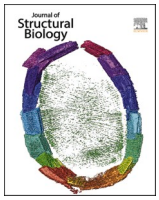
Journal of Structural Biology 216 (2024) 108054

Contents lists available at ScienceDirect

Journal of Structural Biology

journal homepage: www.elsevier.com/locate/yjsbi



Research Article

Exploring the folding landscape of leptin: Insights into threading pathways

Fernando Bruno da Silva [a,b], Jennifer M. Simien [c], Rafael G. Viegas [b,d], Ellinor Haglund [c,*], Vitor Barbanti Pereira Leite [b,*]

[a] Centre of New Technologies, University of Warsaw, Banacha 2c, Warsaw, Poland
[b] Institute of Biosciences, Humanities and Exact Sciences (IBILCE), São Paulo State University (UNESP), São José do Rio Preto, SP, Brazil
[c] Department of Chemistry, University of Hawaii at Manoa, Honolulu, Hawaii 96822, United States
[d] Federal Institute of Education, Science and Technology of São Paulo (IFSP), Catanduva, SP 15.808–305, Brazil

ARTICLE INFO

Keywords:
Topology
Energy landscape
Protein folding
Pierced Lasso Topology
Leptin

ABSTRACT

The discovery of new protein topologies with entanglements and loop-crossings have shown the impact of local amino acid arrangement and global three-dimensional structures. This phenomenon plays a crucial role in understanding how protein structure relates to folding and function, affecting the global stability, and biological activity. Protein entanglements encompassing knots and non-trivial topologies add complexity to their folding free energy landscapes. However, the initial native contacts driving the threading event for entangled proteins remains elusive. The Pierced Lasso Topology (PLT) represents an entangled topology where a covalent linker creates a loop in which the polypeptide backbone is threaded through. Compared to true knotted topologies, PLTs are simpler topologies where the covalent-loop persists in all conformations. In this work, the PLT protein leptin, is used to visualize and differentiate the preference for slipknotting over plugging transition pathways along the folding route. We utilize the Energy Landscape Visualization Method (ELViM), a multidimensional projection technique, to visualize and distinguish early threaded conformations that cannot be observed in an *in vitro* experiment. Critical contacts for the leptin threading mechanisms were identified where the competing pathways are determined by the formation of a hairpin loop in the unfolded basin. Thus, prohibiting the dominant slipknotting pathway. Furthermore, ELViM offers insights into distinct folding pathways associated with slipknotting and plugging providing a novel tool for *de novo* design and *in vitro* experiments with residue specific information of threading events *in silico*.

1. Introduction

In the last 30 years, the role of protein topology has emerged with increased interest as new entanglements and loop crossings are discovered. Protein topology refers to the local arrangement of neighboring amino acids and the global three-dimensional structure of the protein. Topology plays a crucial role in understanding the structure–function relationships and can provide insights into the mechanisms of protein folding and molecular recognition integral for the proteomic diversity responsible for sustaining life.

Connolly *et al.* discovered topological features of entangled protein backbones, where the polypeptide occasionally circles back to form a loop which is then threaded (Connolly et al., 1980). Building on this discovery, the Pierced Lasso Topology (PLT) was discovered in the structure of leptin (Haglund et al., 2012; Haglund, 2015). PLTs are found in proteins containing covalent-loops created by an intramolecular disulfide or isopeptide bond, where the polypeptide backbone is threaded through the macrocycle. The PLT is not unique to leptin and encompasses over 600 proteins represented in all kingdoms of life with 14 conserved biological functions (Niemyska et al., 2016; Simien and Haglund, 2021; Dabrowski-Tumanski et al., 2016). The final entangled topology for disulfide-linked PLTs is formed through three possible mechanisms: (*i*) Plugging where a free terminus crosses the covalent-loop (Sulkowska et al., 2009), (*ii*) Slipknotting where an internal residue of the free polypeptide chain crosses the covalent-loop (Sulkowska et al., 2009), and (*iii*) Belting where the protein folds and the covalent-loop wraps around the threaded element to form the covalent-linker (Simien and Haglund, 2021; Perego and Potestio, 2019). These disulfide-linked PLTs form with the catalysis of an oxidoreductase or at the redox potential *in vitro* and *in vivo*.

In this work, we utilize the founding member of the PLTs, leptin, to study the folding free energy landscape of the oxidized protein. Leptin is

* Corresponding authors.
E-mail addresses: ellinorh@hawaii.edu (E. Haglund), vitor.leite@unesp.br (V.B.P. Leite).

https://doi.org/10.1016/j.jsb.2023.108054
Received 4 October 2023; Received in revised form 27 November 2023; Accepted 4 December 2023
Available online 6 December 2023

a 146 amino acids protein forming a four-helix bundle motif with a single loop-crossing, a so-called L_1 topology (Haglund et al., 2012; Niemyska et al., 2016). Leptin is an ideal protein to study threading as the covalent-loop is 50 amino acids located at the C-terminal. The threaded element is 50 amino acids, leaving 50 amino acids at the N-terminus. Thus, independent of pathway, plugging or slipknotting, there is a 50/50 probability for 50 amino acids to thread through the covalent-loop in either mechanism. However, utilizing the crossing reaction coordinate C (Niemyska et al., 2016) we observe that leptin dominantly slipknots through the covalent-loop 99% of the time (Haglund et al., 2014) (Fig. 1).

In vitro folding experiments show a typical two-state behavior with a linear correlation between the chemical denaturant and the logarithm of the rate for folding (k_f) and unfolding (k_u) depicted as a chevron plot (Fig. 2). The observed rates are an average of both the plugging and slipknotting mechanisms in the ensemble experiments. In the case of leptin, the observed refolding rate is a direct measurement of the dominant slipknotting mechanism, as plugging only is observed 1% of the time. A previous computational study of a leptin variant, the N-loop protein, has been observed to predominantly plug (Haglund et al., 2017). This loop-variant has two amino acid substitutions increasing the size of the loop to 100 amino acids. The chevron plot shows no significant change in the refolding rate of the N-loop protein where the observed rate is dominated by the plugging mechanism (Fig. 2). Thus, distinguishing the two pathways is complex and requires ingenuity. Molecular dynamics (MD) simulations using the crossing reaction coordinate C are able to support our experimental data to detect loop-crossings. However, the C is not able to detect pre-knotted regions that dictate the threading event.

To further investigate the conformational changes and threading events on the folding trajectories of leptin, the Energy Landscape Visualization Method (ELViM) (Oliveira et al., 2019; Oliveira et al., 2014) is utilized. ELViM is a multidimensional projection tool (Ortigossa et al., 2022; Tribello et al., 2019) providing intuitive representations for the energy landscape of biomolecules. Originally designed for visualizing protein folding funnels in lattice models (Oliveira et al., 2014), this method has been successfully extended to more complex biological systems (da Silva et al., 2023; Oliveira et al., 2019, 2021; Sanches et al., 2022a,b). ELViM applies a robust metric to calculate structural distances between pairs of conformations sampled in MD simulations. The dimensionality of the system is reduced to generate an effective two-dimensional representation of the energy landscape that seeks to optimally maintain the original structural distances. The underlying hypothesis of multidimensional projection techniques is that reducing the dimensionality, while retaining relevant information, enables the extraction of significant features that may be concealed in high-dimensional spaces (Glielmo et al., 2021; Ortigossa et al., 2022; Tribello et al., 2019). In this study, we utilized ELViM to investigate the folding landscape of wild-type leptin. Analyzing folding trajectories of the wild-type allows for the detection of different pre-threaded conformation associated with the plugging or slipknotting mechanisms within the same structure. The results show that ELViM projection can discriminate the threading mechanisms, where plugging and slipknotting divides the folding landscape into roughly two distinct pathways. In addition, a new reaction coordinate is proposed aiming to classify the propensities of unthread conformations to fold by plugging or slipknotting. As ELViM optimally retains structural similarity in a local neighborhood, the ELViM projection allows the identification of significant pre-knotted regions. Conformations within these regions were utilized to obtain contact maps to shed light on topological clusters that may control the threading mechanism in leptin.

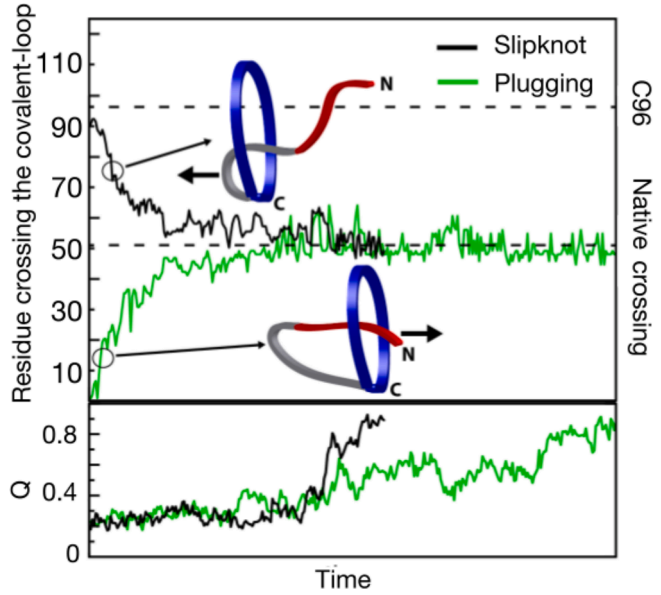


Fig. 1. Plugging and slipknotting trajectories of leptin from (Haglund et al., 2012; Haglund et al., 2014). *In silico* trajectories of wild-type leptin depicting the plugging (green) and slipknotting (black) pathways. The top panel depicts the time for threading versus the residue crossing the covalent-loop. The bottom panel depicts the time versus the reaction coordinate Q (number of nativeness). The crossing coordinate (C) was used to plot the data (Niemyska et al., 2016).

2. Methods

2.1. Structure Based Model (SBM)

To provide a more detailed interpretation about the folding landscape of leptin, coarse-grained molecular dynamics simulations based on the structure-based Cα model (SBM-Cα) taken from (Haglund et al., 2014) were analyzed. The SBM-Cα defines the potential in such a way that the minimum energy value is attributed to the protein's native structure. Native contacts are determined using the shadow contact map algorithm (Noel et al., 2012) with the standard parameters defined by the SMOG server (Noel et al., 2010). The simulations were performed in Gromacs version 4.5.3 (David Van Der Spoel et al., 2005). Additional details about the simulation can be found in (Haglund et al., 2012; Haglund et al., 2014).

2.2. Energy Landscape Visualization Method - ELViM

ELViM can generate intuitive low-dimensional representations of the multidimensional conformational phase-space of biomolecules. This projection relies on an estimate of structural distances, or dissimilarities, between pairs of conformations populating a high-dimensional space. ELViM uses a dissimilarity metric related to the fraction of native contacts, Q_w (Hardin et al., 2003; Samuel et al., 2006). Considering two conformations, k and l, their structural similarity is defined as

$$q_w^{k,l} = \frac{1}{N_{pairs}} \sum_{i,j \in pairs} \exp\left[\frac{-\left(r_{i,j}^k - r_{i,j}^l\right)^2}{2\sigma_{i,j}^2}\right], \quad (1)$$

where $r_{i,j}^k$ is the Euclidean distance between atoms, i and j, in the conformation k. Therefore, the metrics takes the difference of corresponding internal distances in conformations k and l. In this study, we considered only distances between Cα carbons. The smoothing parameter $\sigma_{i,j}$ sets the similarity resolution and is traditionally given by $\sigma_{i,j} = \sigma_0 |i-j|^{\epsilon}$, with $\epsilon = 0.15$ (Hardin et al., 2003; Samuel et al., 2006). Choosing different values of σ_0 may fine-tune the metric resolution to different systems. Here, we adopted $\sigma_0 = 0.8$. The dissimilarity is defined by $\delta^{k,l} = 1 - q_w^{k,l}$, which is unitless and equals zero, for identical conformations, and tends to one for very different conformations. The

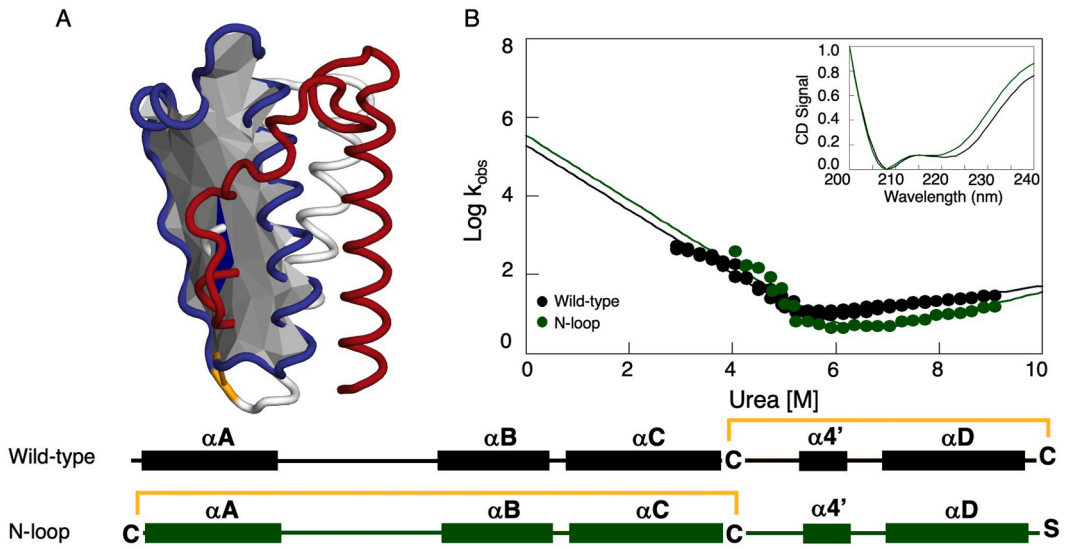


Fig. 2. The structure and folding kinetics of leptin. A) The structure of leptin showing the pierced lasso topology (Gierut et al., 2017). The covalent-loop (blue) is kept intact by a disulfide bond between C96 and C146 (yellow). The covalent-loop is threaded by residue 51–95 (white) leaving the N-terminal residues 50–1 unthreaded (red). B) Chevron plot of wild-type (black) and N-loop leptin (green). The inlet depicts the CD spectrum showing the helical propensity of t.he two variants.

multidimensional projection consists of representing each conformation by a dot on the plane such that their pairwise Euclidean distance, $d^{k,l}$, optimally corresponds to their high-dimensional dissimilarity, $\delta^{k,l}$. For this purpose, ELViM uses the force scheme (Tejada et al., 2003) algorithm to minimize a cost function, $E = \sum_{K,l}|\delta^{k,l} - d^{k,l}|$. The optimization procedure involves iteractively minimizing the cost function E by adjusting the distances between a randomly chosen reference conformation k and all other conformations. For each conformation l, a vector $\vec{r}_{K,l}$, representing the displacement between k and l, is calculated, and the dot l is moved in the direction of $\vec{r}_{K,l}$ that minimizes E by a distance $\Delta|\delta^{k,l} - d^{k,l}|^2$. Δ is a learning rate parameter taking values between 0.4 and 0.08. An iteration is completed when every dot has been selected as reference, and the total number of iterations equals the square root of the number of projected conformations.

2.3. Leptin expression and purification

Leptin was cloned into a pET-3A vector (GenScript, New Jersey, USA), transformed, overexpressed, and purified as previously described (Haglund et al., 2012; Simien et al., 2023). The loop-variant, N-loop, was previously designed as (Haglund et al., 2017). The N-loop variant links the N-terminal to C96, forming a 100 amino acid long covalent loop where part of the C-terminus is threaded through.

2.4. Circular Dichroism (CD)

The CD spectra were collected on a Chirascan V100 spectrometer (Applied Photophysics, Leatherhead, U.K.). The spectra were collected between 200–250 nm with a final protein concentration of 0.1 mg/ml using a 1 mm cuvette in 10 mM phosphate buffer pH 7.4 at 25 °C.

2.5. Kinetic measurements of leptin variants

Kinetic measurements were collected on a SX20 stopped-flow spectrometer (Applied Photophysics, Leatherhead, U.K.). Tryptophan fluorescence was used as the folding reaction probe by 280 nm LED excitation, and 350 nm emission monitored utilizing a 295 nm cutoff filter. Folding reactions were collected using a final protein concentration of 1 mM in 10 mM Mes pH 6.3 at 25 °C. Urea from 0–10 M was used to unfold and refold the protein and data was fitted to:

$$\begin{aligned} \log k_{obs} &= \log\left(k_f + k_u\right) \\ &= \log\left[10^{\left(\log k_f^{H_2O} + m_f\left(Urea\right)\right)} + 10^{\left(\log k_u^{H_2O} + m_u\left(Urea\right)\right)}\right] \end{aligned} \quad (2)$$

where $k_f^{H_2O}$ and $k_u^{H_2O}$ are the extrapolated values of the refolding and unfolding reaction in water, and m_f and m_u are the constants for the exposed surface area (Fersht, 1999). The kinetic data is represented as a chevron plot relating chemical denaturant concentration to the logarithmic observed rate of folding and unfolding.

3. Results

Studying how protein can self-tangle *in vitro* remains a challenge. In this work, we utilize the PLT protein leptin to conduct kinetic experiments to investigate if unfold and unthread can occur. In combination with MD simulations and the visualization tool ELViM, we were able to distinguish key components for the early knot formation and structural relationships within local groups of plugging and slipknotting on the free energy landscape.

3.1. In vitro kinetics of leptin

The kinetics of proteins probes the folding free energy landscape unraveling the complex relationship between protein structure and function. A chevron plot is a graphical representation of folding and unfolding rates *In vitro*. By plotting the logarithm of the observed rates for folding ($\log k_f$) and unfolding ($\log k_u$) against denaturant concentrations, the plot depicts a linear relationship of the change in observed rates (see Fig. 2B). The straight limbs indicated that leptin folds in a two-state manner from the denatured state (D) to the native state (N), without populating intermediate states along the folding pathway. No significant perturbations are observed in folding rate between the two proteins, with $\log k_f^{H_2O}$ 5.17 ± 0.12 and 5.44 ± 0.44 (Fig. 2B). The change in stability, ΔG_{D-N} from 6.82 ± 0.04 kcal/mol and ΔG_{D-N} from 7.38 ± 0.10 kcal/mol for wild-type and N-loop protein respectively, is attributed to the stabilizing effects of connecting helix A, through the introduced disulfide bond, to the four-helix bundle. The change in solvent exposed area is attributed to the change in loop size, where the m_{D-N} for the wild-type is 1.01 and N-loop is 1.73. A CD spectrum was collected to evaluate perturbation of the secondary structure (Fig. 2B). The spectra

coincide well with classical α-helical spectra with two minima at 208 nm and 222 nm, indicating that the secondary structure is conserved.

3.2. Slipknotting versus plugging of leptin

Determination of the threading mechanism is essential for understanding the complex free energy landscapes of proteins with entangled backbones. Folding trajectories were collected for wild-type leptin (Zhang et al., 1997) and analyzed using the crossing reaction coordinate (C) (Fig. 1). First, the reaction coordinate detects the covalent linker and triangulates the minimal surface spanning the covalent-loop. Second, the algorithm detects the crossing points across the minimal surface. Finally, the algorithm reduces the loop crossing from artificial loop crossings (Niemyska et al., 2016). The data shows that leptin can thread using both a plugging and slipknotting mechanism, but predominantly slipknots through the covalent loop (Haglund et al., 2014). ELViM was utilized to visualize and distinguish the different conformational states populated by the two possible pathways (Oliveira et al., 2019).

The ELViM effective phase-space for the wild-type protein, where each dot represents a sampled conformation, is shown in Fig. 3A. The axes are removed for clarity as the structural dissimilarity is represented as pairwise distances. The color assigned to each dot represents the value of the fraction of native contacts, Q. Thus, the folded basin ($Q > 0.9$) is located in the middle right region of the ELViM projection, while unfolded structures are distributed at the outer left region. It is important to note that the points are not homogeneously distributed throughout the projection due to the optimization procedure, which tends to group similar structures into local dense clusters. However, the reaction coordinate Q varies smoothly across the projection, indicating dense pathways connecting the unfolded and folded basins.

To investigate whether the ELViM projection can distinguish between the plugging and slipknotting folding pathways, the projection points corresponding to threaded structures are colored based on the crossing reaction coordinate C (Fig. 3B). Interestingly, the ELViM projection reveals that structures on the plugging pathway ($C \lessapprox 35$, shades of blue) are concentrated in a well-defined region. On the other hand, conformations from the slipknotting pathway ($C \gtrapprox 65$, yellow to red) are spread throughout the inner center of the projection, with the highest density observed at the bottom region, in opposition to the plugging pathway. Thus, the ELViM projection suggests two distinct pathways for leptin based on the threading event from the unfolded unthreaded states to the native structure. While the plugging pathway is more constrained to a specific region, the slipknotting route appears to be accessible from a larger basin in the projection.

Fig. 3C shows representative structures highlighting the folding event where the plugging pathway is depicted at the top and slipknotting at the bottom with index P_i and S_i, respectively. All dots from Fig. 3C can be directly compared in Fig. 3A. Six representative structures are depicted, each of which demonstrates the threading mechanism of plugging and slipknotting conformations. Based on the structures where the N-terminal has crossed the loop (plug) or where a hairpin or loop threads through the closed loop (slipknot), the crossing reaction coordinate is an appropriate reaction coordinate to visualize which structure is being favored. Nevertheless, the unfolded ensemble provides essential information on specific conformations controlling the threading mechanism and the propensity for plugging as opposed to slipknotting. An in-depth exploration of the unfolded ensemble was conducted to clarify the factors that confer advantages to slipknot transitions over plug transitions.

To classify the unfolded ensemble and early threading events, we

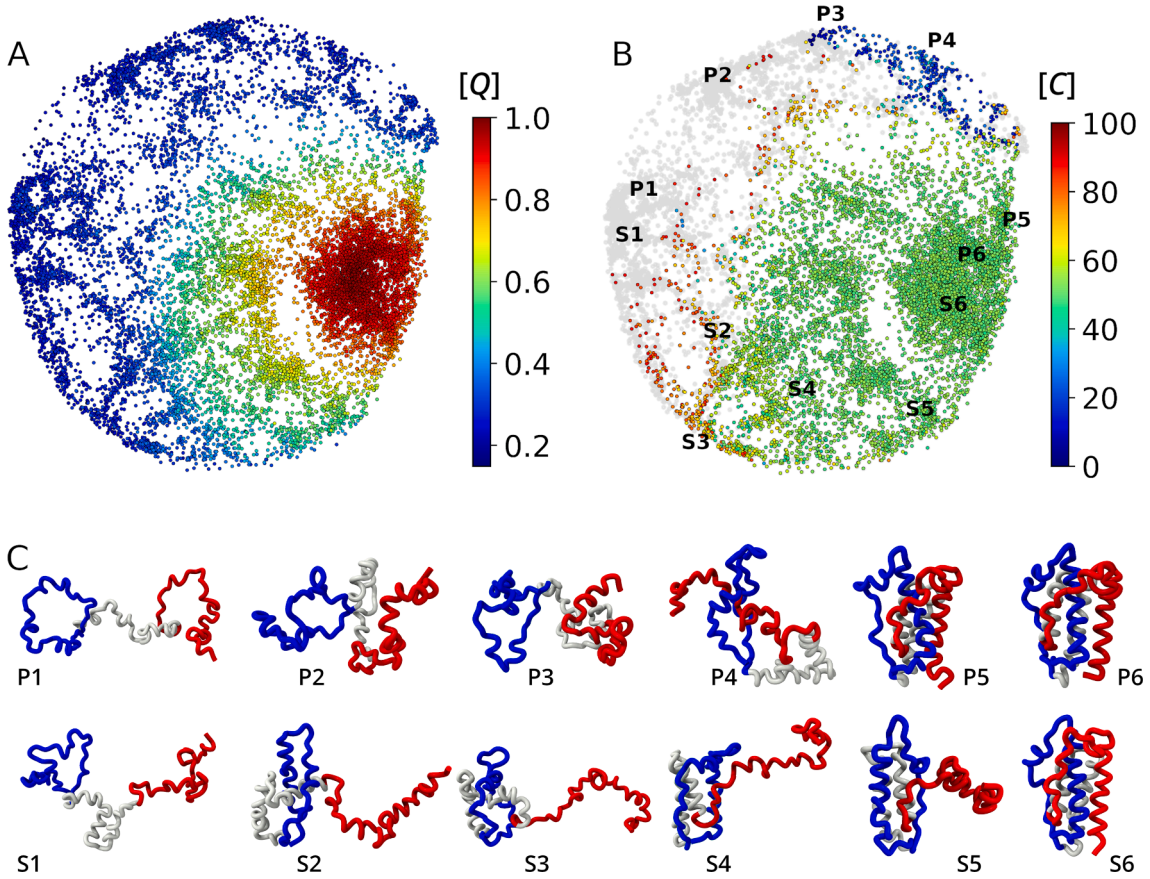


Fig. 3. ELViM effective phase-space for wild-type leptin. Each point corresponds to a conformation state and is colored based on (A) the fraction of native contacts [Q], and (B) the crossing reaction coordinate [C]. Unthreaded structures are depicted in gray in (B). The axes are removed for clarity. (C) Represents conformations for the plugging (from unfolded, P1, to folded, P6, state) and slipknotting (from unfolded, S1, to folded, S6, state) pathways.

defined a new reaction coordinate (Λ) based on selected distances according to the Equation:

$$\Lambda = d_{nl} - (d_{cl} - d_{nc} + 1)/2. \quad (3)$$

The terms d_{nl},d_{cl}, and d_{nc} correspond to distance between the N-terminal residue and center of mass of the covalent-loop, centers of mass of the protein and the covalent-loop, and the N-terminal residue and the center of mass of the protein, respectively. A schematic representation of unthreaded structures and early cross loop conformation, slipknot (i and ii) and plug (iii and iv) are shown in Fig. 4A. When utilizing only two distances, i.e., the distance between the N-terminal and the center of mass of the covalent-loop (d_{nl}), the early cross loop conformation may not be distinguishable, introducing uncertainty in the interpretation of the results (Figs. 4iii)). In Figs. 4Aii) and 4A(iv), based on the d_{nc} and d_{cl}, the unthreaded structures and the early crossing transitions may be distinguished. Ideally, the difference between these terms (d_{cl}-d_{nc}) should be positive and negative for early plugging and slipknotting transitions, respectively.

In Fig. 4B, ELViM projection shows the structures (each dot) in the early loop crossing conformations and threaded structures with $C > 65$ and $C < 35$. The ELViM projections in Fig. 3 are represented by two reaction coordinates, Q and C, in which Q provides information about the folding event and C the threaded conformations. The Λ reaction coordinate is able to fill the gap between these two reaction coordinates, providing information about the early loop crossing conformations. According to Fig. 4, there are two clear regions (red and blue) that include unfolded and partially folded structures. The small region corresponding to the plugged and slipknotted conformations was amplified with the new reaction coordinate, Λ (Fig. 4). The fraction of native contacts (Q) and the crossing reaction coordinate (C) are well-established parameters. However, the motivation to develop the new reaction coordinate (Λ) was to bridge the gap between Q and C to explore the early loop crossing conformations. Furthermore, (Λ) may also elucidate unfolded conformations determining the preference for plugging versus slipknotting transitions observed in leptin.

To investigate the distinctions in contact propensities along the two threading pathways, we designated a region I on the plugging pathway and a region II on the slipknotting pathway, as illustrated in Fig. 5. These regions encompass conformations in the early stages of threading, along with neighboring unthreaded conformations that are structurally similar based on the ELViM metrics. Subsequently, we computed contact maps (with a Cα distance cut-off of 10 Å) for each conformation within our selected regions and determined the contact frequency, defined as the average contact map and denoted by $f_{i,j}$. This contact frequency equals one if the contact between alpha carbons (i,j) is present in all conformations within the selection. By subtracting the contact frequencies observed on the plugging pathway from those on the slipknotting pathway, denoted as $f^{II}_{i,j} - f^{I}_{i,j}$, we aimed to identify critical contacts defining the threading mechanism. For comparison, we included a contact map displaying native contacts, which is depicted in gray in the lower half of the map.

The difference in contact frequency for unthreaded conformations is depicted in Fig. 5A. Contacts depicted in blue are associated with the plugging pathway, while those in red are prevalent on the slipknotting pathway. The most significant difference in the contact frequency for the unthreaded conformations, highlighted in orange, corresponds to native and non-native contacts between residues L45-L49 and K53-L58. These contacts seem to favor the interactions that yields a plugging mechanism. On the other hand, for threaded conformations (Fig. 5B), the difference in the contact frequency shows two very distinct regions, blue and red, indicating different folding routes for the different mechanisms.

4. Discussion

The introduction of threaded topologies adds complexity to the folding free energy landscape while diversifying the biological function and lifetime of proteins. A polypeptide chain may spontaneously self-entangle, on a biologically relevant timescale from milliseconds to seconds, into the active three-dimensional structure. Encoded in its amino acid sequence is not only the final native structure but also the mechanism to achieve the entanglement. Entangled biopolymers remain one of the most challenging interdisciplinary questions requiring a combination of mathematical knot theory, physics, chemistry, and biology.

Analyzing MD simulation trajectories using ELViM, we are able to detect distinct conformations associated with plugging and slipknotting pathways in leptin. Our findings reveal specific characteristics of conformational states in the early stages of plugging ($C < 35$) and the early slipknotting stages ($C > 65$) within the ELViM projection. Con-

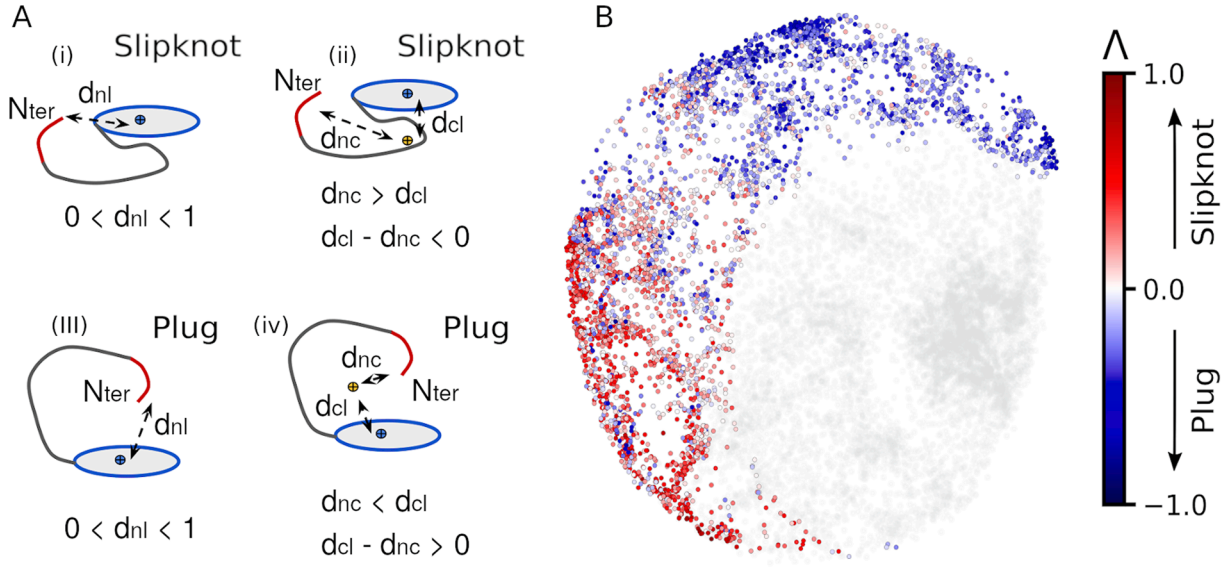


Fig. 4. (A) Schematic representation illustrate the Λ reaction coordinate. Here, d_{nl}, d_{nc}, and d_{cl}, correspond to the distance between the N-terminal (Nter) and the center of mass of the covalent-loop, the N-terminal and the center of mass of the protein, and the center of mass of the protein and the center of mass of the covalent-loop, respectively. (B) The plot depicts unthreaded structures and conformations with an $C > 65$ and $C < 35$ (red and blue) and threaded native conformations with an index of $35 < C < 65$ (gray color). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

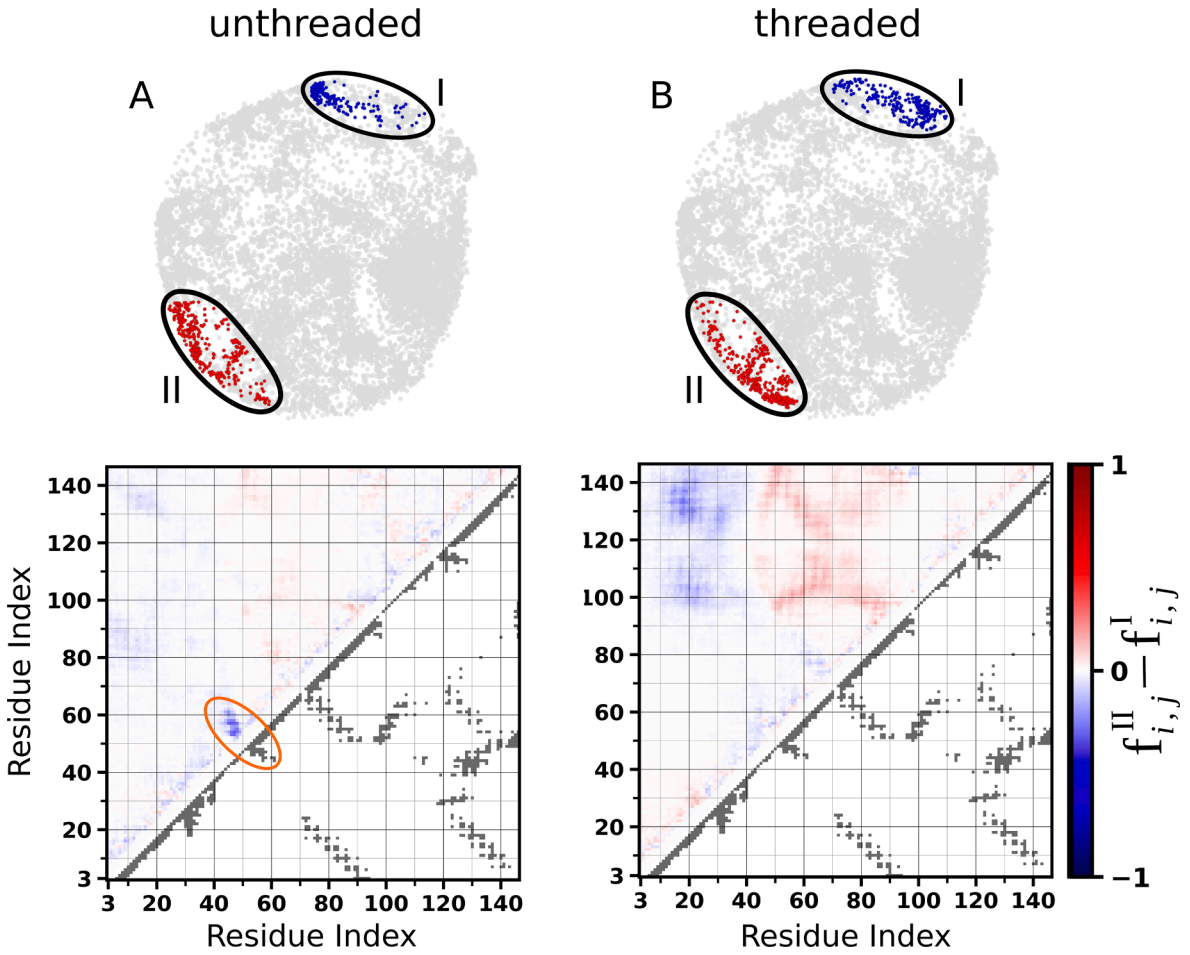


Fig. 5. The ELViM projection for regions I and II for plug (blue) and slipknot (red), respectively. (A) depicts the data for the unthreaded conformations while B) depict the threaded conformations in regions I and II for plug (blue) and slipknot (red), respectively. The bottom panels correspond to the contact frequency difference (upper triangle) and native contact maps (lower triangle). The frequencies $f^{I}_{i,j}$ and $f^{II}_{i,j}$ are calculated by taking the average for all threaded or unthread conformations from regions I and II, respectively. Native contacts are calculated considering only the Cα carbons and a cut-off equal to 10 Å.

formations in the non-knotted and early plugging stages are confined to a specific region of the ELViM projection, while early crossings on the slipknotting pathway exhibit a more dispersed distribution, with the highest density found in a region diametrically opposed to the plugging pathway. To further characterize the two threading pathways, we conducted an analysis using contact maps designed to differentiate between their mechanisms. For this purpose, we selected two distinct regions within the ELViM projection, each encompassing structurally similar unthreaded and early threaded conformations. This selection aimed to minimize noise arising from misplaced data points and facilitate the identification of prevalent contacts that might contribute to the selection of different threading pathways. Our results clearly indicate that the contact frequencies within regions *I* and *II* are initiated by the formation of a hairpin loop, between residues L45-L49 and K53-L59 in the unfolded basin. Thus, promoting the plugging mechanism of the dominant slipknotting pathway. These differences, which are associated with sets of native and non-native contacts, can be experimentally probed by changing their interaction strength, elucidating the dominant folding mechanism.

It is important to note that multidimensional projection techniques can result in some misplacement of data points, primarily due to the inherent mathematical limitations of representing all pairwise distances from the high-dimensional phase-space onto a two-dimensional plane (Ortigossa et al., 2022). Additionally, since the optimization procedure is stochastic, different projections may yield slightly different representations of the effective phase-space. Moreover, the presence of entangled conformations adds complexity, as some conformations may be structurally similar but topologically different. In this context, we assert that the ELViM projection is capable of effectively distinguishing between the two threading pathways, offering valuable atomistic insights into the free energy landscape of threaded proteins.

5. Conclusion

Topological constraints in proteins introduce complexity on the folding pathways and may encompass a threading event. The multidimensional projection tool ELViM is able to visualize and distinguish structures with different topologies with a smooth conformational transition between them. This information is used to separate folding pathways associated with plugging and slipknotting pathways in leptin.

To our knowledge, there are only a few experimental techniques that may be able to detect small populations of separate conformational states such as single molecule Förster resonance energy transfer (FRET) experiments and Paramagnetic Resonance Enhancement (PRE) Nuclear Magnetic Resonance (NMR) experiments. A concern in utilizing these experiments is the induction of detection probes without interring with the threading mechanism. ELViM provides a novel tool for the design of such experiments as it is able to predict where to introduce probes to monitor treading as well as aiding in the analysis providing information on a molecular level that is not possible to detect experimentally. We propose to utilize multiple FRET pairs or NMR PRE *in vitro* experimental techniques as they may be able to probe the proximity of the threaded

element to the covalent-loop. The observed distances from these experiments can be directly compared to the distances from ELViM to quantify the threading landscape. However, *in vitro* experiments will not capture the orientation of the polypeptide chain in the unfolded state. Thus, ELViM is able to support the *in vitro* analysis, providing molecular details describing the orientation and the mechanism of threading. Furthermore, ELViM may be used to enhance the experimental design to specifically determine where to introduce probes for *in vitro* experiments to avoid clashes that may interfere with the threading and folding of the protein.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgements

VBPL is supported by Brazilian agencies FAPESP (Grant 2023/02219–1) and CNPq (Grant 310017/2020–3). The research reported by the Haglund group is supported by the National Science Foundation award number CHE2145906 and the Hawaii Community Foundation award number HCF40846 (013357–00002).

References

Connolly, Michael L., Kuntz, I.D., Crippen, Gordon M., 1980. Linked and threaded loops in proteins. Biopolymers 19 (6), 1167–1182.

Haglund, Ellinor, Sulkowska, Joanna I., He, Zhao, Feng, Gen-Sheng, Jennings, Patricia A., Onuchic, José N., 2012. The unique cysteine knot regulates the pleotropic hormone leptin. PLOS ONE 7(9), 1–13.

Haglund, E., 2015. Engineering covalent loops in proteins can serve as an on/off switch to regulate threaded topologies. J. Phys. Condens. Matter 27 (35), 354107.

Simien, Jennifer Michelle, Haglund, Ellinor, 2021. Topological twists in nature. Trends Biochem. Sci., 46(6):461–471.

Dabrowski-Tumanski, Pawel, Niemyska, Wanda, Pasznik, Pawel, Sulkowska, Joanna I., 2016. LassoProt: server to analyze biopolymers with lassos. Nucl. Acids Res., 44 (W1):W383–W389.

Sulkowska, Joanna I., Sulkowski, Piotr, Onuchic, José, 2009. Dodging the crisis of folding proteins with knots. Biophys. J., 96(3):81a.

Ortigossa, E.S., Dias, F.F., Nascimento, D.C.D., 2022. Getting over High-Dimensionality: How Multidimensional Projection Methods Can Assist Data Science. Appl. Sci. 12 (13), 6799.

Perego, Claudio, Potestio, Raffaello, 2019. Searching the optimal folding routes of a complex lasso protein. Biophys. J. 117 (2), 214–228.

Haglund, Ellinor, Sulkowska, Joanna I., Noel, Jeffrey K., Lammert, Heiko, Onuchic, José N., Jennings, Patricia A., 2014. Pierced lasso bundles are a new class of knot-like motifs. PLoS Comput. Biol., 10(6):1–11.

Haglund, Ellinor, Pilko, Anna, Wollman, Roy, Ann Jennings, Patricia, Onuchic, José Nelson, 2017. Pierced lasso topology controls function in leptin. J. Phys. Chem. B, 121(4), 706–718.

Oliveira, Antonio B., Yang, Huan, Whitford, Paul C., Leite, Vitor B.P., 2019. Distinguishing biomolecular pathways and metastable states. J. Chem. Theory Comput. 15 (11), 6482–6490. PMID: 31618581.

Oliveira, Antonio B., Fatore, Francisco M., Paulovich, Fernando V., Oliveira, Osvaldo N., Leite, Vitor B.P., 2014. Visualization of Protein Folding Funnels in Lattice Models. PLoS ONE 9 (7), e100861.

Sanches, M.N., Parra, R.G., Viegas, R.G., Oliveira, A.B., Wolynes, P.G., Ferreiro, D.U., Leite, V.B., 2022a. Resolving the fine structure in the energy landscapes of repeat proteins. QRB Discov. 3, e7.

Sanches, M.N., Knapp, K., Oliveira Jr, A.B., Wolynes, P.G., Onuchic, J.N., Leite, V.B., 2022b. Examining the ensembles of amyloid-β monomer variants and their propensities to form fibers using an energy landscape visualization method. J. Phys. Chem. B 126 (1), 93–99. PMID: 34968059.

Oliveira, Antonio B., Junior, Xingcheng Lin, Kulkarni, Prakash, Onuchic, José N., Roy, Susmita, Leite, Vitor B.P., 2021. Exploring energy landscapes of intrinsically disordered proteins: Insights into functional mechanisms. J. Chem. Theory Comput. 17 (5), 3178–3187. PMID: 33871257.

Glielmo, Aldo, Husic, Brooke E., Rodriguez, Alex, Clementi, Cecilia, Noé, Frank, Laio, Alessandro, 2021. Unsupervised Learning Methods for Molecular Simulation Data. Chem. Rev. 121 (16), 9722–9758.

da Silva, F.B., Martins de Oliveira, V., de Oliveira Junior, A.B., Contessoto, V., Leite, V.B., 2023. Probing the Energy Landscape of Spectrin R15 and R16 and the Effects of Non-native Interactions. J. Phys. Chem. B 127 (6), 1291–1300.

Gierut, Aleksandra M., Niemyska, Wanda, Dabrowski-Tumanski, Pawel, Sułkowski, Piotr, Sulkowska, Joanna I, 2017. Pylasso: a pymol plugin to identify lassos. Bioinform. 33 (23), 3819–3821.

Hardin, Corey, Eastwood, Michael P., Prentiss, Michael C., Luthey-Schulten, Zadia, Wolynes, Peter G., 2003. Associative memory Hamiltonians for structure prediction without homology: alpha/beta proteins. Proc. Natl. Acad. Sci. U.S.A. 100 (4), 1679–1684.

Niemyska, Wanda, Dabrowski-Tumanski, Pawel, Kadlof, Michal, Haglund, Ellinor, Sułkowski, Piotr, Sulkowska, Joanna I., 2016. Complex lasso: new entangled motifs in proteins. Sci. Rep. 6 (1), 36895.

Noel, Jeffrey K., Whitford, Paul C., Onuchic, José N., 2012. The shadow map: a general contact definition for capturing the dynamics of biomolecular folding and function. J. Phys. Chem. B 116 (29), 8692–8702.

Jeffrey K Noel, Paul C Whitford, Karissa Y Sanbonmatsu, and José N Onuchic. Smog@ctbp: simplified deployment of structure-based models in gromacs. Nucleic Acids Res., 38(suppl_2):W657–W661, 2010.

Van Der Spoel, David, Lindahl, Erik, Hess, Berk, Groenhof, Gerrit, Mark, Alan E., Berendsen, Herman J.C., 2005. Gromacs: fast, flexible, and free. J. Comput. Chem., 26(16):1701–1718.

Tejada, Eduardo, Minghim, Rosane, Nonato, Luis Gustavo, 2003. On improved projection techniques to support visual exploration of multi-dimensional data sets. Inf. Vis. 2 (4), 218–231.

Simien, Jennifer M., Orellana, Grace E., Phan, Hoa T.N., Yao, Hu., Kurth, Emily A., Ruf, Christine, Kricek, Franz, Wang, Qian, Smrcka, Alan V., Haglund, Ellinor, 2023. A small contribution to a large system: The leptin receptor complex. J. Phys. Chem. B 127 (11), 2457–2465.

Samuel S. Cho, Yaakov Levy, and Peter G. Wolynes. P versus Q: Structural reaction coordinates capture protein folding on smooth landscapes. Proc. Natl. Acad. Sci. U.S.A., 103(3):586–591, 2006.

Fersht, Alan, 1999. Structure and Mechanism in Protein Science: a guide to enzyme catalysis and protein folding. W.H. Freeman and Company, New York.

Tribello, Gareth, A., Gasparotto, Piero, 2019. Using dimensionality reduction to analyze protein trajectories. Front. Mol. Biosci. 6, 1–11.

Zhang, F., Basinski, M.B., Beals, J.M., Briggs, S.L., Churgay, L.M., Clawson, D.K., DiMarchi, R.D., Furman, T.C., Hale, J.E., Hsiung, H.M., Schoner, B.E., Smith, D.P., Zhang, X.Y., Wery, J.P., Schevitz, R.W., 1997. Crystal structure of the obese protein leptin-e100. Nature 387 (6629), 206–209.