC1 should use an external partner as a redox source.[15]

Deregulation of the catalytic activity of hVKORC1, usually prompted by naturally occurring mutations, is responsible for different diseases including haemorrhaging disorders[16] and arterial calcifications,[17] whereas hVKORC1 polymorphisms are associated with resistance to vitamin K agonists (VKAs).[18] Analysis of hVKORC1 structure–function relationships may provide a better understanding of its activation process and guide therapeutic strategies aimed at neutralizing enzyme dysfunctions.

As for many membrane proteins, VKOR structure remains very much a matter of debate. In the absence of hVKORC1 structural data, homology models were built from the crystallographic structure of *Synechococcus* sp. VKOR, a bacterial homolog of hVKORC1, in which the 280-amino acid sequence is organised into five transmembrane helices.[10] This homology modelling has produced divergent topologies of the human protein, including a three-helix (3-H TM) a four-helix (4-H TM) and a five-helix (5-H TM) transmembrane model[12,19–22] (Fig. 1). The 3-H TM model suggested by Stafford[23,24] was improved by adding a charged residue distribution that affects TM orientation.[25,26] The 4-H TM topology model generated by a recent biophysical study of VKORC1 and VKORC1L1 could be more

*[a]Centre de Mathématiques et de Leurs Applications (CMLA), ENS Cachan, CNRS, Université Paris-Saclay, 61 avenue du Président Wilson, Cachan, 94235 France. E-mail: Luba.Tchertanov@ens-cachan.fr*

*[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 *et al.*

‡ Electronic supplementary information (ESI) available: 11 figures, 1 table and methods details. See DOI: 10.1039/c7ra07463h
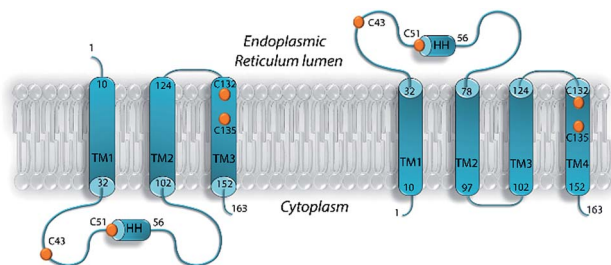
Fig. 1 Proposed topological models for VKOR. A three-helix transmembrane domain (3-H TMD) model with the N- and C-terminus located in the endoplasmic reticulum (ER) lumen and in the cytoplasm respectively (left). A four-helix transmembrane domain (4-H TMD) model with the N- and C-terminus located in the cytoplasm (right). Positions of the first and the last residues of the transmembrane (TM) helices and a short helix (HH) are shown. The conserved cysteine residues are denoted as orange balls.

physiologically relevant[27] but establishing a reliable 3D model for hVKORC1 remains a challenge.

VKORC1 contributes to vitamin K transformation within the endoplasmic reticulum.[28] It has been proposed that the redox partner of VKORC1 is a protein disulphide isomerase or thioredoxin-like protein, namely TMX, TMX4, and ERp18.[9,29]

A thiol/disulphide-based mechanism of vitamin K reduction proposed in ref. 8 and 30 postulates a sequence of chemical transformations involving the four cysteine residues of hVKORC1 – C43, C51, C132 and C135 – in particular, their reversible protonation/deprotonation of the sulphur atom (S) followed by forming/breaking of the disulphide bonds. These events are associated with hVKORC1 conformational reorganisation leading to transition from the passive (inactive) to the active states through transient intermediate enzymatic products.[31] There are different combinations of protonation states for each cysteine residue coded for the nine distinct forms of VKORC1 – the extreme sets in which all cysteine residues are either deprotonated and linked by S–S bonds (oxidised form) or protonated (SH, reduced form) and the intermediate sets formed by combination of protonated and deprotonated cysteine residues. The endoplasmic reticulum location of VKORC1 suggests that once integrated into a membrane, the protein folding is stabilized by the two disulphide bonds, maintaining a fully folded (inactive) enzyme state. This enzyme state can be activated by external factors, such as binding of thioredoxin(s) initiating a reduction in hVKORC1 protein by thiol–disulphide exchange.[9,32] Because of electron-transfer, the covalently linked cysteine residues of hVKORC1 switch to enzymatically active items, specifying intermediate (or meta-stable) states of the protein that become capable of transforming vitamin K. In accordance with a multi-step procedure for vitamin K modification, the biochemical and structural data show four or five intermediate states.[11] Such an abundance of enzymatic transformations and the approximate assignment of the cysteine residue's states/functions indicate that modelling of the activation process and definition of the enzymatically active and biologically relevant hVKORC1 states is a *bona fide enigma*.

Considering the questions addressed on VKORC1 topology and on the assignment of the biologically consistent states, the theoretical description of hVKORC1 structural features should be (i) maximally disassociated from the uncertain/contradictory literature data, and (ii) based on methods yielding pragmatic and sine *qua non* releases.

In the present work, we have designed an hVKORC1 3D atomistic model to be used in MD simulation. To avoid *a priori* assignment of the sulphur bridge, we performed this simulation on a fully reduced form of enzyme. We do not state that the protein dynamics of this model, which are described using the classical mechanics approximation, will reproduce exactly the functional states of protein that are mandatorily associated with proton exchange and disulphide bridge formation/disruption. However, the MD conformations generated over multiple short simulations (100 ns) and the extended trajectories (the microsecond time scale), could be sufficient to predict the functionally relevant enzymatic states of hVKORC1. We distinguished four conformations viewed as plausible mimics of the intermediate states of enzyme. The predicted conformations were further optimised and subjected to MD simulations producing the functionally related enzymatic states, which were assigned by probing their recognition properties with respect to vitamin K in its quinone and hydroxyquinone forms. Two states were identified as contributing to two-steps vitamin K transformation. The state highly selective for native vitamin K was further validated through analyses of its free energy of binding with the vitamin K agonists (VKAs) that showed a high correlation with the experimental inhibiting effect.

## Results and discussion

### 1. Modelling of hVKORC1

After searching the Protein Data Bank (PDB),[33] we retrieved four reports describing the structure of VKOR from *Synechococcus* sp. (bVKOR, 283 amino acids, aas) (ESI, Table S1‡). The alignment of the sequences revealed a bVKOR segment (13–181 aas) having the best identity/homology (25/36%) with hVKORC1 (163 aas) (ESI, Fig. S2A‡). Since the homology is weak between sequences, we first used a set of independent bioinformatics methods for prediction of the secondary structure composition in proteins. These methods are designed on distinct concepts and algorithms and they correctly forecasted all structural elements of the experimentally characterized bVKOR (4NV5, resolution of 2.80 Å)[11] (ESI, Fig. S1‡), thus validating the accuracy of the prediction methods. These methods were applied to predict the secondary structure of the target sequence, hVKORC1. Apparently, the overall structure of hVKORC1 consisted of one short and four elongated helixes (ESI, Fig. S2B‡). Secondly, we built a homology model for hVKORC1 using 4NV5 as a template. The homology model denotes the four-helix structure of the VKORC1 transmembrane domain, which is consistent with the predicted secondary structure composition and the experimental data supporting the 4-H TM topology.[27] In our model, the first and the second transmembrane helices, respectively TM1 and TM2, are linked by the extended loop ($L_2^1$) with a small horizontal helix (HH) in the middle, while the other helices are
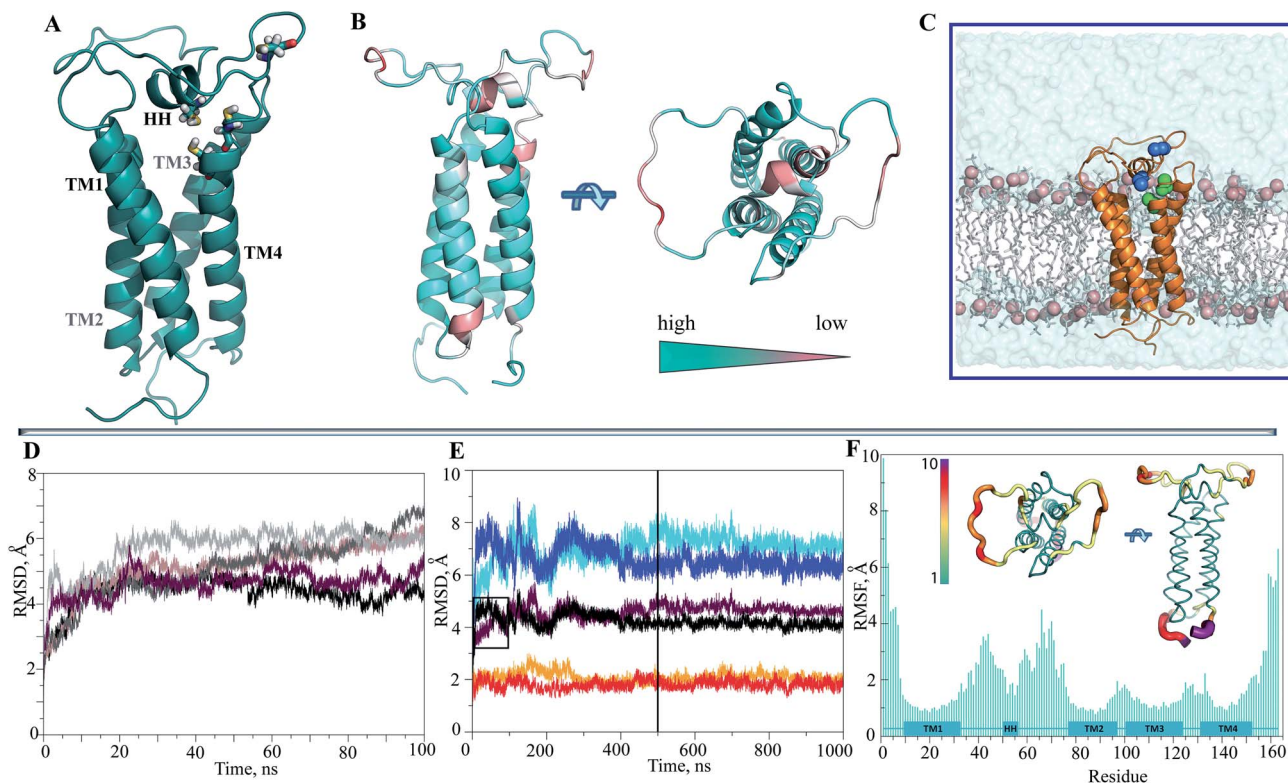
Fig. 2 Modelling and molecular dynamics simulations of hVKORC1. (A) The homology model of hVKORC1 is pictured as cartoon. The cysteine residues (C43, C51, C132 and C135) are shown in sticks coloured by the atoms type. (B) The per domain quality estimation of the hVKORC1 model that shown in two orthogonal projections. The quality of prediction is specified by color ranged from deep teal (high) to light salmon (low). (C) The simulated system is composed of the hVKORC1 model embedded into membrane and completed with water molecules. (D) The RMSDs from the initial coordinates ($t = 0$ ns) are computed on the $C\alpha$ atoms of 100 ns MD trajectories (replicas 1–5 are differed by color). (E) The RMSDs from the initial coordinates ($t = 0$ ns) computed from two extended trajectories, of 500 μs ($\mathbf{1} \rightarrow \mathbf{1'}$ and $\mathbf{3} \rightarrow \mathbf{3'}$, delimited by vertical line) and of 1 μs ($\mathbf{1} \rightarrow \mathbf{1''}$ and $\mathbf{3} \rightarrow \mathbf{3''}$) on the all $C\alpha$ atoms (brown/black), on the $C\alpha$ atoms from the TM helices (orange/red) and on the $C\alpha$ atoms from the extended luminal loop (cyan/blue). A small rectangle borders the 100 ns part of trajectories $\mathbf{1'}$ and $\mathbf{3'}$. (F) The RMSFs are computed on the $C\alpha$ atoms over the 1 μs simulation ($\mathbf{3''}$) of hVKORC1. In inserts: the average conformation of hVKORC1 ($\mathbf{3''}$) is presented as tubes in two perpendicular projections. The tube size is proportional to the by-residue atomic fluctuations computed on the $C\alpha$ atoms. The high fluctuation regions are specified by color ranged from green, the most rigid, to violet, the most mobile.

linked by the short loops in a head-to-tail manner, as TM2–TM3 and TM3–TM4 (Fig. 2A). The four-helix compacted transmembrane bundle is stabilized by (i) the coil-coiled dimeric structural motif composed of a *HxxHCxC* heptad repeat (hydrophobic residues H, charged residues C and random residues x) – [24]LYALHVK[30] and [94]LGCLRTR[100] – from TM1 and TM2 respectively, and (ii) the mutual hydrophobic interactions between four TMs (ESI, Fig. S3‡).

The two cysteine residues, C43 and C51, are positioned at $L_{\frac{1}{2}}$ and the two others, C132 and C135, at helix TM4. The feasibility of the prediction was encouraging, with the per domain estimation indicating a low quality secondary structure prediction for only several residues at the extremity of TM1 and TM3, for the HH, and for the distal segments of $L_{\frac{1}{2}}$ (Fig. 2B).

Assignment of the transmembrane fraction in template and target was achieved by the two-prediction consensus associated with the secondary structure and the membrane topology contents respectively (ESI, Fig. S1 and S2B‡).

In accordance with the 4-H TM topology, our *in silico* prediction showed that the extended loop $L_{\frac{1}{2}}$ and the short

TM3–TM4 linker were located in ER lumen, whereas the N- and C-terminals together with the loop linker TM2–TM3 were positioned in the cytoplasm. The transmembrane topology assignment was used for an appropriate embedding of the hVKORC1 structural model in phospholipid bilayers, mimicking the ER membrane (Fig. 2C). Instead of phosphatidylethanolamine (PE) used in a bilayer composition in the MD simulation of hVKORC1,[34] we generated the membrane with phosphatidylcholine (PC) molecules, as this is most appropriate for reticulum content.[35]

## 2.  Structure and dynamics of hVKORC1

Since the simulated system, which is composed of the hVKORC1 homology model embedded in the membrane and completed with water molecules (Fig. 2C), may be too large to expect statistical convergence within a 100 ns MD simulation, we first performed five repeats (replicas 1–5) to probe the conformational variability of protein using the random starting velocities. Such probing may be pertinent when scanning a large conformation diversity of the fully reduced model. The
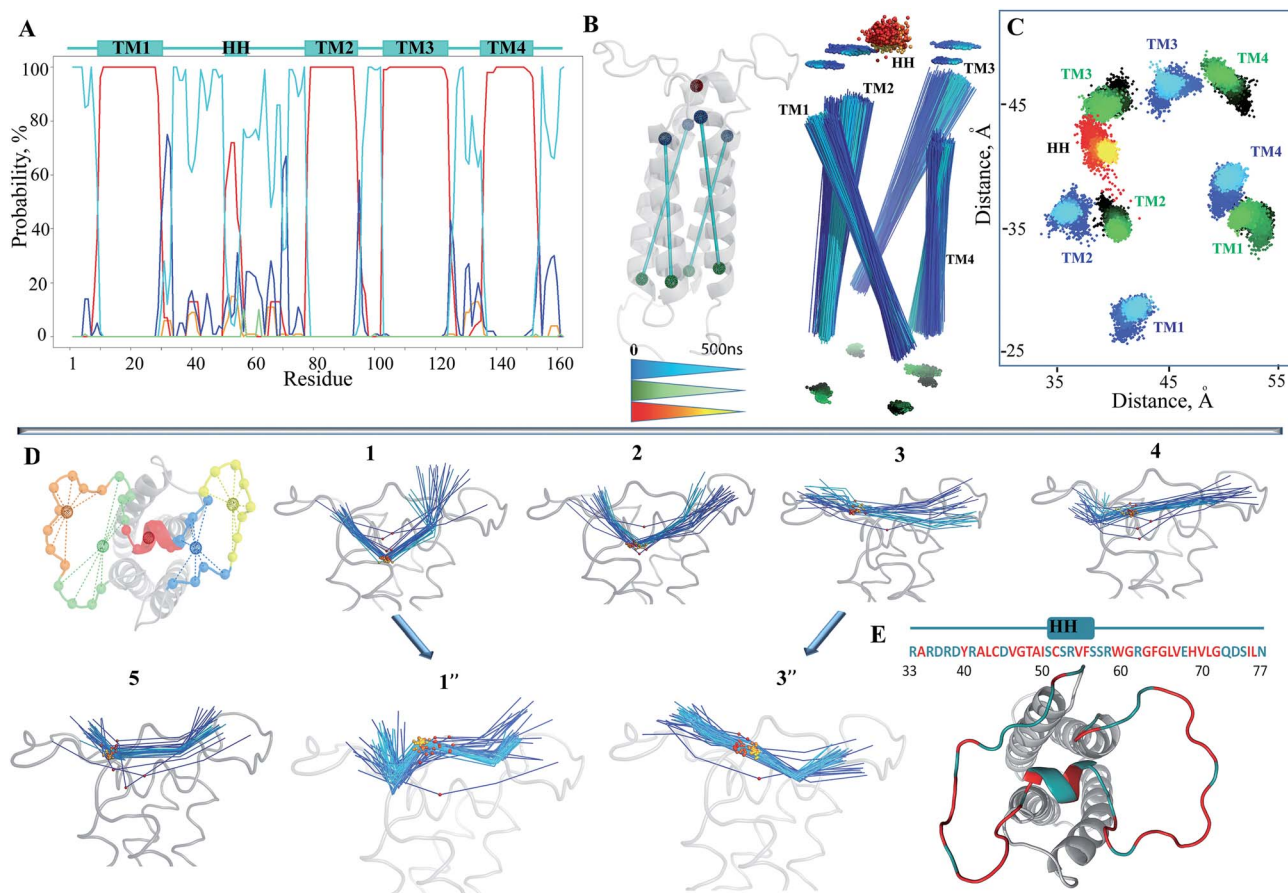
**Fig. 3** Structural features of hVKORC1. (A) The secondary structure evolution over MD simulations (all trajectories were merged). The $\alpha$-helix, $3_{10}$ helix, $\beta$-bridge, turn and loop are shown in red, orange, light-green, blue and cyan, respectively, and referred to the predicted helices (top). (B–C) Drift of helices, monitored over the extended (1 $\mu$s) simulations, is shown for **3″**. 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 (B) and as the points projected on the $x$–$y$ plane (C) coloured from blue ($t = 0$) to cyan ($t = 1$ $\mu$s) for the top of TM helices; from black ($t = 0$) to green ($t = 1$ $\mu$s) for the bottom; and from red ($t = 0$) to yellow ($t = 1$ $\mu$s) for HH. (D) Drift of the two $L_{\frac{1}{2}}$ sub-fragments, R33–S50 and S57–N77, respect to HH was monitored for each simulation separately. Centroids were defined on HH and on each sub-fragment of the $L_{\frac{1}{2}}$, subdivided into the proximal and distal parts in respect to HH as shown at the left. Coordinates of each centroid (five in total) were computed for each MD simulation time and presented as the lines connecting the two sequent centroids projected on the $y$–$z$ plane, and presented as the color-coded lines or points superimposed into the hVKORC1 conformation at $t = 0$ (grey). (E) Hydrophobic/hydrophilic (red/cyan) content of $L_{\frac{1}{2}}$ is shown on the sequence and on the structure.

root mean square deviations (RMSDs), computed using the positions of C$\alpha$ atoms relative to the initial structure ($t = 0$ ns) showed the poor stability of this model within the simulation time (Fig. 2D). Two randomly chosen MD simulations (**1** and **3**) were prolonged to 500 ns (**1′** and **3′**) and further to 1$\mu$s (**1″** and **3″**). It was supposed that the extended simulations of hVKORC1 would allow either the equilibrated state(s) of protein to be reached or the conformations mimicking the functional states of enzyme to be produced. In these extended trajectories (**1″/3″**) the RMSDs (all C$\alpha$ atoms, brown/black) showed large variations (4.5–6.5 Å) until ~300–400 ns and further, demonstrating their fine stability (Fig. 2E). The largest contribution to the ample RMSDs was gathered by the extended luminal loop, $L_{\frac{1}{2}}$, (blue/cyan) while the transmembrane helices (orange/red) had no considerable impact.

To quantify the extent to which the homology model represents a realistic and stable structure in the timescales studied,

we assessed the protein secondary structure with DSSP.[36] Analysis performed on the individual MD trajectories of different lengths and on the merged data, indicated well-defined, long-lived transmembrane helices (Fig. 3A, ESI, Fig. S4A‡), supporting a correctness of the four-helix model. However, $L_{\frac{1}{2}}$ demonstrated some structural variability over MD simulations. In particular, the fold of sequence R33–N77 ($L_{\frac{1}{2}}$) was interpreted as the two loops linked either by a short $\alpha$-helix (70–90%) or by a $3_{10}$-helix/turn. The tiny turns and the short $\alpha$- or $3_{10}$-helices observed in the $L_{\frac{1}{2}}$ and in the loops linking each pair of transmembrane helices were only partially stable.

The root mean square fluctuations (RMSFs) computed for each residue from either the short (100 ns) or the prolonged (1 $\mu$s) simulations were comparable (ESI, Fig. S5‡). As expected, the highest fluctuations were principally observed for the residues of $L_{\frac{1}{2}}$, of the N- and C-terminals and also, but essentially less so, of the loop-linkers, TM2–TM3 and TM3–TM4 (Fig. 2F).
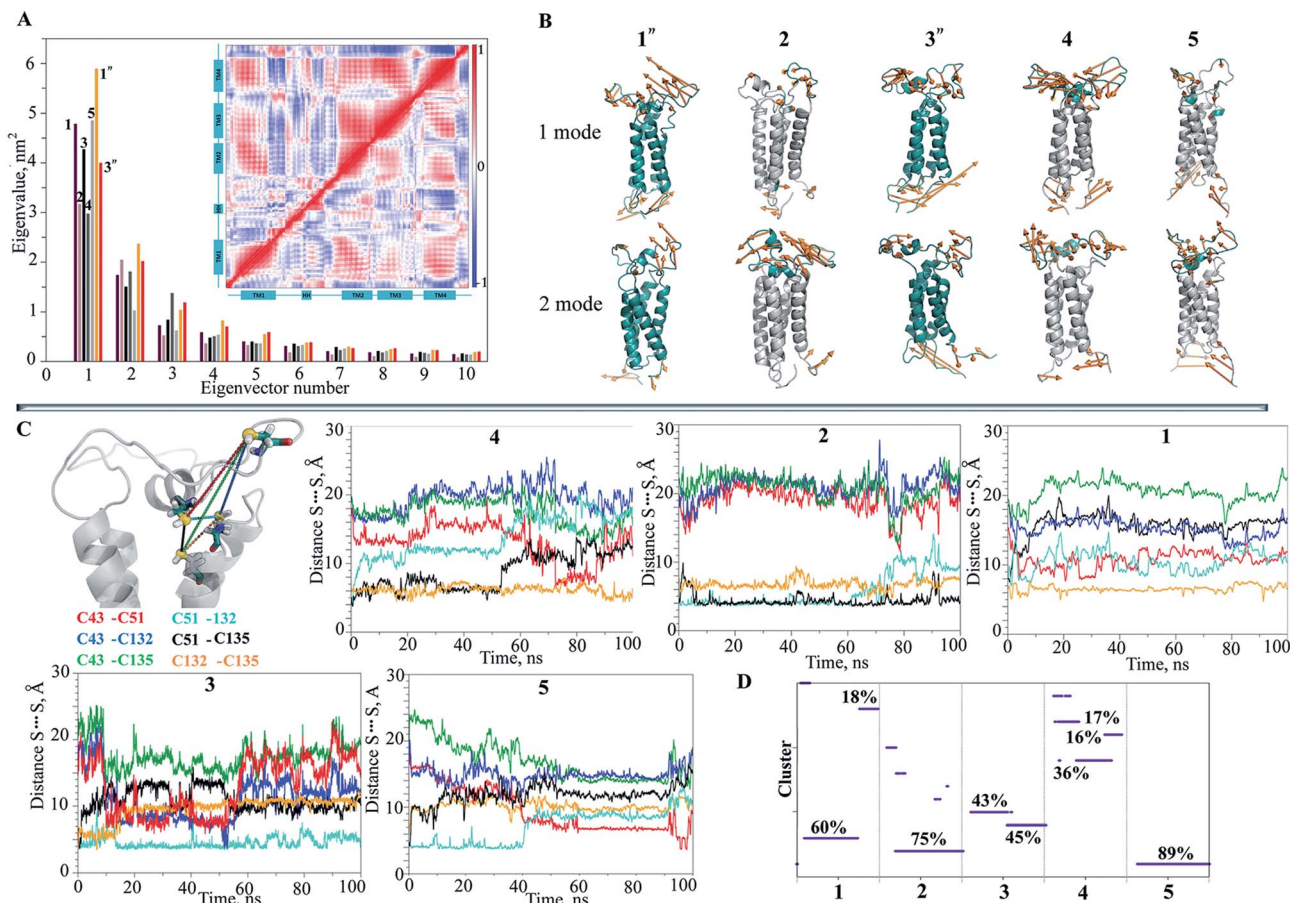
Fig. 4 Structural and dynamical features of hVKORC1. (A) Principal component analysis (PCA) was performed on the backbone atoms from the 100 ns (brown, dark salmon, black, grey dark and grey light bar graph for replicas **1–5** respectively) and the 1 μs MD trajectories (orange and red for **1″** and **3″** respectively). In insert: inter-residue cross-correlations maps resulting from PCA of two extended (1 μs) trajectories, **1″** (the upper half-matrix) and **3″** (the lower half-matrix). Correlated and anti-correlated motions between atom pairs are presented as color gradient of red (positive) and blue (negative) respectively. (B) Slow motions (modes 1 and 2, computed for each trajectory) are illustrated as orange arrows on the cartoon representation of protein coloured in deep teal for extended simulations (**1″** and **3″**) and in grey for short simulations (**2**, **4** and **5**) in which the most moving fragments are highlighted in deep teal. (C) Interatomic S···S distances between the sulphur atoms from cysteine residues in hVKORC1 (left) were monitored over each short (100 ns) MD simulations. The curves characterizing the S···S distances between different pairs of cysteine residues are differentiated by color: C43–C51 is in red, C43–C132 is in blue, C43–C135 is in green, C51–C132 is in turquoise, C51–C135 is in black and C132–C135 is in orange as showed on the first image, together with the axes labels. (D) The clusters of conformations defined with cutoff of 2.0 Å and their population (in %) calculated for each short (100 ns) simulation.

To identify the source of the high RMSD values, we first analysed the structural drift of the folded TM and HH helices, and secondly that of the coiled $L_2^1$. Surprisingly, within the same TM helix the movement of residues at the extremities of the TM helices, either close to the cytoplasm or to the ER lumen, was not equivalent during the simulations. This difference was in the path direction and the absolute value of displacement, resulting in a bent TMs drift (Fig. 3B and C; ESI, Fig. S4B‡).

The small HH of $L_2^1$ is well conserved over the simulations and to a first approximation may be viewed as the internally stable structural element in the studied time-scale that is able to move with respect to the membrane helices rather than change a structure fold *per se*. The HH is displaced largely in a plane defined by the apexes of the TMs and along their axes (Fig. 3C). This ample alteration of HH position may simultaneously depend on the TM helices drift and on the large movements of

the two coiled sub-fragments of $L_2^1$, encompassing R33–S50 and S57–N77 respectively (Fig. 3E). The drift of these sub-fragments showed that the luminal part of hVKORC1 was running from a 'cave-like' to a 'plateau-like' configuration (Fig. 3D), showing that the relative positions of the two loop sub-fragments and of the HH and TM helices varied from proximal to distant.

The slow motions of hVKORC1 were characterized by the principal component analysis (PCA). Ten PCA modes were sufficient to describe ∼80–90% of the total fluctuations of the hVKORC1 backbone (Fig. 4A). The two first modes illustrate large atomic motions of the residues from the loop $L_2^1$, the loop-linkers TM2–TM3 and TM3–TM4, and the N- and C-terminals. Nevertheless, the direction of the motions and their amplitudes presented by each corresponding mode are not equivalent in distinct trajectories (Fig. 4B).
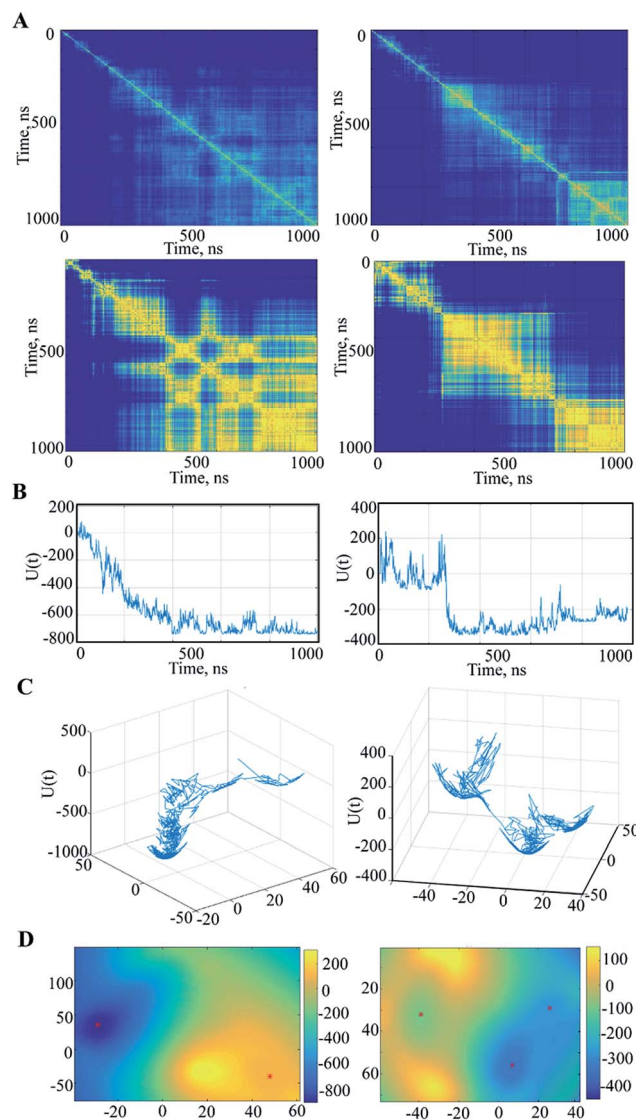
**Fig. 5** Spectral analysis of the MD simulation data. (A) Similarity matrices (pairwise distance) between all observed conformations $\{x_i, i = 1, …, N\}$ computed for all C$\alpha$-atoms of $\mathbf{1}''$ and $\mathbf{3}''$ trajectories before (top) and after the PCA (bottom), in which each vector $x_i$ was replaced by the vector $\xi_i$ of a dimension 2. The color-coded distance between each vector, yellow for the similar conformations and blue for the distant. (B) The temporal evolution of the variation $\delta U(\xi_i) \doteq U(\xi_i) - U(\xi_0)$ with respect to the energy of the initial conformation at $t = 0$. (C) The trace of the trajectories of the conformations after a 2D PCA reduction on the energy surface for trajectories $\mathbf{1}''$ and $\mathbf{3}''$ plotted with respect to the two first PCA components. (D) The energy landscapes for the two trajectories using the spline surface representation.

Similar to the RMSF values, the PCA denotes the greatest mobility of L$\frac{1}{2}$ with respect to the other domains. The PCA cross-correlation patterns of the short and extended trajectories are generally comparable and demonstrate the highly coupled motions between distant structural subunits (insert in Fig. 4A; ESI, Fig. S6‡). The strong correlation of the TM helices motion, apparently, is associated with their drift, while the large dynamic flexibility of the extra-membrane loop L$\frac{1}{2}$ exposed to an aqueous environment demonstrates a rich conformational variability in the hVKORC1 model.

Since the distance between sulphur atoms (S⋯S) in the cysteine residues is the indicative metric that distinguishes conformations between favouring or frustrating the formation of a disulphide bond, S–S, the six distances between each pair of sulphur atoms were monitored over each MD trajectory (Fig. 4C). Despite the same initial structural model at $t = 0$, in which three S-atoms (from C51, C132 and C135) are adjacent (all interatomic S⋯S distances of 4–6 Å), and the sole S-atom (from C43) is largely distant (at 20 Å) in this S-cluster, the model evolution over the different MD trajectories was not equivalent. Each S⋯S distance is initially either short or large and then this varied significantly during the simulation with respect to the primary value. Delimiting the range of 4–6 Å as favourable for the establishment of the S–S bond, and considering an occurrence of such a contact during the simulations, we suggest that the pairs C43–C51, C51–C132, C132–135 and C51–C135 are the candidates for the S–S covalent bonding. Conversely, the systematically observed large distance between C43 and C135, and between C43 and C132, indicates a low probability of interaction between these pairs of residues. The multiple MD conformations have a short S⋯S distance between two pairs of cysteine residues simultaneously (C43–C51 and C132–C135) and this may encourage the formation of the two S–S bridges.

To focus our analysis on the protein backbone conformational changes, we decided to perform the clustering calculations. We found that (i) the MD conformations of each trajectory were grouped (cutoff of 2 Å) in one (**2**, **5**), two (**1**, **3**) or more (**4**) clusters (subsets) and (ii) the subsets are mutually exclusive (the portioning effect) (Fig. 4D), giving large conformational samples to be analyzed.

We further searched the conformational space explored by protein in the 1 μs MD simulations using an estimation of the population (density) of the conformations in the low dimensional space (the first PCA mode). The similarity matrix $W$ (see Methods and References herein) computed from the raw conformations or after a PCA reduction, is structured into blocks, each consisting of a set of close conformations (Fig. 5A).

The different blocks are plausibly associated with metastable or intermediate states. The blocks show distinct patterns in the analysed trajectories that are partitioned (after the 200–300 ns) into the segregate blocks in $\mathbf{3}''$ and are arranged into the extended rosette-like crowds in $\mathbf{1}''$, which may indicate revisited conformations. Note first that the PCA dimension reduction does not really degrade the cluster patterns but results in highly contrasted similarity content. Second, for computation and display purposes, this PCA reduction can be helpful for the data analysis. Each vector $x_i$ is replaced by a vector $\xi_i$ of very small dimension $r$ by a PCA (typically, $r = 2$ or 3).

The local occupation density $\pi(x)$ at a given scale $\sigma$ is estimated as described in Methods, inducing a free energy landscape $U(x)$ that reflects the similarity matrix composed of blocks corresponding to time intervals where the energy is relatively stable. This property is emphasized by the slow variable profiles, as shown in ESI (Fig. S7 and S8‡) for the trajectory $\mathbf{3}''$. A slow variable is a one-dimensional function $q(x_i)$ that is related to the underlying Langevin dynamics and thus to the similarity
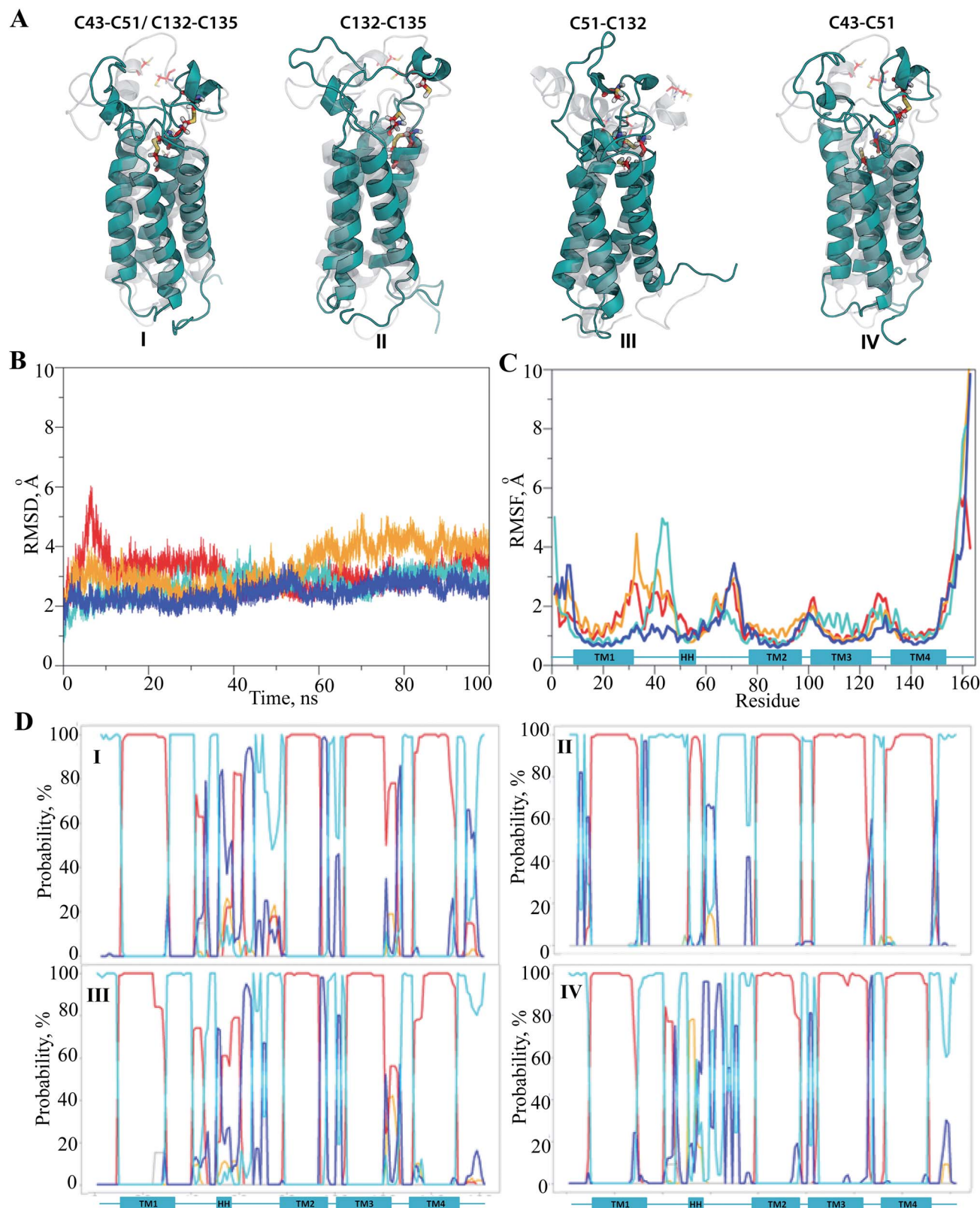
**Fig. 6** Molecular dynamics simulations of the hVKORC1 models $M^I$–$M^{IV}$. (A) The optimized and refined by MD simulations structural models (deep teal) with S–S of 2.1 Å are superposed with conformations identified from MD simulations of fully protonated model (grey). For each model, numerated by Roman, the pairs of cysteine residues forming S–S bond are denoted at the top. Protein is depicted as cartoon, the cysteine residues (C43, C51, C132 and C135) are shown as sticks coloured by the atoms type. (B) The RMSDs from the initial coordinates ($t = 0$ ns) are computed on the Cα atoms over MD trajectories (100 ns) of $M^I$–$M^{IV}$ (blue, orange, deep teal and red, respectively). (C) The RMSFs are computed on the Cα atoms over the 100 ns simulations of $M^I$–$M^{IV}$ (blue, orange, deep teal and red, respectively). (D) The secondary structure evolution over MD simulations of $M^I$–$M^{IV}$. 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.

matrix $W$, which gives a synthetic and simplified view of the blocks in $W$, as is the case in spectral clustering.[37,38]

For the free energy profile the temporal evolutions of energy $U$ are plotted as curves $t_i \rightarrow U(x_i)$ or $t_i \rightarrow U(\xi_i)$ depending on whether one considers PCA dimension reduction or not. However, it is interesting to represent the free energy landscape $U(x)$ as a hypersurface parametrized using the conformation $x$. For visualization we use a 2D reduction to display the free energy landscape as a 2D surface in 3D as $\xi \rightarrow U(\xi)$, and display the evolution of the dynamics as a random walk in this landscape. Doing so, the trace of the path of the random walk on the energy landscape is plotted (Fig. 5C) and reveals a sampled version of the main visited basins. The energy landscape is useful for computation of the energy minima locations, as shown in Fig. 5D. In fact, these minima are extracted from a continuous smooth representation of the sampled surface, which is computed using a spline model.[39]

When applied to the trajectory $3''$, the energy landscape shows three well-defined and highly populated local minima that probably correspond to the metastable states of hVKORC1, and are linked by saddles connecting two neighbouring minima that describe a transition between the different states (Fig. 5C and D). The observed dynamics of these multiple-basin landscapes involves long period motions within a basin followed by fast transition to another basin, which is an event described as a 'jumping among minima' that has been observed in a study of molecular dynamics using other methods.[40] The height of the energy barriers is proportional to the timescale that is necessary to stochastically convert between states. The energy landscape of $1''$ trajectory, with a meaningless minimum at the beginning of a simulation, is better described as a fall of leaves to the ground reaching the global minimum.

In summary, based on the prediction of secondary structure and comparative (homology) modelling we have obtained a 3D model of hVKORC1 consisting of four transmembrane helices that corresponds to the 4-H TM topology. The MD simulations of this model in the fully reduced state and integrated into a membrane were analysed with the statistical methods (RMSDs, RMSFs, PCA, cross-correlation, clustering), which gave evidence on features of the hVKORC1 dynamics that can be related to its functions. In particular, the ability of the extended luminal loop to undergo ample movements reflects the conformational plasticity of hVKORC1 enabling us to provide suitable sites for substrate recognition and binding. The binding pocket for thioredoxin proteins, which are the hVKORC1 reducers, may be constituted by numerous hydrophobic residues (60%) in the loop (Fig. 3E). The correlated movements of the transmembrane helices and the luminal loop indicate the global conformational adaptability of hVKORC1, which is relevant to the proton transfer processes followed by forming/breaking of the disulphide bonds, and consequently, to the hVKORC1 conformational reorganisation that leads to different transient enzymatic states, from inactive to active. Observing the variation in distances between the sulphur atoms from the cysteine residues indicates a great conformational variability in hVKORC1 associated with these multiple states.

We selected pairs of cysteine residues viewed as candidates for establishing disulphide bonds. The stochastic MD trajectory representation by the similarity matrix, energy profile and energy surface, provided the highly populated minima that are composed of conformations showing short S⋯S contact between two pairs of cysteine residues.

The hVKORC1 conformations having a short C51–C132 contact that may mimic the active state of hVKORC1 described in the literature[31] and observed in most crystallographic structures of the bacterial homologue, bVKOR. The hVKORC1 conformation showing a short S⋯S distance between the pairs of cysteine residues C132–135 is also observed in crystallographic structure 3KP9 of bVKOR.[10]

## 3. Modelling the enzymatically relevant states of hVKORC1

The hVKORC1 conformations that have a short S⋯S distance between the pairs of cysteine residues were locally modified to adjust the predicted state of hVKORC1. Specifically, the hydrogen atoms connected to at the sulphur atoms that exhibited a short S⋯S distance were removed and the S⋯S distance was constrained to 2.1 Å, which is typical for an S–S bond. The resulting models were optimised and made the subject of standard MD simulations. These models may represent four intermediate states of hVKORC1 – the state stabilised with two pairs of covalently bonded sulphur atoms from cysteine residues C43–C51 and C132–C135 ($\mathbf{M^I}$), and the states stabilized by a lone S–S bond linking either C132–C135 ($\mathbf{M^{II}}$) or C51–C132 ($\mathbf{M^{III}}$) or C43–C51 ($\mathbf{M^{IV}}$) (Fig. 6A).

It was suggested that the MD trajectories of 100 ns would be sufficient to achieve the equilibrated states since the models' $\mathbf{M^I}$–$\mathbf{M^{IV}}$ were stiffened by the one or two S–S bridges with respect to the fully-protonated initial model of hVKORC1 with four cysteine residues in reduced form (SH). The RMSDs, which were computed for the Cα atom positions relative to the initial structure ($t = 0$ ns), show the excellent ($\mathbf{M^I}$) and satisfactory stability ($\mathbf{M^{II}}$–$\mathbf{M^{IV}}$) within the simulation time (Fig. 6B). The RMSFs showed that the largest contribution to the ample fluctuations in all simulated models was gathered by the extended luminal loop, $L\frac{1}{2}$ and the inter-helices loops, which is similar to the fully-protonated model (Fig. 6C and 2F). However, the values of these fluctuations were considerably diminished. The lowest values of RMSFs were observed in the MD simulation of $\mathbf{M^I}$, where only the S57–N77 loop fluctuations up to 3.5 Å were present. In models containing a single S–S bond fluctuation values up to 5 Å were observed, varying in the order: $\mathbf{M^{II}} > \mathbf{M^{IV}} > \mathbf{M^{III}} > \mathbf{M^I}$.

Analysis of the protein secondary structure content (DSSP)[36] indicated well-defined, long-lived transmembrane helices, which was strongly conserved in all models (Fig. 6D; ESI, Fig. S9‡). In contrast, $L\frac{1}{2}$ demonstrated some structural variability that was similar to the fully-protonated model. Nevertheless, the overall conformational variability was diminished for each model, $\mathbf{M^I}$–$\mathbf{M^{IV}}$, as shown by the reduced number of clusters composed of similar conformations (within a cut-off of 2 Å). Indeed, only one highly populated cluster was observed in the conformational set of $\mathbf{M^I}$, $\mathbf{M^{III}}$ and $\mathbf{M^{IV}}$ with populations of
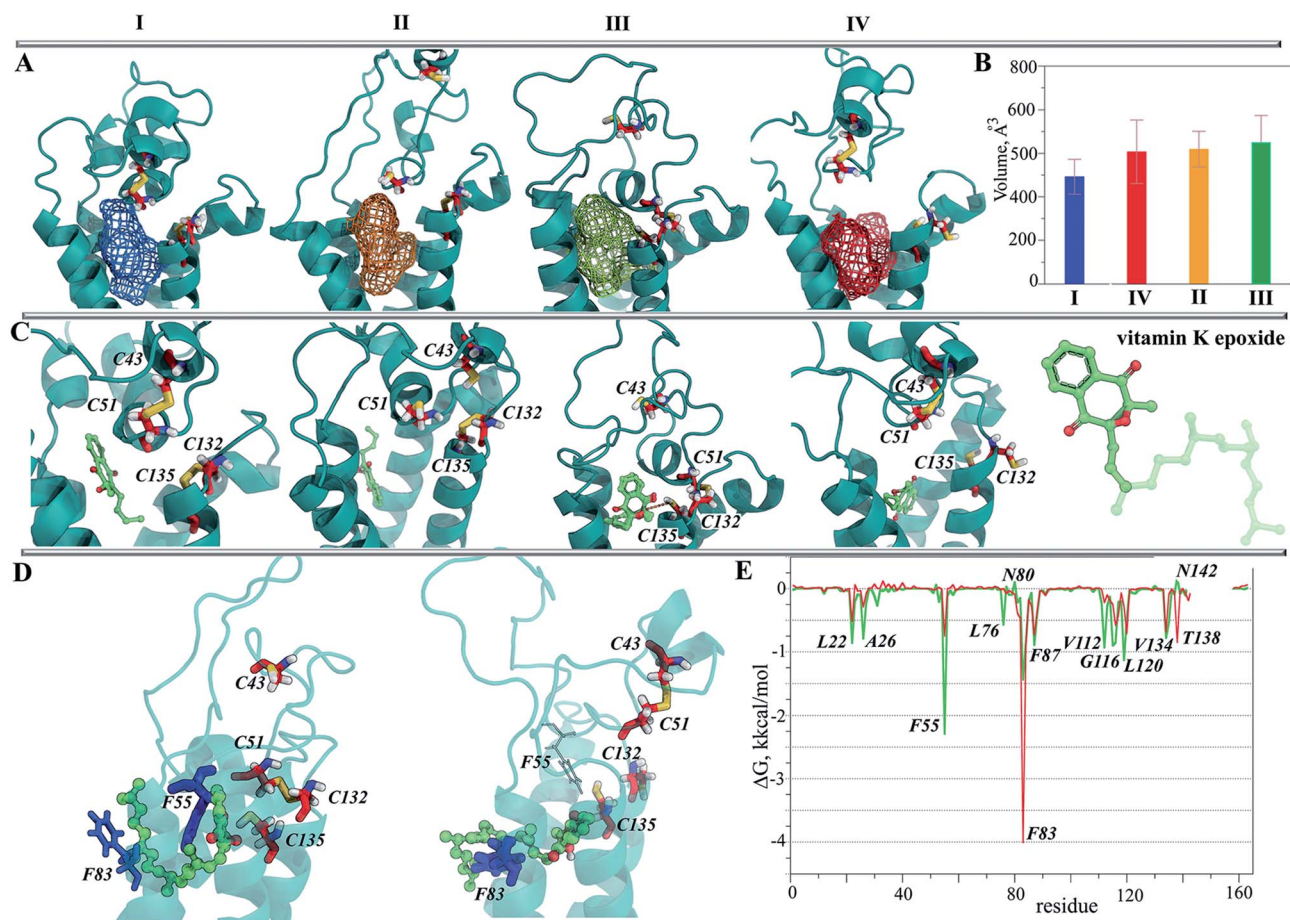
**Fig. 7** Representation of the hVKORC1 states and their binding properties in respect to vitamin K. (A) Pocket (in mesh) detected at each representative conformation $M^I-M^{IV}$ (left) and (B) its volume (right). (C) The best docking poses of the cleaved vitamin K epoxide found for each representative conformation obtained from MD simulations of $M^I-M^{IV}$. Vitamin K epoxide with the cleaved chain in opaque is showed in right. (D) The complexes vitK$^{EP}$·$M^{III}$ (left) and vitK$^{2-OH}$·$M^{IV}$ (right) equilibrated by MD simulations. (E) The per residues contribution to the binding free energy ($\Delta G$) in vitK$^{EP}$·$T^{III}$ (green) and vitK$^{2-OH}$·$T^{IV}$ (red) complexes. The labeled residues are characterized by the higher $\Delta G$ values. The protein is shown as cartoon with cysteine residues in sticks; vitamin K is shown in ball-and-sticks.

76, 73 and 56%, respectively, as well as in $M^{II}$ where the conformations are grouped in three distinct clusters with populations of 37, 24 and 18% (ESI, Fig. S10‡).

The representative conformations extracted from each of the most populated clusters are different (i) in conformation of L-loop, (ii) in folding of HH and (iii) in position of HH relative

**Table 1** The vitamin K docking into hVKORC1. Each form of vitamin K was docked (AutoDock) on every target. Docking poses were characterized by a number of poses, their population ([in %], separated by slash) and scores (separated by slash, in kcal mol$^{-1}$). Poses with population more than 10% (in black) were analysed. The best score values are denoted in bold

| Form | | | |
|---|---|---|---|
| vitK$^{EP}$ | vitK$^{2-OH}$ | vitK$^{3-OH}$ | vitK$^{QU}$ |
| State | | | |
| $T^I$ | | | |
| 2 [58/42] −7.0/−6.8 | 7[3/54/22/4/9/1/7] −7.2/−6.9 | 3[57/5/38] −7.3/−6.8 | 5[60/17/1/19/3] −7.0/−7.0/−6.9 |
| $T^{II}$ | | | |
| 5[42/25/4/16/13] −6.9/−6.9/−6.7/−6.6 | 9[57/25/1/1/3/2/2/2/7] −6.7/−6.4 | 7[11/21/2/5/2/3/56] −6.7/−6.7/−5.7 | 6[52/20/4/1/14/9] −6.8/−6.7/−6.5 |
| $T^{III}$ | | | |
| 2[**88**/12] −6.2/−7.7 | 5[50/27/10/12/1] −6.3/−6.2/−6.1/−6.0 | 3[33/61/6] −6.7/−6.3 | 3[54/42/4] −6.5/−6.2 |
| $T^{IV}$ | | | |
| 4[74/14/10/2] −7.4/−7.2/−7.2 | 5[9/**82**/7/1/1] −7.6 | 5[41/44/8/6/1] −7.7/−7.4 | 6[48/36/12/1/2/1] **−8.2**/−7.6/−7.4 |

to the TM helices. The small helix HH was observed in fragment L58–L60 in $\mathbf{M^{II}}$, while in $\mathbf{M^{I}}$–$\mathbf{M^{III}}$, the regularly folded structure included the residues that preceded and followed this region (Fig. 7A). The most folded and compacted L-loop is detected in $\mathbf{M^{I}}$. In $\mathbf{M^{III}}$ the HH is oriented horizontally with respect to TMs that is positioned in close proximity to the transmembrane domain. In the other models the HH axis is coincident with the TMs helices, but its relative position with respect to TMs varies from proximal (in $\mathbf{M^{I}}$) to distant (in $\mathbf{M^{IV}}$).

The representative conformation from each of the most populated clusters was characterized with *Fpocket*[41] in order to localize a surface pocket that could be a binding site for vitamin K. In all conformations, we found that the pocket is positioned between the TM helices in the proximity of the endoplasmic reticulum lumen, and that the pocket's volume varies from 450 $\mathring{A}^3$ ($\mathbf{M^{I}}$) to ~530 $\mathring{A}^3$ ($\mathbf{M^{III}}$) (Fig. 7B). It is interesting to note that during the simulations we observed the lipids being inserted into the pocket, indicating the hydrophobic nature of residues forming in the pocket's cavity.

## 4. Recognition of vitamin K by hVKORC1

It was supposed that the representative conformation for each identified cluster is the target ($\mathbf{T}$) for the vitamin K epoxide or its degraded form, namely vitamin K hydroxyquinone (ligand, $\mathbf{L}$). First, to identify the appropriate target of the native form, the vitamin K epoxide ($vK^{EP}$) was docked (AutoDock[42]) with each target, $\mathbf{T^{I}}$–$\mathbf{T^{IV}}$. The docking of the full-length vitamin K produced mutual poses of the ligand that probably arose from high flexibility of the aliphatic chain. Since only the naphthoquinone fragment of vitamin K is functional,[43] the aliphatic chain was cut at the fourth carbon atom and the cleaved vitamin K was re-docked with the targets. The number of vitamin K poses within each target was diminished, but the docking scores were much closer and were not clearly distinguishable from the $\mathbf{L}\cdot\mathbf{T}$ complexes (Table 1). Nevertheless, a visual inspection of the ligand poses with an occurrence greater than 10% permits us to distinguish between the apt and irrelevant $\mathbf{L}\cdot\mathbf{T}$ complexes.

In $\mathbf{T^{I}}$, which has two S–S bridges, the two alternative poses of the substrate in the binding pocket gave a comparable score and population. Similarly, in $\mathbf{T^{II}}$ the substrate is localized in four clusters composed of the randomly oriented molecule. In contrast, the poses of $vK^{EP}$ in $\mathbf{T^{III}}$ form the unique highly populated cluster (88%), which qualifies the $vK^{EP}\cdot\mathbf{T^{III}}$ complex

as very plausible (Fig. 7C). The visual inspection of the docking solution showed that vitamin K is localized in the proximity of the protonated residues C43 and C135 of $\mathbf{T^{III}}$ with the epoxide fragment oriented towards these residues. The vitamin K docking with $\mathbf{T^{IV}}$ showed the most populated poses of a substrate in a position remote from the protonated cysteine residues.

Consequently, from these docking trials we have defined $\mathbf{T^{III}}$ as the most probable target of a vitamin K epoxide. This target may represent the active (or activated) state of hVRORC1 that recognizes specifically $vK^{EP}$ and performs the first step in vitamin K transformation, which is the opening of the epoxide cycle of vitamin K by a cleavage in the C–O bond.

To better study the enzymatic steps leading to vitamin K transformation, we probed a vitamin K hydroxyquinone ($vK^{2-OH}$ and $vK^{3-OH}$) docking with the targets $\mathbf{T^{III}}$ and $\mathbf{T^{IV}}$. The docking results indicated that (i) $vK2^{-OH}$ formed the more populated docking poses with all inspected modified targets with respect to $vK^{3-OH}$; (ii) the best pose of $vK^{2-OH}$ was obtained by the docking with $\mathbf{T^{IV}}$. The vitamin K quinone ($vitK^{Q}$) was randomly positioned in the pocket forming four principle clusters.

## 5. The vitamin K affinity to hVKORC1

Since the complexes $vK^{EP}\cdot\mathbf{T^{III}}$ and $vK^{2-OH}\cdot\mathbf{T^{IV}}$ were potentially plausible, as was found from the docking trials (scores, poses and their population), we studied their stability and compared the free binding energy that maintains these complexes. First, we reconstituted the full-length vitamin K by a consequent growing of its aliphatic chain cleaved prior to the docking. The two complexes with the reconstituted vitamin K were further studied using MD simulations.

The RMSDs calculated on the backbone atoms from $\mathbf{T^{III}}$ and $\mathbf{T^{IV}}$ showed that their values were reduced relative to the free targets and were stable during the 100 ns MD simulations (ESI, Fig. S11A‡). The pattern of the RMSFs generally corresponds to that in free targets (ESI, Fig. S11B‡). Analysis of the $vK^{EP}\cdot\mathbf{T^{III}}$ conformations at the end of the simulations (80–100 ns) revealed that vitamin K is stably bound to the hydrophobic groove of the target formed with residues of TMs and HH (Fig. 7D; ESI, PDB Files S1, S1-1, S1-2‡). The naphtoquinone fragment of vitamin K in $vK^{EP}\cdot\mathbf{T^{III}}$ is located in the proximity of C135, with a distance of 4.0 $\mathring{A}$ between the carbon atom of the epoxide fragment of vitamin K and the sulphur atom of C135,

**Table 2** The free energy of binding in the vitamin K·targets complexes. The free binding energy ($\Delta G_{bind}$) and contributions of electrostatic ($\Delta G_{elect}$), van der Waals ($\Delta G_{vdw}$) and solvation energies ($\Delta G_{pol}$) and ($\Delta G_{non-pol}$) were calculated using MM-PBSA for the complexes formed by targets $\mathbf{T^{III}}$ and $\mathbf{T^{IV}}$ with vitamin K epoxide ($vK^{EP}$), vitamin K hydroxyquinone ($vK^{2-OH}$), and VKAs, acenocoumarol (A), difenacoum (D), phenprocoumon (P) and warfarin (W). The solvation energy is decomposed into polar ($\Delta G_{pol}$) and non-polar ($\Delta G_{non-pol}$) parts. All energy values are in kcal mol$^{-1}$

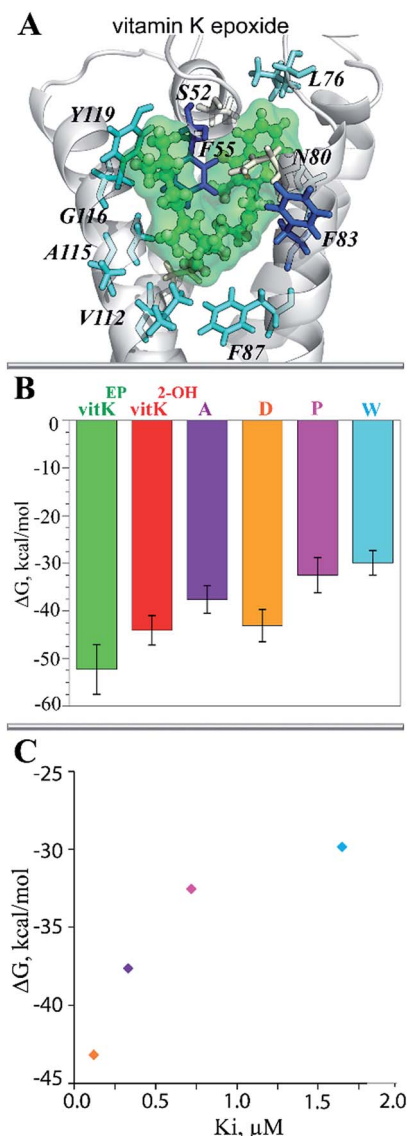| Ligand·target | $\Delta G_{elect}$ | $\Delta G_{vdw}$ | $\Delta G_{pol}$ | $\Delta G_{non-pol}$ | $\Delta G_{bind}$ |
|---|---|---|---|---|---|
| $vK^{EP}\cdot\mathbf{T^{II}}$ | $-1.1 \pm 0.3$ | $-53.5 \pm 5.2$ | $-8.1 \pm 0.8$ | $-5.7 \pm 0.4$ | $-52.3 \pm 5.2$ |
| $vK^{2-OH}\cdot\mathbf{T^{IV}}$ | $-1.0 \pm 0.4$ | $-47.6 \pm 3.6$ | $-9.9 \pm 2.1$ | $-5.4 \pm 0.4$ | $-44.1 \pm 3.1$ |
| $A\cdot\mathbf{T^{III}}$ | $-0.9 \pm 0.4$ | $-42.5 \pm 2.9$ | $-10.4 \pm 1.3$ | $-4.5 \pm 0.2$ | $-37.6 \pm 2.9$ |
| $D\cdot\mathbf{T^{III}}$ | $-0.5 \pm 0.2$ | $-46.9 \pm 2.7$ | $-9.3 \pm 2.3$ | $-5.2 \pm 0.3$ | $-43.2 \pm 3.4$ |
| $P\cdot\mathbf{T^{III}}$ | $-0.7 \pm 0.3$ | $-34.3 \pm 3.4$ | $-6.3 \pm 1.0$ | $-3.9 \pm 0.2$ | $-32.5 \pm 3.7$ |
| $W\cdot\mathbf{T^{III}}$ | $-0.7 \pm 0.4$ | $-34.6 \pm 2.1$ | $-9.5 \pm 1.6$ | $-4.1 \pm 0.2$ | $-29.9 \pm 2.6$ |

**Fig. 8** The ligands binding to hVKORC1. (A) Vitamin K epoxide stabilized by van-der-Waals interactions with hydrophobic residues of $\mathbf{T^{III}}$. The protein is viewed as cartoon with cysteine residues in sticks; vitamin K is shown in ball-and-sticks. (B) The binding free energy ($\Delta G$) calculated by MM/PBSA in complexes formed by vitamin K ant its targets, $vK^{EP} \cdot \mathbf{T^{III}}$ and $vK^{2\text{-}OH} \cdot \mathbf{T^{IV}}$, and in complexes inhibitor·$\mathbf{T^{III}}$. (C) The correlation between the computed $\Delta G$ and the measured inhibiting constants, $K_i$. Each inhibitor is denoted by color: acenocoumarol, A (violet), difenacoum, D (orange), phenprocoumon, P (purple) and warfarin, W (cyan).

which corresponds well to the suggested association between the vitamin K epoxide and VKORC1.[43] The aliphatic chain of vitamin K protrudes into the membrane in the direction perpendicular to the naphthoquinone plane, between TM2 and TM3. In the $vK^{2\text{-}OH} \cdot \mathbf{T^{IV}}$ complex the partially transformed molecule is still bound in the binding site, but is displaced from the position observed in $vK^{EP} \cdot \mathbf{T^{III}}$.

The binding free energy, $\Delta G$, estimated by the MM/PBSA approach[44] differs considerably, with values of −52.3 and −44.1 kcal mol$^{-1}$ for $vK^{EP} \cdot \mathbf{T^{III}}$ and $vK^{2\text{-}OH} \cdot \mathbf{T^{IV}}$ respectively

(Table 2 and Fig. 8B). The decomposition of $\Delta G$ for the different terms contributing to the binding affinity of vitamin K helps to identify the primary forces driving such a binding. The van-der-Waals energy ($\Delta G_{vdw}$) shows more variation (−53.5 kcal mol$^{-1}$ for $vK^{EP} \cdot \mathbf{T^{III}}$ and −47.6 kcal mol$^{-1}$ for $vK^{2\text{-}OH} \cdot \mathbf{T^{IV}}$) than for the other terms, and appears to be the principal force contributing to the difference in the final binding energy values ($\Delta G_{bind}$) in the studied complexes (Table 2). The electrostatic energy ($\Delta G_{el}$) and the solvation energy, represented by polar ($\Delta G_{pol}$) and non-polar ($\Delta G_{non\text{-}pol}$) parts, show only minor variations between two complexes.

To identify the key molecular determinants responsible for the binding of vitamin K to the targets, we calculated the per residue energy contribution of all the protein residues to the overall binding energy. In both complexes, $vK^{EP} \cdot \mathbf{T^{III}}$ and $vK^{2\text{-}OH} \cdot \mathbf{T^{IV}}$, the free energy binding value is only influenced by the hydrophobic residues demonstrating a favourable contribution to the vitamin K affinity (Fig. 7E and 8A). The largest contribution to vitamin K epoxide affinity was from F55 interacting with the naphthoquinone fragment within the binding site and by F86 interfacing the aliphatic chain protruding into the membrane. In the $vK^{2\text{-}OH} \cdot \mathbf{T^{IV}}$ complex the vitamin K hydroxyquinone is mainly bound by the highly increased interaction with F86, while the interaction with F55 is reduced with respect to $vK^{EP} \cdot \mathbf{T^{III}}$. The other residues, L22, A26, F87, V112, A115, G116, L120 and V134, contribute positively to the substrate affinity in both forms, though their impact is different. The role of F55 as a key residue binding either the substrate or warfarin was recently reported.[45]

This simple analysis of the free binding energy has shown that (i) the hydrophobic interactions are key to vitamin K recognition by hVKORC1, (ii) $\mathbf{T^{III}}$ provides the most favourable hydrophobic interactions, which are greater than the target $\mathbf{T^{IV}}$ and (iii) F55 contributes most to affinity of the vitamin K epoxide in the binding pocket of $\mathbf{T^{III}}$, while F86 preserves the binding of vitamin K hydroxyquinone to $\mathbf{T^{IV}}$.

## 6. Models-based concept of the hVKORC1 enzymatic cycle

By considering the distinct recognition properties of the generated models of hVKORC1 with respect to vitamin K in its different forms, we attempted to assemble these models in a logical scheme that presents a plausible enzymatic cycle of the vitamin K transformation. Initially, in the absence of external factors hVKORC1 is apparently folded in the fully deprotonated state and is stabilized by two S–S covalent bonds. This state, represented by the $\mathbf{T^I}$ model that non-selectively recognized vitamin K and that might correspond to the inactive state of hVKORC1. Upcoming protons (delivered by a thioredoxin through the inter-protein proton exchange) perturb the hVKORC1 by initiating a break of S–S bridge between the L-loop cysteine residues, C43 and C51, and their protonation producing the state described by model $\mathbf{T^{II}}$. Further proton-transfer (intra-protein) promotes a reduction in the cysteine residues, C132 and C135, which is accompanied by conformational adaptation of hVKORC1 to a state favourable to vitamin K binding in the pocket in close proximity to the activated cysteine
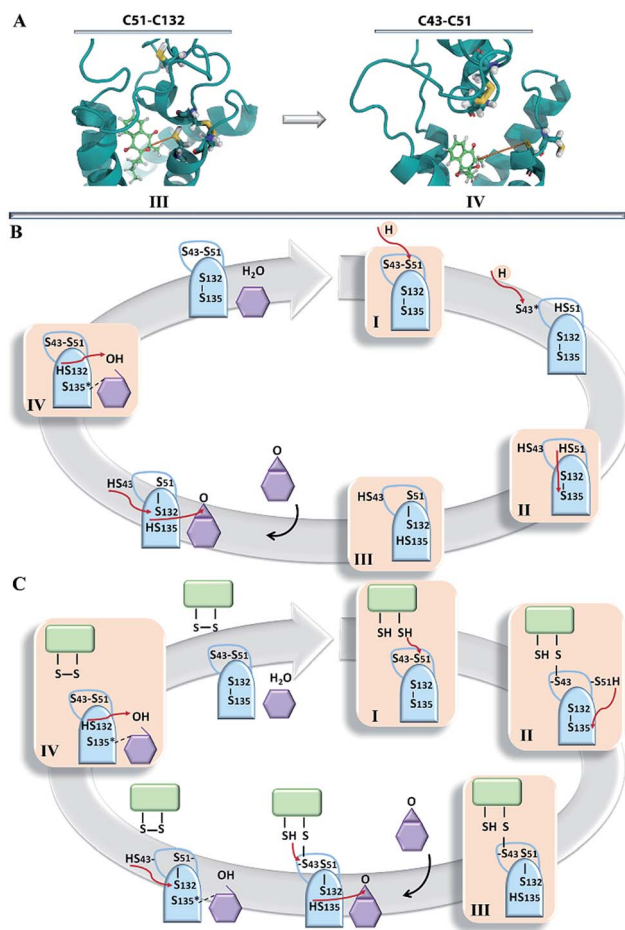
**Fig. 9** Role of the metastable states of hVKORC1 in the hVKORC1 enzymatic process. (A) Vitamin K epoxide (left) and vitamin K hydroxyquinone (right) stabilized by van-der-Waals interactions with hydrophobic residues of T$^{III}$ and T$^{IV}$ respectively. (B) The single-protein-based reduction cycle and (C) the two-proteins-based reduction process of vitamin K. hVKORC1, vitamin K and thioredoxin are sketched in blue, violet and green respectively. The hVKORC1 states predicted from MD simulations are countered in salmon. The proton path and the point of vitamin K insert are showed by red and black arrows, respectively.

residues (Fig. 7C). Such a step is shown by the target T$^{III}$ demonstrating a highly selective binding with the vitamin K epoxide. This step yields the hydroxyquinone form of vitamin K, with the opened epoxy group that is still bound to the binding site of the newly transformed hVKORC1 and target T$^{IV}$ that is ready to attack the vitamin K again. The targets T$^{III}$ and T$^{IV}$ may represent two active states of VKORC1, which successively transform the vitamin K through the epoxide opening (T$^{III}$, active-1) followed by the OH group removal (T$^{IV}$, active-2). The fully transformed vitamin K is not retained by the enzyme, which is further transformed to the initial oxidase state.

Recapitulating our results, we postulated two plausible sketches of the hVKORC1 enzymatic process, which were conceived entirely from the predicted models and their ability to bind vitamin K. The first hypothesis suggested a direct involvement of thioredoxin in the reversible covalent

interaction of vitamin K with hVKORC1 (Fig. 9C), and this is derived from the enzymatic mechanism specified for the bacterial homolog bVKOR.[11] According to the second hypothesis, the thioredoxin protein initiates the reduction in hVKORC1 as a proton donor (Fig. 9B). This simple and comprehensive sketch may reflect the most rational and probably the most credible catalytic cycle of the hVKORC1 enzymatic mechanism.

## 7. Inhibitor affinity to hVKORC1

Since target T$^{III}$ is highly selective for native vitamin K binding, we postulated that this state is a target of the vitamin K agonists (VKAs). The well-known VKAs, namely acenocoumarol (A), difenacoum (D), phenprocoumon (P) and warfarin (W) were docked with T$^{III}$. As the AVKs do not contain the aliphatic chain, the docking trials produced a highly populated cluster, which demonstrated the unique pose of each inhibitor. Each inhibitor is localized in the binding pocket of the target that takes the position of the naphthoquinone fragment of the vitamin K (Fig. 10B). Each inhibitor·target complex was further studied using MD simulations (100 ns). Similar to vK$^{EP}$·T$^{III}$, the RMSDs of hVKORC1 in each inhibitor·target complex were stable during the MD simulations (ESI, Fig. S11‡) and each inhibitor was stably bound to the hydrophobic groove of the target (Fig. 10D; ESI, PDB Files S2, S2-1, S2-2‡). The binding free energy ($\Delta G$) was estimated using the MM/PBSA method and gave values of $-37.6 \pm 2.9$, $-43.2 \pm 3.4$, $-32.5 \pm 3.7$ and $-29.9 \pm 2.6$ kcal for A, D, P and W respectively (Fig. 8B and Table 2), which are significantly smaller than for the vitamin K epoxide. The decomposition of the binding free energy for the different terms contributing to inhibitor affinity showed that the primary forces driving inhibitor binding are also the van-der-Waals interactions. The calculated per residue energy contribution of all target residues indicated that the free energy binding value is only influenced by the hydrophobic residues, which demonstrate a favourable contribution to the inhibitor affinity, principally from F55 and V134, similarly to the natural substrate binding (Fig. 10C). F55 was previously reported as a key residue for warfarin binding.[45]

To validate the calculated binding free energy values, the experimental measurements were performed to determine the inhibiting effect of these molecules (acenocoumarol, difenacoum, phenprocoumon and warfarin) on hVKORC1 activity. These molecules show a wide range of hVKORC1 inhibition effects. The obtained $K_i$ values indicate the following order of hVKORC1 activity: D > A > P > W. The lowest value of the binding free energy ($\Delta G$) is shown by difenacoum (D), which is the inhibitor that has the best inhibition effect, while the highest value of $\Delta G$ was displayed by warfarin (W), having the smallest $K_i$ (Fig. 8C).

Despite the limitations of the MM/PBSA method compared to more accurate approaches such as alchemical free energy perturbation molecular dynamics (FEP/MD),[46] we have obtained a qualitative characterisation of VKAs binding with hVKORC1 that is in good correlation with the experimentally measured inhibiting effect. We have found that hydrophobic interactions
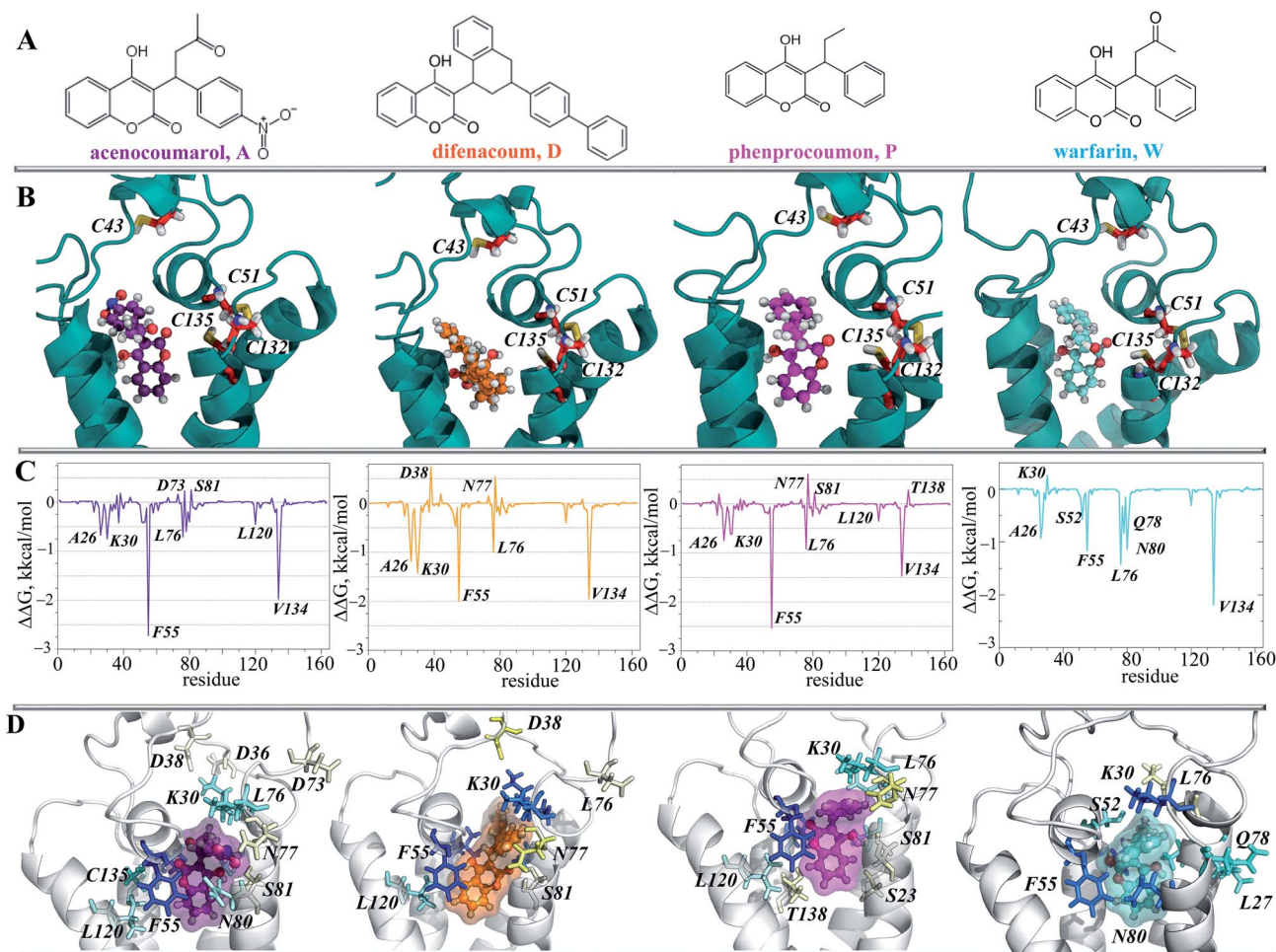
Fig. 10 Inhibitors binding to hVKORC1. (A) Structural formula of acenocoumarol (A), difenacoum (D), phenprocoumon (P) and varfarin (W). (B) Docking poses of inhibitors in the binding pocket of hVKORC1. The protein is viewed as cartoon with cysteine residues in sticks; each inhibitor is shown in ball-and-sticks. (C) The per residues contribution to the binding energy ($\Delta G$) of $T^{III}$ to acenocoumarol (1st column), difenacoum (2nd column), phenprocoumon (3rd column) and warfarin (4th column). (D) AVKs, (A, D, P and W) stabilized by van-der-Waals interactions with hydrophobic residues of $T^{III}$. Protein is depicted as cartoon; inhibitors are showed in ball-and-sticks with their surface; the hydrophobic residues are drawing in sticks with a color coded for the strength of interaction, from a strong (in dark) to a weak (white), and the contribution to the binding free energy, from a positive (blue) to a negative (yellow).

are key to hVKORC1 inhibition by VKAs as revealed by MD simulations and MM/PBSA free energy calculations.

## Conclusions and perspectives

As suggested in our working hypothesis, the hVKORC1 model with all protonated cysteine residues is a paradigmatic specimen used to predict the enzymatic states of a protein through exploration of its conformational space using MD simulations. Distinct and independent statistical techniques showed the great conformational variability of the protein. Estimation of the energy profile and capture of the protein conformations at local minima provided a set of pro-models mimicking the biologically relevant states.

The generated models of the functional metastable states of hVKORC1 and their validation through *in silico* screening led to the concept of two plausible mechanisms for enzymatic reactions. These results suggest several additional questions, the

most important being the real enzymatic machinery of hVKORC1. What is the exact role of thioredoxin(s) in initiating the hVKORC1 reduction? *i.e.* does it behave as a proton donor or as a co-partner participating in covalent binding with VKORC1? Additional experimentally based studies will be required to select the most appropriate hypothesis. Furthermore, structural models depicting distinct metastable states of protein represent the different steps in the catalytic cycle. The validation of these states using a docking of vitamin K in its epoxide and hydroquinone forms have qualified two hVKORC1 conformations as the active states (active-I and active-II), and these two states may be pertinent targets for the development of novel highly selective and state-specific molecules. The use of model $T^{IV}$ as a target is more questionable as this model may require *a prior* presence of vitamin K in the active site. Theoretically, such a state may be targeted by a competitive ligand that would prevent the inactive state folding required for the next

enzymatic cycle. In this context, the other enigmatic question that remains unanswered is, do the cysteines exert the same role in the active site and the redox-active?

The hydrophobic interactions between the substrate and the targets were identified as the main factor contributing to such a difference, and the alternation of these interactions leads to selective target sensitivity to a given form of vitamin K. The state, which was identified as contributing to the first step in vitamin K transformation, was validated through analysis of hVKORC1's binding energy with three vitamin K agonists. The obtained *in silico* data was found to be highly correlated with data measured *in vitro*, and as such, this validated our theoretical models.

# Methods

## 1. Structural modelling

**The protein modelling.** The hVKORC1 sequence (*Q9BQB6*) was retrieved from UniprotKB (http://www.uniprot.org/uniprot). PREDATOR, GOR IV, PSIPRED, PROF and JPRED were used for prediction of the secondary structure of hVKORC1. HMMTOP, OCTOPUS, SPOCTOPUS, PHOBIUS, POLYPHOBIUS, PROF, PHILIUS and SCAMPI were used for estimation of the transmembrane part of protein (ESI, Methods‡). The crystallographic structures of homologous proteins (ESI, Table S1‡) were found by a BLAST search. The sequence alignment of hVKORC1 (*Q9BQB6*, target, 163 residues) and bVKOR (*Q2JJF6*, template, 283 residue) was used to identify the bVKOR segment having the best identity/homology with hVKORC1. Structure of bVKOR (4NV5),[11] containing this segment, was chosen as a template for homology modelling of hVKORC1. The hVKORC1 model was generated with MODELLER.[47] The missing residues[49–53] were modelled with the '*loopmodel*'. The ten-step iterations of loop refinement produced ten models of the full-length target protein. The most relevant structural model was chosen based on MODELLER scores. The four cysteine residues (C43, C51, C132 and C135) were considered in a protonated state (SH). No constraint was applied on cysteine–cysteine distances. This model (**M**) represents the fully protonated hVKORC1. For a modelling of the partially protonated or fully deprotonated protein, the MD conformations of the fully protonated hVKORC1 were used. The conformations that have a short S···S distance between the pairs of cysteine residues were locally modified to adjust the predicted state of hVKORC1. Specifically, the hydrogen atoms connected to the sulphur atoms that exhibited a short S···S distance were removed and the S···S distance was constrained to 2.1 Å, which is typical for an S–S bond. The resulting models – **M**$^{\text{I}}$, stabilised with two pairs of covalently bonded sulphur atoms from residues C43–C51 and C132–C135, **M**$^{\text{II}}$, **M**$^{\text{III}}$ and **M**$^{\text{IV}}$, stabilized by a lone S–S bond linking C132–C135, C51–C132 and C43–C51 respectively, were optimised and minimized.

**Modeling of the systems.** A lipid bilayer was used to mimic the membrane environment in which hVKORC1 exist. This lipid bilayer, initially composed of 200 1,2-dilauroyl-*sn*-glycero-3-phosphocho-line (DLPC) lipids, had been prepared with *Mem-Builder*[48] resulting in an area per lipids of 63 Å$^2$ and a bilayer

thickness and 31 Å, consistent with experimentally measured values (*cf.* 64 Å$^2$ for the area per lipid and 30 Å for the bilayer thickness[49]). This lipid bilayer had been completed with 4000 water molecules (TIP3P[50]) and pre-equilibrated during 1.5 ns of MD. To setup the protein-membrane systems, each structural model of hVKORC1 was embedded in the equilibrated and hydrated membrane using the *g_membed* tool[51,52] from GROMACS package. The protein coordinates were scaled down by a 100 factor in the membrane surface plane. 8 lipids and 52 water molecules that overlapped with the protein model were removed. Protein coordinates were finally restored during a short (1 ns) MD simulation. Cl$^-$ ions were placed randomly to neutralize the system. The total number of atoms in the system (protein, DLPC lipids, water molecules and counter ions) was 58 241 and the box size was 80 × 80 × 87 Å$^3$. The 3D models **M**, **M**$^{\text{I}}$, **M**$^{\text{II}}$, **M**$^{\text{III}}$ and **M**$^{\text{IV}}$ were used for a modelling of each protein-membrane system.

## 2. Molecular dynamics simulations

**Set-up of the system.** The set-up of molecular dynamics (MD) simulations were performed using GROMACS 4.6.5 program package.[53] The CHARMM-36 force field[54] modified to support DLPC lipids simulations[49] was employed. The 50 000-step energy minimization was performed on each full system using a steepest descent energy minimization. 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 temperature of protein, membrane and solvent (water and ions) was separately coupled to the velocity rescale thermostat,[55] a modified Berendsen thermostat,[56] 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 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[57] was applied on the protein, lipids and solvent (water and ions) separately, with a relaxation time of 0.5 ps, to keep temperature constant at 310 K. Water and ions were allowed to move freely during equilibration.

**Production of MD trajectories.** For the system containing the model **M**, five 100 ns trajectories of MD simulations were carried out with different starting velocities. Two trajectories were extended to 1 μs. For each system containing the models **M**$^{\text{I}}$–**M**$^{\text{IV}}$, two MD trajectories of 100 ns were performed. A time step of 2 fs was used to integrate the equations of motion based on the Leap-Frog algorithm.[58] Coordinates files were recorded every 5 ps. Neighbour searching was performed by the Verlet algorithm.[59] The particle mesh Ewald (PME) method[60] was used to treat long-range electrostatic interactions at every step. The van der Waals interactions were modeled using a 6–12 Lennard-Jones potential. Electrostatic interactions were computed using a particle mesh Ewald algorithm. The list for the electrostatic interactions was updated every 5 steps. The Nose–Hoover thermostat[57] was used to keep simulation temperature constant by coupling (τ = 0.5 ps) the protein, lipids and solvent (water and ions) separately to a temperature bath of 310 K. Pressure

was kept constant (1 bar) by Parrinello–Rahman coupling[55] and semi-isotropic coordinate scaling ($\tau = 2$ ps). Coordinates files were recorded every 5 ps.

All simulations were performed on HP computer TURING, IBM Blue Gene/Q at IDRIS, using parallel 1.6 GHz 64-bit POW-ERPC A2 node cards.

## 3. Standard analysis of MD trajectories

Standard GROMACS routines were utilized in the analysis of MD trajectories (RMSD, RMSF, PCA and clustering calculation). Visual inspection of conformations was made with PyMOL[61] and VMD.[62] The secondary structure of the protein (every 20 ps) was analysed with DSSP.[36] Cross-correlation matrices were calculated with Bio3d R package.[63]

The RSMD and RMSF values were calculated on the $C\alpha$ atoms using the initial model (at $t = 0$ ns as a reference). RSMD, RMSF, PCA, cluster analysis, correlation matrices and structural drift were calculated after least-square fitting the protein snapshots using transmembrane region as the reference, thus removing the rigid-body motion from the analysis. Transmembrane region was assessed using the secondary structure analysis detected that TM1, TM2, TM3 and TM4 consist of G9-H28, N77-L93, A102-I123 and C132-R151 respectively.

**Principal component analysis.** The collective motions of hVKORC1 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 ($x(t)$). 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]

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

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

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

$$C = V\Lambda V^{\mathrm{T}} \tag{2}$$

This results in a diagonal matrix $\Lambda$ 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]

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

where $i$ and $j$ are two atoms $C\alpha$; $\Delta r_i$ and $\Delta r_j$ are displacement vectors of $i$ and $j$; and $\Delta r^{\mathrm{T}}$ 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 $C\alpha$ as a reference before performing cross-correlations analysis.

**Clustering.** A cluster analysis was performed on each MD trajectory of models I–IV considering either the all simulation or the productive (20–100 ns) using an ensemble-based approach[64] with Gromos,[65] after snapshots fitting using the transmembrane domain $C\alpha$ as a reference. The algorithm extracts representative MD conformations from a trajectory by clustering the recorded snapshots according to their $C\alpha$-atoms RMSDs. The procedure for each trajectory is described as follows: (i) a reference structure is randomly picked up in the MD conformational ensemble and all conformations distant by less than an arbitrary cutoff $r$ are popped out of the ensemble; this step is repeated until no conformation remains in the ensemble, providing a set of reference structures distant from at least $r$; (ii) the MD conformations are grouped into $n$ reference clusters based on their RMSDs from each reference structure using the RMSD cutoff $r$ of 2 Å.

**The drift monitoring.** The helices drift analysis was performed using the centroids ($C_i$) defined on the $C\alpha$ atoms for each group of amino acids (aas): $C_0$ (51–56 aas) for HH and two centroids for each TM helix extremity – $C_1$ (9–12 aas) and $C_{1'}$ (25–28 aas) for TM1, $C_2$ (77–80 aas) and $C_{2'}$ (90–93 aas) for TM2, $C_3$ (102–105 aas) and $C_{3'}$ (120–123 aas) for TM3, $C_4$ (132–135 aas) and $C_{4'}$ (148–151 aas) for TM4. The $L_2^1$-loop motion were analysed using the centroids defined on HH ($C_1$, 51–56 aas) and on the loop residues 47–50 and 35–38 ($C_2$), 39–46 ($C_3$), 57–60 and 69–72 ($C_4$), and 61–68 ($C_5$). Positions of these centroids were monitored over MD simulations, and their coordinates were projected on the $x$–$z$ and $y$–$z$ planes.

## 4. Advanced statistical analysis

We performed a statistical analysis of each observed MD trajectory $\Xi = \{x_1, ..., x_N\}$ using an estimation of the density of the conformations in a low dimensional space using the first modes arising from the PCA. The local occupation density $\pi(x)$ at a given scale $\sigma$ is estimated using a Parzen Window density estimator (implemented in MATLAB) with a Gaussian kernel,[66,67]

$$\pi(x) \propto \sum_{\{i=1\}}^{N} \exp\left( -\frac{\|x - x_i\|^2}{2\sigma^2} \right),$$

inducing a free energy landscape

$$U(x) = -k_{\mathrm{B}}T \log(\pi(x)) + C.$$

This energy should not be identified with the potential (force field) driving the MD simulation, but should be considered as

a statistical proxy used to analyse the region of high occupation linked with potential meta-stable states visited by the dynamics. Introducing the similarity matrix[68,69] between conformations

$$W_{i,j} = \exp\left(-\frac{\|x_i - x_j\|^2}{2\sigma^2}\right), \text{ we have } \pi(x_i) \propto W_{i,.} \doteq \sum_{\{j=1\}}^{N} W_{i,j}.$$

However, beyond the occupation density some of the dynamical aspects can be empirically analysed through the estimation of a random walk on the simulated data that is compatible with the occupation density. Let us consider a random walk $n \rightarrow X_n$ on the MD conformations such that the transition probability $P_{ij}$ between $i$ and $j$ is given by

$$P_{i,j} = P(X_{n+1} = j | X_n = i) \propto \frac{W_{i,j}^{\alpha}}{W_{i,.}^{\alpha} W_{j,.}^{\alpha}}, \tag{4}$$

with $\alpha = \frac{1}{2}$. After the time change $t_n = n\sigma^2 k_B T$, this random walk approximates a continuous diffusion $t \rightarrow X_t$ following a Langevin diffusion $dX_t = -\nabla U(X_t)dt + \sqrt{2k_B T}dB_t$ giving an occupation density $\pi(x)$ at equilibrium consistent with the observations.[38,68,69] The spectral analysis of the associated Laplacian $L_{ij} = \delta_{ij} - P_{ij}$ gives interesting slow variables for the estimated dynamics that are strongly linked with the metastable states.

**Slow variables.** A companion matrix of the Markov kernel $P$ is the normalized graph Laplacian $L = I - P$, which is a generalization of the standard Laplacian on a grid. Its eigenvectors have a key position in spectral clustering. They are approximately indicator variables of clusters in $\Xi$ and especially when the eigenvalues of $P$ are close to 1. In this case, these variables are called slow variables because their profiles are useful to identify energy basins, the passage of the random walk from one basin to another is doing at long time scales, hence the term slow. The first two slow variables for the trajectory **3″** with and without PCA, are shown in ESI (Fig. S7 and S8,‡ respectively). Since slow variables are indicator variables of clusters, we can use this property to compare for each variable its profiles before and after PCA, by searching the changes of sign of these profiles.

## 5. Molecular docking, MD simulations and analysis of the complexes

**Ligands modelling.** The vitamin K coordinates were retrieved from the crystallographic structure 2WJN, containing menaquinone 7.[70] The structural models of vitamin K in four forms – epoxide (vitK$^{EP}$), quinone (vitK$^{QU}$) and hydroquinone with the hydroxy group at the carbon atom in position 2 (vK$^{2\text{-OH}}$) and 3 (vK$^{3\text{-OH}}$) – was prepared from menaquinone 7 structure, using UCSF Chimera v1.10.1.[71] The constructed vitamin K models were cleaved at the fourth carbon atom of the aliphatic chain. The coordinates of acenocoumarol (A) were retrieved from ref. 72. The structure of difenacoum (D), phenprocoumon (P) and warfarin (W) were modeled from 3D structure of acenocoumarol using UCSF Chimera v1.10.1. All models of vitamin K and of inhibitors were minimized and used as input for Swissparam,[73] generating topology files for each ligand.

**Targets selection.** For each simulated system (where the protein is $\mathbf{M^I}$, $\mathbf{M^{II}}$, $\mathbf{M^{III}}$, $\mathbf{M^{IV}}$), the representative conformations

were extracted by a fast clustering, based on convergence analysis of the concatenated trajectories.[65] We selected the more-represented reference structure for each simulated model, and defined the targets $\mathbf{T^I}$, $\mathbf{T^{II}}$, $\mathbf{T^{III}}$, $\mathbf{T^{IV}}$.

**Pockets detection.** The proteins pockets were detected using *MDpocket*, a software based on the *fpocket* algorithm.[41] For each MD conformation, *MDpocket* associates the α-spheres (*i.e.*, spheres contacting with four protein atoms without any other atom within the sphere) to a grid point. The pockets maps were produced by iterating this step over all conformers employing the default parameters. *MDpocket* requires the superposition of each conformer on a reference structure prior to any calculation to avoid artefacts which may occurred over the grid-based process. The selected pockets were further analyzed in a second *MDpocket* run. For each selected pocket, the grid points with an occupancy time of at least 0.5 (*i.e.*, grid points associated with α-spheres at least 50% of the simulation time) were retained to extract the pocket's metrics, such as volume, over the MD trajectory.

**Molecular docking.** The docking calculations were performed by using the AutoDock,[42] a hybrid search method that applies a Lamarckian genetic algorithm (GA). Exploration of the binding site is based on a global search that uses a GA followed by an adaptive local search method derived from the optimization algorithm of Solis and Wets.[74] The graphical user interface of AutoDock 4.2 was used for preparation of ligand and receptor files. Grid maps of interaction energies for various atom types were carried out with a grid box of dimension $20 \times 20 \times 23$ Å centered on the previously detected pockets. Calculations were performed with a population size of 150, number of energy evaluations of $5 \times 10^6$, maximum number of generations of 27 000 and mutation and crossover rates of 0.02 and 0.8, respectively. The number of runs was set to 100 to explore a large number of poses of the highest affinity, and the Solis and Wets algorithm was used to relax the best 10% of the obtained conformations. The binding pocket coordinates were explicitly defined in the input file.

The binding affinity of ligand-hVKORC1 complexes was expressed in terms of number of clusters, poses population and docking scores. For each ligand conformation, the most populated poses (>10%) were saved and analyzed.

**MD simulations.** The systems composed of protein (protein is a ligand-hVKORC1 complex) the DLPC lipids, water molecules and counter ions, were constructed and equilibrated as described above. For each system including a ligand-hVKORC1 complex, two MD trajectories each of 100 ns were performed.

The six pdf files supplied contain the atomic coordinates of the ligand–hVKORC1 complexes (ligand = vitK$^{EP}$ and warfarin, W) at $t = 0$ and $t = 100$ ns for each trajectory.

**Free energy calculation.** Standard MM-PBSA (Molecular Mechanics – Poisson Boltzmann Surface Area) method[44] has been used for calculation of thermodynamic parameters and free energy of binding. In MM-PBSA, the free energy difference of binding ($\Delta G_{bind}$) between a ligand (L) and a receptor (R) to form a complex RL is defined as:

$$\Delta G_{bind} = \Delta H - T\Delta S \cong \Delta E_{MM} + \Delta G_{sol} - T\Delta S \tag{5}$$

$$\Delta E_{MM} = \Delta E_{int} + \Delta E_{elect} + \Delta E_{vdw} \tag{6}$$

where $\Delta E_{MM}$, $\Delta G_{sol}$ and $-T\Delta S$ are the changes of the gas phase MM energy, the solvation free energy decomposed into polar ($\Delta G_{pol}$) and non-polar ($\Delta G_{non-pol}$) parts, and the conformational entropy upon binding, respectively. The solute and solvent dielectric coefficients were set to 16 and 80, respectively.

$\Delta E_{MM}$ includes $\Delta E_{int}$ (covalent bonding including the bond, angle and dihedral energies), electrostatic $\Delta E_{elect}$ and van der Waals $\Delta E_{vdw}$ energies. The polar contribution was calculated using PB model, while the non-polar free energy estimated by solvent accessible surface area (SASA).

$\Delta G_{bind}$ was computed for each generated ligand·target complex over the merged trajectories containing 400 frames (and also for the individual MD simulation replicas over a trajectory containing 200 frames). We used the *g_mmpbsa* module of GROMACS, which combines subroutines from GROMACS and APBS[75] to calculate the enthalpy contribution ($\Delta H$) to the Gibbs free energy and the per-residue energy decomposition. The energy components $\Delta E_{MM}$ and the polar and non-polar components of $\Delta G_{sol}$ were calculated in the bound and unbound form, and subsequently their contribution to the binding energy $\Delta R_x^{BE}$ of residue $x$ (ref. 76) was calculated as follows:

$$\Delta R_x^{BE} = \sum_{i=0}^{n} \left( A_i^{bound} - A_i^{free} \right) \tag{7}$$

where $A_i^{bound}$ and $A_i^{free}$ are the energy of $i$th atom from $x$ residue in bound and free forms, respectively, and $n$ is the total number of atoms in the residue.

The *g_mmpbsa* module does not include the calculation of entropic terms and therefore it gives the relative binding energy. Since we have used the analogous compounds showing the same binding mode, the entropy contribution was neglected.[44]

### 6. Measuring the dissociation constants of inhibitors

Recombinant hVKORC1 protein was expressed in *P. pastoris* and microsomal fractions containing recombinant proteins were prepared as described previously.[18] VKOR activity was performed as described previously.[77] To evaluate the inhibiting effect of acenocoumarol, difenacoum and phenprocoumon on VKOR activity, $K_i$ were determined after addition of anticoagulant to the standard reaction (*i.e.*, 200 mM Hepes buffer (pH 7.4) containing 150 mM KCl, 1 mM dithiothreitol, 0.75 to 2 g L$^{-1}$ of total proteins containing membrane VKORC1) in the presence of increasing amounts of vitamin K1 epoxide (from 0.003 to 0.2 mM) using anticoagulant concentrations from about 0.05 to 20 × $K_i$. Data were fitted by non-linear regression to the non-competitive inhibition model[4] using the R-fit program. $K_i$ values were obtained from at least three separate experiments performed on three different batches of protein.

$$v = \frac{V_{max}[S]}{\alpha K_m + \alpha'[S]} = \frac{\left(1/\alpha'\right)V_{max}[S]}{(\alpha/\alpha')K_m + [S]} \tag{8}$$

where $\alpha = 1 + \dfrac{[I]}{K_i}$ and $\alpha' = 1 + \dfrac{[I]}{K'_i}$.

## Conflicts of interest

There are no conflicts to declare.

## Acknowledgements

## References

1 D. Y. Jin, J. K. Tie and D. W. Stafford, The conversion of vitamin K epoxide to vitamin K quinone and vitamin K quinone to vitamin K hydroquinone uses the same active site cysteines, *Biochemistry*, 2007, **46**(24), 7279–7283.

2 R. Wallin, N. Wajih and S. M. Hutson, VKORC1: a warfarin-sensitive enzyme in vitamin K metabolism and biosynthesis of vitamin K-dependent blood coagulation factors, *Vitam. Horm.*, 2008, **78**, 227–246.

3 J. Stenflo, P. Fernlund, W. Egan and P. Roepstorff, Vitamin K dependent modifications of glutamic acid residues in prothrombin, *Proc. Natl. Acad. Sci. U. S. A.*, 1974, **71**(7), 2730–2733.

4 B. A. Willems, C. Vermeer, C. P. Reutelingsperger and L. J. Schurgers, The realm of vitamin K dependent proteins: shifting from coagulation toward calcification, *Mol. Nutr. Food Res.*, 2014, **58**(8), 1620–1635.

5 G. Karsenty and M. Ferron, The contribution of bone to whole-organism physiology, *Nature*, 2012, **481**(7381), 314–320.

6 K. L. Berkner, Vitamin K-dependent carboxylation, *Vitam. Horm.*, 2008, **78**, 131–156.

7 J. Oldenburg, M. Marinova, C. Muller-Reible and M. Watzka, The vitamin K cycle, *Vitam. Horm.*, 2008, **78**, 35–62.

8 L. Goodstadt and C. P. Ponting, Vitamin K epoxide reductase: homology, active site and catalytic mechanism, *Trends Biochem. Sci.*, 2004, **29**(6), 289–292.

9 N. Wajih, S. M. Hutson and R. Wallin, 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, *J. Biol. Chem.*, 2007, **282**(4), 2626–2635.

10 W. Li, S. Schulman, R. J. Dutton, D. Boyd, J. Beckwith and T. A. Rapoport, Structure of a bacterial homologue of vitamin K epoxide reductase, *Nature*, 2010, **463**(7280), 507–512.

11 S. Liu, W. Cheng, G. R. Fowle, G. Shen and W. Li, Structures of an intramembrane vitamin K epoxide reductase homolog reveal control mechanisms for electron transfer, *Nat. Commun.*, 2014, **5**, 3110.

12 K. J. Czogalla, M. Watzka and J. Oldenburg, Structural Modeling Insights into Human VKORC1 Phenotypes, *Nutrients*, 2015, **7**(8), 6837–6851.

13 S. Rost, A. Fregin, M. Hunerberg, C. G. Bevans, C. R. Muller and J. Oldenburg, Site-directed mutagenesis of coumarin-type anticoagulant-sensitive VKORC1: evidence that highly conserved amino acids define structural requirements for enzymatic activity and inhibition by warfarin, *Thromb. Haemostasis*, 2005, **94**(4), 780–786.

14 N. Wajih, D. C. Sane, S. M. Hutson and R. Wallin, Engineering of a recombinant vitamin K-dependent gamma-carboxylation system with enhanced gamma-carboxyglutamic acid forming capacity: evidence for a functional CXXC redox center in the system, *J. Biol. Chem.*, 2005, **280**(11), 10540–10547.

15 J. K. Tie and D. W. Stafford, Structural and functional insights into enzymes of the vitamin K cycle, *J. Thromb. Haemostasis*, 2016, **14**(2), 236–247.

16 B. W. Weston and P. E. Monahan, Familial deficiency of vitamin K-dependent clotting factors, *Haemophilia*, 2008, **14**(6), 1209–1213.

17 M. Teichert, L. E. Visser, R. H. van Schaik, A. Hofman, A. G. Uitterlinden, P. A. De Smet, *et al.*, Vitamin K epoxide reductase complex subunit 1 (VKORC1) polymorphism and aortic calcification: the Rotterdam Study, *Arterioscler., Thromb., Vasc. Biol.*, 2008, **28**(4), 771–776.

18 A. Hodroge, B. Matagrin, C. Moreau, I. Fourel, A. Hammed, E. Benoit, *et al.*, VKORC1 mutations detected in patients resistant to vitamin K antagonists are not all associated with a resistant VKOR activity, *J. Thromb. Haemostasis*, 2012, **10**(12), 2535–2543.

19 S. Rost, A. Fregin, V. Ivaskevicius, E. Conzelmann, K. Hortnagel, H. J. Pelz, *et al.*, Mutations in VKORC1 cause warfarin resistance and multiple coagulation factor deficiency type 2, *Nature*, 2004, **427**(6974), 537–541.

20 T. Li, C. Y. Chang, D. Y. Jin, P. J. Lin, A. Khvorova and D. W. Stafford, Identification of the gene for vitamin K epoxide reductase, *Nature*, 2004, **427**(6974), 541–544.

21 W. D. Van Horn, Structural and functional insights into human vitamin K epoxide reductase and vitamin K epoxide reductase-like1, *Crit. Rev. Biochem. Mol. Biol.*, 2013, **48**(4), 357–372.

22 K. J. Czogalla, A. Biswas, A. C. Wendeln, P. Westhofen, C. R. Muller, M. Watzka, *et al.*, Human VKORC1 mutations cause variable degrees of 4-hydroxycoumarin resistance and affect putative warfarin binding interfaces, *Blood*, 2013, **122**(15), 2743–2750.

23 J. K. Tie, C. Nicchitta, H. G. von and D. W. Stafford, Membrane topology mapping of vitamin K epoxide reductase by in vitro translation/cotranslocation, *J. Biol. Chem.*, 2005, **280**(16), 16410–16416.

24 J. K. Tie, D. Y. Jin and D. W. Stafford, Human vitamin K epoxide reductase and its bacterial homologue have different membrane topologies and reaction mechanisms, *J. Biol. Chem.*, 2012, **287**(41), 33945–33955.

25 H. G. von, Control of topology and mode of assembly of a polytopic membrane protein by positively charged residues, *Nature*, 1989, **341**(6241), 456–458.

26 E. Hartmann, T. A. Rapoport and H. F. Lodish, Predicting the orientation of eukaryotic membrane-spanning proteins, *Proc. Natl. Acad. Sci. U. S. A.*, 1989 Aug, **86**(15), 5786–5790.

27 Z. Cao, L. M. van, L. J. Mitchell, M. A. Pringle, K. Inaba and N. J. Bulleid, The membrane topology of vitamin K epoxide reductase is conserved between human isoforms and the bacterial enzyme, *Biochem. J.*, 2016, **473**(7), 851–858.

28 L. A. Rutkevich and D. B. Williams, Vitamin K epoxide reductase contributes to protein disulfide formation and redox homeostasis within the endoplasmic reticulum, *Mol. Biol. Cell*, 2012, **23**(11), 2017–2027.

29 S. Schulman, B. Wang, W. Li and T. A. Rapoport, Vitamin K epoxide reductase prefers ER membrane-anchored thioredoxin-like redox partners, *Proc. Natl. Acad. Sci. U. S. A.*, 2010, **107**(34), 15027–15032.

30 P. Dowd, S. W. Ham, S. Naganathan and R. Hershline, The mechanism of action of vitamin K, *Annu. Rev. Nutr.*, 1995, **15**, 419–440.

31 M. A. Rishavy, A. Usubalieva, K. W. Hallgren and K. L. Berkner, 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, *J. Biol. Chem.*, 2011, **286**(9), 7267–7278.

32 B. A. Soute, M. M. Groenen-van Dooren, A. Holmgren, J. Lundstrom and C. Vermeer, Stimulation of the dithiol-dependent reductases in the vitamin K cycle by the thioredoxin system. Strong synergistic effects with protein disulphide-isomerase, *Biochem. J.*, 1992, **281**(1), 255–259.

33 H. M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, *et al.*, The Protein Data Bank, *Nucleic Acids Res.*, 2000, **28**(1), 235–242.

34 S. Wu, J. K. Tie, D. W. Stafford and L. G. Pedersen, Membrane topology for human vitamin K epoxide reductase, *J. Thromb. Haemostasis*, 2014, **12**(1), 112–114.

35 D. E. Warschawski, A. A. Arnold, M. Beaugrand, A. Gravel, E. Chartrand and I. Marcotte, Choosing membrane mimetics for NMR structural studies of transmembrane proteins, *Biochim. Biophys. Acta*, 2011, **1808**(8), 1957–1974.

36 W. Kabsch and C. Sander, Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features, *Biopolymers*, 1983, **22**(12), 2577–2637.

37 U. Von Luxbourg, A tutorial on spectral Clustering. Statistics and Computing, 2007, vol. 17(4), arXiv:0711.0189.

38 B. Nadler, S. Lafon, R. R. Coifman and I. G. Kevrekidis, Diffusion maps, spectral clustering and eigenfunctions of Fokker-Planck operators, *Neural Information Processing Systems*, 2005, p. 18.

39 B. Chalmond, Modeling and Inverse Problem in Image Analysis, *Applied Mathematical Sciences*, 2003, p. 155.

40 R. Harada and A. Kitao, Parallel Cascade Selection Molecular Dynamics (PaCS-MD) to generate conformational transition pathway, *J. Chem. Phys.*, 2013, **139**(3), 035103.

41 G. V. Le, P. Schmidtke and P. Tuffery, Fpocket: an open source platform for ligand pocket detection, *BMC Bioinf.*, 2009, **10**, 168.

42 G. M. Morris, R. Huey, W. Lindstrom, M. F. Sanner, R. K. Belew, D. S. Goodsell, *et al.*, AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility, *J. Comput. Chem.*, 2009, **30**(16), 2785–2791.

43 B. Matagrin, A. Hodroge, A. Montagut-Romans, J. Andru, I. Fourel, S. Besse, *et al.*, New insights into the catalytic mechanism of vitamin K epoxide reductase (VKORC1) – The catalytic properties of the major mutations of rVKORC1 explain the biological cost associated to mutations, *FEBS Open Bio*, 2013, **3**, 144–150.

44 T. Hou, J. Wang, Y. Li and W. Wang, Assessing the performance of the MM/PBSA and MM/GBSA methods. 1. The accuracy of binding free energy calculations based on molecular dynamics simulations, *J. Chem. Inf. Model.*, 2011, **51**(1), 69–82.

45 K. J. Czogalla, A. Biswas, K. Honing, V. Hornung, K. Liphardt, M. Watzka, *et al.*, Warfarin and vitamin K compete for binding to Phe55 in human VKOR, *Nat. Struct. Mol. Biol.*, 2017, **24**(1), 77–85.

46 W. Jiang and B. Roux, Free Energy Perturbation Hamiltonian Replica-Exchange Molecular Dynamics (FEP/H-REMD) for Absolute Ligand Binding Free Energy Calculations, *J. Chem. Theory Comput.*, 2010, **6**(9), 2559–2565.

47 B. Webb and A. Sali, Comparative Protein Structure Modeling Using MODELLER, *Curr Protoc Bioinformatics*, 2016, **86**, 2.

48 M. M. Ghahremanpour, S. S. Arab, S. B. Aghazadeh, J. Zhang and D. Van Der Spoel, MemBuilder: a web-based graphical interface to build heterogeneously mixed membrane bilayers for the GROMACS biomolecular simulation program, *Bioinformatics*, 2014, **30**(3), 439–441.

49 J. P. Jambeck and A. P. Lyubartsev, Derivation and systematic validation of a refined all-atom force field for phosphatidylcholine lipids, *J. Phys. Chem. B*, 2012, **116**(10), 3164–3179.

50 D. J. Price and C. L. Brooks III, A modified TIP3P water potential for simulation with Ewald summation, *J. Chem. Phys.*, 2004, **121**(20), 10096–10103.

51 M. G. Wolf, M. Hoefling, C. Aponte-Santamaria, H. Grubmuller and G. Groenhof, g_membed: Efficient insertion of a membrane protein into an equilibrated lipid bilayer with minimal perturbation, *J. Comput. Chem.*, 2010, **31**(11), 2169–2174.

52 S. O. Yesylevskyy, ProtSqueeze: simple and effective automated tool for setting up membrane protein simulations, *J. Chem. Inf. Model.*, 2007, **47**(5), 1986–1994.

53 D. Van Der Spoel, E. Lindahl, B. Hess, G. Groenhof, A. E. Mark and H. J. Berendsen, GROMACS: fast, flexible, and free, *J. Comput. Chem.*, 2005, **26**(16), 1701–1718.

54 R. B. Best, X. Zhu, J. Shim, P. E. Lopes, J. Mittal, M. Feig, *et al.*, Optimization of the additive CHARMM all-atom protein force field targeting improved sampling of the backbone phi, psi and side-chain chi(1) and chi(2) dihedral angles, *J. Chem. Theory Comput.*, 2012, **8**(9), 3257–3273.

55 G. Bussi, D. Donadio and M. Parrinello, Canonical sampling through velocity rescaling, *J. Chem. Phys.*, 2007, **126**(1), 014101.

56 W. F. van Gunsteren and H. J. Berendsen, Computer simulation as a tool for tracing the conformational differences between proteins in solution and in the crystalline state, *J. Mol. Biol.*, 1984, **176**(4), 559–564.

57 W. G. Hoover, Canonical dynamics: equilibrium phase-space distributions, *Phys. Rev. A Gen. Phys.*, 1985, **31**(3), 1695–1697.

58 V. J. de, R. M. Scheek, W. F. van Gunsteren, H. J. Berendsen, R. Kaptein and J. Thomason, Combined procedure of distance geometry and restrained molecular dynamics techniques for protein structure determination from nuclear magnetic resonance data: application to the DNA binding domain of lac repressor from Escherichia coli, *Proteins: Struct., Funct., Genet.*, 1988, **3**(4), 209–218.

59 L. Verlet, Computer "Experiments" on Classical Fluids. I. Thermodynamical Properties of Lennard-Jones Molecules, *Phys. Rev.*, 1967, **159**(1), 98–103.

60 D. M. York, A. Wlodawer, L. G. Pedersen and T. A. Darden, Atomic-level accuracy in simulations of large protein crystals, *Proc. Natl. Acad. Sci. U. S. A.*, 1994, **91**(18), 8715–8718.

61 S. Yuan, H. C. Chan, S. Filipek and H. Vogel, PyMOL and Inkscape Bridge the Data and the Data Visualization, *Structure*, 2016, **24**(12), 2041–2042.

62 W. Humphrey, A. Dalke and K. Schulten, VMD: visual molecular dynamics, *J. Mol. Graphics*, 1996, **14**(1), 33–38.

63 B. J. Grant, A. P. Rodrigues, K. M. ElSawy, J. A. McCammon and L. S. Caves, Bio3d: an R package for the comparative analysis of protein structures, *Bioinformatics*, 2006, **22**(21), 2695–2696.

64 E. Lyman and D. M. Zuckerman, Ensemble-based convergence analysis of biomolecular trajectories, *Biophys. J.*, 2006, **91**(1), 164–172.

65 X. Daura, W. F. van Gunsteren and A. E. Mark, Folding-unfolding thermodynamics of a beta-heptapeptide from equilibrium simulations, *Proteins: Struct., Funct., Genet.*, 1999, **34**(3), 269–280.

66 E. Parzen, On Estimation of a Probability Density Function and Mode, *Ann. Math. Stat.*, 1962, **33**(3), 1065–1076.

67 B. W. Silverman, Density Estimation for Statistics and Data Analysis. *Monographs on Statistics and Applied Probability*, Chapman and Hall, London, 1986, vol. 48.

68 R. R. Coifman, I. G. Kevrekidis, S. Lafon, M. Maggioni and B. Nadler, Diffusion Maps, Reduction Coordinates, and Low Dimensional Representation of Stochastic Systems, *Multiscale Model. Simul.*, 2008, **7**, 842–864.

69 B. Nadler, S. Lafon, R. R. Coifman and I. G. Kevrekidis, Diffusion maps, spectral clustering and reaction coordinates of dynamical systems, *Appl. Comput. Harmon. Anal.*, 2006, **21**, 113–127.

70 A. B. Wohri, W. Y. Wahlgren, E. Malmerberg, L. C. Johansson, R. Neutze and G. Katona, Lipidic sponge phase crystal structure of a photosynthetic reaction center reveals lipids on the protein surface, *Biochemistry*, 2009, **48**(41), 9831–9838.

71 E. F. Pettersen, T. D. Goddard, C. C. Huang, G. S. Couch, D. M. Greenblatt, E. C. Meng, *et al.*, UCSF Chimera– a visualization system for exploratory research and analysis, *J. Comput. Chem.*, 2004, **25**(13), 1605–1612.

72 D. Elenkova, B. Morgenstern, I. Manolov and M. Milanova, Synthesis, crystal structure and physico-chemical properties of 3,3'-[(4-hydroxyphenyl)methyl] bis-(4-hydroxy-2*H*-chromen-2-one), *Acta Chim. Slov.*, 2014, **61**(4), 718–728.

73 V. Zoete, M. A. Cuendet, A. Grosdidier and O. Michielin, SwissParam: a fast force field generation tool for small organic molecules, *J. Comput. Chem.*, 2011, **32**(11), 2359–2368.

74 F. J. Solis and R. J.-B. Wets, Minimization by Random Search Techniques, *Math. Oper. Res.*, 1981, **6**, 19–30.

75 N. A. Baker, D. Sept, S. Joseph, M. J. Holst and J. A. McCammon, Electrostatics of nanosystems: application to microtubules and the ribosome, *Proc. Natl. Acad. Sci. U. S. A.*, 2001, **98**(18), 10037–10041.

76 R. Kumari, R. Kumar and A. Lynn, g_mmpbsa–a GROMACS tool for high-throughput MM-PBSA calculations, *J. Chem. Inf. Model.*, 2014, **54**(7), 1951–1962.

77 A. Hodroge, C. Longin-Sauvageon, I. Fourel, E. Benoit and V. Lattard, Biochemical characterization of spontaneous mutants of rat VKORC1 involved in the resistance to antivitamin K anticoagulants, *Arch. Biochem. Biophys.*, 2011, **515**(1–2), 14–20.

This journal is © The Royal Society of Chemistry 2017