Bifurcated Hydrogen Bonds and the Fold Switching of Lymphotactin

**Prabir Khatua**, **Alan J Ray**, **Ulrich H. E. Hansmann**

Dept. of Chemistry & Biochemistry, University of Oklahoma, Norman, OK 73019, USA

Abstract

Lymphotactin (Ltn) exists under physiological conditions in an equilibrium between two interconverting structures with distinct biological functions. Using Replica-Exchange-with-Tunneling we study the conversion between the two folds. Unlike previously proposed, we find that the fold switching does not require unfolding of Lymphotactin, but proceeds through a series of intermediates that remain partially structured. This process relies on two bifurcated hydrogen bonds that connect the β_2 and β_3 strands and ease the transition between the hydrogen bond pattern by which the central three-stranded β-sheet in the two forms differs.

Graphical Abstract
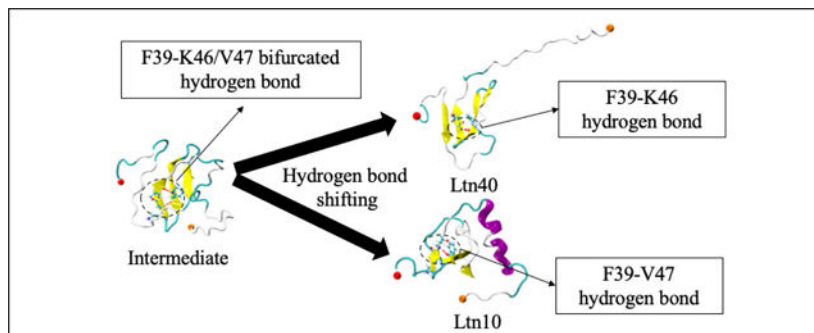
1   INTRODUCTION

Exhibiting a diverse spectrum of functions, ranging from transport of molecules to catalysis of biochemical reactions, proteins play a crucial role in the molecular machinery of cells. Protein function is determined by the three-dimensional structure. In the now classical model of protein folding the sequence of amino acids encodes an energy landscape that funnels folding pathways into a unique native state.[1,2] While this mechanism describes folding of many proteins, it needs to be modified for intrinsically disordered[3,4] or metamorphic proteins[5,6] where the sequence encodes not only a single native fold but an ensemble of structures, allowing a single protein to take multiple functions. Hence, for

uhansmann@ou.edu.

understanding the role of intrinsically disordered and metamorphic proteins in cells, and the regulation of their function, it is necessary to establish the underlying multi-funnel energy landscape that leads to this ensemble of diverse structures, and to comprehend the mechanism by which these structures convert into each other.

This task can be most easily tackled for metamorphic proteins such as the 93-residue protein Lymphotactin (Ltn) which are observed in two well-defined structures. As a chemokine Ltn belongs to a family of signaling proteins whose primary function is to direct immune response leukocytes toward areas of inflammation.[7,8] Ltn has only one of the two disulfide bonds otherwise found in chemokines, allowing it to adopt and switch between two well-defined native folds with distinct functions. The first one (Ltn10) is a typical chemokine-like fold with a three-stranded β-sheet attached to a C-terminal α-helix (PDB ID: 2HDM;[9] see Figure 1(a)). When in its second form, Ltn40, the protein forms dimers and has a four-stranded β-sheet in a dimeric β-sandwich (PDB ID: 2JP1;[10] see Figures 1(b-c)). The two forms have distinct and complementary functions: Ltn10 activates the XCR1 receptor on the cell surface, while Ltn40 binds to heparin, a polysaccharide component of the extracellular matrix.[5,11] By being able to assume both motifs and perform disparate and complementing functions, Lymphotactin bridges the two main functional aspects of chemokine physiology: activation of specific G-protein-coupled receptors (GPCRs) leading to chemotaxis,[12] and establishing a signaling gradient toward the target location by binding with extracellular matrix glycosaminoglycans (GAGs). Under physiological conditions, both forms rapidly interconvert and are equally populated. However, the presence of several basic amino acids (nine Arg and six Lys residues) makes Lymphotactin sensitive to solution conditions and temperature. For example, Ltn10 is the dominant conformation at lower temperature (10°C) and high salt concentration (200 mM NaCl), while the alternative fold Ltn40 is dominant at 40°C and no salt. This shift of the equilibrium with temperature and ionic strength was investigated in CHARMM simulations[13,14] in which an accumulation of sodium ions around the charged residues of the helical region increased with decreasing temperature. These computational results are in agreement with experimental observations that high salt concentration and low temperature stabilize the chemokine fold.[15]

While the structure and function of the two Lymphotactin motifs have been studied in detail,[9–11,15,16] the mechanism of interconversion is still unclear. Unlike other metamorphic proteins such as mitotic arrest deficiency 2 protein (Mad2),[17–19] the fold switch requires a complete reorganization of core residues[5] making the conversion especially difficult to study in experiments. Volkman's group[20] examined the kinetic rates of the process by stopped-flow fluorescence. Their results suggest that the conversion process involves large-scale unfolding with a disruption of all stabilizing hydrogen bonds. However, such a mechanism is difficult to reconcile with the surprisingly low barrier separating Ltn10 and Ltn40 that rather suggests a conversion pathway going through intermediates with conserved local contacts (such as in three β-strands β_1, β_2 and β_3) found in both folds and only encountering minimal disruptions of bonds. Unfortunately, such crucial but transient intermediates are hard to resolve on the short-time scales by which Ltn10 and Ltn40 convert into each other, and may have been below the temporal resolution of the Volkman's experiments.[20] On the other hand, the interconversion time scales are still too long to be studied with sufficient statistics in constant temperature all-atom molecular dynamics simulations. Computational studies of

fold switching have to rely instead on either enhanced sampling techniques[21,22] or structure-based models (also called Go-models).[23] They were used, for instance, to study the fold switching in the transcription factor RfaH.[22,23] A two-funneled Go-model was also used in a recent computational study of Camilloni and Ludovico[24] to probe the fold-switching of Lymphotactin. While this study reported the presence of structured intermediates for the fold-switching, it is not clear how far the presence of the intermediates and the barrier heights reflect the details of the construction of this model rather than the physics of the system.

In this study, we aim to resolve the experimental discrepancy by studying the Lymphotactin conversion process in all-atom simulations relying on a physical force field. As it is difficult to obtain from regular molecular dynamics sufficient statistics, we utilize an enhanced sampling technique developed in our lab, that was designed specifically for the investigation of such switching processes. Our technique, Replica-Exchange-with-Tunneling (RET), [21,22,25,26] allows us to observe the interconversion process with sufficient detail to characterize important intermediates. This in turn enables us to propose a conversion mechanism that is consistent with the experimentally observed low barrier separating Ltn10 and Ltn40. While the experimentally observed equilibrium is between Ltn10 monomers and Ltn40 homodimers, dimerization and conversion are separate processes, with the transition between Ltn10 monomers and Ltn40 monomers being the rate-limiting process.[20] For this reason, we consider only the conversion of monomers, but one should keep in mind that the Le Chatelier's principle would imply a shift of the equilibrium between Ltn10 and Ltn40 monomers toward the Ltn40 form if also the subsequent dimerization is considered. Our analysis indicates that the fold switch in Lymphotactin monomers occurs along a series of only partially unfolded intermediates, with the breaking and reformation of secondary structure relying on the presence of two bifurcated backbone hydrogen bonds that connect the β_2 and β_3 strands found in both motifs. We conjecture that these bifurcated hydrogen bonds are essential for fold switching, as they allow a repositioning of the β-strand forming residues without the need to cross high energy barriers.

2 MATERIALS AND METHODS

2.1 Replica-Exchange-with-Tunneling

In order to understand the mechanism of fold switching in Lymphotactin by way of computer simulations, one has to sample the free energy landscape of the protein with high accuracy. However, the accessible time scales in all-atom molecular dynamics simulations in explicit solvent are even for small proteins with less than 100 residues only of order $\approx \mu s$, and therefore insufficient to obtain sufficient statistics. We have proposed in earlier work[21,22,25,26] a variant of the Hamilton Replica Exchange method[27,28] as a way to overcome this sampling problem in studies of conformational transitions. Our approach relies on two ingredients. First, a ladder of replicas is set up, where on each replica a "physical" model is coupled with a structure-based model. On one side of the ladder the structure-based model biases the physical system toward the Ltn10 state, on the other side toward Ltn40. The strength of the coupling (biasing) on each replica is controlled by a parameter λ which is maximal at the two ends, and zero for the central replica, where the

physical model is therefore not biased by one of the structure-based models. Exchange moves between neighboring replicas induce a walk along the ladder by which the Lymphotactin configuration changes from one motif into the other. When these exchange moves are accepted or rejected with the criterium commonly used in Replica Exchange Sampling, the correct distribution according to the given λ-value will be sampled on each replica. Hence, on the central replica, at $\lambda = 0$, the correct and unbiased distribution of the physical model of Lymphotactin will be sampled. While formally correct, this sometimes also called Multi Scale Essential Sampling (MSES)[29,30] is of limited use in studies of large systems, as the acceptance probability for exchange moves becomes vanishingly small. Hence, the second ingredient of our approach is to replace the canonical acceptance criterium with a new one that allows the system to "tunnel" through the unfavorable "transition state" generated by the exchange move. This tunneling is achieved by re-scaling the velocities of atoms in the two configurations in such a way that the total energy at a given λ value is the same before and after the exchange. The two replicas evolve then by microcanonical molecular dynamics, exchanging potential and kinetic energy, symbolized by the gradual color change in the schematic diagram of Figure 2. After a short time (a few ps) the velocity distribution of each of the two replicas will approach the one that would be expected at the given temperature. At this point, the potential energies of the two configurations are compared with the corresponding energies before the exchange move, and either accepted or rejected. We have coined this approach as Replica-Exchange-with-Tunneling (RET), and described it and its limitations in detail in our earlier works.[21,25] In previous work,[21,22,26] we could show that our above described approach leads indeed to an enhanced sampling of transition events and an improved statistics in the sampled energy landscape (which is generated from the unbiased central replica where $\lambda = 0$). Being specifically designed for simulating conversions between known structures, RET proved to be more efficient than other enhanced sampling technique in our previous studies,[21,22,26] however, the improvement is a quantitative one, not a qualitative one. As in other generalized ensemble techniques[31,32] the system evolves in a RET simulation by an artificial kinetics which does not necessarily lead to physical trajectories, see also our discussion in the result section. Hence, kinetic information has to be obtained in an indirect way by extracting it from the free energy landscape by transition path theory,[33] Markov state model (MSM) analysis[34–36] or similar approaches. For instance, upper bounds on the transition rates can in principle be deduced from the height of the free energy barriers using Kramers' theory as proposed in Ref.[37]

2.2 Simulation Set-up

In the present work we use our enhanced sampling method to study the conversion between Ltn10 and Ltn40 monomers. For this purpose, we simulate our system with an energy function

$$E_{pot} = E_{phy}(q_{phy}) + E_{go}(q_{go}) + \lambda E_{\lambda}(q_{phy}, q_{go}) \tag{1}$$

where E_{pot} is the total potential energy of the system, $E_{phy}(q_{phy})$ and $E_{go}(q_{go})$ are the potential energies of the physical model and Go-model, respectively, and $E_{\lambda}(q_{phy}, q_{go})$ describes the coupling between the two models.

Interactions in the physical model are described by the CHARMM36m force field[38] in combination with TIP3P explicit solvent,[39] with an acetyl group cap on the N-terminus and a methylamine group cap for the C-terminus. The protein is then solvated with 9639 water molecules in a cubic box of length 67.5 Å. Each system is neutralized by adding 8 chloride (Cl⁻) ions. By choosing the box size comparable to the end-to-end distance of a polymer in a good solvent, we try to allow for the possibility that conversion between Ltn10 and Ltn40 requires unfolding of the protein (as was proposed in earlier work[20]), while at the same time minimizing computational costs. Re-scaling the masses of the all-atom physical models by 14.49 is required to match the temperature scales of the two models as the Go-models do not include hydrogen atoms, i.e., have a smaller number of degrees of freedom. The initial configurations (taken from the Protein Data Bank) are randomized for 1 ns in high temperature molecular dynamics simulations at 1500 K. After visual inspection to exclude unphysical geometries, we have cooled the system down to the target temperature of 310 K by performing an additional simulation of 1 ns duration. The resulting configuration has been further minimized to generate the start configuration for the actual RET run.

The two structure-based (Go-)models (one biasing toward Ltn10, the other toward Ltn40) were generated using the SMOG-Server[40] at http://smog-server.org. While the wild-type form of Lymphotactin consists of 93 residues, most of the C-terminus tail is disordered in both conformations, 30% for Ltn10 and 42% for Ltn40.[10] As the set-up of a structure-based energy function is meaningless for such unstructured regions, we have not considered this tail. Instead, we have restricted our simulations to a 75 residue fragment which describes the parts of Lymphotactin that are structured in at least one of the two folds. The length of this fragment matches the one of the NMR resolved Ltn10 forms (PDB ID: 2HDM[9]), but is larger than the Ltn40 form (PDB ID: 2JP1[10]) for which only 60 residues are resolved. For the generation of the SMOG parameter the remaining 15 residues were assumed to be in a random configuration and added using the PyMOL[41] mutagenesis tool.

The biasing energy is defined as[29,30]

$$
E_\lambda = \begin{cases}
\frac{1}{2}\left(\Delta^2(i,j)\right) & -ds < \Delta(i,j) < ds \\[2mm]
A + \dfrac{B}{\Delta^S(i,j)} + f_{max}\Delta(i,j) & \Delta(i,j) > ds \\[2mm]
A + \dfrac{B}{\Delta^S(i,j)}(-1)^S - f_{max}\Delta(i,j) & \Delta(i,j) < -ds
\end{cases}
\tag{2}
$$

where ds marks the region in which E_λ is a quadratic function of $\Delta(ij) = \delta_{phy}(ij) - \delta_{go}(ij)$, i.e., of the difference of the distances between C_α-atoms i and j as measured in the two models. Guided by previous work we chose $ds = 0.3$ nm. The control parameter f_{max} sets the maximum force when $\Delta(i,j) \rightarrow \infty$, and S determines how fast this value is realized. As discussed in earlier studies,[29,30] it is convenient to choose $S = 1$ and $f_{max} = 0$. The parameter ds determines the region in which the E_λ is assumed to be a quadratic function. The parameters A and B ensure continuity of E_λ and its first derivative at $(i,j) = \pm ds$, and are given by

$$A = \left(\frac{1}{2} + \frac{1}{S}\right)ds^2 - \left(\frac{1}{S} + 1\right)f_{max}ds \quad \text{and} \quad B = \left(\frac{f_{max} - ds}{S}\right)ds^{S+1}. \tag{3}$$

The 24 replica systems were prepared with a λ distribution of $\lambda = 0.1$, 0.09, 0.08, 0.07, 0.06, 0.05, 0.04, 0.03, 0.02, 0.01, 0.009, 0, 0, 0.009, 0.01, 0.02, 0.03, 0.04, 0.05, 0.06, 0.07, 0.08, 0.09, 0.1. Here the E_λ term biases replica 0–10 toward the Ltn10 motif, and replica 13–23 toward Ltn40. In order to simplify our programming, we use two replicas (with indices 11 and 12) to represent the case where the physical model is not biased by any structure-based biasing-term, i.e., where $\lambda = 0$. One of the two replicas exchanges configurations with the neighboring replica in the Ltn10 branch, and the other with the neighbor replica of the Ltn40 branch. Since at $\lambda = 0$ physical and Go-model are independent, the configuration of the physical model can be easily exchanged between the two replicas.

Our simulations rely on an in-house implementation of the above described approach into the Gromacs 4.6.5 package[42] that is available from the authors and in Github (github.com/orgs/hansmannlab). The equations of motion are integrated with the Velocity Verlet algorithm,[43] with hydrogen bonds constrained by the LINCS algorithm,[44] using a time step of 2 fs. The van der Waals and electrostatic cutoffs are set to 1.2 nm. Note that instead of simulating each of the replicas at a constant temperature, the temperature of the replicas is changed in steps of 0.01 K between 310 and 310.23 K due to the way our code has been implemented into Gromacs, where the temperature of the replicas is maintained by using the v-rescale thermostat.[45] Choosing the length of the microcanonical segment in the RET move as 1 ps, a 100 ns trajectory was generated. Thermodynamic quantities were calculated solely from replicas with $\lambda = 0$, i.e., without bias from the two structure-based models.

2.3 Observables

The free energy as a function of order parameter, λ, is defined as

$$\Delta G(\lambda) = -k_B T[\ln\rho(\lambda) - \ln\rho_{max}], \tag{4}$$

where k_B and T denote the Boltzmann's constant and temperature, respectively. ρ is an estimate of the probability density function calculated from a histogram of the data, while ρ_{max} is the maximum of the density. The second term ensures that $G = 0$ for the lowest free energy minimum. Free energy values reported in this work are calculated at a temperature of 310 K.

The configurations as obtained from the equilibrated trajectories that correspond to unbiased trajectories ($\lambda = 0$) are characterized based on secondary structure pattern. The pattern is calculated using the STRIDE algorithm[46] as implemented in VMD software[47] and characteristic backbone hydrogen bonding pattern specific to each of the Ltn native forms (see Table 1). Ltn has four characteristic β-strands, β_0 (residue 10–15), β_1 (residue 25–31), β_2 (residue 34–41), β_3 (residue 44–51) and a C-terminal helix, H (residue 54–66). Based on the secondary structural pattern of these five regions and backbone hydrogen bonding pattern in $\beta_1 - \beta_2 - \beta_3$ region, we define three variables (B0, H, and B123) using following

logical expression, which eventually helps us characterize the configurations and distinguish the pattern among them.

- **B**0: If there are at least two residues in β-strand in β_0 region, then the value of B0 is 1 i.e., β_0 exists; otherwise it is zero.

- **H**: If there are at least three residues in helix in C-terminal helix region, then the value of H is 1 i.e., C-terminal helix exists; otherwise it is zero.

- **B**123: If there is no characteristic hydrogen bond (Ltn10-like or Ltn40-like) in $\beta_1 - \beta_2 - \beta_3$ region and there is less than two residues in β-strand in $\beta_1 - \beta_2$ or $\beta_2 - \beta_3$ regions, B123 will be assigned to have a value of zero. If this condition is not satisfied, B123 will have hydrogen bonding pattern as that of Ltn10-like or Ltn40-like or Mixed (i.e., substantial existence of both type of hydrogen bonds) depending on the conditions: if the difference in number of characteristic Ltn10-like and Ltn40-like hydrogen bonds (see Table 1) is greater than or equal to two, B123 is Ltn10-like. Similarly, if the difference in number of characteristic Ltn40-like and Ltn10-like hydrogen bonds is greater than or equal to two, B123 is Ltn40-like. All other remaining configurations will be considered to have mixed hydrogen bonding pattern.

One quantity by which we have measured similarity of a given configuration to one of the two native Lymphotactin folds is the fraction of specific native contacts $Q_{spec, 1}(X)$, defined

$$Q_{spec, 1}(X) = \frac{1}{N} \sum_{(i, j)} \frac{1}{1 + \exp\left[\beta\left(r_{ij}(X) - \min\left(\lambda r_{ij}^1, r_{ij}^2\right)\right)\right]} \tag{5}$$

Here, 1 denotes one native fold of Lymphotactin (Ltn10 or Ltn40), while 2 represents the alternative one. Only contacts specific to either of the native folds are considered, i.e., contacts that are found in both native folds are excluded. Here, we define a contact in the native structure by the requirement that two backbone atoms on distinct residues are within 4.5 Å. Thus, N is the number of such contact pairs (specific to the one form of Lymphotactin native structure) of (i,j) backbone atoms i and j belonging to residues θ_i and θ_j. To avoid the contacts forming by adjacent residues, we have considered only the residues where $|\theta_i - \theta_j| > 3$. $r_{ij}(X)$ is the distance between the atoms i and j in conformation X, while r_{ij}^1, r_{ij}^2 are that distance in the native fold 1 and 2, respectively. β is used to smooth the distribution of the values and considered to be 5 Å^{-1}. The fluctuation of the contact formation is controlled by the minimum value of λ times r_{ij}^1 and r_{ij}^2, where λ is taken to be 1.8. By introducing such a minimum value between the above mentioned two quantities, we set a range of fluctuation for each of the specific native contacts considered so that one could differentiate between the set of specific native contacts with respect to the two folds, and thus measures the similarity of a given configuration with respect to only one of the two native folds.

The transition pathway between Ltn10 and Ltn40 is derived from the free energy landscape projected on suitable coordinates by calculating the minimum energy pathway with the

MEPSA software.[48] This user-friendly software utilized the graph-theory based Dijkstra's algorithm[49] to construct a minimum energy pathway between two given minima, identifying the barriers among different minima along such a pathway. Hence, the perspective here is one of transition state theory. Unlike competing approaches such as transition path sampling, [50–52] string method,[53,54] kinetic network model,[55] Traveling-salesman based Automated Path Searching (TAPS),[56] etc., it requires fewer resources but may not in all cases identify the optimal pathway.

3 RESULTS AND DISCUSSION

Previous investigations into the conversion mechanism between the two Lymphotactin motifs were hampered by the computational difficulties of sampling the protein's free energy landscape with sufficient statistics. In regular molecular dynamics simulations, the protein will spend most of the time in one of the basins of attraction - either exploring configurations similar to Ltn10, or, in the other case, Ltn40-like configurations - with transitions between the two basins being rare events. We argue that these computational difficulties can be overcome with our variant of replica exchange sampling which was designed specifically for investigation of switching between two well-defined states. In order to demonstrate that our approach allows indeed a more accurate sampling of the free energy landscape by raising the rate of transitions between the two main basins, we show in Figure 3(a) the walk of a typical realization of our system along the ladder of replicas. At start time ($t = 0$) the physical system sits on a replica where it is biased with $\lambda = 0.009$ toward the Ltn40 form. During the 100 ns of simulation this realization of Lymphotactin walks numerous times between the two end-points of the ladder. On the one end, the physical system will be biased maximally with coupling parameter $\lambda_{max} = 0.1$ toward the Ltn40 structure, and on the other end with a maximal $\lambda_{max} = 0.1$ toward Ltn10. The average exchange rate between neighboring replicas is ∼ 47%. In Figure 3(a) we show that this walk through λ-space induces indeed interconversion between the two forms. For this purpose, we characterize the state of a given configuration by the C_α distances between two specific residue pairs. In the Ltn10 structure the two residues T15 and A49 form a hydrophobic contact, while they are separated by a large distance in Ltn40. The opposite relation is found for the pair L14 and L45, which form hydrophobic contacts in Ltn40 but are far away from each other in the Ltn10 structure. Both distances are shown in Figure 3(b) and evolve in anti-correlated fashion over the 100 ns of simulation. If the system is on one side of the ladder and biased toward Ltn10, the distance between T15 and A49 (shown in red) will have small values and the distance between L14 and L45 (drawn in black) have large values, while the opposite is true once the system is on replica where the bias from the structure-based model is toward the Ltn40 structure.

A measure for the efficiency of our method, and a lower limit on the number of independent configurations sampled at the $\lambda = 0$ replicas, is the number of walks along the whole ladder, from the replica with maximal bias toward Ltn10 to the one with maximal bias toward Ltn40, and back. Such a walk along the whole ladder of replicas is called by us a tunneling event, and inversely related to the average time needed to cross the ladder (termed by us the tunneling time). The higher the number of tunneling events, and the shorter the tunneling time, the more efficient will be our approach. However, calculation of the number of

tunneling events or the tunneling time gives only meaningful results after the system has approached equilibrium. This convergence of the simulation is checked by calculating the free energy as a function of the two distances introduced above, and comparing it for two non-overlapping different time intervals. Resemblance of the data for these two intervals in Figure 4 suggests that the simulation has converged after 20 ns, and therefore, we use the last 80 ns of the simulated trajectories for our analysis. In this time span, we find a total of 31 tunneling events with an average tunneling time of ~ 33 ns. We have demonstrated in earlier work[21,22,25,26] for smaller systems that the number of tunneling events is in our approach much higher than the ones found in regular Hamilton replica exchange simulations with comparable number of replicas and λ distribution. In the latter case, often not a single tunneling event could be detected. As Lymphotactin is larger than the previously studied systems, and the sampling difficulties increase exponentially with system size, we expect that for Lymphotactin the improvement over regular Hamilton replica exchange is even higher than seen in our previous work.

The increased efficiency of our approach, leading to 31 tunneling events, gives us confidence in the free energy landscape found at $\lambda = 0$, i.e., at a replica where the "physical" model of our system is not biased toward either Ltn10 or Ltn40. In order to measure the frequency of these two motifs for the unbiased replica, we define a configuration as Ltn10-like if the C_a distances between T15 and A49, d_{C_a} (T15-A49), is less than 8 Å and and that between L14 and L45, d_{C_a} (L14-L45), greater than 12 Å. This definition was derived from visual inspection of the landscape as obtained for the replica with maximal bias toward Ltn10. Guided in a similar way by a visual inspection of the landscape as obtained for the replica with maximal bias toward Ltn40, we define a configuration as Ltn40-like if d_{C_a} (L14-L45) is less than 8 Å and d_{C_a} (T15-A49) greater than 12 Å. Using the above definitions we find that on the unbiased replica about 17% of configurations are Ltn10-like, while 34% are Ltn40-like. This suggests that Ltn40 is the most stable form and about 50% of the configurations sampled in our simulation do not represent either fold. Visual inspection and secondary structure analysis indicate that most of those configurations are intermediates on the pathway between the two opposite folds. The slightly higher population of Ltn40 over Ltn10 is in accord with the experimental study,[15] where under physiological conditions 46% of the configurations are Ltn10-like, and 54% Ltn40-like. On the other hand, configurations that do not belong to either of the two motifs are not observed with the large frequency seen in our simulations. This difference can be explained by the low temporal resolution of the experiments which makes it difficult to characterize short-lived intermediates, i.e., the experimentally reported frequencies for Ltn10 and Ltn40-like configurations are relative frequencies resulting from the well-resolved signals corresponding to 12 residues collected from a two-dimensional ^{15}N-^{1}H HSQC spectrum.

In order to explore the interconversion pathway, we show in Figure 5(a) the free energy landscape projected on the two characteristic distances introduced earlier: d_{C_a} (L14-L45) (measuring similarity with Ltn40) and d_{C_a} (T15-A49) (quantifying similarity to Ltn10). For calculating the landscape, we use only data sampled at the $\lambda = 0$ replicas, where the physical model is not biased by any Go-model term. The so-drawn landscape is characterized by two prominent basins corresponding to either Ltn10-like or Ltn40-like configurations. Conversion events can be described by pathways connecting the two basins in the landscape.

However, not all possible pathways are equally likely. Take as an example the pathway represented by a black line in Figure 5(a). This path is obtained by projecting onto the landscape the configurations sampled along the walk in λ space during a certain tunneling event. This pathway describes a transition between the two Lymphotactin folds that requires crossing an energy barrier much higher than that reported in the experiments, indicating that this is not true pathway. On the other hand, we can obtain a thermodynamically reasonable pathway by using the MEPSA software[48] to construct the minimum energy pathway, which we have drawn as a white line in the landscape in Figure 5(a). The algorithm used to determine such a minimum energy pathway in the MEPSA software[48] and the advantages of this software have been discussed in Section 2.3. Unlike the tunneling pathway, this pathway does not go through regions of the landscape characterized by unfolded configurations, but instead proceeds through a series of basins, with an effective energy barrier similar to that reported in experiments. This indicates that the interconversion process does not proceed by unfolding of the Ltn10 or Ltn40 structure but rather involves a series of intermediates or transition states.

By construction the minimum energy pathway does not connect specific configurations but bins in the landscape each containing a certain number of configurations sampled throughout the simulations. These configurations can be characterized according to presence (or lack) of a C-terminal helix (found in the Ltn10 structure), the N-terminal β_0 strand (found in the Ltn40-structure), and the hydrogen bonding in the β_1 to β_3 region, which offers another way of distinguishing between the Ltn10 and Ltn40-like configurations. The procedure by which we attribute these three traits to a configuration is described in the method section. The frequency with which the three traits are observed allows one to identify five distinct regions along the pathway that correlate with the basins and barriers of the landscape. These segments are labeled as A to E in Figure 5(a). A similar division is not possible for the pathway, derived from the tunneling event.

The difference between the two possible pathways does not depend on the specific coordinates on which the landscape is projected. This can be seen from Figure 5(b) where we project the free energy landscape on the fraction of specific native contacts with respect to each of the two folds, and overlay again both paths on the landscape. The pathway derived from the tunneling event is again not consistent with the landscape. Hence, the tunneling events in our RET approach cannot be used to derive a conversion mechanism as they rely on an artificial dynamics, designed to increase sampling efficiency. On the other hand, while the configurations in the minimum energy pathway calculated for the new landscape will differ from the one calculated for the other landscape, we can again identify the same five regions. We remark that the radius of gyration (a measure for the compactness of configurations) of the central part, made up of $\beta_1 - \beta_2 - \beta_3$ in both Ltn10 and Ltn40, differs little along the pathway, which implies that the Lymphotactin configurations do not unfold and refold while assuming this pathway during the interconversion process.

The qualitative agreement between the minimum energy pathways found for the two landscapes suggests that these pathways describe indeed the conversion process. Hence, in order to determine the mechanism and to establish the separating free energy barrier, we have analyzed in more detail the pathway shown in Figure 5(a). When starting from the

Ltn40 basin of region A, the β_0 strand gets dissolved when reaching basin B which has about 1.5 kcal/mol higher free energy than the Ltn40 fold of basin A. Upon crossing this barrier, the Lymphotactin configuration evolves further through a number of intermediates with little difference in free energy, forming the flat-floored valley of region C. Using visual inspection and the STRIDE algorithm[46] as implemented in VMD software[47] for secondary structure analysis, we observe a re-arrangement of backbone hydrogen bonds, until, when entering region D the hydrogen bond pattern of the remaining three β-strands (β_1, β_2, and β_3) becomes similar to that of the Ltn10 fold. The Ltn10 basin (region E) has again about 1.5 kcal/mol lower free energy value than the transition state D and is reached when the helix slowly begins to form at the C-terminus (within the residues 54 to 66). A schematic diagram explaining the mechanism is shown in Figure 6.

From the above discussed conversion process, we have estimated a total energy barrier of ~ 2 kcal/mol for Ltn40→Ltn10 conversion, while slightly a lower value of ~ 1.5 kcal/mol is the barrier for the reverse process. While the higher energy barrier for the Ltn40→Ltn10 conversion, implying a slower conversion rate than the reverse one, is in agreement with the experimentally reported results,[11,20,57] the energies are higher than the ones measured in the experiments: 0.5 kcal/mol and 0.2 kcal/mol, respectively. This divergence may reflect again the difference between our computational setting (considering a transition between monomer structures) and the experimental setting that studies the transition from Ltn10 monomers to Ltn40 dimers. As the dimerization is the fast process, the Le Chatelier's principle will predict a shift of the equilibrium toward the Ltn40 motif, effectively lowering the barrier.

Unlike an earlier proposal,[20] our conversion mechanism does not require unfolding of the Ltn monomers. Instead it assumes partially conserved contacts as have also been reported in another previous computational study.[24] Our mechanism relies on a sequence of local changes with the main assumption that the intermediates keep part of the local ordering. Especially, we assume that the central three-stranded β-sheet (β_1, β_2, and β_3) is preserved. This sheet is found in both motifs, but with a shift in the backbone hydrogen pattern by one residue,[10] see Table 1. Note that residues on the central β_2 strand do not get shifted.

Instead, they change their hydrogen bond forming partners on the adjacent strands (β_1 or β_3). Our proposed conversion mechanism therefore has to explain how this shift between these residues can proceed without the need of a complete unfolding. A more thorough analysis of the pathway shows that the unfolding of the β-sheet and the resulting high barriers are avoided by some residues forming bifurcated/bridged hydrogen bonds with two consecutive residues. For example, while in the Ltn40 configurations of basin A, residue F39 forms a backbone hydrogen bond with residue K46, and in the Ltn10 configurations of basin E backbone hydrogen bonds with residue V47. However, we observe that F39 (located on strand β_2) can also form simultaneously hydrogen bonds with residues K46 and V47 on strand β_3. In this case, the amide nitrogen of F39 only participates in forming hydrogen bonds and thus acts as a bifurcated donor, or both carbonyl oxygen and amide nitrogen of F39 participate in forming those hydrogen bonds, helping in forming bridged hydrogen bonds. Similarly, we find that in a similar manner residue T41 can also form hydrogen bonds with both residues G44 and L45. All six residues on the central β_2 strand (see Table 1) can form such bifurcated hydrogen bonds that bridge between Ltn10-like and Ltn40-like

hydrogen bonding, but for most residues these bifurcated hydrogen bonds appear with a relative frequency of less than 5%, that is, less than 5% of all hydrogen bonds connecting a residue on strand β_2 with a partner residue on either β_1 or β_3 are bifurcated hydrogen bonds.

The exceptions are the already mentioned bifurcated hydrogen bonds F39-K46/V47, which is observed with a relative frequency of about 24%, and T41-G44/L45, which appears with a relative frequency of about 11%. We remark that we also find similar frequencies for the minimum free energy path of the free energy landscape in Figure 5(b). Both bifurcated hydrogen bonds connect residues located at the start of the sheet formed by the β_2 and β_3 strand. In the Ltn10 motif, residue F39 forms a hydrogen bond with V47, and T41 one with L45. Transitioning to the Ltn40 hydrogen bonding of F39 with K46 and T41 with G44 will be eased by transient formation of the bifurcated hydrogen bonds at these locations as they avoid the energetic costs of dissolving and reforming hydrogen bonds. Correspondingly, the F39-K46/V47 is observed with highest relative frequency in the transition regions B (about 37%) and D (about 64%), and with about 16% relative frequency in the intermediate region C. The corresponding frequencies are lower for the T41-G44/L45 bifurcated hydrogen bond, which appears in 8% (11%) in the transition region B (D), and with a relative frequency of 14% in the intermediate region C.

We conjecture that formation of the bifurcated/bridged hydrogen bonds F39-K46/V47 and T41-G44/L45 at the turn region between the β_2 and β_3 strands is crucial for enabling the fold switch as it disturbs the geometry of the sheet and initiates a wave of successive re-arrangement of hydrogen bonds in the three-stranded β-sheet which avoids large energy barrier that would otherwise arise from breaking and forming hydrogen bonds. Our conjecture could be tested in principle by mutation experiments where the mutated side chain would form contacts that lead to repositioning of the backbone atoms that would restrict formation of such bifurcated hydrogen bonds. Another possibility would be the use of deuterium which would also alter the relative frequency of bifurcated hydrogen bonds.

Interestingly, both bifurcated hydrogen bond pairs are also seen with substantial frequency in the region E (dominated by Ltn10-like configurations), where the F39-K46/V47 bond is observed in about 49% of the configurations, and the T41-G44/L45 one in about 16% of configurations. Hence, in Ltn10-like configurations the hydrogen bond pairs F39-V47 and T41-L45 are easily replaced by the corresponding bifurcated hydrogen bonds. This is likely because the extension from the hydrogen bond F39-V47 to the bifurcated hydrogen bond F39-K46/V47 involves rotation of the side chains of the three residues that reduces the hydrophobic solvent accessible surface area (SASA) by about 26 $\mathring{A}^2$, and increases the exposure of the charged K46 by about 7 $\mathring{A}^2$. Both are energetically favorable changes. The effects are less pronounced for the bifurcated hydrogen bond T41-G44/L45 where there is only a small reduction in hydrophobic SASA ≈ 7 $\mathring{A}^2$. On the other hand, in region A (where Ltn40-like configurations are found) only the T41-G44/L45 bifurcated hydrogen bond is observed with substantial, but much smaller, relative frequency (about 9%). Here, formation of a bifurcated hydrogen bond T41-G44/L45 would not change the solvent accessible surface area of the three involved residues, while the bifurcated hydrogen bond F39-K46/V47 would lead to a reduction of about 45 $\mathring{A}^2$ of solvent exposure for the charged K46 that could not be compensated by the favorable loss of hydrophobic SASA of about 18 $\mathring{A}^2$.

However, we remark that unlike to the Ltn10 configuration, where the β_3 strand is stabilized by four contacts with the N-terminal helix, no such β_3-stabilizing contacts with the β_0 strand exist in the Ltn40 configuration. Hence, the barrier for unraveling the $\beta_2 - \beta_3$ hydrogen bonding is likely lower in the Ltn40 configurations than in the Ltn10 configurations where instead formation of the two bifurcated hydrogen bonds circumvents the otherwise higher barrier.

Differences in the number of stabilizing side chain contacts are also the reason why we observe bifurcated hydrogen bonds with appreciable frequency only between β_2 and β_3, but not between β_2 and β_1. In the Ltn10 configuration, four side chain contacts (A49-W55, A49-V56, D50-V56, P51-V56) with the helix stabilize the β_3 strand, but ten such side chain contacts (L24-W55, L24-C59, L24-M63, L24-K66, T26-M63, Y27-V60, Y27-R61, Y27-M63, Y27-D64, I29-V60) that stabilize the β_1 strand. The difference in number of stabilizing side chain contacts is even larger for the Ltn40 configuration. In this motif, there are no side chain contacts with the β_0 strand that could stabilize the β_3 strand, but four such contacts stabilizing the β_1 strand. As formation of bifurcated hydrogen bonds requires repositioning of backbone atoms that in turn depends on suitable rotation of side chain atoms, such bifurcated hydrogen bonds are less likely between β_1 and β_2 than between β_2 and the more flexible β_3.

4 CONCLUSIONS

Using a variant of Replica-Exchange-with-Tunneling (RET), we have probed the interconversion of a metamorphic protein that switches biological function by altering its three dimensional structure between two well-defined native forms, Ltn10 and Ltn40. While Ltn10 is monomer with three-stranded β-sheet ending with a C-terminal helix, Ltn40 exists as a dimer with all β-sheet arrangements. In order to ease the numerical difficulties, we have only considered the conversion for a monomer. This is justified as conversion and dimerization are separate events, with the conversion the time-limiting process.[20] Our investigation relies on the use of RET, an enhanced sampling method developed in our group, that has enabled us to sample the free energy landscape of the protein with high precision. We find relative population frequencies that are consistent with experimental measurements, but our simulations predict a larger population of intermediate configurations than reported in the experiments. We reason that our method allows us to identify intermediates that due to their short-life time are difficult to observe in experiments. Analyzing the free energy landscape allows us to identify a conversion mechanism that relies on passage through a number of distinct structural intermediates, and involves breaking and reformation of the β_0-strand and the C-terminal helix and a re-arrangement of hydrogen bonds in the central three-stranded β-sheet made from β_1, β_2 and β_3. The associated high costs of breaking and forming hydrogen bonds are avoided by formation of bifurcated hydrogen bonds that naturally bridge between the characteristic hydrogen bond pattern in the three β-sheets common in both motifs. These pattern differs in both forms by being shifted by one residue. We surmise that formation of these bifurcated hydrogen bonds facilitates the switch between these two patterns, guiding in this way the conversion between the two motifs.

Acknowledgement

References

(1). Leopold PE; Montal M; Onuchic JN Protein Folding Funnels: A Kinetic Approach to the Sequence-Structure Relationship. Proc. Natl. Acad. Sci. U. S. A 1992, 89, 8721–8725. [PubMed: 1528885]

(2). Onuchic JN; Wolynes PG Theory of Protein Folding. Curr. Opin. Struct. Biol 2004, 14, 70–75. [PubMed: 15102452]

(3). Dunker AK; Lawson JD; Brown CJ; Williams RM; Romero JS, Pedro Oh; Oldfield CJ; Campen AM; Ratliff CM; Hipps KW; Ausio J et al. Intrinsically Disordered Protein. J. Mol. Graph. Model 2001, 19, 26–69. [PubMed 11381529]

(4). Dyson HJ; Wright PE Intrinsically Unstructured Proteins and Their Functions. Nat. Rev. Mol. Cell Biol 2005, 6, 197–208. [PubMed: 15738986]

(5). Murzin AG Metamorphic Proteins. Science 2008, 320, 1725–1726. [PubMed: 18583598]

(6). Bryan PN; Orban J Proteins that Switch Folds. Curr. Opin. Struct. Biol 2010, 20, 482–488. [PubMed: 20591649]

(7). Kelner GS; Kennedy J; Bacon KB; Kleyensteuber S; Largaespada DA; Jenkins NA; Copeland NG; Bazan JF; Moore KW; Schall TJ et al. Lymphotactin: A Cytokine that Represents a New Class of Chemokine. Science 1994, 266, 1395–1399. [PubMed: 7973732]

(8). Rossi D; Zlotnik A The Biology of Chemokines and Their Receptors. Annu. Rev. Immunol 2000, 18, 217–242. [PubMed: 10837058]

(9). Tuinstra RL; Peterson FC; Elgin ES; Pelzek AJ; Volkman BF An Engineered Second Disulfide Bond Restricts Lymphotactin/XCL1 to a Chemokine-like Conformation with XCR1 Agonist Activity. Biochemistry 2007, 46, 2564–2573. [PubMed: 17302442]

(10). Tuinstra RL; Peterson FC; Kutlesa S; Elgin ES; Kron MA; Volkman BF Interconversion between Two Unrelated Protein Folds in the Lymphotactin Native State. Proc. Natl. Acad. Sci. U. S. A 2008, 105, 5057–5062. [PubMed: 18364395]

(11). Volkman BF; Liu TY; Peterson FC Lymphotactin Structural Dynamics. Methods Enzymol. 2009, 461, 51–70. [PubMed: 19480914]

(12). Thelen M; Stein JV How Chemokines Invite Leukocytes to Dance. Nature Immunol. 2008, 9, 953–959. [PubMed: 18711432]

(13). Formaneck MS; Ma L; Cui Q Effects of Temperature and Salt Concentration on the Structural Stability of Human Lymphotactin: Insights from Molecular Simulations. J. Am. Chem. Soc 2006, 128, 9506–9517. [PubMed: 16848488]

(14). Ma L; Cui Q The Temperature Dependence of Salt-Protein Association Is Sequence Specific. Biochemistry 2006, 45, 14466–14472. [PubMed: 17128985]

(15). Kulo lu ES; McCaslin DR; Markley JL; Volkman BF Structural Rearrangement of Human Lymphotactin, a C Chemokine, under Physiological Solution Conditions. J. Biol. Chem 2002, 277, 17863–17870. [PubMed: 11889129]

(16). Peterson FC; Elgin ES; Nelson TJ; Zhang F; Hoeger TJ; Linhardt RJ; Volkman BF Identification and Characterization of a Glycosaminoglycan Recognition Element of the C Chemokine Lymphotactin. J. Biol. Chem 2004, 279, 12598–12604. [PubMed: 14707146]

(17). Luo X; Tang Z; Xia G; Wassmann K; Matsumoto T; Rizo J; Yu H The Mad2 Spindle Checkpoint Protein has Two Distinct Natively Folded States. Nat. Struct. Mol. Biol 2004, 11, 338–345. [PubMed: 15024386]

(18). Mapelli M; Massimiliano L; Santaguida S; Musacchio A The Mad2 Conformational Dimer: Structure and Implications for the Spindle Assembly Checkpoint. Cell 2007, 131, 730–743. [PubMed: 18022367]

(19). Luo X; Yu H Protein Metamorphosis: the Two-State Behavior of Mad2. Structure 2008, 16, 1616–1625. [PubMed: 19000814]

(20). Tyler RC; Murray NJ; Peterson FC; Volkman BF Native-State Interconversion of a Metamorphic Protein Requires Global Unfolding. Biochemistry 2011, 50, 7077–7079. [PubMed: 21776971]

(21). Bernhardt NA; Xi W; Wang W; Hansmann UH Simulating Protein Fold Switching by Replica Exchange with Tunneling. J. Chem. Theory Comput 2016, 12, 5656–5666. [PubMed: 27767301]

(22). Bernhardt NA; Hansmann UH Multifunnel Landscape of the Fold-Switching Protein RfaH-CTD. J. Phys. Chem. B 2018, 122, 1600–1607. [PubMed: 29323497]

(23). Ramírez-Sarmiento CA; Noel JK; Valenzuela SL; Artsimovitch I Interdomain Contacts Control Native State Switching of RfaH on a Dual-Funneled Landscape. PLoS Comput. Biol 2015, 11, e1004379. [PubMed: 26230837]

(24). Camilloni C; Sutto L Lymphotactin: How a Protein can Adopt Two Folds. J. Chem. Phys 2009, 131, 245105.

(25). Ya ar F; Bernhardt NA; Hansmann UH Replica-Exchange-with-Tunneling for Fast Exploration of Protein Landscapes. J. Chem. Phys 2015, 143, 224102.

(26). Zhang H; Xi W; Hansmann UH; Wei Y Fibril–Barrel Transitions in Cylindrin Amyloids. J. Chem. Theory Comput 2017, 13, 3936–3944. [PubMed: 28671829]

(27). Fukunishi H; Watanabe O; Takada S On the Hamiltonian Replica Exchange Method for Efficient Sampling of Biomolecular Systems: Application to Protein Structure Prediction. J. Chem. Phys 2002, 116, 9058–9067.

(28). Kwak W; Hansmann UH Efficient Sampling of Protein Structures by Model Hopping. Phys. Rev. Lett 2005, 95, 138102.

(29). Moritsugu K; Terada T; Kidera A Scalable Free Energy Calculation of Proteins via Multiscale Essential Sampling. J. Chem. Phys 2010, 133, 224105.

(30). Zhang W; Chen J Accelerate Sampling in Atomistic Energy Landscapes Using Topology-Based Coarse-Grained Models. *J. Chem. Theory Comput.* 2014, 10, 918–923 [PubMed: 26580171]

(31). Hansmann UH; Okamoto Y Generalized-Ensemble Monte Carlo Method for Systems with Rough Energy Landscape. Phys. Rev. E 1997, 56, 2228.

(32). Okamoto Y Generalized-Ensemble Algorithms: Enhanced Sampling Techniques for Monte Carlo and Molecular Dynamics Simulations. J. Mol. Graph. Model 2004, 22, 425–439. [PubMed: 15099838]

(33). Vanden-Eijnden E, et al. Transition-Path Theory and Path-Finding Algorithms for the Study of Rare Events. Annu. Rev. Phys. Chem 2010, 61, 391–420. [PubMed: 18999998]

(34). Shukla D; Hernández, C. X.; Weber, J. K.; Pande, V. S. Markov State Models Provide Insights into Dynamic Modulation of Protein Function. Acc. Chem. Res 2015, 48, 414–422. [PubMed: 25625937]

(35). Kelley NW; Vishal V; Krafft GA; Pande VS Simulating Oligomerization at Experimental Concentrations and Long Timescales: A Markov State Model Approach. J. Chem. Phys 2008, 129, 214707.

(36). Wu H; Paul F; Wehmeyer C; Noé F Multiensemble Markov Models of Molecular Thermodynamics and Kinetics. Proc. Natl. Acad. Sci. U. S. A 2016, 113, E3221–E3230. [PubMed: 27226302]

(37). Caldarulo E; Barducci A; Wüthrich, K.; Parrinello, M. Prion Protein $\beta2$–$\alpha2$ Loop Conformational Landscape. Proc. Natl. Acad. Sci. U. S. A 2017, 114, 9617–9622. [PubMed: 28827331]

(38). Best RB; Zhu X; Shim J; Lopes PE; Mittal J; Feig M; MacKerell AD Jr Optimization of the Additive CHARMM All-Atom Protein Force Field Targeting Improved Sampling of the Backbone φ, ψ and Side-chain $\chi1$ and $\chi2$ Dihedral Angles. J. Chem. Theory Comput 2012, 8, 3257–3273. [PubMed: 23341755]

(39). Jorgensen WL; Chandrasekhar J; Madura JD; Impey RW; Klein ML Comparison of Simple Potential Functions for Simulating Liquid Water. J. Chem. Phys 1983, 79, 926–935.

(40). Noel JK; Whitford PC; Sanbonmatsu KY; Onuchic J. N. SMOG@ ctbp: Simplified Deployment of Structure-Based Models in GROMACS. Nucleic Acids Res 2010, 38, W657–W661. [PubMed: 20525782]

(41). Schrödinger, LLC,

(42). Hess B; Kutzner C; Van Der Spoel D; Lindahl E GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation. J. Chem. Theory Comput 2008, 4, 435–447. [PubMed: 26620784]

(43). Swope WC; Andersen HC; Berens PH; Wilson KR A Computer Simulation Method for the Calculation of Equilibrium Constants for the Formation of Physical Clusters of Molecules: Application to Small Water Clusters. J. Chem. Phys 1982, 76, 637–649.

(44). Hess B P-LINCS: A Parallel Linear Constraint Solver for Molecular Simulation. J. Chem. Theory Comput 2008, 4, 116–122. [PubMed: 26619985]

(45). Bussi G; Donadio D; Parrinello M Canonical Sampling through Velocity Rescaling. J. Chem. Phys 2007, 126, 014101.

(46). Frishman D; Argos P Knowledge-Based Protein Secondary Structure Assignment. Proteins 1995, 23, 566–579. [PubMed: 8749853]

(47). Humphrey W; Dalke A; Schulten K VMD: Visual Molecular Dynamics. J. Mol. Graphics 1996, 14, 33–38.

(48). Marcos-Alcalde I; Setoain J; Mendieta-Moreno JI; Mendieta J; Gómez-Puertas P MEPSA: Minimum Energy Pathway Analysis for Energy Landscapes. Bioinformatics 2015, 31, 3853–3855. [PubMed: 26231428]

(49). Dijkstra EW A Note on Two Problems in Connexion with Graphs. Numerische mathematik 1959, 1, 269–271.

(50). Bolhuis PG; Chandler D; Dellago C; Geissler PL Transition Path Sampling: Throwing Ropes over Rough Mountain Passes, in the Dark. Annu. Rev. Phys. Chem 2002, 53, 291–318. [PubMed: 11972010]

(51). Dellago C; Bolhuis PG; Csajka FS; Chandler D Transition Path Sampling and the Calculation of Rate Constants. J. Chem. Phys 1998, 108, 1964–1977.

(52). Dellago C; Bolhuis PG; Chandler D Efficient Transition Path Sampling: Application to Lennard-Jones Cluster Rearrangements. J. Chem. Phys 1998, 108, 9236–9245.

(53). Weinan E; Ren W; Vanden-Eijnden E String Method for the Study of Rare Events. Phys. Rev. B 2002, 66, 052301.

(54). Maragliano L; Fischer A; Vanden-Eijnden E; Ciccotti G String Method in Collective Variables: Minimum Free Energy Paths and Isocommittor Surfaces. J. Chem. Phys 2006, 125, 024106.

(55). Unarta IC; Zhu L; Tse CKM; Cheung PP-H; Yu J; Huang X Molecular Mechanisms of RNA Polymerase II Transcription Elongation Elucidated by Kinetic Network Models. Curr. Opin. Struct. Biol 2018, 49, 54–62. [PubMed: 29414512]

(56). Zhu L; Sheong FK; Cao S; Liu S; Unarta IC; Huang X TAPS: A Traveling-Salesman Based Automated Path Searching Method for Functional Conformational Changes of Biological Macromolecules. J. Chem. Phys 2019, 150, 124105. [PubMed: 30927873]

(57). Tyler RC; Wieting JC; Peterson FC; Volkman BF Electrostatic Optimization of the Conformational Energy Landscape in a Metamorphic Protein. Biochemistry 2012, 51, 9067–9075. [PubMed: 23102260]
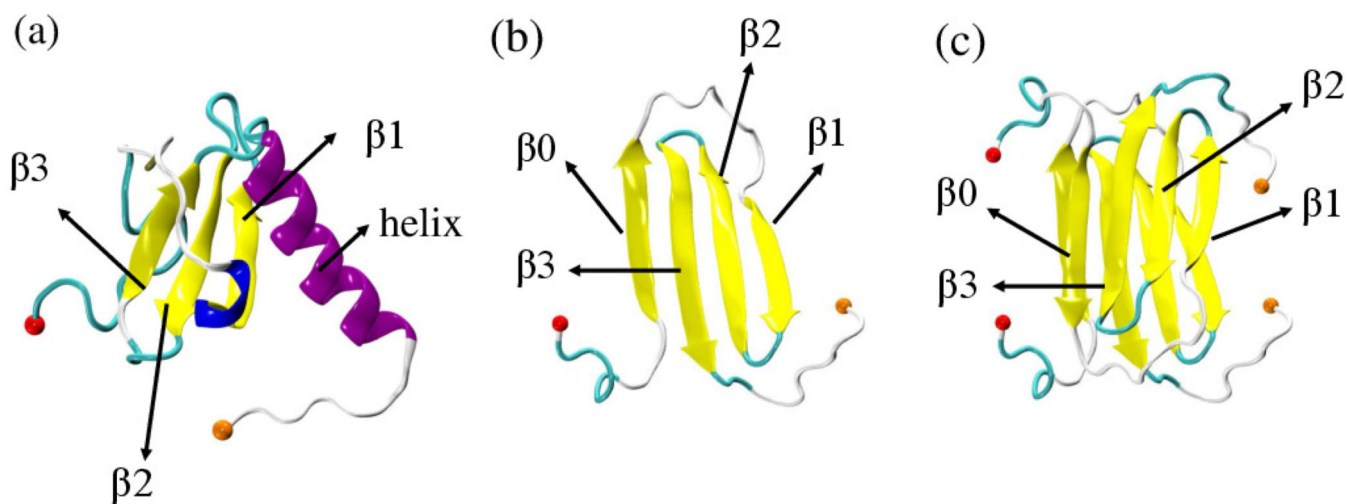
Figure 1:
Lymphotactin chains can take two distinct structures, both deposited in the Protein Data Bank: (a) Ltn10 (PDB-ID: 2HDM) and Ltn40. The Ltn40 monomer is shown in (b) and derived from the experimentally observed dimer (PDB-ID: 2JP1) shown in (c). Labels 30 identify the secondary structural elements, and the N-terminal and C-terminal C_a atoms are drawn as spheres in red and orange, respectively.
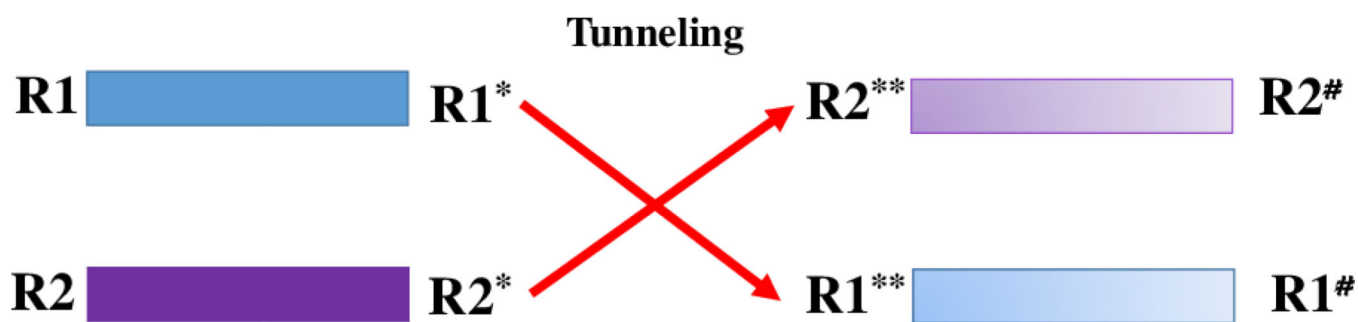
Tunneling

Figure 2:
A schematic diagram of the RET approach. The new configuration R2** on replica 1 has after re-scaling of velocities the same total energy as the old configuration R1*. While evolving to the state R2$^\#$ potential and kinetic energies interconvert, symbolically noted by the shift in color. In a similar way the configurations R1** and R2* on replica 2 have the same energy, and R1** evolves to R1$^\#$.
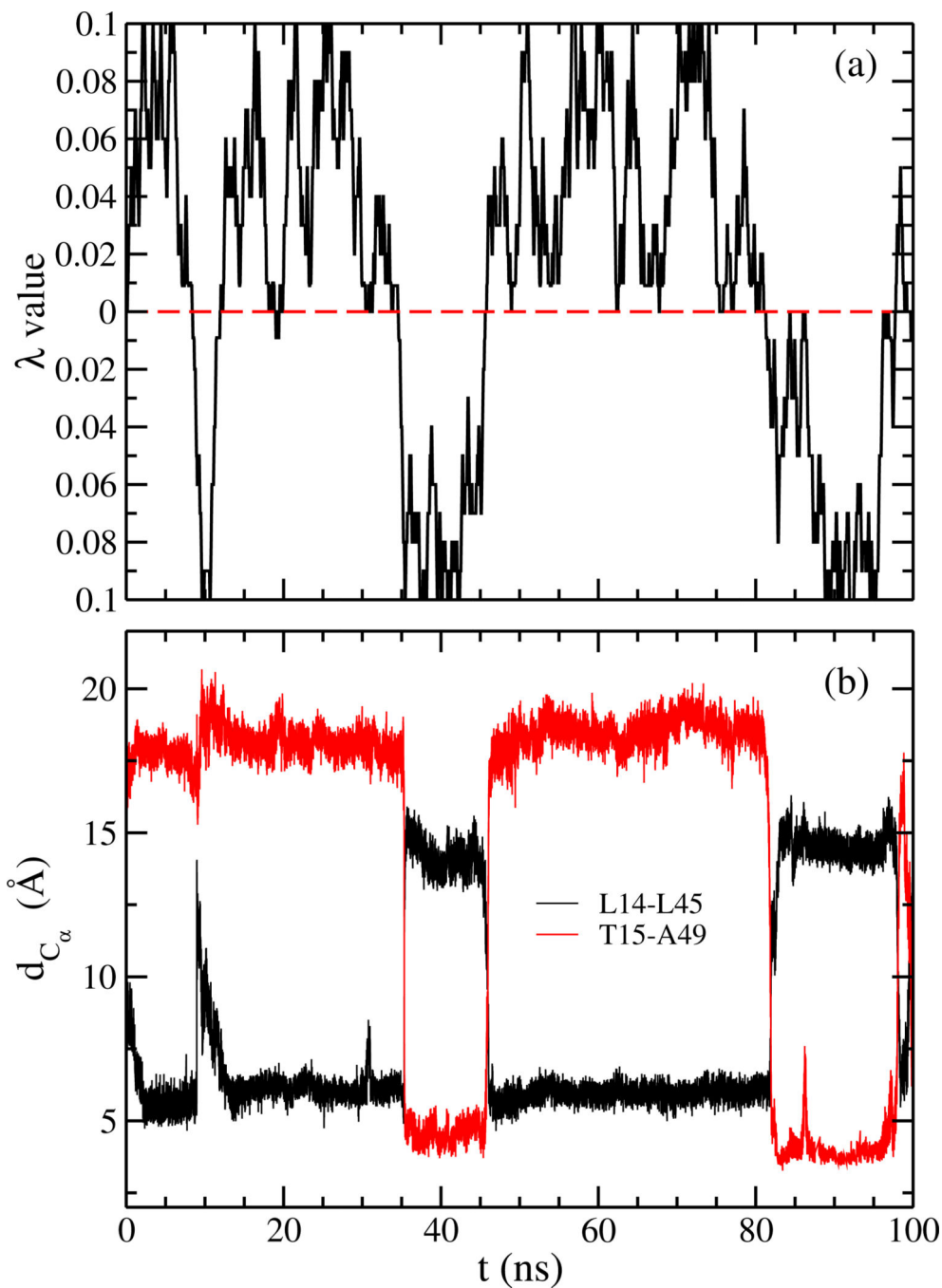
Figure 3:
(a) A typical example of a replica walking through λ space starting from a replica, where the physical model is initially biased toward Ltn40 with $\lambda = 0.009$. While the system walks between replicas with bias toward Ltn10 and such with bias toward Ltn40, its configuration changes accordingly. This can be seen in (b) where we show the corresponding time evolution of C_a distances (d_{C_a}) between L14 and L45 (black) and that between T15 and A49 (red). The first distance is a measure for the similarity with Ltn40, and the second distance one for the similarity with Ltn10.
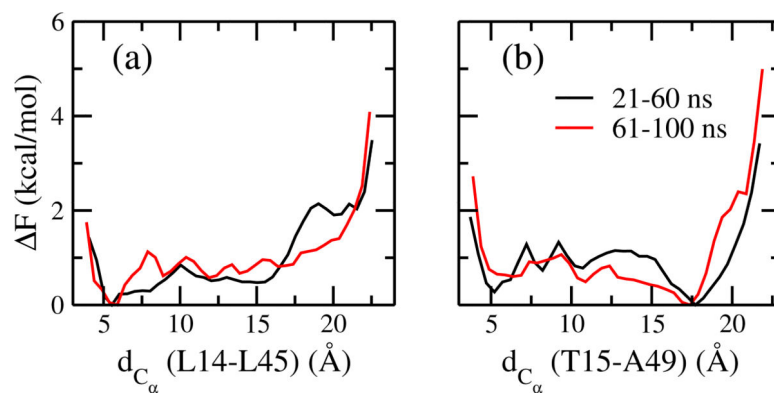
Figure 4:
Free energy (ΔF) as a function of the two C_a distances (d_{Ca}) between (a) L14 and L45 (a measure for the similarity to Ltn40) and (b) between T15 and A49 (a measure for the similarity to Ltn10), as obtained from the unbiased replica where $\lambda = 0$. Shown are values as measured for different segments of the simulation.
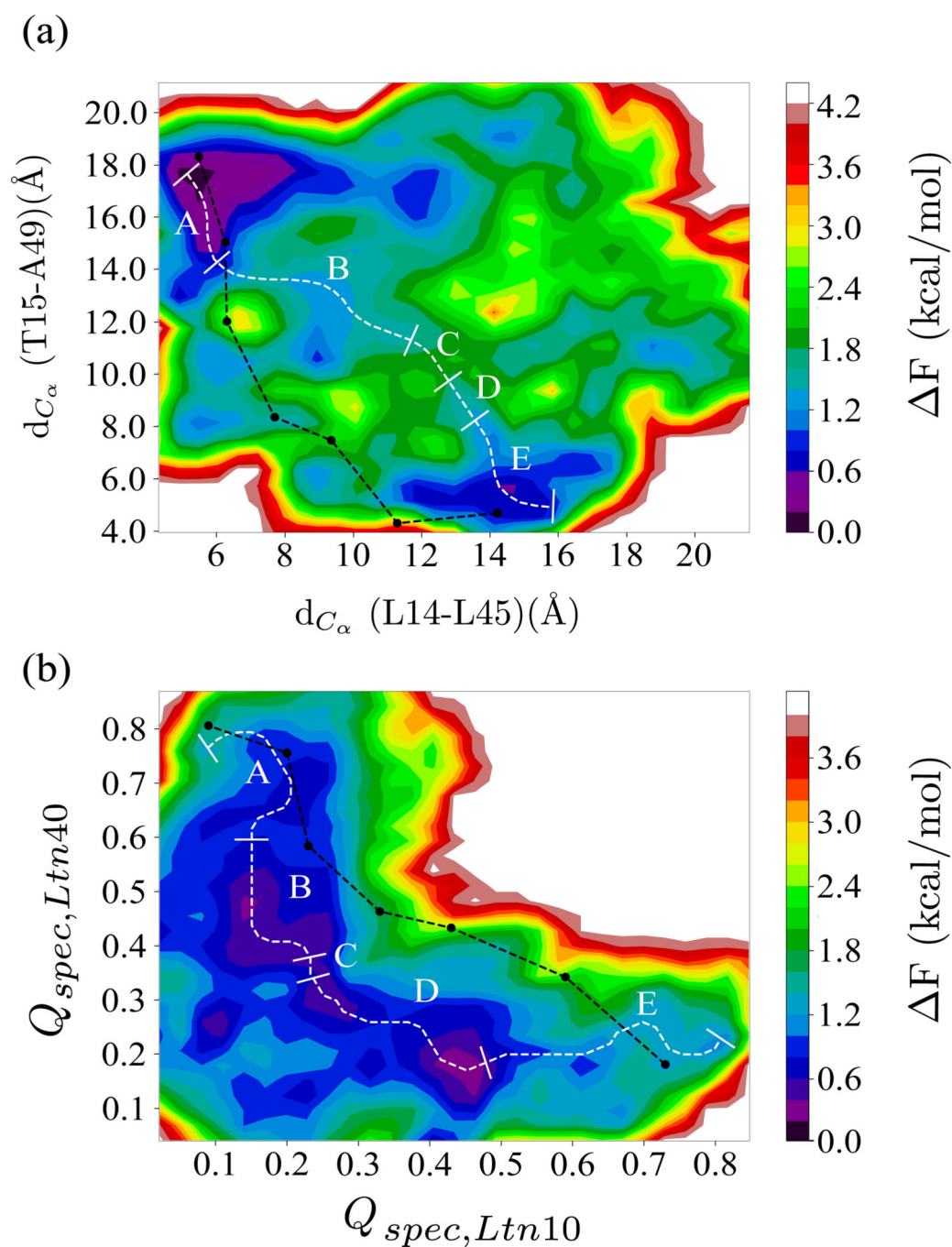
(a)

(b)

Figure 5:
Free energy landscape as obtained from our RET simulation at replicas where the physical models are not biased by any Go-term. The landscape is in (a) projected on C_a distances between L14 and L45 ((d_{C_a} (L14-L45)), a measure for similarity to the Ltn40 structure) and between T15 and A49 ((d_{C_a} (T15-A49), a measure for the similarity to Ltn10). The black line shows a typical pathway as obtained during a tunneling event, i.e., from a walk in λ-space between the structure strongly biasing toward Ltn10 and that toward Ltn40. On the other side, the minimum energy pathway as obtained from MEPSA software[48] is drawn in

white. The labels A-E mark five distinct regions that can be identified along this pathway (discussed in the text). For comparison, we show in (b) the same landscape, but projected on the fraction of specific native contacts (defined in Eqn. 5) with respect to Ltn10 ($Q_{spec,Ltn10}$) and Ltn40 ($Q_{spec,Ltn40}$) native structure. Both the minimum energy and the tunneling pathways are shown again, using the same color coding as in (a).
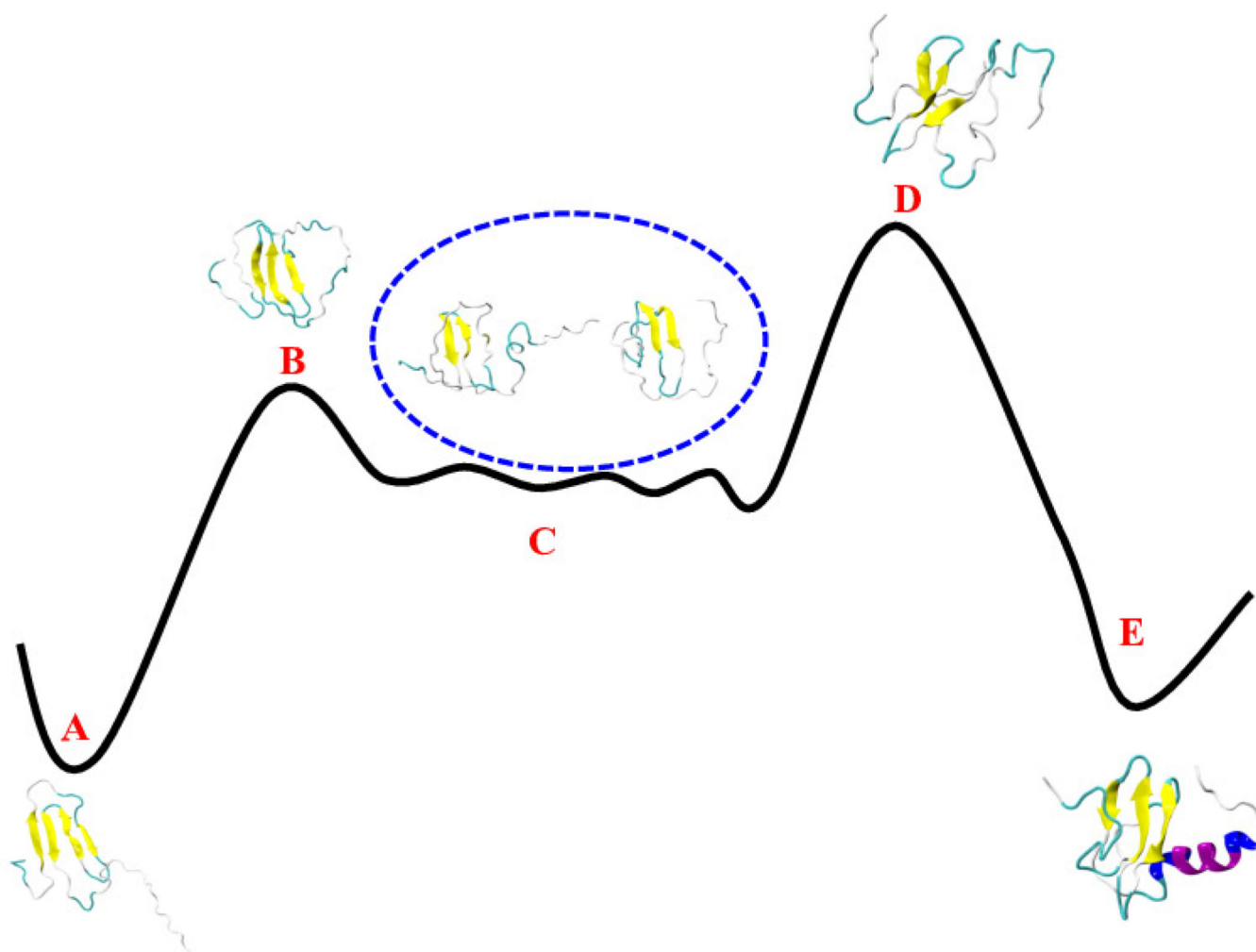
Figure 6:
A schematic diagram of the interconversion mechanism between Ltn40 and Ltn10 with representative structures.

Table 1:

Characteristic Backbone Hydrogen Bonding Pattern among Different β-Strands for Ltn10 and Ltn40 Lymphotactin Native Forms as Obtained from Our Study and Reported in the NMR Structure.[10]

$\beta_1 \leftrightarrow \beta_2$		$\beta_2 \leftrightarrow \beta_3$	
Ltn10	**Ltn40**	**Ltn10**	**Ltn40**
T26 ↔ I40	K25 ↔ I40	–	R35 ↔ D50
T28 ↔ I38	Y27 ↔ I38	V37 ↔ A49	V37 ↔ C48
T30 ↔ A36	I29 ↔ A36	F39 ↔ V47	F39 ↔ K46
–	E31 ↔ L34	T41 ↔ L45	T41 ↔ G44