



OPEN ACCESS

EDITED BY

Lu Shaoyong,
Shanghai Jiao Tong University, China

REVIEWED BY

Qin Xu,
Shanghai Jiao Tong University, China
Ruo-Xu Gu,
Shanghai Jiao Tong University, China
Durba Sengupta,
National Chemical Laboratory (CSIR),
India
Jinan Wang,
University of Kansas, United States

*CORRESPONDENCE

Harel Weinstein,
✉ haw2002@med.cornell.edu

RECEIVED 30 March 2023

ACCEPTED 04 May 2023

PUBLISHED 18 May 2023

CITATION

Xie H and Weinstein H (2023).
Allosterically coupled conformational dynamics in solution prepare the sterol transfer protein StarD4 to release its cargo upon interaction with target membranes.
Front. Mol. Biosci. 10:1197154.
doi: 10.3389/fmolb.2023.1197154

COPYRIGHT

© 2023 Xie and Weinstein. This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](#). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Allosterically coupled conformational dynamics in solution prepare the sterol transfer protein StarD4 to release its cargo upon interaction with target membranes

Hengyi Xie¹ and Harel Weinstein^{1,2*}

¹Department of Physiology and Biophysics, Weill Cornell Medicine, New York, NY, United States, ²Institute for Computational Biomedicine, Weill Cornell Medicine, New York, NY, United States

Complex mechanisms regulate the cellular distribution of cholesterol, a critical component of eukaryote membranes involved in regulation of membrane protein functions directly and through the physiochemical properties of membranes. StarD4, a member of the steroidogenic acute regulator-related lipid-transfer (StART) domain (StARD)-containing protein family, is a highly efficient sterol-specific transfer protein involved in cholesterol homeostasis. Its mechanism of cargo loading and release remains unknown despite recent insights into the key role of phosphatidylinositol phosphates in modulating its interactions with target membranes. We have used large-scale atomistic Molecular dynamics (MD) simulations to study how the dynamics of cholesterol bound to the StarD4 protein can affect interaction with target membranes, and cargo delivery. We identify the two major cholesterol (CHL) binding modes in the hydrophobic pocket of StarD4, one near S136&S147 (the Ser-mode), and another closer to the putative release gate located near W171, R92&Y117 (the Trp-mode). We show that conformational changes of StarD4 associated directly with the transition between these binding modes facilitate the opening of the gate. To understand the dynamics of this connection we apply a machine-learning algorithm for the detection of rare events in MD trajectories (RED), which reveals the structural motifs involved in the opening of a front gate and a back corridor in the StarD4 structure occurring together with the spontaneous transition of CHL from the Ser-mode of binding to the Trp-mode. Further analysis of MD trajectory data with the information-theory based NbIT method reveals the allosteric network connecting the CHL binding site to the functionally important structural components of the gate and corridor. Mutations of residues in the allosteric network are shown to affect the performance of the allosteric connection. These findings outline an allosteric mechanism which prepares the CHL-bound StarD4 to release and deliver the cargo when it is bound to the target membrane.

KEYWORDS

molecular dynamics (MD) simulations, cholesterol traffic and distribution in cells, N-body information theory analysis of MD trajectories, detection of rare events in MD trajectories, allosteric network coordination, machine learning analysis of MD trajectories, ligand-induced conformational changes, allosteric channel

Introduction

Cholesterol is a critical component of mammalian cell membranes, involved in the regulation of membrane protein function both through direct protein-CHL interactions and through the effect of CHL on the physiochemical properties of the host membranes. It is heterogeneously distributed among cellular organelles: the plasma membrane (PM) and the endocytic recycling compartment (ERC) are major pools of the total cellular cholesterol, whereas the endoplasmic reticulum (ER), where the cholesterol concentration is sensed and regulated, and the

cellular sterol homoeostasis is maintained (Liscum and Munn, 1999; Radhakrishnan et al., 2008; Iaea and Maxfield, 2015), contains only a low amount of cellular cholesterol (0.1%–2%).

Subcellular distribution of CHL occurs through both vesicular and non-vesicular transport mechanisms, the latter accounting for 70% of the cholesterol transport (Iaea and Maxfield, 2015; Hao et al., 2002; Iaea et al., 2017). Rapid non-vesicular transport requires lipid transfer proteins that provide the hydrophobic environment needed to accommodate lipid transferring across the aqueous phase (Prinz, 2007; Wong et al., 2019). One major family of sterol transfer proteins is

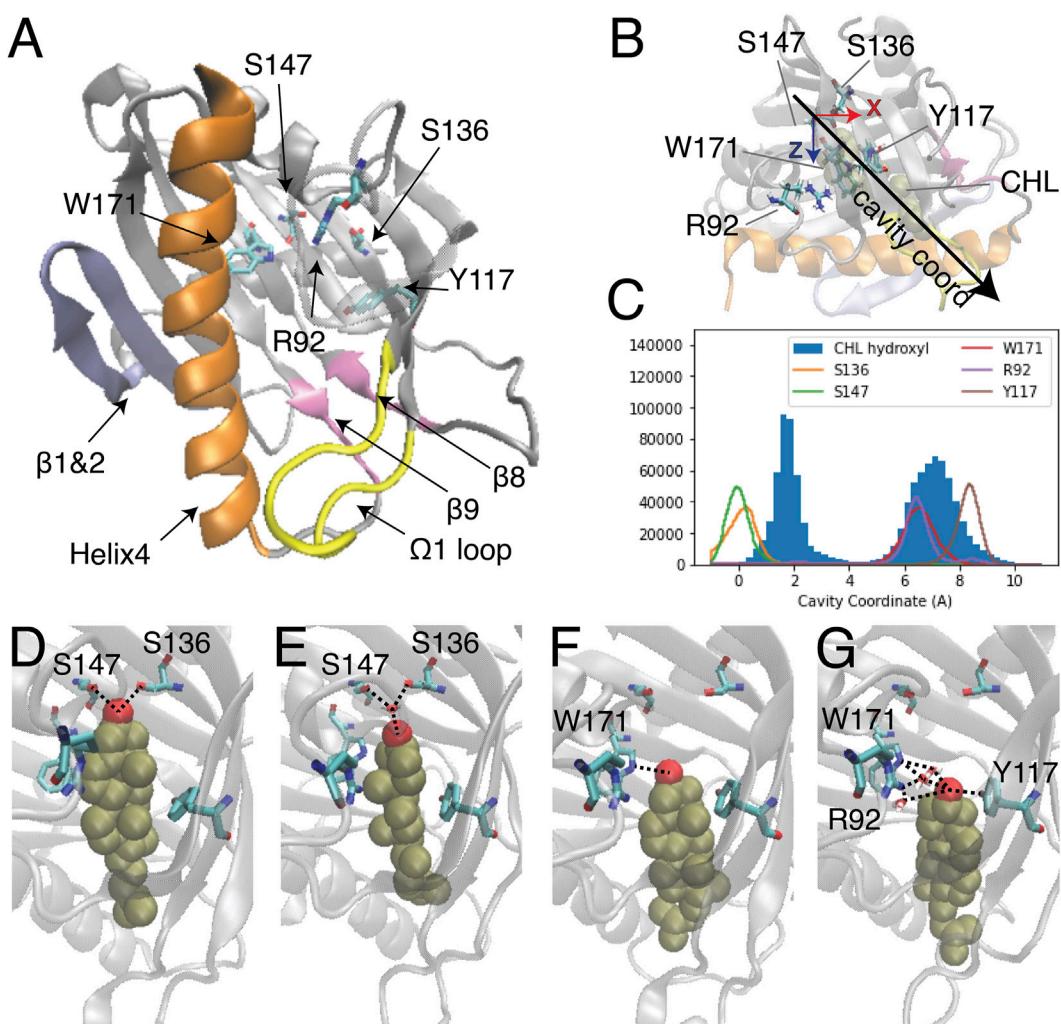


FIGURE 1

Two modes of CHL binding in the hydrophobic pocket of StarD4. The binding of cholesterol was modeled computationally in the crystal structure of StarD4 (PDB ID: 1jss) (A) The α/β helix-grip structure of the STARD domain in the crystal structure of StarD4. StarD4 is rendered in gray, with the C-terminal Helix (Helix4) in orange, the loop between β 5 and β 6 (Ω 1 loop) in yellow, the β 1&2 sheets in blue, and the β 8 loop and β 9 loop in pink. Cholesterol binding sites, residues S136, S147, W171, R92 and Y117, are rendered in "licorice" which draws the atoms as spheres and the bonds as cylinders, with oxygen colored in red, nitrogen in blue and carbons in cyan (hydrogens are not shown). (B, C) The "cavity coordinate" of CHL (rendered in VDW in tan color) and StarD4 residues reveal two modes of interaction. (B) Definition of the "cavity coordinate": With the CHL-StarD4 complex rotated first to align the H4 along the x coordinate and position the Center of Mass (CoM) of the protein right above the CoM of H4, the "cavity coordinate" is defined by taking the center of mass of S136&S147 as the origin and measuring the distance along the direction of the angle bisector of x- and z-axes. (C) Frequency histogram of the cavity coordinate of the hydroxyl oxygen in CHL (filled), and of the sidechain oxygen or nitrogen atoms in S136, S147, W171, R92 and Y117 (hollow). (D–G) Binding modes of CHL in the CHL-StarD4 complex: (D) the "Ser-binding mode"; (E) the "water-bridged Ser-binding mode"; (F) the "Trp-binding mode"; (G) "water-bridged Trp-binding mode". StarD4 is shown in the same representation as in (A). CHL is rendered in VDW with the oxygen in red, and other atoms in tan color. Water molecules directly involved in hydrogen bonds with the CHL and binding-sites are indicated with dashed lines.

the (START) domain family of the steroidogenic acute regulatory protein (StAR)-related lipid-transfer. A START domain (STARD) is composed of ~210 amino acids that fold into an α/β helix-grip structure to create an internal hydrophobic cavity for lipid binding and recognition as illustrated in **Figure 1A** (Mathieu et al., 2002; Letourneau et al., 2015). The mammalian STARD protein family comprises 15 proteins, subdivided into six subfamilies based on domain architecture and ligand specificity (Alpy and Tomasetto, 2005; Mesmin et al., 2011; Wong et al., 2019; Clark, 2020). The characteristic of the STARD4 subfamily is a single soluble sterol-binding START domain. The StarD4 protein is well-adapted to bind and transport sterol (Romanowski et al., 2002; Iaea et al., 2015) and is widely expressed in multiple tissues (Soccio et al., 2002; Elbadawy et al., 2011; Rodriguez-Agudo et al., 2011). StarD4 knockdown was shown to result in a decrease of (i) ER cholesterol concentration, (ii) acyl-CoA:cholesterol acyl-transferase 1 (ACAT1) activity, and (iii) cellular cholesterol ester concentration (Garbarino et al., 2012). The non-vesicular sterol transport kinetics between PM, ERC and ER was reduced by 33% (Iaea et al., 2020). Conversely, StarD4 overexpression resulted in facilitated intracellular cholesterol ester accumulation in a ACAT1 dependent manner (Rodriguez-Agudo et al., 2008; Mesmin et al., 2011).

Despite extensive studies of STARD protein function, no structure of a START domain complexed with a sterol has been resolved. Thus, the ligand binding modes of the START domain remain undetermined, and the molecular mechanism that underlies the lipid specificity and determines the lipid uptake and release pathways, is still unclear. Studies of sterol binding in START domains that employed primarily docking and short MD simulation have identified the cholesterol binding pocket and residues likely to be involved in ligand binding (**Figure 1A**) (Tsujishita and andHurley, 2000; Mathieu et al., 2002; Murcia et al., 2006; Thorsell et al., 2011; Tan et al., 2019). In these studies, sterol was reported to favor a binding mode where the CHL inserts into the depths of the hydrophobic cavity and the hydroxyl group forms interactions with polar side chains or backbone carbonyls of residues at the bottom of the pocket. In **Figure 1A** residues S136 and S147 are labeled in the binding pocket because they correspond to the predicted CHL binding sites in StarD3 (S362 and R351) and StarD5 (S132).

The recently determined structures of LAM proteins in the StARkin superfamily show that the sterol hydroxyl group makes hydrogen bonds with residues located in the upper two-thirds of the hydrophobic tunnel, and the sterol hydrocarbon tail is exposed to the solvent through a partially open lid (Horenkamp et al., 2018; Jentsch et al., 2018; Tong et al., 2018). Structural comparison between yeast LAM and human StarD4 also suggested that conformational changes at Helix4 and Ω 1 loop (the loop between β 5 and β 6) (**Figure 1A**) are required for the apo-StarD4 crystal structure to accommodate a cholesterol ligand at the same binding site as LAM protein (Tan et al., 2019).

In the present study we have used extensive (~0.4 milliseconds) MD simulation trajectories to explore the binding modes of cholesterol in the pocket of StarD4 in solution in order to (1) assess the structural relationships between the *apo* and *holo* states of the protein, (2) identify the dynamic

rearrangements required to accommodate the binding of CHL, and (3) the relation of these dynamic changes to the formation of the protein state required for its functionally productive membrane interactions. The long trajectories revealed a rich dynamic landscape of the protein structure in which the bound CHL adopts positions and configurations suggesting preparations for its CHL trafficking functions. Such function-related conformational changes of the StarD4 protein and its complex with CHL were revealed from the analysis of the MD trajectories with a machine learning-based Rare Event Detection (RED) protocol (Plante and Weinstein, 2021). To reveal the mechanisms underlying the function-related conformational changes we explored the allosteric pathways connecting the CHL repositioning in the binding site, with the identified conformational changes by applying the N-body Information Theory (NbIT) analysis (LeVine et al., 2014; LeVine and Weinstein, 2014) to the MD trajectories. We show here that the allosteric mechanism resulting from this analysis involves the coupling between the cholesterol translocation dynamics in the binding site and configurational changes of specific regions of the structure that are involved in the interaction of StarD4 with the membrane (Zhang et al., 2022). The detailed structure-based information about the conformational changes and the allosteric channel enabled the development of specific testable hypotheses for mutations that we used here to probe the molecular mechanisms of function of a CHL-loaded StarD4 diffusing in the aqueous medium of the cytosol, and mutations that affect the molecular rearrangements that prepare the pathway for sterol release when the protein is embedding in a target membrane as we have shown (Zhang et al., 2022).

Results

MD simulations reveal two modes of cholesterol binding in StarD4

The binding modes of CHL in the pocket identified previously (Mathieu et al., 2002; Clark, 2020) were explored with extensive atomistic MD simulations comprised of 12 statistically independent replicates of 33.6 μ s each for a total of 403 μ s, starting from the crystal structure of StarD4 (PDBid: 1JSS) with a CHL molecule docked in the binding site using the Schrodinger Induced Fit Docking protocol (Sherman et al., 2006a; Sherman et al., 2006b) (see Methods for details). The simulations produced two different modes of CHL binding in the hydrophobic pocket (**Figure 1B**), one with the hydroxyl group located near Ser136/Ser147 (termed “Ser-binding mode”), the other with the CHL OH near Trp171 (“Trp-binding mode”). In the “Ser-binding mode”, the cholesterol hydroxyl group forms hydrogen bonds with the Ser136 and Ser147 sidechains either directly, or through a water bridge (**Figures 1C, D**). In the “Trp-binding mode”, the cholesterol hydroxyl group interacts either directly with Trp171, or resides between Trp171, Arg92 and Tyr117 with which it interacts through H-bonds mediated by a water molecule (**Figures 1E, F**). Spontaneous transitions of cholesterol between binding modes are observed in more than half of the trajectories (**Supplementary Figure S1**).

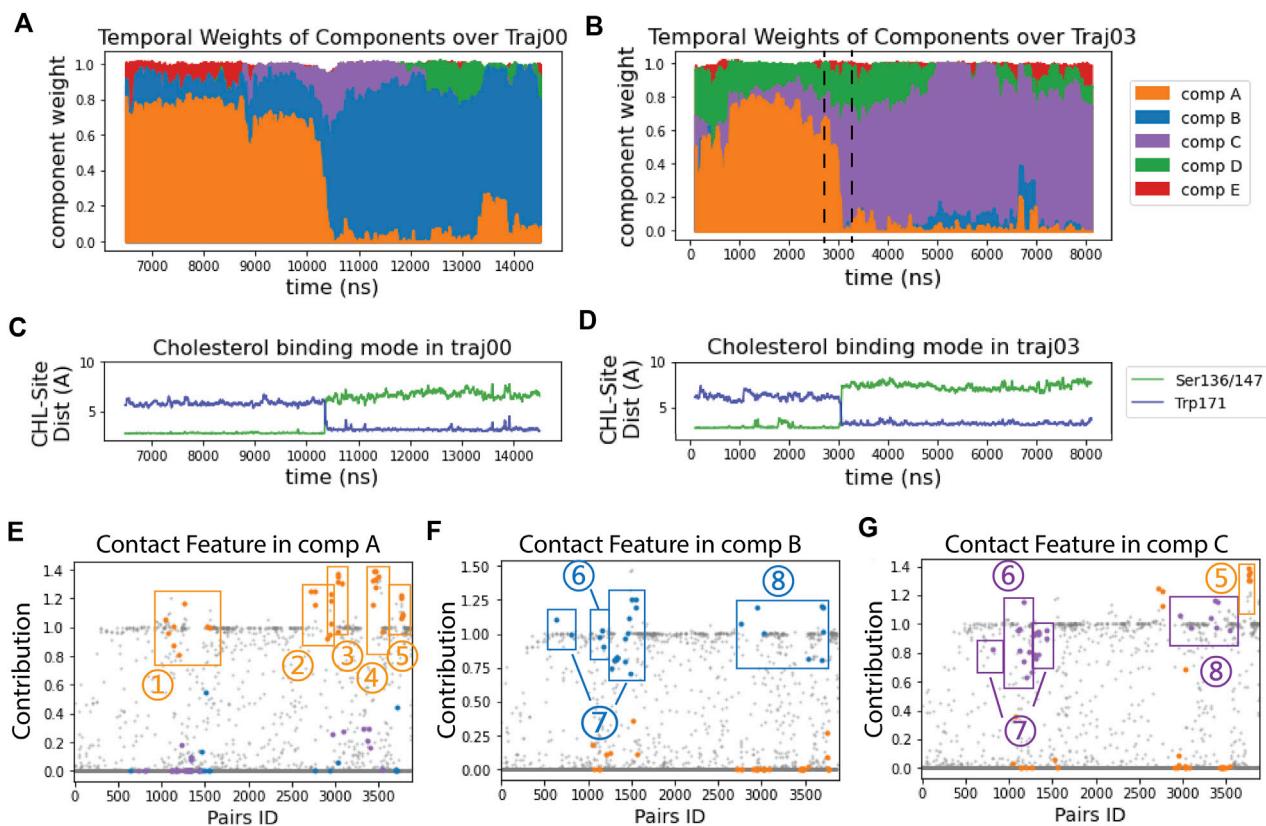


FIGURE 2

RED analysis revealed a set of simultaneous conformational changes concurrent with the translocation of CHL. (A, B) The time evolution of the RED-detected normalized temporal weight of structural components over the simulation time of trajectory 0 (A) and 3 (B). Color code: comp (A) (orange), comp (B) (blue), comp (C) (purple), comp (D) (green), comp (E) (red). At each time point, the color column represents the stacked values of the weights for all the components beneath. For example, at time = 2700 ns in traj03 [left dashed line in (B)] the weight of comp A is 0.67, comp C is 0.12, comp D is 0.22, and comp B&E are 0; at time = 3300 ns (right dashed line), the weight of comp A is 0.06, comp B is 0, comp C is 0.66, comp D is 0.23 and comp E is 0.05. (C, D) The time evolution of the cholesterol binding mode, represented by the change in the distances between CHL and the Ser binding site (in green) and to the Trp site (blue) for: (C) in trajectory 0; (D) in trajectory 3. (E–G) Spatial arrays of components A, B, and C show the contact features characterized by the structure-differentiating contact pairs (SDCPs). Along the X-axis there are 3,880 data points, each representing one residue pair. The Y coordinate of each data point shows the contribution of the residue pair to the contact feature of the component (see Methods). The SDCPs (summarized in [Supplementary Table S1](#)) are grouped by secondary structure location and highlighted in colors. Each E–G shows a conformational state for which the determinant conformational features are encoded by the contact patterns of structure-differentiating contact pairs (SDCPs). For example, in comp B (in (F)) residue pair groups 1–5 lost contact, while groups 6–8 have formed contacts. The text defines the structural elements containing the pairs collected in the SDCP groups. Together, these groups contain the SDCPs that define the structural feature in the time segment when comp B is prevalent in determining the conformational state. Note that the contributions in the RED spatial component analysis (Figures 2E–G) are not constrained to values lower than 1, because they are relative values indicating the contribution of residue pairs in comparison to the constructive residue pairs. Constructive residue pairs are those remaining in contact in all conformational states, and thus they are assigned a constant contribution value of 1 in every component. In contrast, the structure-differentiating contact pairs have highly varying contribution values across the different components. In some components, these pairs may contribute more strongly than the constructive residue pairs, resulting in contribution values > 1.

Transitions between the two modes of cholesterol binding are concurrent with conformational changes of the StarD4 protein

To reveal the sequence of conformational events that accompany the transitions between the Ser and Trp binding modes, we applied the recently developed Rare Event Detection (RED) protocol (Plante and Weinstein, 2021) as described in Methods. The mechanistically important state-to-state transitions of StarD4 in response to the translocation of ligand cholesterol (see ref. 31 for a discussion of the relation between rare events and function) were extracted from the dynamic information in an

ensemble of 6 trajectory stretches (4–8 μ s each, 34.4 μ s total) from different trajectory replicas in which a stable cholesterol transition from one binding mode to another was observed. This ensemble of 6 trajectories served as the input training data for the RED protocol in identifying the rearrangement events shared among the different independent replicas.

The RED algorithm decomposed the trajectories into 5 components based on the evolution of the residue-residue contact map over time (see Methods). As described in ref (Plante and Weinstein, 2021), each component represents a state of StarD4 in which a particular conformation becomes dominant, at a particular time in the trajectory. This time-ordered series of events in which different components dominate the structural changes

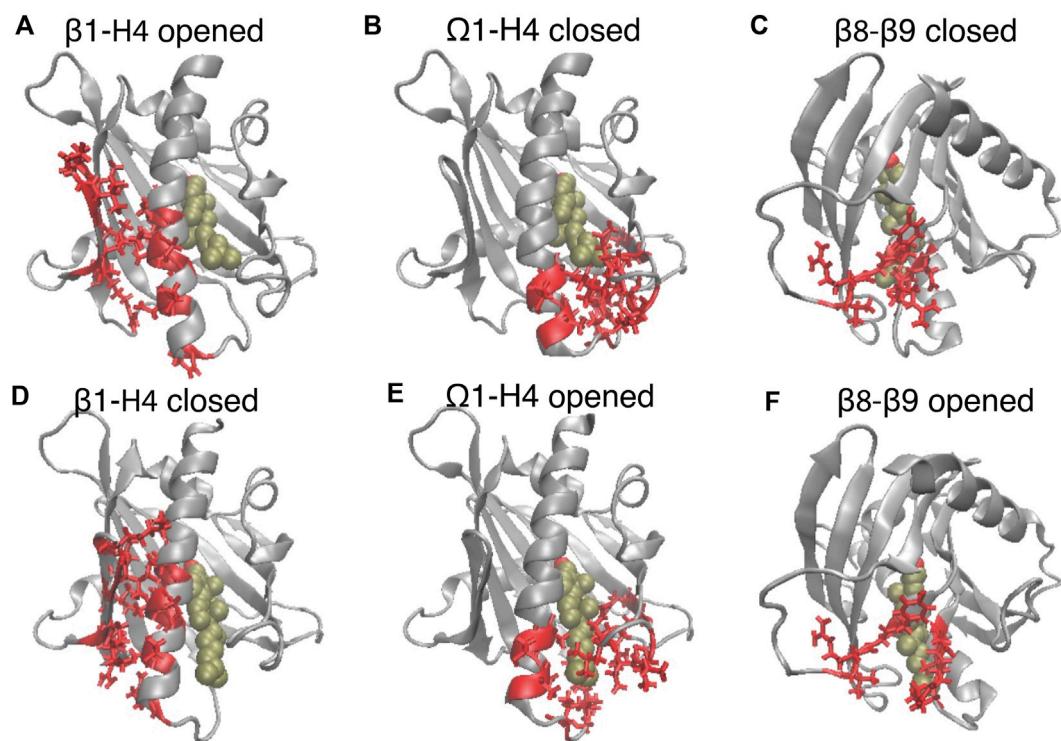


FIGURE 3

The translocation of CHL is concurrent with the opening of the gate and corridor to the hydrophobic pocket. The elements of motif movements are compared in representative conformation in the closed state (A–C) (corresponding to the **left** dashed line in Figure 2B), and in the opened state (corresponding to the **right** dashed line in Figure 2B). StarD4 is shown in gray, and the CHL is rendered the same way as in Figure 1. The SDCPs are shown in red with licorice representations: (A, D) the contact forming SDCPs (group 6, summarized in Supplementary Table S1) between β 1-H4; (B, E) the contact breaking SDCPs (group 2,3) between β 9, H4 and Ω 1; (C, F) The contact breaking SDCPs (group 4) between β 8- β 9.

encoded in the trajectory, is shown in Figures 2A,B. Several components identified by the RED algorithm are observed to be dominant concurrently with the repositioning of cholesterol in the binding site (Figures 2A–D). This is remarkable, because the information provided to the RED algorithm consists only of the contact matrices for each frame in the trajectory, without direct information about CHL and its position. The correspondence between the binding modes and particular RED components is illustrated in the time frames by the correspondence of the times when component A (“comp A” in Figures 2A,B) is dominant, and CHL is in the Ser-binding mode of CHL (Figures 2C,D). Moreover, as the dominance of the A component wanes and is replaced by that of comp-s B and C, CHL is in the Trp-binding mode. Thus, in every independent trajectory the transition events from comp A to either comp B or comp C coincide with the translocation of cholesterol (Traj0&3 shown in Figures 2A–D, and Traj1,4,6,7 in Supplementary Figure S2A). This suggests a mechanism of coupled dynamics in the cholesterol-StarD4 complex, connecting the structural rearrangements of the protein frame with the transition of CHL in the binding pocket. The states associated with the transition of CHL to the Trp-binding mode are of particular interest because in this position the CHL is near a putative “exit gate” from the binding pocket.

We propose that the specific structural rearrangements of the protein frame constitute a preparatory step for a subsequent

functional docking of the loaded StarD4 to the membrane for the release of its cargo. A key indication for the likelihood of this mechanism is the consistency in the different simulation replicas of the conformational changes coupled to the transitions between CHL binding modes.

RED analysis reveals the specific conformational changes that occur simultaneously: The conformational changes that are coupled to the CHL translocation toward the vicinity of the “gate” are identified from comparisons of the spatial arrays in the RED analysis. For this comparison we identify structure-differentiating contact pairs (SDCPs) that underlie the conformational changes that occur simultaneously during the rare event of transition from one component to the next. While numerous residue contacts included in a component may change from “on” (i.e., in contact) to “off” (i.e., no longer in contact; see protocol in Methods) an SDCP must satisfy two simultaneous criteria: the pair makes a substantial contribution to the contact feature in one component and a minimal contribution in the subsequent component in the transition. Both the substantial contribution, and the minimal contributions, are determinant elements of a particular component’s conformational features, because the change in structure depends on both the formation of new contacts and the release of contacts from a previous conformation. Thus, they collectively determine the conformational change characteristic to the rare dynamic

event which makes them essential for interpreting the RED results.

Comparison of the time axis in panels A and B of **Figure 2** with that in panels C and D shows the transition between components A and B, and A and C occurs at the same time as the CHL transitions from the Ser-binding mode to the Trp-binding mode. The structural motifs involved in this transition are identified by the quantification of contributions to the components shown in **Figure 2E–G**, and defined structurally in **Figure 3**.

The SDCPs in **Figures 2E–G** are grouped (from 1 to 8) by their location in a particular secondary structure motif. Specifically, group 1 includes contacts between $\beta 1$ $\beta 2$ & the C-terminal Helix (H4); group 2 includes contacts between $\Omega 1$ & H4 (see **Figures 3B, E**); group 3 includes contacts between $\Omega 1$ & $\beta 9$ (**Figures 3B, E**); group 4 has the contacts between $\beta 9$ & $\beta 8$ (**Figures 3C, F**), and group 5 includes the interactions in the first turn of H4 (**Supplementary Figure S4A**).

The complete list of SDCPs is given in **Supplementary Table S1**, which summarizes the structural characteristics of the system at different stages in the trajectory. The structural analysis shows that SDCPs in groups 1 to 5 are in contact in comp A in which the CHL is in the Ser-bound mode (orange, **Figure 2E**), but most of them break contact during the transition from comp A to comp B or C, and thus have low feature values in comp B and comp C where CHL is in the Trp-binding mode (orange, **Figure 2F, G**).

Another set of SDCPs, 6 to 8, have low values in comp A (blue and purple in **Figure 2E**) but high values in comp B and C (blue in **Figure 2F** and purple in **Figure 2G**). They are found to represent contacts formed only after the repositioning of the CHL. More generally, comp B and comp C share a highly similar (but not identical) set of structural specifics represented in the salient SDCPs in groups 6 to 8. The highly similar pattern of motif interactions represented by these shared SDCPs include the following: in group 6: interactions between $\beta 1$ $\beta 2$ & H4 (**Figures 3A, D**); in group 7: interactions between $\beta 2$ & $\beta 3$ & $\beta 9$; in group 8: interactions of $\beta 8$ or $\beta 9$ with $\Omega 1$ or H4. Thus, the summary in **Supplementary Table S1** shows that groups 6-to-8 in comp B (B6-B8) and groups C6-C8 in comp C share most of the key secondary structural elements and their conformations.

The major difference between comp B and comp C is the partial unfolding of the N-terminal head of H4 (group 5; residues 199–201) which loses some contacts in comp B but remains structured in comp C. Together, the structural features of components B or C compared to those of comp A reveal the specific conformational changes that are captured as “rare events”.

The main rare events identified by the transition from a conformation in which one component is dominant to a new conformation with a different dominant component, are the A→B and A→C transitions which are concurrent with the CHL translocation (**Figures 2A–D**). The comparison of the structural features in comp A and in comp-s B&C shows that the conformational changes concurrent with the CHL translocation are the opening the access to the hydrophobic pocket. These conformational changes include (1) the opening of a front gate by the distancing of the H4 N-terminus from the $\Omega 1$ -loop and strengthening interactions with $\beta 1$ (see **Figures 3A, B, D, E**, and the detailed residue interaction data in **Figures 8A,B**); the opening of a back corridor between the $\beta 9$ tail and the $\beta 7\beta 8$ -loop (**Figures 3C, F**).

The predominant changes in the conformational space of StarD4 emerge from concurrent dynamics of CHL binding mode switching and gate opening

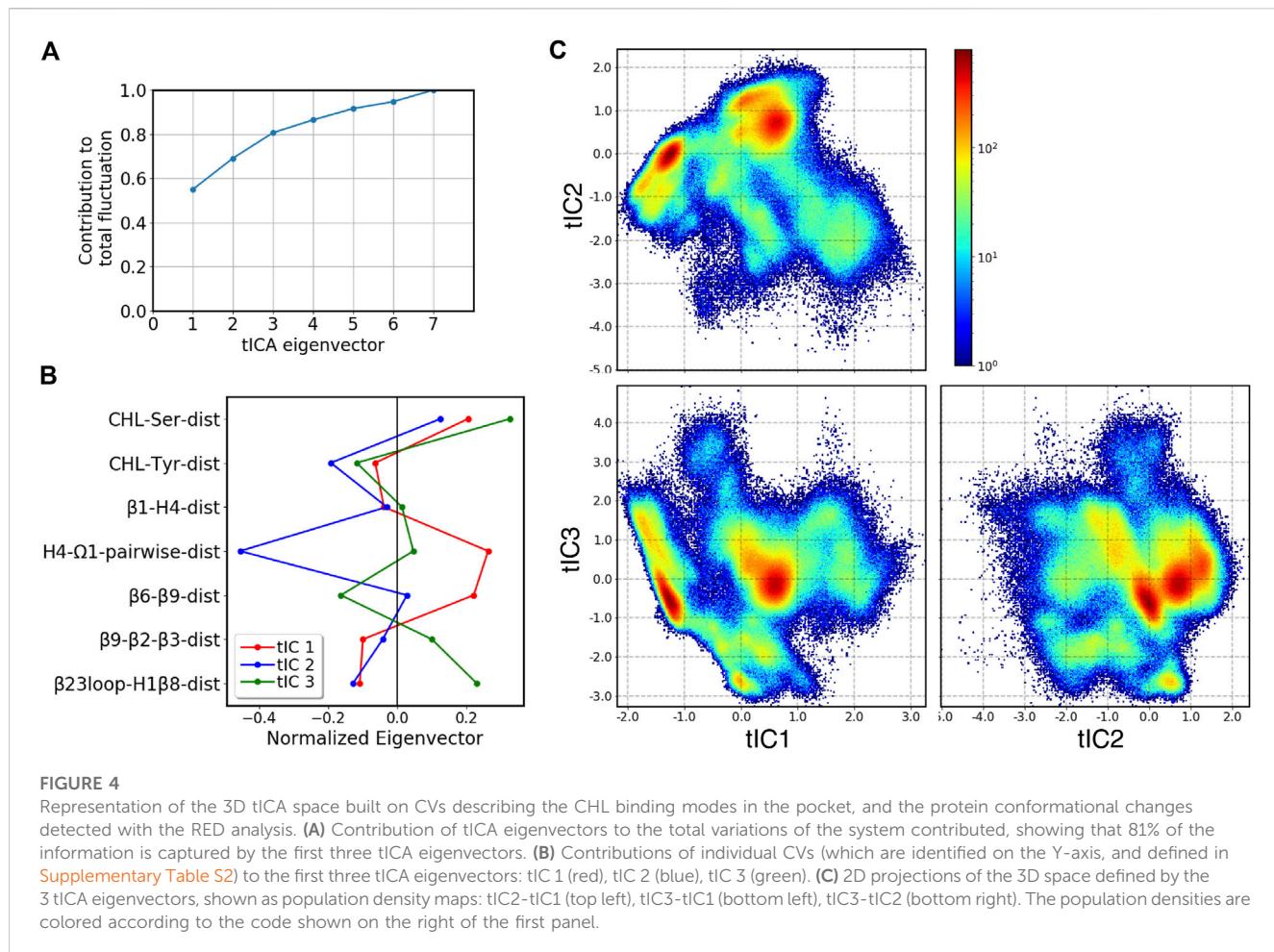
To facilitate analysis of the metastable states and transition pathways encoded in the MD trajectory data, we built the conformational landscape of the cholesterol-bound StarD4 by applying the dimensionality reduction approach tICA: time-structure based independent component analysis. To describe the dynamics of cholesterol binding modes and of the protein conformational changes (see Methods and **Supplementary Table S2**) we chose a set of 7 collective variables (CVs) to define the tICA space. The analysis showed that the first 3 tICA vectors describe 81% of the total dynamics of the system in the 403 μ s production run trajectories (**Figures 4A,B**). The projections on the space spanned by these three vectors are shown in **Figure 4C**.

To create a Markov State Model (MSM) and analyze the transitions among the metastable states visited by the StarD4 protein, the 3D tICA space was first discretized into 200 microstates using k-means clustering (see Methods and specifics in **Supplementary Figure S5**). The 200 microstates were then segregated according to similarity in kinetic properties, resulting in the identification of 9 macrostates. The 5 macrostates with high population are shown in **Figure 5A**, and representative conformations obtained from structural analysis are shown in **Figures 5B–D** and **Supplementary Figures S6, S8**.

The structural interpretation of macrostates 0–4 in **Figures 5B–D** and **Supplementary Figures S6, S8** indicated that the differences reflect the dynamics of gate, corridor, and cholesterol we detected in the stable transition events, as indicated by the data in **Figure 5E** showing the relevant CVs, and underscoring that these differences prevail throughout the conformational space.

Specifically, macrostate 1 in **Figure 5B** represents conformational states where cholesterol is bound in the Ser-site mode (with low “CHL-Ser-dist” and high “CHL-Tyr-dist”, seen in the first and second panel of **Figure 5E**), StarD4 is in the crystal-like conformations with a closed H4- $\Omega 1$ gate (high “ $\beta 1$ -H4-dist”, low “H4- $\Omega 1$ -dist” in the third and fourth panel of **Figure 5E**) in which the $\beta 8$ - $\beta 9$ corridor is closed (shown by low “ $\beta 6$ - $\beta 9$ -dist” and high “ $\beta 9$ - $\beta 2$ - $\beta 3$ -dist” & “ $\beta 2$ 3loop-H1 $\beta 8$ -dist” in the fifth to seventh panel in **Figure 5E**). Macrostate 0 also represents conformational states with Ser-binding CHL in crystal-like StarD4 with gates closed, but with a conformation of the $\beta 2$ 3 loop that is extended away (**Supplementary Figure S6A**) (high “ $\beta 2$ 3loop-H1 $\beta 8$ -dist”, seventh panel, **Figure 5E**) which may hinder the opening of $\beta 8$ - $\beta 9$ corridor. These Ser-binding/gates-closed conformations in Macrostates 1 and 0 contribute 30% of the populations in the conformation space.

The most populated macrostate is Macrostate 3, with 47% of the population. It contains conformational states in which CHL is in the Trp-binding mode with both the H4- $\Omega 1$ gate and $\beta 8$ - $\beta 9$ corridor open, and with $\beta 9$ leaning towards $\beta 2$ & $\beta 3$ which rearrange into narrower conformations (**Figures 5D, E**). Notably, all these observed rearrangements are consistent with the conformational changes detected by the RED analysis of the corresponding RED event. In Macrostate 4 the cholesterol is in a Trp-binding binding mode or even further down towards the H4- $\Omega 1$ gate that opens wider (**Figure 5E**, **Supplementary Figure S6B**). As indicated by the analysis, Macrostates 0,1,3,4 constitute 88% of the conformational space containing the



concurrent dynamic rearrangements of the cholesterol binding mode, the opening H4-Ω1 gate, and the β8-β9 corridor.

The other structural motif movements that were most important in the Comp A to Comp B/C transition event detected with RED are also evidenced by the comparisons of CVs related to the transition between the Ser- and Trp-binding modes of CHL. These are quantified in [Supplementary Figure S7](#) and include the unfolding in H4 as well as the straightening of H4, the narrowing of β2 and β3, and residue interactions between β1-H4 and disassociations between β8-β9. Thus, the conformational changes of the StarD4 protein that we found to occur simultaneously with the transitions between CHL binding modes, yield conformational states that are specific to the position of CHL in the binding pocket.

The time sequence of cholesterol translocation in the binding site is embedded in the MSM transition pathways starting from Macrostate 1 (crystal-like StarD4 with Ser-binding cholesterol) and ending in Macrostate 3 (gates-opened StarD4 with Trp-binding cholesterol). The fluxes along the most probable pathways (contributing ~95% of the transition flux) are summarized in [Table 1](#). Each of the top two most probable paths contributes ~30% of the total flux. One is the direct transition from Macrostate 1 to Macrostate 3 which encompasses the simultaneous translocation of CHL and conformational change of StarD4, and the other involves an intermediate, the less populated Macrostate 2 in which CHL is in Ser-binding modes but with both H4-Ω1 gate and β8-β9 corridor open ([Figures 5C, E](#)). These results are

consistent with the structural analysis in a report ([Tan et al., 2019](#)) showing that conformational changes at the H4-Ω1 gate are required for cholesterol to bind lower in the hydrophobic pocket, due to steric hindrance.

The less populated intermediate macrostates are summarized in [Supplementary Figure S8](#). They indicate the flexibility of the StarD4-cholesterol complex and support the relation between CHL position and the gate and corridor dynamics. Thus, with CHL in the Ser-binding mode, the gates-closed crystal-like conformations represent 30% of the population, much more than the ones with either the H4-Ω1 gate or the β8-β9 corridor open (10%). Most likely this is due to the energy cost of exposing the StarD4 hydrophobic pocket and cholesterol's hydrophobic tail to water. In contrast, when CHL is in the Trp-binding mode the H4-Ω1 gate is always open, and the conformations with both gate and corridor open amount to 58% compared to those with β8-β9 corridor closed, which are very unlikely (1.7%).

The allosteric pathway that couples cholesterol translocation in the hydrophobic pocket to the opening of access gates

Having identified a set of concurrent conformational changes in several structural motifs that include the cholesterol binding site, we

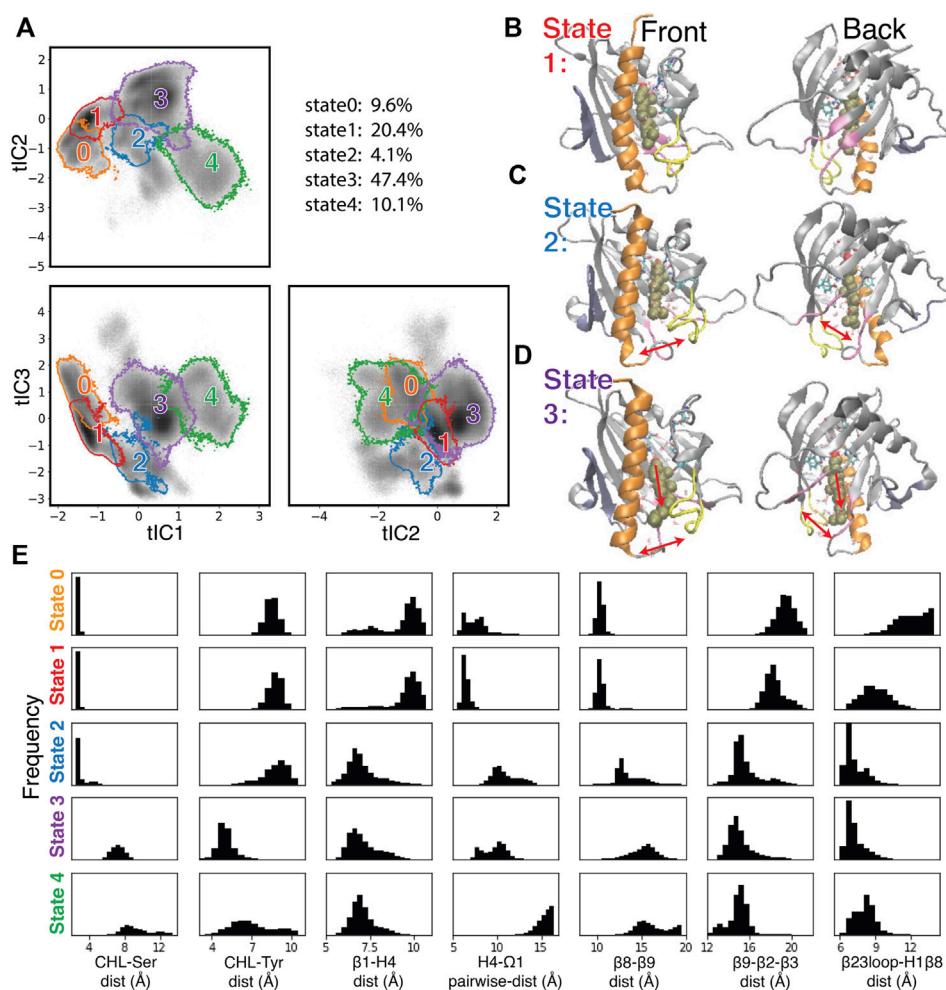


FIGURE 5

Structural elements in the conformational changes of the CHL translocation coupled to the opening of the H4-Ω1 gate and β8-β9 corridor. (A) Macrostates 0–4 are outlined in the 2D projections of the 3D conformational space: the tlC2-tlC1 plane (upper left), tlC3-tlC1 plane (lower left), tlC3-tlC2 plane (lower right). The gray shade represents the population density in the conformational space as shown in Figure 4C. The populations (% of total frames) are listed at the upper right for each microstate. (B–D) Structural models representing macrostates 1 to 3 viewed from front (left column) and back (right column). The StarD4-CHL complex is rendered in the same way as in Figure 1. Red arrows in the second and third rows (C, D) indicate the conformational changes of the H4-Ω1 gate and the β8-β9 corridor, as well as the movement of CHL. (E) Structural characteristics of macrostates 0 to 4 shown represented by probability density histograms of the characteristic CV values that determine the tlCA space, as defined in Supplementary Table S2.

TABLE 1 Top transition pathways in the cholesterol translocation process from Ser-binding state (State 1) to Trp-binding state (State 2).

Path	Norm Flux	Cumulative Flux
1 → 2 → 3	0.3269	0.3269
1 → 3	0.2984	0.6253
1 → 0 → 7 → 3	0.1151	0.7404
1 → 0 → 2 → 3	0.0826	0.8230
1 → 0 → 3	0.0464	0.8754
1 → 5 → 3	0.0414	0.9168
1 → 0 → 2 → 4 → 3	0.0258	0.9426
1 → 5 → 2 → 4 → 3	0.0140	0.9567

applied NbIT analysis (see (LeVine et al., 2014; LeVine and Weinstein, 2014) and Methods) to reveal the allosteric network that connects them. NbIT analysis calculates the normalized coordination information (NCI), which describes how much information of the *receiver* system can be gained from information about the *transmitter* system (see (LeVine and Weinstein, 2014) and Method for details). Here we applied the NCI analysis to the trajectories from the StarD4-cholesterol complex simulations to quantify the information transmission between motifs that emerged from the RED analysis described above. The residue index in Method identifies the components of the structural motifs used in this analysis summarized in Table 2. In Supplementary Table S3 we report the structural motifs that were used as negative controls for which the NCI quantification is expected to reveal only background-level sharing of information with the cholesterol binding site.

The results from the NCI calculation show that the $\beta 1$, $\beta 2\beta 3$, H4 head, and $\Omega 1$ loop motifs are strongly coordinated with the cholesterol binding site (S136, S147, W171, R92, Y117), while weaker coordination is found with the $\beta 9$ and the $\beta 7\beta 8$ loop regions. Notably, the H4 head and $\Omega 1$ loop constitute the H4- $\Omega 1$ gate, and the $\beta 9$ and the $\beta 7\beta 8$ loop constitute the $\beta 8$ - $\beta 9$ corridor of the hydrophobic pocket. The stronger coordination of the CHL binding site by the H4- $\Omega 1$ gate than $\beta 8$ - $\beta 9$ corridor is in accordance with the structural analysis in the previous section showing that the conformational changes and interaction changes at the H4- $\Omega 1$ gate always accompany CHL translocation to the Trp-binding state (Figures 3, 5E).

Because $\beta 1$ and $\beta 2\beta 3$ are also strong coordinators of the binding site despite the long distance separating these motifs, we calculated the *normalized mutual coordination information* (NMCI) to detect the coordination channel for information transmission from $\beta 1$ to the CHL binding site. The calculated NMCI shows that $\beta 2\beta 3$, H4, and $\beta 9$, all share the information with the binding site, which identifies them as components of the *coordination channel* shown in Figure 6A. These results show how the coordination of cholesterol binding mode by $\beta 1$ is mediated through the interaction of $\beta 1$ with $\beta 2\beta 3$ and H4, which stabilizes the $\beta 9$ and H4 in the gate opening conformation, as summarized in Figure 6B.

The probability and stability of cholesterol binding in the Trp-binding mode are reduced in the W171A mutant StarD4

The release pathway of CHL in the yeast sterol transport protein Osh4 (Singh et al., 2009) was shown to comprise step-wise interactions of the CHL hydroxyl with a sequence of

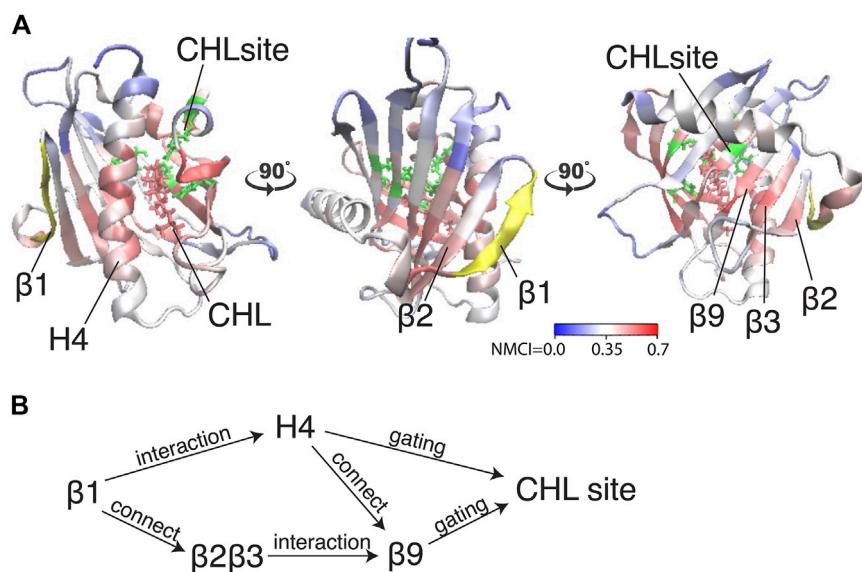
hydrophilic residues inside the hydrophobic pocket. In StarD4 we observed a similar sequence of CHL occupying binding sites along the pocket, composed of the S136&S147-site (Figure 1D), a water-mediated S136&S147-site (Figure 1E), the W171-site (Figure 1F), and a water-mediated W171, R92 & Y117 site (Figure 1G). Given the coordination between CHL translocation in the binding site and specific changes in the conformation of the StarD4 protein identified from the detailed analysis of MD simulation trajectories we report, we probed the impact of impairing such a binding site by introducing a W171A mutation. We used a Free Energy Perturbation (FEP) protocol to introduce the mutation in representative conformations of the two most populated Macrostates (Figure 5): State 1 with CHL in a Ser-binding mode, and State 3 where CHL interacts with W171 and is also involved in the water bridged interactions with R92 and Y117. From each binding mode we launched 24 replicate runs of 2,720ns/each, resulting in an ensemble of simulations totaling 131 μ s for the W171A-StarD4 mutant.

In simulations starting from the Ser-binding state, we find that the W171A mutation eliminated the translocation of CHL away from the Ser-sites, as the ligand is either in the Ser-binding mode or in the water-bridged Ser-binding mode. In the simulations started from the Trp-binding-like state where the CHL is interacting with R92 and Y117 (through a water-bridge), the W171A mutation destabilizes this binding mode and the CHL tends to translocate from the Trp-binding-like state back to the Ser-binding site. The backward transitions are observed in 8 out of the 24 replicas, with the Ser-binding conformation contributing 21% of the population in the ensemble started from the Trp-binding-like mode (Supplementary Figure S9). In comparison, stable backward transitions were observed only 2 times in the 403 μ s simulations of the WT StarD4

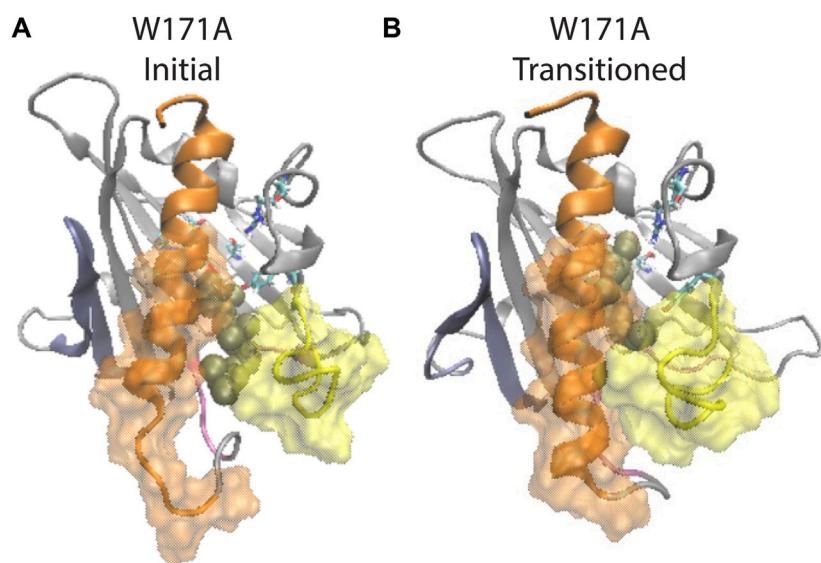
TABLE 2 Normalized coordination information between sites in cholesterol-StarD4 system.

		Transmitter:									
		$\beta 1$	$\beta 2\beta 3$	$\beta 2\beta 3$ loop	H4 head	$\Omega 1$ loop	$\beta 9$	$\beta 7\beta 8$ loop-near $\beta 9$	$\beta 7\beta 8$ loop-mid	$\beta 7\beta 8$ loop-near $\beta 6$	CHLsite
Receiver	Transmitter										
Receiver	$\beta 1$	19.48	22.1%	8.2%	16.8%	7.7%	14.0%	6.2%	5.7%	6.8%	12.2%
	$\beta 2\beta 3$	30.6%	14.68	31.1%	24.3%	13.5%	20.8%	7.7%	6.5%	7.3%	18.4%
	$\beta 2\beta 3$ loop	10.0%	35.4%	10.11	9.4%	7.0%	11.0%	4.9%	3.9%	3.8%	9.3%
	H4head	26.5%	27.9%	9.9%	16.50	10.6%	25.1%	7.2%	6.1%	5.9%	16.9%
	$\Omega 1$ loop	8.8%	12.5%	6.6%	12.7%	7.49	11.8%	6.0%	6.6%	6.4%	18.5%
	$\beta 9$	16.0%	20.6%	12.6%	22.0%	8.8%	6.45	7.1%	6.8%	6.3%	13.6%
	$\beta 7\beta 8$ loop-near $\beta 9$	4.9%	4.8%	2.7%	3.3%	3.4%	6.1%	8.90	32.9%	18.8%	10.9%
	$\beta 7\beta 8$ loop-mid	3.8%	4.6%	2.8%	3.1%	1.9%	5.3%	27.4%	13.62	38.1%	5.9%
	$\beta 7\beta 8$ loop-near $\beta 6$	4.5%	4.2%	3.4%	3.5%	1.5%	5.2%	18.4%	38.0%	9.57	8.2%
	CHLsite	35.4%	36.0%	19.7%	29.4%	29.8%	21.4%	22.6%	19.1%	31.6%	2.13

*The normalized coordination information (NCI) values are presented for NbIT, calculations with the *Transmitter* residues defining the columns acting and *Receiver* residues defining the rows. On the diagonal, the total correlation (TC) of the site is shown in gray.

**FIGURE 6**

The allosteric coordination channel between $\beta 1$ and the CHL binding site involves Helix4, and the $\beta 2\beta 3$ and $\beta 9$ sheets. **(A)** StarD4 is rendered in cartoon, with residues colored according to the normalized mutual coordination information (NMCI) value calculated with the NbtI information theory-based analysis (LeVine et al., 2014; LeVine and Weinstein, 2014). The quantification scale is at the lower right. The transmitter motif ($\beta 1$ residues 47–52) is colored in yellow, and the receiver motif (CHL binding site residues S136, S147, W171, R92 and Y117) is labeled in green cartoon on the secondary structure with side chains in green licorice. **(B)** An allosteric model of information transmission from $\beta 1$ to the cholesterol binding site.

**FIGURE 7**

$\beta 1$ and H4 conformational changes are allosterically coupled with the translocation of CHL binding in the StarD4 W171A mutant. **(A)** Starting position of the CHL in the simulation of the W171A StarD4 structural model. CHL resides in the water-bridged binding site with R92 and Y117. **(B)** Shows that the CHL has transitioned back up the hydrophobic binding site to the Ser-binding site in which the gate and corridor are closed. StarD4 is rendered in gray, with the H4 in orange, the $\Omega 1$ loop in yellow, the $\beta 182$ sheets in blue, and the $\beta 8$ loop and $\beta 9$ loop in pink. Residues S136 and S147 are rendered in "licorice" which draws the atoms as spheres and the bonds as cylinders, with oxygen colored in red, nitrogen in blue, carbons in cyan and hydrogens in white. CHL is rendered in tan color, in the VDW representation. The surface of residues 198 to 212 of H4 is rendered in transparent orange, and the surface of residues 121 to 130 on $\Omega 1$ loop is rendered in transparent yellow.

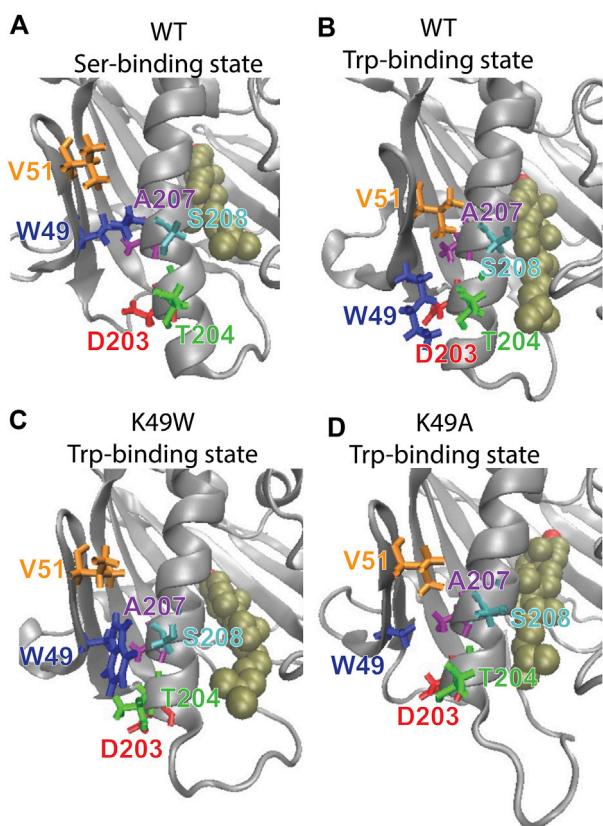


FIGURE 8

The $\beta 1$ -H4 interactions that are allosterically coupled with CHL translocation in WT are destabilized in K49 mutants. Comparison of detailed changes in H4 and $\beta 1$ residue interaction modes in (A) WT StarD4 in the initial Ser-binding state; (B) WT StarD4 after transition to the Trp-binding state; (C) K49W StarD4 after transition to the Trp-binding state; (D) K49A StarD4 after transition to the Trp-binding state. The StarD4 and CHL are rendered the same way as in Figure 3. Residues K/W/A49, V51, D203, T204, A207 and S208 are rendered in "licorice" in blue, orange, red, green, purple and cyan, respectively.

Mutations of K49 disrupt the allosteric network and slow down the dynamics of cholesterol transitions in the binding site

The mutation was chosen based on the results from analysis with the RED algorithm which revealed that the transitions of CHL in the binding site are coupled to a shift of the K49-S208 interaction to a K49-T204 configuration, and the establishment of a V51-A207 contact (Supplementary Table S1). According to the allosteric model created with NbIT, $\beta 1$ is an allosteric transmitter that coordinates the dynamics of bound CHL through its interaction with H4 and $\beta 2$. The H4- $\beta 1$ interaction involved in H4 repositioning is stabilized by both electrostatic and hydrophobic interactions. Thus, when the H4- $\beta 1$ gate is closed the K49 sidechain relocates near S208 and its hydrocarbon chain interacts with A207 (Figure 8A). When the H4- $\beta 1$ gate opens, the K49 sidechain shifts closer to T204, enabling V51 to form a contact with A207 and remove this hydrophobic residue from an unfavorable aqueous environment (Figures 8A, B). The NbIT coordination contribution analysis confirms that K49 is contributing the most to the allosteric interaction between $\beta 1$ and the CHL binding site residues. Therefore, we mutated K49 expecting changes of the $\beta 1$ -H4 interaction mode that weaken the allosteric coordination.

Mutations K49A and K49W were introduced with the same FEP protocol as above into the StarD4-CHL complex in Microstates 1 (Ser-binding mode) and 3 (Trp-binding mode). Then, simulations were carried out for 24 replicas from each state for 2,320ns per replica of the K49A-StarD4 (111 μ s total), and 2,720ns per replica for K49W-StarD4 (131 μ s total). When started from the Ser-binding mode, the CHL transitioned to the Trp-binding site in 4 out of 24 trajectories for K49W-StarD4 (containing 7% of the population, Supplementary Figure S10A), and in only 1 out of 24 trajectories for K49A-StarD4 (containing 2% of the population, Supplementary Figure S10B). These results suggest that the dynamics of CHL are significantly slower in the mutants compared to the WT StarD4 in which 6 out of 12 trajectories sampled the transition within the first 2720ns. No backward transition from a Trp-binding state to a Ser-binding state was observed in the K49A and K49W mutant constructs. These results are consistent with decreased CHL dynamics in the K49W and in the K49A mutants.

To study the conformational changes and allosteric effect, adaptive sampling was carried out to expand the sampling along the reaction pathway of the K49A mutation for another 48 μ s in 24 replicas (Supplementary Figure S10C). The impact of the mutation on the allosteric mechanism of conformational changes in StarD4 is evident in the change of contact probability between residues in $\beta 1$ and H4 as shown in Table 3.

The resulting structures, compared in Figure 8, show how the K49W and K49A mutations changed the mode of interaction between $\beta 1$ and H4. In the WT StarD4, K49 establishes both hydrophobic and hydrophilic contact with A207, S208 and T204, and its translocation opens space for the formation of a V51-A207 contact pair (Figures 8A, B). With the K49W mutation, the large W49 forms stable contacts with both S208 and T204 but hinders sterically the V51-A207 interaction (Figure 8C; Table 3). The K49A mutation, on the other hand, resulted in a decreased contact rate with both S208 and T204, eliminated the hydrophilic interactions, and destabilized the V51-A207 contact (Figure 8D; Table 3).

TABLE 3 Contact probability* between β 1 and H4 in WT, K49W and K49A StarD4 WT.

Contact probability	K49-T204 (%)	K49-S208 (%)	V51-A207 (%)
Ser-binding	13.6	68.8	14.8
Trp-binding	83.3	13.1	75.4
K49W			
Contact probability	W49-T204 (%)	W49-S208 (%)	V51-A207 (%)
Ser-binding	51.0	77.1	1.3
Trp-binding	91.4	93.2	0.3
K49A			
Contact probability	A49-T204 (%)	A49-S208 (%)	V51-A207 (%)
Ser-binding	2.8	10.9	3.2
Trp-binding	76.5	17.7	39.6

*Defined as the probability of the specified residue pairs having at least one atom in each at a distance $<3.5\text{\AA}$. The Probability is expressed as the percent occurrences of such contacts.

The interactions between β 1 and H4 are weakened overall in both mutants but the interaction modes differ. The probability of contact is decreasing gradually from WT to K49W to K49A, and the electrostatic interaction is also gradually removed in the same order. Consistent with the weakening of interaction and the decrease of CHL dynamics, the NbIT results showed a decrease in coordination information (CI values) in the allosteric network between the β 1, H4 and β 9 motifs the K49W mutant (Supplementary Table S6), and a loss of coordination effect from β 1 to the CHL site in the K49A mutant (Supplementary Table S7)—as expected from our results from the analysis of WT StarD4.

Discussion

The function of StarD4 in vectorial transport of sterols between preferred organelles is reported to be regulated by anionic lipids (Zhang et al., 2022). To learn about the molecular mechanism that underlies the StarD4 sterol transport, and the role of phosphatidylinositol phosphate (PIP) anionic lipids in regulating its function, we explored the structural and dynamic properties of the protein and the binding modes of cholesterol (CHL). Our findings show that CHL can shuttle between two modes of stable binding in the hydrophobic pocket of StarD4, interacting as shown in Figure 1 either with Ser136 and Ser147 which are located at the upper end of the hydrophobic pocket, or with Trp171, Arg92 and Tyr117 which are lower in the hydrophobic pocket and thereby place the CHL closer to the exit gate (see also Supplementary Figures S11A, B). An even lower position of the bound CHL was observed in the simulations (Figure 1G), but it is visited only rarely because it triggers incipient water penetration from the aqueous environment which leads to energetically unfavorable interactions with the hydrophobic tail of CHL. Together, the three modes of CHL binding outline an energetically favorable path for the sterol in the functional steps of uptake and release when StarD4 is embedded in the membrane environment appropriate for these functions.

Indeed, our recent findings from a combined experimental/computational study (Zhang et al., 2022) showed that StarD4 can

preferentially extract sterol from liposome membranes containing certain PIPs (especially, PI(4,5)P₂ and to a lesser degree PI(3,5)P₂), while enhancement of transport was less effective when the same PIPs were present in the acceptor membranes. Moreover, we showed that StarD4 recognizes membrane-specific PIPs through specific interaction with the geometry of the PIP headgroup as well as the surrounding membrane environment.

Our key mechanistic finding of the connection between the movement of CHL among its binding modes in the hydrophobic pocket, and the movements of specific structural motifs of StarD4, emerged from the analysis of the very long MD trajectories with the time-resolved rare event detection algorithm RED (Plante and Weinstein, 2021). These trajectory analyses showed that the connected conformational changes at the β 9 C-terminal and H4 N-terminal portions result in the opening of a gate between H4 and Ω 1-loop, and a corridor between β 9 and β 788-loop. From the time-resolved RED analysis we were able to ascertain that these gate-opening/gate-closing conformational changes occur simultaneously with the transitions between CHL binding modes, with the opening corresponding to the position of CHL in the binding pocket near Trp171, Arg92 and Tyr117.

The opening of β 8- β 9 corridor and the partial unraveling of Helix4 N-terminal head (res199-201) in the Trp-binding state make the lower end of the hydrophobic pocket accessible to the external environment. Although being observed here directly for the first time, these function-related conformational changes are consistent with results from a previous study using a membrane penetration assay to study the environment of M206 (Iaea et al., 2015). M206 is a residue located at the lower half of the hydrophobic pocket adjacent to the gate. In the crystal structure (PDBid:1JSS), it is facing inwards to the hydrophobic environment. However, in the absence of membrane, the iodoacetamide-NBD (IANBD) probe at the M206C site gave a signal similar to the L124C site immersed in water, which suggests that the M206 at the inner side of H4 is accessible to the external water environment. With the introduction of liposomes, the probe at the M206C site was found to be inserted into the nonpolar environment of the membrane. The demonstrated accessibility of M206 to the external water or membrane

environments supports the inference that the Trp-binding state with opened gates and local unraveling in H4 represent an intermediate state of StarD4 in an aqueous environment that is also likely to be an intermediate state along the CHL release/uptake pathways when StarD4 becomes embedded in a target membrane.

That conformational changes of the H4-Ω1gate are related to sterol release has long been proposed for StarD-protein family members (Murcia et al., 2006; Iaea et al., 2015; Letourneau et al., 2016; Tan et al., 2019), but the allosteric path and exact conformational changes remained unknown. While the Ω1 loop was suggested to be essential for the opening of the gate, the structural comparison analysis shown in (Letourneau et al., 2016; Tan et al., 2019) suggests as well a change in H4—as seen in our study—likely as a result of steric interaction with the CHL at the Trp-binding site.

To understand the dynamics of the mechanism observed in the RED analysis, we employed the information-theory-based analysis NbIT (LeVine et al., 2014; LeVine and Weinstein, 2014) to quantify the allosteric interactions that connect the changes in CHL binding modes to the function-related conformational changes. The results identified the paths of allosteric connectivity and showed that the gating element H4 is indeed a strong allosteric coordinator of the cholesterol binding mode dynamics. The coordination is achieved dynamically as cholesterol moves to a binding site lower in the pocket, by (1) the straightening of the H4 helix that makes space for the movement of the CHL, and (2) H4 dissociating from the interaction with the Ω1 loop which opens the gate further. These coordinated changes trigger yet more conformational rearrangement as shown in macrostates 4 and 6 (Supplementary Figures S6B, 8D). Notably, this coordinated set of CHL translocation in the binding pocket and rearrangements of structural motifs result in a conformation similar to the structures of cholesterol-binding LAM protein complexes. There, the bound sterol is located close to the tunnel entrance rather than residing at the bottom of the hydrophobic cavity, and as the tunnel entrance opens slightly at the Ω1-loop, the ligand sterol is exposed to water (Horenkamp et al., 2018; Jentsch et al., 2018; Tong et al., 2018). This similarity is illustrated in the structural superposition of StarD4 and LAM2 (Supplementary Figures S11C, D) where the location of the LAM ligand close to the entrance is shown to correspond to that of the CHL in the Trp-binding mode of StarD4, adjacent to the water network between W171, R92 and Y117 (Zhu and Weng, 2005).

Notably, the allosteric network revealed by the NbIT analysis in the StarD4-cholesterol complex shows that the allosteric coordination between the cholesterol binding and the peripheral motifs includes the β1-3 and β9 sheets. The β1 and β2 sheets are known to be essential for StarD4 membrane-binding that is mediated by their basic residues patch which interacts with anionic lipids (Iaea et al., 2015; Zhang et al., 2022). Thus, we have shown recently (Zhang et al., 2022) that upon membrane-embedding of StarD4 the β1 and β2 sheets bind PIP2 in a subtype-specific manner, which is likely to contribute to the recognition of target organelles based on the prevalent PIP2 subtype composition of their membranes. Also, membrane-embedded StarD4 adopts an orientation where the β9 and β7β8-loop motifs that are part of the allosteric network (Figure 6), are interacting with lipid heads in the membrane (Zhang et al., 2022).

Together, the results from the RED and NbIT analyses of the extensive set of long MD trajectories of cholesterol-bound

complexes of WT StarD4 and the three mutants have revealed a well-defined allosteric mechanism in CHL-bound StarD4. The conformational rearrangements determined with the RED algorithm to occur simultaneously with the relocation of the CHL from the Ser-binding mode to the Trp-binding mode in the hydrophobic pocket are proposed to constitute a preparation for CHL release into a membrane environment. The mechanism of this preparation is defined in specific detail by the NbIT-quantified allosteric network of interactions between the CHL binding site and the peripheral structural motifs involved in cholesterol translocation. As β1 was reported to be an essential membrane-binding motif (Iaea et al., 2015; Zhang et al., 2022), we propose that when the bound StarD4 interacts with the membrane, the movement of CHL between binding modes propagate through the allosteric pathway to change the conformations of the peripheral motifs of StarD4 to enable delivery. We are currently evaluating this hypothesis in membrane-embedded StarD4 systems described in Zhang et al. (2022).

Materials and methods

Modeling of the cholesterol StarD4 complex

The structure of mouse apo-StarD4 was taken from the X-ray crystal structure (PDB ID: 1jss), which includes residues 24 to 222 of the protein (Romanowski et al., 2002). Lys223 and Ala224 were introduced using Modeller 9.23 software (Sali, 1995), resulting in the conformation of mouse StarD4 22–224 segment with N-terminus acetylation. StarD4 complexes with cholesterol (CHL) were obtained by docking the ligand in the hydrophobic pocket of StarD4 using Schrodinger's Induced Fit Docking protocol (Sherman et al., 2006a; Sherman et al., 2006b). The common features shared by the top 15 docking results are: the interaction of the cholesterol hydroxyl head that forms 2 or 3 hydrogen bonds with the sidechains of Ser136 & Ser147 (Figure 1A) and with the backbone carbonyl of Cys148, and the sequestering of the cholesterol hydrocarbon tail inside the hydrophobic pocket, away from the water environment. This top docking mode was used as the initial structure for the MD simulations of the StarD4-CHL complex solvated in 0.15M K⁺Cl⁻ ionic aqueous solution (~32,600 atoms).

Long unbiased simulation of cholesterol-bound StarD4 in water

The cholesterol-bound StarD4 system was equilibrated first in MD simulations with NAMD version 2.12 (Phillips et al., 2005) using a multi-stage protocol. In the first stage the backbone of StarD4 and the heavy atoms of the ligand were harmonically constrained and gradually released in three steps of 0.5ns each, with restraining force constants changing from 1 kcal/(mol·Å²) to 0.5, and 0.1, respectively. This stage was followed by 6ns unbiased MD simulation using NAMD. In the third (production) stage, the velocities of all the atoms were reset, and the system was simulated with openMM software (Eastman et al., 2017) in 12 independent replicates, each for 33.6μs, resulting in a cumulative simulation time of 403μs (>0.4 milliseconds). The OpenMM simulations were

conducted in NPT ensemble ($T = 310\text{K}$, $p = 1\text{ atm}$) using a 4fs integration time-step and a Monte Carlo Membrane Barostat.

Mutant constructs of the StarD4-CHL complex

Mutations were introduced into the StarD4-CHL complex with the free-energy perturbation protocol in NAMD version 2.12 (Phillips et al., 2005). Accordingly, hybrid StarD4 models containing overlapped residues (e.g., overlapped K49 and A49 to introduce the K49A mutation) were constructed using representative conformations obtained in the most popular Ser-binding state 1 and Trp-binding state 3. Then, the mutations are introduced by gradual annihilation of the interactions between the WT residue and its surrounding, achieved by decreasing the coupling parameters from 1 to 0 in 10 steps, and concomitantly increasing the interactions between the mutant residue and its surrounding from 0 to 1 in 10 steps. The simulation was run for 5ns per window. After the last window, the annihilated original residues are removed from the hybrid StarD4 models, and the mutant StarD4 is equilibrated for another 10ns. For production runs, 24 replicas are made for each system and sampled in long unbiased simulation with openMM (Eastman et al., 2017).

Rare event detection (RED) protocol

Rare events in MD trajectory data were detected as described with our RED protocol (Plante and Weinstein, 2021) which utilizes a Non-negative Matrix Factorization (NMF) algorithm implemented in the Scikit-Learn python package (Fabian Pedregosa et al., 2011). The unsupervised machine learning algorithm learns to decompose a trajectory into a set of components that represent structural motifs that move simultaneously, and reports their appearance as dominant structural characteristics during a particular period in the course of the trajectory. NMF is a machine learning technique used to decompose high-dimensional non-negative data. It has been widely applied in computational biology research, where it has proven to be a powerful tool for tasks such as the molecular pattern discovery, class comparison and prediction, and functional characterization of genes as described in Plante and Weinstein (2021) and Lee and Seung (1999). Briefly summarized, the event detection method in the RED protocol uses the NMF algorithm to learn a sparse, parts-based representation of the data (Lee and Seung, 1999). For trajectory analyses the input array (I) is a contact matrix constructed as described below from the data in each frame of the trajectory. By construction, this array encodes the information about the protein structure as it evolves over the trajectory time. The NMF algorithm decomposes the I array into a “spatial” array and a “temporal” array, which together are responsible for the integrated detection of the temporally defined rare events that provide the mechanistic information. Thus, the columns of the “spatial” output matrix produced by the NMF analysis are termed “components”, as each of them represents a set of conformational features that change simultaneously even if they are in different structural motifs. As the determinant features of a conformational state evolve over trajectory time, the different

components containing them dominate the structural characteristics of the molecule at different times. Thus, the corresponding temporal array represents the contribution of each component along the trajectory, and thus identifies the time period where a particular component dominates the structural characteristics of the protein (Plante and Weinstein, 2021). A simple illustration provided in the full description of the method (Plante and Weinstein, 2021) refers to the analysis of a conformational transition in the unfolding of a fully folded alpha helical segment. The analysis results in one component that is the folded structure, and a second one that is the unfolded structure. Each frame of the simulation trajectory will be a mixing of these archetypes, with the folded component dominating the trajectory frames at the time segments before the transition, and the unfolded component dominating those after the transition.

Building the time-evolution contact map for the input array (I): A contact map is constructed between every pair of residues for each frame in the trajectory, with contact marked as 1 if any atom of one residue is $<3.5\text{\AA}$ away from any atom in another residue. For the 201 residues of StarD4 this yields a tensor of 0 and 1 values with dimensions $(t, 201, 201)$, where t represents the number of frames. Subsequently, the tensor is rearranged to a matrix of size 201^2 columns and t rows, where 201^2 is the total number of residue pairs. The matrix columns, representing all the residue pairs, are further trimmed by excluding the contact pairs that remain invariant throughout the entire simulation, which yields a I matrix with $n_{\text{pairs}} = 3,880$ columns. Each column represents a contact pair that has altered its contact state at least once during the entire simulation, encapsulating protein dynamics information for further analysis.

To focus on the conformational changes around the time of cholesterol relocation between its binding modes in the hydrophobic pocket, we excerpted 6 trajectory segments from the 6 independent runs as indicated by the time intervals (**traj0**, from 6.5 to $14.5\mu\text{s}$; **traj1**, 0.1– $4.1\mu\text{s}$; **traj3**, 0.1– $8.1\mu\text{s}$; **traj4**, 0.1– $6.5\mu\text{s}$; **traj6**, 0.1– $4.1\mu\text{s}$; **traj7**, 0.1– $4.1\mu\text{s}$; $34.4\mu\text{s}$ total). In order to obtain results that are generalizable among trajectories (which decreases the sensitivity to incidental structural fluctuations and benefits the reproducibility of RED analysis), the contact matrices of the 6 segments are inputted at the same time as the training data for the NMF algorithm. This input of multiple trajectories is constructed by concatenating the data matrices along the time-axis. This contact map is recorded every 0.8ns, which yields an I matrix with $t = 43,000$ rows. Smoothing is applied to the contact matrix across the time dimension by averaging the array in a sliding window of 20 nanoseconds (25 frames).

Details of the protocol: The algorithm decomposes the input contact matrix described above into multiple components, where each component has a structural state described by a “spatial” row array with length n_{pairs} (here $n_{\text{pairs}} = 3,880$), and a corresponding “temporal” column array of length t (here $t = 43,000$). The number of components (c) is a super-parameter chosen by the user according to the number of independent processes in the system and was set to be $c = 5$ to study generalizable features shared in independent runs. Thus, the ensemble of the “temporal” column arrays is a temporal matrix of size (t, c) in this study, and the ensemble of the “spatial” row arrays is a spatial matrix of size $(c, n_{\text{pairs}} = 3,880)$, where the multiplication of the temporal matrix and the spatial matrix is fitted to resemble the time-evolution contact map data matrix. Each row in

the $c = 5$ rows of components in the spatial matrix is an array that encodes the determinant conformational features that evolve concurrently over time. The corresponding column in the temporal matrix represents the contribution of these features in determining the conformational state of the protein at each time. It and identifies the trajectory times at which this particular set of features dominates the structural characteristics of the protein by showing it to have the highest “weight” in this period of trajectory time.

Interpretation of the results: As shown in [Figures 2E–G](#), the spatial arrays scatter plots show that there are groups of pairs that form a horizontal line, suggesting that these pairs are constantly contributing similarly to every component. These groups of pairs that are salient in every component by being in contact throughout the StarD4 simulation, and contribute similarly to each component, are used for normalization of the spatial arrays. Thus, each spatial array is divided by a normalization factor that equals to the mean value of the constructive residue pairs contributing to this particular component, thereby placing the contact features of these constitutive residue pairs at 1 in [Figures 2E–G](#); [Supplementary Figure S2B](#). To preserve the relative weight between components, each corresponding temporal array is multiplied by the same normalization factor derived from the spatial array of this particular component, so that the multiplication result of spatial and temporal matrix remains the same and resembles the input contact matrix.

This normalization allows for direct comparison between components. Subtracting the spatial array of comp A from comp B, as represented in ([Supplementary Figure S3](#)), cancels out the contributions of the constitutive residue pairs so that pairs with positive contributions are those newly established in the event corresponding to the A→B transition, whereas the negatively contributing pairs are those breaking contact during this A→B event ([Supplementary Figure S3](#)). These pairs are salient in the transition as they contribute to the structural differences between components and are termed *structure differentiating contact pairs* (SDCPs). The SDCPs that are more salient than 4 times the standard deviation, are summarized in [Supplementary Table S1](#) and are grouped based on the motif movement they represent.

Dimensionality reduction with the “time-structure based independent component analysis” (tICA) approach

The MD trajectory frames were projected onto a space of 7 collective variables described in [Supplementary Table S2](#), with the “time-structure based independent component analysis” (tICA) method. The tICA method identifies the slowest reaction coordinates of a system in the dynamic dataset ([Molgedey and Schuster, 1994](#); [Naritomi and Fuchigami, 2011](#); [Perez-Hernandez et al., 2013](#)) by solving for the generalized eigenvectors and eigenvalues of the time-lagged covariance matrix:

$$C^x(\tau)V_i = \lambda_i C^x(0)V_i$$

in which the $C^x(\tau)$ is the time-lagged covariance matrix $C^x(\tau) = \langle X(t)X^T(t+\tau) \rangle_t$ between collective variables, $C^x(0) =$

$\langle X(t)X^T(t) \rangle_t$ is the regular covariance matrix, $X(t)$ is the data vector at time t , τ is the lag-time, V_i is the i^{th} tICA eigenvector, and λ_i is the i^{th} tICA eigenvalue. In this study, the stride of the trajectories is 0.4 ns/frame, and the lag-time τ for the covariance matrix construction is 10ns (25 frames), thus ignoring any fluctuations faster than 10ns which are not relevant to the slow processes in the system. The first 1μs in every independent run are discarded to equilibrate the systems.

Construction of the Markov State Model (MSM) and assignment of macrostates

1. Construction of the Markov State Model

Markov State Model (MSM) is a powerful tool to study the kinetics of equilibrium state-transition processes in protein function or activation pathways ([Beauchamp et al., 2012](#); [Kohlhoff et al., 2014](#); [Shukla et al., 2014](#)). The implementation and validation of MSMs are well documented ([Noe and Fischer, 2008](#); [Pande et al., 2010](#); [Prinz et al., 2011](#)). To construct MSMs, the 3D conformational space of the first three tIC vectors was discretized into microstates. The transition count matrixes between the microstates were then recorded in a transition count matrix, which is then symmetrized by averaging with its transpose matrix in order to satisfy detailed balance and local equilibrium ([Beauchamp et al., 2011](#); [Prinz et al., 2011](#)). Finally, the transition probability matrixes (TPMs) were built by normalizing the transition count matrix by the transitions departing from the same microstate.

2. Selection of parameters for markov model construction

The choice of super parameters: the lag-time of the transition, and the discretization of the conformational space (i.e., number of microstates), are optimized to ensure the quality of MSM ([Beauchamp et al., 2011](#); [Prinz et al., 2011](#)). The lag-time is optimized with implied timescale test:

$$\tau_i = \frac{\tau'}{\ln(\lambda_i)}$$

in which τ' is the lag-time used for building the transition matrix, λ_i is the i^{th} eigenvector of the TPM, and τ_i is the implied timescale corresponding to the i^{th} relaxation mode of the system. In this study, the lag-time of 120 ns defined as shown in [Supplementary Figure S5A](#) ensures the best available Markovian behavior with the minimal loss of data. The best discretization of the conformational space is chosen by comparing the resulted MSM with the generalized matrix Rayleigh quotient (GMRQ) method ([Beauchamp et al., 2011](#); [McGibbon and Pande, 2015](#)), where the MSM is trained on a randomly picked training set comprised of half the set of trajectories, and then cross-validated on the test set which includes the other half of the trajectory set. The method tests how well the eigenvectors of the training set diagonalize the TPM of the test set, scored as:

$$R = \text{Tr}(V^T S T V (V^T S V)^{-1})$$

in which Tr denotes trace of the inner matrix product, V contains the first n eigenvectors of the training set, S is the diagonal matrix

composed of equilibrium populations of the test set (i.e., a diagonal matrix composed by the first eigenvector of the test set TPM), and T is the test set TPM. The best MSM parameters yield the highest GMRQ score. In this study, the 200 k-means microstates were chosen so as to provide a good balance between the consistency shown by GMRQ result (Supplementary Figure S5B) and the coverage of the 3D conformational space.

3. Interpretation of MSM, and macrostate assignment

To extract information from the MSM, TPM is decomposed into its eigenvalues and eigenvectors. The largest eigenvalue is 1, and its corresponding first eigenvector represents the equilibrium state population of the system. The remaining eigenvectors represent the population flow between microstates. The second largest eigenvalue corresponds to the eigenvector representing the population flow contributing to the slowest relaxation process of the system. In turn, the next largest eigenvalues and eigenvector correspond to the next slowest relaxation process, and so on. Here, the 200 microstates were grouped into 9 macrostates according to their kinetics similarities learnt from MSM, using the using the Robust Perron Cluster Analysis (PCCA+) algorithm (Deuflhard and Weber, 2005).

Transition path theory (TPT) analysis

The transition paths among the 9 macrostates were obtained from the TPT analysis (Berezhkovskii et al., 2009), which is based on the flux matrix between states whose components are defined as:

$$J_{ij} = \pi_i (1 - P_{target,i}) T_{ij} P_{target,j}$$

where π_i is the equilibrium population of state i , T_{ij} is the transition probability from state i to state j , and $P_{target,i}$ is the probability of visiting the target state i as the first state during the transition before visiting other states. The most probable transition pathways between specific states was obtained by iteratively finding the next pathway with the highest flux in the flux matrix using a graph theory algorithm, the Dijkstra algorithm (Dijkstra, 1959), implemented in the MSM builder software package (Beauchamp et al., 2011).

Quantifying allosteric effect using N-body information theory (NbIT)

1. NbIT analysis focused on multiple motifs in the StarD4-CHL complex

The information theory-based NbIT analysis was used to measure the coordination between distanced motifs and to quantify the allosteric effect (LeVine et al., 2014; LeVine and Weinstein, 2014). For the calculation of the coordination information we chose the following structural motifs: **CHLsite**: S136 S147 W171 R92 Y117; **β1**: res R46 V47 A48 K49 K50 V51 K52; **β2β3**: res R58 K59 P60 Y67 L68 Y69; **β2β3loop**: E62 E63 F64 N65; **H4head**: Q199 S200 A201 D203 T204 A207; **Ω1loop**:

L124 N125 I126; **β9**: D192 R194 G195; **β7β8loop-nearβ9**: V162 R163 G164; **β7β8loop-mid**: T157 R158 P159 E160; **β7β8loop-nearβ6**: E153 W154 S155 E156; **H4tail**: R218 K219 G220 L221; **β3tail**: G73 V74 M75 D76; **β8β9loop**: S179 P180 S181 Q182. The allosteric effect was then quantified from the coordination information as described in ref 32.

Briefly, the first step in the calculation of the coordination information was the alignment of the trajectories to a reference structure—we used the StarD4 crystal structure (PDB ID: 1JSS). The alignment of snapshot (per 0.4ns) was carried out on the alpha carbons of residues in β3-9 and H4, excluding the motifs have been found to undergo conformational changes (70–76, 79–86, 103–108, 113–118, 131–140, 146–150, 170–175, 184–189, 211–220). With the aligned trajectory, the entropy in each motif was calculated analytically through the differential entropy:

$$H(X) = \frac{1}{2} \ln (|2\pi e C(X)|)$$

in which $C(X)$ is the covariance matrix describing all variables in X , and X is the displacement array describing the position of every atom in the motif by its distance to the ref conformation. Then, the *Total Correlation* (TC) describing the total amount of information share in a set of motifs was calculated as:

$$TC(X_1, \dots, X_N) = \sum_i^N H(X_i) - H(X_1, \dots, X_N).$$

And the *Coordination Information* (CI) describing the amount of information shared in the receptors that is also shared with another motif that works as the *transmitter*:

$$CI(\{X_1, \dots, X_N\}, X_m) = TC(X_1, \dots, X_N) - TC(X_1, \dots, X_N | X_m)$$

Here $TC(X_1, \dots, X_N | X_m)$ is the conditional total correlation between $\{X_1, \dots, X_N\}$ conditioning on X_m :

$$TC(X_1, \dots, X_N | X_m) = \sum_i^N (H(X_i, X_m) - H(X_m)) \\ - (H(X_1, \dots, X_N, X_m) - H(X_m)).$$

The *Coordination Information* was then normalized by the *Total Correlation* of the *receiver*, to represent the portion of dynamics of the receptor that is allosterically coupled with the *transmitter*.

$$\overline{CI}(\{X_1, \dots, X_N\}, X_m) = \frac{CI(\{X_1, \dots, X_N\}, X_m)}{TC(X_1, \dots, X_N)} * 100\%.$$

In Table 2 and the Supplementary Figures S3, S5, S6, S7, the *Total Correlation* (TC) within each motif is displayed along the diagonal, while the *Normalized Coordination Information* is presented in the off-diagonal elements, with residues on the top (columns) acting as the *Transmitter* and residues on the left (rows) being the coordinated *Receiver*.

3. Coordination channel analysis based on mutual coordination information

To define the allosteric channels that mediate the coordination information (Figure 6A), we calculated the *mutual coordination information* that measures the amount of coordination information shared between the *receiver* and the *transmitter* motifs that is also shared with another structural element:

$$MCI(\{X_1, \dots, X_N\}, X_m, X_n) = CI(\{X_1, \dots, X_N\}, X_m) + CI(\{X_1, \dots, X_N\}, X_n) - CI(\{X_1, \dots, X_N\}, X_m, X_n)$$

Here X_m, X_n is the union set constituted by the residues in the *Transmitter* X_m and the *channel* X_n . The *mutual coordination information* was then normalized by the *Coordination Information* between the receiver and the *transmitter* to obtain the NMCI values used in the determination of the allosteric pathway in [Figure 6B](#)

$$NMCI(\{X_1, \dots, X_N\}, X_m, X_n) = \frac{MCI(\{X_1, \dots, X_N\}, X_m, X_n)}{CI(\{X_1, \dots, X_N\}, X_m)}.$$

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: Computational data presented in the manuscript will be managed in full accordance with the institution's Data Management and Sharing policy of Weill Cornell Medical College which is in full compliance with the NIH requirements. The raw data used to reach the inferences and conclusions are available upon reasonable request from (HW). Atomistic MD simulations were carried out with OpenMM 7.5. Computational analysis was carried out using a combination of python scripts, and in-house scripts available on GitHub <https://github.com/weinsteinlab/>.

Author contributions

Conceptualization: HW; Data curation: HW and HX; Formal analysis: HW and HX; Investigation: HX; Methodology: HW and HX; Resources: HW; Software: HX; Supervision: HW; Visualization: HX; Writing—Original draft preparation: HX, HW; Writing—Review and editing: HX, HW; Writing—Final draft review: HX and HW. All authors contributed to the article and approved the submitted version.

Acknowledgments

We gratefully acknowledge helpful discussions with Drs. Ambrose Plante, Ekaterina D. Kots, and Prof. George Khelashvili. HX gratefully acknowledges their helpful

References

Alpy, F., and Tomasetto, C. (2005). Give lipids a START: The StAR-related lipid transfer (START) domain in mammals. *J. Cell Sci.* 118, 2791–2801. doi:10.1242/jcs.02485

Beauchamp, K. A., Bowman, G. R., Lane, T. J., Maibaum, L., Haque, I. S., and Pande, V. S. (2011). MSMBuilder2: Modeling conformational dynamics at the picosecond to millisecond scale. *J. Chem. Theory Comput.* 7, 3412–3419. doi:10.1021/ct200463m

Beauchamp, K. A., McGibbon, R., Lin, Y. S., and Pande, V. S. (2012). Simple few-state models reveal hidden complexity in protein folding. *Proc. Natl. Acad. Sci. U. S. A.* 109, 17807–17813. doi:10.1073/pnas.1201810109

Berezhkovskii, A., Hummer, G., and Szabo, A. (2009). Reactive flux and folding pathways in network models of coarse-grained protein dynamics. *J. Chem. Phys.* 130, 205102. doi:10.1063/1.3139063

Clark, B. J. (2020). The START-domain proteins in intracellular lipid transport and beyond. *Mol. Cell Endocrinol.* 504, 110704. doi:10.1016/j.mce.2020.110704

Deuffhard, P., and Weber, M. (2005). Robust Perron cluster analysis in conformation dynamics linear algebra and its applications. *Linear Algebra its Appl.* 398, 161–184. doi:10.1016/j.laa.2004.10.026

Dijkstra, E. W. (1959). A note on two problems in connexion with graphs. *Numer. Math.* 1, 269–271. doi:10.1007/bf01386390

Eastman, P., Swails, J., Chodera, J. D., McGibbon, R. T., Zhao, Y., Beauchamp, K. A., et al. (2017). OpenMM 7: Rapid development of high performance algorithms for molecular dynamics. *PLoS Comput. Biol.* 13, e1005659. doi:10.1371/journal.pcbi.1005659

guidance regarding the principles, applicability, implementation, and interpretation of the RED, NbIT and TPT analysis. We are grateful for discussions with Prof. Frederick R. Maxfield and his lab members. Support from the 1923 Foundation for the project “How Needed Molecular Precision is Achieved for Addressing, Pickup, and Delivery in the Trafficking of Cholesterol Among Cell Membranes” is gratefully acknowledged. The computational resources and technical help at the Center for Computational Innovations (CCI) at the Rensselaer Polytechnic Institute (RPI), and the efficient and sustained access to the AiMOS supercomputer at CCI generously awarded through the COVID-19 High Performance Computing Consortium, are gratefully acknowledged. We are grateful for the computational resources under Projects BIP225 and BIP109 at the Oak Ridge Leadership Computing Facility, which is a DOE Office of Science User Facility supported under Contract DE-AC05-00OR22725, and for the in-house computational resources of the David A. Cofrin Center for Biomedical Information in the Institute for Computational Biomedicine at Weill Cornell Medical College.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fmolb.2023.1197154/full#supplementary-material>

Elbadawy, H. M., Borthwick, F., Wright, C., Martin, P. E., and Graham, A. (2011). Cytosolic StAR-related lipid transfer domain 4 (STARD4) protein influences keratinocyte lipid phenotype and differentiation status. *Br. J. Dermatol.* 164, 628–632. doi:10.1111/j.1365-2133.2010.10102.x

Fabian Pedregosa, G. V., Gramfort, A., Michel, V., Bertrand, T., Grisel, O., and Blondel, M. (2011). Scikit-learn. *Mach. Learn. Python JMLR* 12, 2825–2830. doi:10.5555/1953048.2078195

Garbarino, J., Pan, M., Chin, H. F., Lund, F. W., Maxfield, F. R., and andBreslow, J. L. (2012). STARD4 knockdown in HepG2 cells disrupts cholesterol trafficking associated with the plasma membrane, ER, and ERC. *J. Lipid Res.* 53, 2716–2725. doi:10.1194/jlr.M032227

Hao, M., Lin, S. X., Karylowski, O. J., Wustner, D., McGraw, T. E., and Maxfield, F. R. (2002). Vesicular and non-vesicular sterol transport in living cells. The endocytic recycling compartment is a major sterol storage organelle. *J. Biol. Chem.* 277, 609–617. doi:10.1074/jbc.M108861200

Horenkamp, F. A., Valverde, D. P., Nunnari, J., and Reinisch, K. M. (2018). Molecular basis for sterol transport by START-like lipid transfer domains. *EMBO J.* 37, e98002. doi:10.1525/embj.201798002

Iaea, D. B., Dikiy, I., Kiburu, I., Eliezer, D., and Maxfield, F. R. (2015). STARD4 membrane interactions and sterol binding. *STARD4 Membr. Interact. Sterol Bind. Biochem.* 54, 4623–4636. doi:10.1021/acs.biochem.5b00618

Iaea, D. B., Mao, S., Lund, F. W., and Maxfield, F. R. (2017). Role of STARD4 in sterol transport between the endocytic recycling compartment and the plasma membrane. *Mol. Biol. Cell* 28, 1111–1122. doi:10.1091/mbc.E16-07-0499

Iaea, D. B., and Maxfield, F. R. (2015). Cholesterol trafficking and distribution. *Biochem.* 57, 43–55. doi:10.1042/bse0570043

Iaea, D. B., Spahr, Z. R., Singh, R. K., Chan, R. B., Zhou, B., Bareja, R., et al. (2020). Stable reduction of STARD4 alters cholesterol regulation and lipid homeostasis. *Biochim. Biophys. Acta Mol. Cell Biol. Lipids* 1865, 158609. doi:10.1016/j.bbapli.2020.158609

Jentsch, J. A., Kiburu, I., Pandey, K., Timme, M., Ramlall, T., Levkau, B., et al. (2018). Structural basis of sterol binding and transport by a yeast STARkin domain. *J. Biol. Chem.* 293, 5522–5531. doi:10.1074/jbc.RA118.001881

Kohlhoff, K. J., Shukla, D., Lawrenz, M., Bowman, G. R., Konerding, D. E., Belov, D., et al. (2014). Cloud-based simulations on Google Exacycle reveal ligand modulation of GPCR activation pathways. *Nat. Chem.* 6, 15–21. doi:10.1038/nchem.1821

Lee, D. D., and Seung, H. S. (1999). Learning the parts of objects by non-negative matrix factorization. *Nature* 401, 788–791. doi:10.1038/44565

Letourneau, D., Bedard, M., Cabana, J., Lefebvre, A., LeHoux, J. G., and Lavigne, P. (2016). STARD6 on steroids: Solution structure, multiple timescale backbone dynamics and ligand binding mechanism. *Sci. Rep.* 6, 28486. doi:10.1038/srep28486

Letourneau, D., Lefebvre, A., Lavigne, P., and LeHoux, J. G. (2015). The binding site specificity of STARD4 subfamily: Breaking the cholesterol paradigm. *Mol. Cell Endocrinol.* 408, 53–61. doi:10.1016/j.mce.2014.12.016

LeVine, M., Perez-Aguilar, J. M., and Weinstein, H. (2014). N-Body information theory (NbIT) analysis of rigid-body dynamics in intracellular loop 2 of the 5-HT 2A receptor. *Proceedings International Work-Conference on Bioinformatics and Biomedical Engineering (IWBBIO- 2014)* 2, 1190–1200.

LeVine, M. V., and Weinstein, H. (2014). NbIT--a new information theory-based analysis of allosteric mechanisms reveals residues that underlie function in the leucine transporter LeuT. *PLoS Comput. Biol.* 10, e1003603. doi:10.1371/journal.pcbi.1003603

Liscum, L., and Munn, N. J. (1999). Intracellular cholesterol transport. *Biochim. Biophys. Acta* 1438, 19–37. doi:10.1016/s1388-1981(99)00043-8

Mathieu, A. P., Fleury, A., Ducharme, L., Lavigne, P., and LeHoux, J. G. (2002). Insights into steroidogenic acute regulatory protein (StAR)-dependent cholesterol transfer in mitochondria: Evidence from molecular modeling and structure-based thermodynamics supporting the existence of partially unfolded states of StAR. *J. Mol. Endocrinol.* 29, 327–345. doi:10.1677/jme.0.0290327

McGibbon, R. T., and andPande, V. S. (2015). Variational cross-validation of slow dynamical modes in molecular kinetics. *J. Chem. Phys.* 142, 124105. doi:10.1063/1.4916292

Mesmin, B., Pipalia, N. H., Lund, F. W., Ramlall, T. F., Sokolov, A., Eliezer, D., et al. (2011). STARD4 abundance regulates sterol transport and sensing. *Mol. Biol. Cell* 22, 4004–4015. doi:10.1091/mbc.E11-04-0372

Molgedey, L., and Schuster, H. G. (1994). Separation of a mixture of independent signals using time delayed correlations. *Phys. Rev. Lett.* 72, 3634–3637. doi:10.1103/PhysRevLett.72.3634

Murcia, M., Faraldo-Gomez, J. D., Maxfield, F. R., and Roux, B. (2006). Modeling the structure of the START domains of MLN64 and StAR proteins in complex with cholesterol. *J. Lipid Res.* 47, 2614–2630. doi:10.1194/jlr.M600232-JLR200

Naritomi, Y., and Fuchigami, S. (2011). Slow dynamics in protein fluctuations revealed by time-structure based independent component analysis: The case of domain motions. *J. Chem. Phys.* 134, 065101. doi:10.1063/1.3554380

Noe, F., and andFischer, S. (2008). Transition networks for modeling the kinetics of conformational change in macromolecules. *Curr. Opin. Struct. Biol.* 18, 154–162. doi:10.1016/j.sbi.2008.01.008

Pande, V. S., Beauchamp, K., and andBowman, G. R. (2010). Everything you wanted to know about Markov State Models but were afraid to ask. *Methods* 52, 99–105. doi:10.1016/j.ymeth.2010.06.002

Perez-Hernandez, G., Paul, F., Giorgino, T., De Fabritiis, G., and Noe, F. (2013). Identification of slow molecular order parameters for Markov model construction. *J. Chem. Phys.* 139, 015102. doi:10.1063/1.4811489

Phillips, J. C., Braun, R., Wang, W., Gumbart, J., Tajkhorshid, E., Villa, E., et al. (2005). Scalable molecular dynamics with NAMD. *NAMD J. Comput. Chem.* 26, 1781–1802. doi:10.1002/jcc.20289

Plante, A., and andWeinstein, H. (2021). Ligand-dependent conformational transitions in molecular dynamics trajectories of GPCRs revealed by a new machine learning rare event detection protocol. *Molecules* 26, 3059. doi:10.3390/molecules26103059

Prinz, J. H., Wu, H., Sarich, M., Keller, B., Senne, M., Held, M., et al. (2011). Markov models of molecular kinetics: Generation and validation. *J. Chem. Phys.* 134, 174105. doi:10.1063/1.3565032

Prinz, W. A. (2007). Non-vesicular sterol transport in cells. *Prog. Lipid Res.* 46, 297–314. doi:10.1016/j.plipres.2007.06.002

Radhakrishnan, A., Goldstein, J. L., McDonald, J. G., and Brown, M. S. (2008). Switch-like control of SREBP-2 transport triggered by small changes in ER cholesterol: A delicate balance. *Cell Metab.* 8, 512–521. doi:10.1016/j.cmet.2008.10.008

Rodriguez-Agudo, D., Calderon-Dominguez, M., Ren, S., Marques, D., Redford, K., Medina-Torres, M. A., et al. (2011). Subcellular localization and regulation of StarD4 protein in macrophages and fibroblasts. *Biochim. Biophys. Acta* 1811, 597–606. doi:10.1016/j.bbapli.2011.06.028

Rodriguez-Agudo, D., Ren, S., Wong, E., Marques, D., Redford, K., Gil, G., et al. (2008). Intracellular cholesterol transporter StarD4 binds free cholesterol and increases cholestryler ester formation. *J. Lipid Res.* 49, 1409–1419. doi:10.1194/jlr.M700537-JLR200

Romanowski, M. J., Soccio, R. E., Breslow, J. L., and Burley, S. K. (2002). Crystal structure of the *Mus musculus* cholesterol-regulated START protein 4 (StarD4) containing a StAR-related lipid transfer domain. *Proc. Natl. Acad. Sci. U. S. A.* 99, 6949–6954. doi:10.1073/pnas.052140699

Sali, A. (1995). Comparative protein modeling by satisfaction of spatial restraints. *Mol. Med. Today* 1, 270–277. doi:10.1016/s1357-4310(95)91170-7

Sherman, W., Beard, H. S., and Farid, R. (2006). Use of an induced fit receptor structure in virtual screening. *Screen. Chem. Biol. Drug Des.* 67, 83–84. doi:10.1111/j.1747-0285.2005.00327.x

Sherman, W., Day, T., Jacobson, M. P., Friesner, R. A., and Farid, R. (2006). Novel procedure for modeling ligand/receptor induced fit effects. *J. Med. Chem.* 49, 534–553. doi:10.1021/jm050540c

Shukla, D., Meng, Y., Roux, B., and Pande, V. S. (2014). Activation pathway of Src kinase reveals intermediate states as targets for drug design. *Nat. Commun.* 5, 3397. doi:10.1038/ncomms4397

Singh, R. P., Brooks, B. R., and Klauda, J. B. (2009). Binding and release of cholesterol in the Osh4 protein of yeast. *Proteins* 75, 468–477. doi:10.1002/prot.22263

Soccio, R. E., Adams, R. M., Romanowski, M. J., Sehayek, E., Burley, S. K., and Breslow, J. L. (2002). The cholesterol-regulated StarD4 gene encodes a StAR-related lipid transfer protein with two closely related homologues, StarD5 and StarD6. *StarD5 StarD6 Proc Natl Acad Sci U. S. A.* 99, 6943–6948. doi:10.1073/pnas.052143799

Tan, L., Tong, J., Chun, C., and Im, Y. J. (2019). Structural analysis of human sterol transfer protein STARD4. *Biochem. Biophys. Res. Commun.* 520, 466–472. doi:10.1016/j.bbrc.2019.10.054

Thorsell, A. G., Lee, W. H., Persson, C., Siponen, M. I., Nilsson, M., Busam, R. D., et al. (2011). Comparative structural analysis of lipid binding START domains. *PLoS One* 6, e19521. doi:10.1371/journal.pone.0019521

Tong, J., Manik, M. K., and Im, Y. J. (2018). Structural basis of sterol recognition and nonvesicular transport by lipid transfer proteins anchored at membrane contact sites. *Proc. Natl. Acad. Sci. U. S. A.* 115, E856–E865. doi:10.1073/pnas.1719709115

Tsujishita, Y., and andHurley, J. H. (2000). Structure and lipid transport mechanism of a StAR-related domain. *Nat. Struct. Biol.* 7, 408–414. doi:10.1038/75192

Wong, L. H., Gatta, A. T., and Levine, T. P. (2019). Lipid transfer proteins: The lipid commute via shuttles, bridges and tubes. *Nat. Rev. Mol. Cell Biol.* 20, 85–101. doi:10.1038/s41580-018-0071-5

Zhang, X., Xie, H., Iaea, D., Khelashvili, G., Weinstein, H., and Maxfield, F. R. (2022). Phosphatidylinositol phosphates modulate interactions between the StarD4 sterol trafficking protein and lipid membranes. *J. Biol. Chem.* 298, 102058. doi:10.1016/j.jbc.2022.102058

Zhu, J., and Weng, Z. (2005). Fast: A novel protein structure alignment algorithm. *Proteins* 58, 618–627. doi:10.1002/prot.20331