

# UC Riverside

## UC Riverside Previously Published Works

### Title

Mechanistic Dissection of Conformational Transitions in a Bicyclic Peptide via Molecular Modeling and Deep Learning

### Permalink

<https://escholarship.org/uc/item/6ts3c919>

### Journal

Journal of Chemical Theory and Computation

### ISSN

1549-9618

### Authors

Hung, Ta I

Venkatesan, Raghu

Chang, Chia-en A

### Publication Date

2026-09-15

### DOI

10.1021/acs.jctc.6c01233

### Copyright Information

This work is made available under the terms of a Creative Commons Attribution License, available at <https://creativecommons.org/licenses/by/4.0/>

Peer reviewed

# Mechanistic Dissection of Conformational Transitions of Bicyclic Peptide via Molecular Modeling and Deep Learning

Ta I Hung <sup>#a</sup>, Raghu Venkatesan <sup>#a</sup>, and Chia-en Chang<sup>\*a</sup>.

<sup>a</sup> Department of Chemistry, University of California, Riverside, CA92521

Email: chiaenc@ucr.edu

<sup>#</sup>These authors contributed equally to the project.

## Abstract

Molecular conformations play a critical role in determining molecular properties, such as membrane permeability, binding affinity, and ultimately therapeutic efficacy. Experimental and computational approaches can characterize conformations and provide insight into why certain conformations are thermodynamically preferred over others. However, examining static conformations alone may not fully explain why subtle differences, such as a single LEU-to-ILE mutation in a bicyclic peptide, can produce markedly distinct conformational ensembles. Furthermore, analyzing transition pathways between conformations reveals the mechanisms that shape these ensembles. Here, we introduce a deep learning model, termed ICoN-v1, trained on molecular dynamics (MD) simulation data to learn the underlying physics that governs cyclic peptide conformational dynamics. We examined cyclic hexapeptides with Nuclear Magnetic Resonance (NMR)-determined structures, and MYC-targeting bicyclic peptides, which are stereo-diversified or have a single LEU-to-ILE mutation. By following minimum-energy pathways in the latent space constructed by ICoN-v1, we efficiently generated smooth conformational transition paths. These pathways reveal sequential sets of concerted backbone and side-chain torsional rotations moving between energy minima. Notably, smooth transition pathways that are absent from MD output can be observed using ICoN-v1. Our results identify various sets of concerted torsional motions that are nonlinearly combined during conformational transitions and reveal the key residues governing each stage of the transition, thereby elucidating how the observed conformations are generated and informing molecular design.

## Introduction

Understanding how molecules reorganize their conformations is fundamental to decoding how molecules function, especially for cyclic peptides, whose transitions between distinct low-energy states often require highly coordinated, concerted torsional rotations and transient intermediate states<sup>1-3</sup>. These conformational transitions govern molecular flexibility and enable cyclic peptides to adopt diverse biological properties including membrane permeability, binding affinity, and therapeutic efficacy<sup>4-7</sup>. Even single-point mutations can substantially alter these pathways, reshape the conformational free-energy landscape and transition pathways, ultimately affecting binding and biological activity. More than 40 cyclic peptide therapeutics were approved over the past two decades highlight an important accelerating trajectory in advancing this class of molecules<sup>8,9</sup>. The mechanistic details of how cyclic peptides reorganize their torsions during target recognition remain poorly understood. Their structural flexibility enables adoption of multiple non-planar conformations that facilitate strong and selective interactions with traditionally “undruggable” protein targets<sup>10,11</sup>. Furthermore, cyclization reduces conformational entropy penalties upon binding and partially bound conformations can retain water molecules to reduce desolvation penalties<sup>4</sup>. In addition, cyclic peptides can enhance membrane permeability by forming intramolecular hydrogen bonds (H-bonds) that shield polar residues<sup>2,3,12</sup>. Their large and structured surface areas enable high binding specificity and affinity to their protein targets<sup>11</sup>. However, capturing these collective dynamics is challenging for both experimental and computational methods, limiting rational design of potent cyclic peptides<sup>13</sup>. A deeper understanding of residue-level motions, concerted torsion angle transitions, and conformational dynamics is crucial to uncover their underlying thermodynamic properties and guide rational design of next-generation therapeutics and uncover their underlying thermodynamic properties.

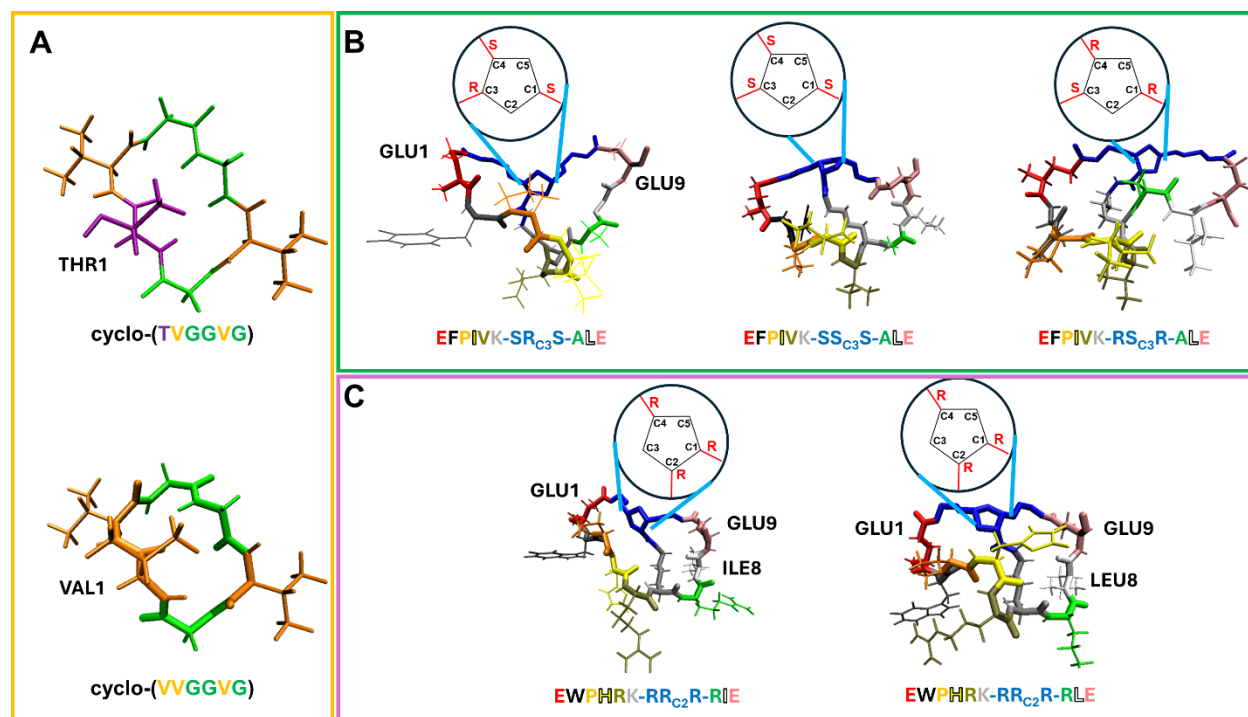
An experimentally determined structure of a cyclic peptide offers atomic-level structural details but provides limited temporal insights into their motions. For instance, X-ray crystallography typically reveals static conformations. Nuclear magnetic resonance (NMR) spectroscopy offers the ability to observe conformational ensembles in solutions, but it usually provides incomplete representation of conformations<sup>12,14</sup>. To overcome these limitations, computational approaches, particularly molecular dynamics (MD) simulations, have been widely used to explore protein

conformations. In addition to classical MD simulations, enhanced sampling techniques sample molecular conformations more efficiently<sup>15,16</sup>. However, the resulting trajectories do not inherently reveal and explain the relationship between peptide sequence and conformational transitions or the underlying physics. A conformational transition in a bicyclic peptide cannot usually be achieved by rotating a single torsion angle. Instead, multiple torsion angles must rotate simultaneously and sequentially. However, determining when and how an individual residue contributes to the conformational transition is difficult. Therefore, a more advanced approach is needed to provide mechanistic understanding to efficiently design cyclic peptides.

Artificial intelligence, in particular deep learning (DL), offers an alternative for modeling conformational transitions. Generative models with coarse-grained or protein backbone-only representations utilize a neural network trained on conformations obtained from MD simulations. These models can learn the probability distribution of the training set and allow for sampling novel conformations from the learned distribution, bypassing the sampling bottleneck posed by energy barriers of molecular simulation<sup>17–20</sup>. However, use of coarse-grained or protein backbone-only representations limit models' ability to resolve critical side chains, functional-group interactions, and hydrogen bond networks that determine the conformational dynamics. Our previous study introduced the Internal Coordinate Net (ICoN) model, which leverages vector features to present Bond-Angle-Torsion (BAT) coordinates with a fully connected neural network architecture to achieve efficient conformational sampling for an intrinsically disordered amyloid- $\beta$ 1–42 monomer at full atomistic detail<sup>21</sup>. Because bicyclic peptides have inherently complex coordinated torsional rotations, existing DL approaches that fail to sophisticatedly incorporate side-chains rotations in the models cannot accurately capture their concerted torsional rotations for dissecting mechanisms of the conformational transition of bicyclic peptides.

In this study, we applied a newly developed, physics-based deep learning model, ICoN version 1 (ICoN-v1), to identify transient conformations of cyclic peptides and characterize their conformational transition pathway. Using classical MD trajectories as the training dataset, ICoN-v1 combines transformer architectures with physics-based loss functions derived from molecular mechanics (MM) energy terms to efficiently generate atomistic conformations along smooth and

physically interpretable transition pathways between local energy minima in latent space. Analysis of these pathways revealed that intramolecular interactions and coordinated torsional rotations are key drivers of ring puckering and conformational reorganization, providing mechanistic insight into how mutations reshape conformational landscapes and transition mechanisms. Here, we first investigate conformational transition pathways for two cyclic hexapeptides, cyclo-(TVGGVG) and cyclo-(VVGGVG), with conformations derived from reported NMR and MD simulations<sup>22</sup>, to explain why a single-point mutation yields markedly different ring conformations. We then focus on revealing fundamental chemical principles governing conformational transitions for five MYC-targeting bicyclic peptides, reported in our previous publication<sup>23</sup> (**Figure 1**). Although our previous study showed that the different stereo-structures and peptide sequences strongly affect bicyclic peptides conformations and MYC binding, how chirality and individual residues govern distinct conformations remained unknown<sup>23</sup>. By using ICoN-v1 to generate smooth and physically interpretable transition pathways, we identified energetic contributions of individual residues and the coordinated torsional rotations that drive diverse conformational transitions. This identification explained the effects of chirality and peptide sequences that lead to distinct conformational transitions. The detailed mechanistic insights explain how these molecules move and inform rational design of cyclic peptides.



**Figure 1.** Representative conformation of seven cyclic peptides: **(A)** cyclo-(TVGGVG) and cyclo-(VVGGVG). **(B)** EFPIVK-**S**R<sub>c3</sub>**S**-ALE, EFPIVK-**S**S<sub>c3</sub>**S**-ALE, and EFPIVK-**R**S<sub>c3</sub>**R**-ALE with same amino acid sequence but epimeric at C1, C3 and C4 positions. **(C)** EWPHRK-RR<sub>c2</sub>R-RLE and EWPHRK-RR<sub>c2</sub>R-RIE with single point mutation from LEU to ILE at residue 8. The detailed 2D structures of bi-cyclic peptide with the stereochemistry (S and R) at each chiral center are shown in **Figure S1**.

## Methods

### Molecular Structure Representation

#### *Bond-Angle-Torsion (BAT) Internal Coordinate*

All conformations (peptides and linker) are defined using the BAT internal coordinate system, which is also known as a Z-matrix representation. Unlike Cartesian coordinates, BAT coordinates accurately capture the circular nature of molecular dihedral rotations<sup>24</sup>. In fact, BAT is the only coordinate system that can accurately describe the molecular motions when our simulation is performed using MM force field (See detail in Supplementary Information for MD simulation protocol).

BAT is defined as sets of root atoms, and the positions of other atoms depend on the previously defined atoms bonded in sequence. For instance, Atom 1 is at the origin; atom 2 is defined by a

bond length ( $b_2$ ) and placed on the z-axis; and atom 3 is defined by a bond length ( $b_3$ ), bond angle ( $a_3$ ) and placed on the x-z plane, and thus we eliminate six degrees of freedom. The total degrees of freedom of a molecule with  $N$  atoms, represented with internal coordinates, is  $3N-6$ . Atoms 1–3 are termed root atoms for this molecule. Atoms  $i > 3$  are defined by  $(b_i, a_i, \phi_i)$ , where  $\phi$  are torsion angles. For example, atom 4 is represented by  $(b_4, a_4, \phi_4)$  (**Figure 2B**). The internal coordinates are then constructed in a tree. Starting from root atoms, all subsequent atoms are defined sequentially.

Bonds and angles do not significantly contribute to the change of protein conformation. Torsional rotations dominate the conformational transition, and thus only torsions will be used for training the model. This significantly reduces the number of features from  $3N-6$  to roughly  $N-2$  while preserving complete structural and dynamic information.

To accurately perform mathematical operations using BAT internal coordinates, we needed to address the discontinuity issue at  $(\pm 180^\circ)$  for the torsional rotations. For instance, dihedral rotation from  $179^\circ$  to  $-179^\circ$  is actually a  $2^\circ$  shift in angular space. However, the basic arithmetic operation will result in a large rotation,  $179^\circ - (-179^\circ) = 358^\circ$ <sup>25</sup>. To address the discontinuity issue, we introduced angle-aware arithmetic by mapping each torsion angle  $\theta$  on the unit circle:  $(\cos \theta, \sin \theta)$ .

### *Primary Torsion Angles*

Because atoms whose torsions share a single bond rotation move together (Figure S2), assigning primary torsion angles allows for a more accurate description of torsional rotation and also reduces the number of features from roughly  $N-2$  to  $(N-2)/3$ . Primary Torsion Angles are constructed automatically by using breadth-first search (BFS) algorithm. For each rotatable bond, the dihedral defined by four heavy atoms is chosen as the primary torsion, while all remaining dihedrals sharing the same central bond are treated as secondary torsions which are excluded in the input features for training.

As illustrated in Figure S2, using the peptide EFPIVK-SR<sub>C3</sub>S-ALE as example, the root atoms (C<sub>1</sub>, C<sub>2</sub>, C<sub>3</sub>, and C<sub>4</sub>) are circled in purple. The first rotatable bond is C<sub>4</sub>–C<sub>2</sub>, which defines two dihedral angles: N<sub>1</sub>–C<sub>4</sub>–C<sub>2</sub>–C<sub>1</sub> and O<sub>1</sub>–C<sub>4</sub>–C<sub>2</sub>–C<sub>1</sub>. Since N<sub>1</sub>–C<sub>4</sub>–C<sub>2</sub>–C<sub>1</sub> follows the heavy-atom backbone traversal, it is selected as the primary torsion (circled in orange), whereas O<sub>1</sub>–C<sub>4</sub>–C<sub>2</sub>–C<sub>1</sub> is classified as a secondary torsion (circled in blue). The BFS search then proceeds to the next bond, N<sub>1</sub>–C<sub>4</sub>, where two dihedrals are identified: C<sub>5</sub>–N<sub>1</sub>–C<sub>4</sub>–C<sub>2</sub> and H<sub>1</sub>–N<sub>1</sub>–C<sub>4</sub>–C<sub>2</sub>. The heavy-atom dihedral C<sub>5</sub>–N<sub>1</sub>–C<sub>4</sub>–C<sub>2</sub> is selected as the primary torsion (circled in orange), while H<sub>1</sub>–N<sub>1</sub>–C<sub>4</sub>–C<sub>2</sub> is classified as secondary (circled in blue). Continuing to the bond C<sub>5</sub>–N<sub>1</sub>, three dihedrals are found: C<sub>6</sub>–C<sub>5</sub>–N<sub>1</sub>–C<sub>4</sub>, H<sub>2</sub>–C<sub>5</sub>–N<sub>1</sub>–C<sub>4</sub>, and H<sub>3</sub>–C<sub>5</sub>–N<sub>1</sub>–C<sub>4</sub>. Again, the heavy-atom dihedral C<sub>6</sub>–C<sub>5</sub>–N<sub>1</sub>–C<sub>4</sub> is designated as the primary torsion (circled in orange), whereas the hydrogen-containing dihedrals, H<sub>2or3</sub>–C<sub>5</sub>–N<sub>1</sub>–C<sub>4</sub> are treated as secondary torsions (circled in blue). This

procedure is fully automatic without requiring any user-defined selection of primary torsion angles.

## Model Architecture

Our ICoN-v1 model is built using a Transformer architecture<sup>26</sup>. First, the input embedding layer linearly projects each torsion angle,  $(\sin(\theta), \cos(\theta))$  pair, from 2 to 512 dimensions (**d\_model=512**), creating a rich feature space for subsequent processing. A positional encoding is added to these embeddings using fixed sine/cosine functions, enabling the model to distinguish torsions by their position along the molecular chain and to learn structure-dependent correlations that arise from molecular topology and connectivity. The encoder consists of eight multi-head self-attention blocks (**n\_heads=8**) enabling each primary torsion angle to be represented as a function of its correlations with all other torsions. Six layers (**n\_layers=6**) were used for both the encoder and decoder with **GELU** activation function. The encoder output is projected into a compact three-dimensional latent space (**d\_latent=3**). The decoder section mirrors the encoder structure in reverse that regenerates a 512 dimensional embedding for every primary torsion angle, followed by a final linear layer that converts these rich embeddings back to predicted  $(\sin(\theta), \cos(\theta))$  pairs.

## Loss Function

To improve the accuracy of model prediction and efficiently learn physics, four losses: Torsion, Cartesian, Energy, and Interpolation Energy loss are added in multiple stages of the training progress. A simple Mean Square Error (MSE) loss function is used to assess the accuracy of prediction during the training progress by comparing the model input and output.

$$MSE = \frac{\sum_{i=1}^n (input - predicted)^2}{N} \quad (1)$$

**Torsion Loss:** During the first training epoch, only the torsion loss is used. Because torsion loss converges quickly, it provides a stable foundation for subsequent training stages.

**Cartesian Loss:** Cartesian loss is used throughout the rest of the training process to ensure the reconstruction accuracy of the model for all atoms. Computing Cartesian loss is relatively expensive, as full molecular conformations must first be reconstructed from BAT internal coordinates, and this reconstruction is sequential, one atom at a time.

**Energy Loss:** After reconstructing predicted conformations, we compute their MM potential energy terms, bond, angle, torsion, van der Waals (vdW), and electrostatic (Elec), and the solvation energy term using the Generalized Born (GB) model to avoid unrealistic Elec attraction observed

in vacuum. Unlike torsion or Cartesian loss, which measure reconstruction error against reference MD frames, the energy itself is used as the loss, which explicitly drives the network toward energy minimum. This loss penalizes unphysical clashes and stretching, encouraging the model to learn conformations that are physically realistic. Since energy evaluation must be performed for every predicted conformation at each iteration and epoch, this step is computationally expensive. Moreover, because energy values are on a different scale than torsion or Cartesian losses, we apply an adaptive weighting scheme to balance their contributions (**see Adaptive Weight section in SI**).

**Interpolated Energy Loss:** To further enforce physical realism in latent space, we apply random spherical linear interpolation (SLERP) between pairs of latent vectors (**see Random Spherical Linear Interpolation (SLERP) section**). Along each interpolation path, random points are sampled, decoded into conformations, and their corresponding energy terms (bond, angle, torsion, vdW, Elec, GB) are computed. Unlike Torsion or Cartesian loss, the interpolated energy itself is directly minimized as the loss, explicitly encouraging the network to generate low-energy conformations along latent-space trajectories. This loss promotes a latent space in which interpolations correspond to physically meaningful, energetically favorable conformational transitions.

### *Random Spherical Linear Interpolation (SLERP)*

The training dataset cannot cover the entire latent space. The empty space does not contain reference conformations to compute the Cartesian loss. Our model implements random spherical linear interpolation (SLERP) between points in latent space and randomly selects a point on the spherical path shown in equations 2 and 3. Each dihedral angle is represented by its sine and cosine components, corresponding to a point on the unit circle. Motivated by this periodic representation, SLERP is employed to generate smooth intermediate latent samples along an angular trajectory. Since MEPs are generally nonlinear, this interpolation strategy is expected to provide more physically realistic intermediate conformations than simple Euclidean linear interpolation. During training, latent representations from pairs of conformations within each training batch are selected. For each pair, SLERP is used to generate a trajectory consisting of 1,000 interpolated latent points. One point is randomly sampled from this trajectory and decoded into dihedral angles, which are subsequently reconstructed into Cartesian coordinates. These reconstructed conformations are then used to evaluate the potential energy, which contributes directly to the training loss. In other words, the network will minimize the interpolated loss/energy during the training process. This process serves as regularization as high-energy regions discovered during interpolation provide learning signals that guide the network to avoid unnatural conformational regions, ultimately leading to a more physically realistic latent space representation.

$$\omega = \arccos(\hat{z}_i \cdot \hat{z}_j) \quad (2)$$

$$interpolation\ path = \frac{\sin((1-t)\omega)}{\sin(\omega)} \vec{z}_i + \frac{\sin(t\omega)}{\sin(\omega)} \vec{z}_j \quad (3)$$

$\vec{z}_i$  represent the first point and  $\vec{z}_j$  represent the last point in the latent space.  $\omega$  is the angle between  $\vec{z}_i$  and  $\vec{z}_j$ .  $t$  is the parameter to move along the spherical path between  $\vec{z}_i$  and  $\vec{z}_j$  where  $0 \leq t \leq 1$ .

### *Latent Space Sampling through Minimum Energy Pathway (MEP)*

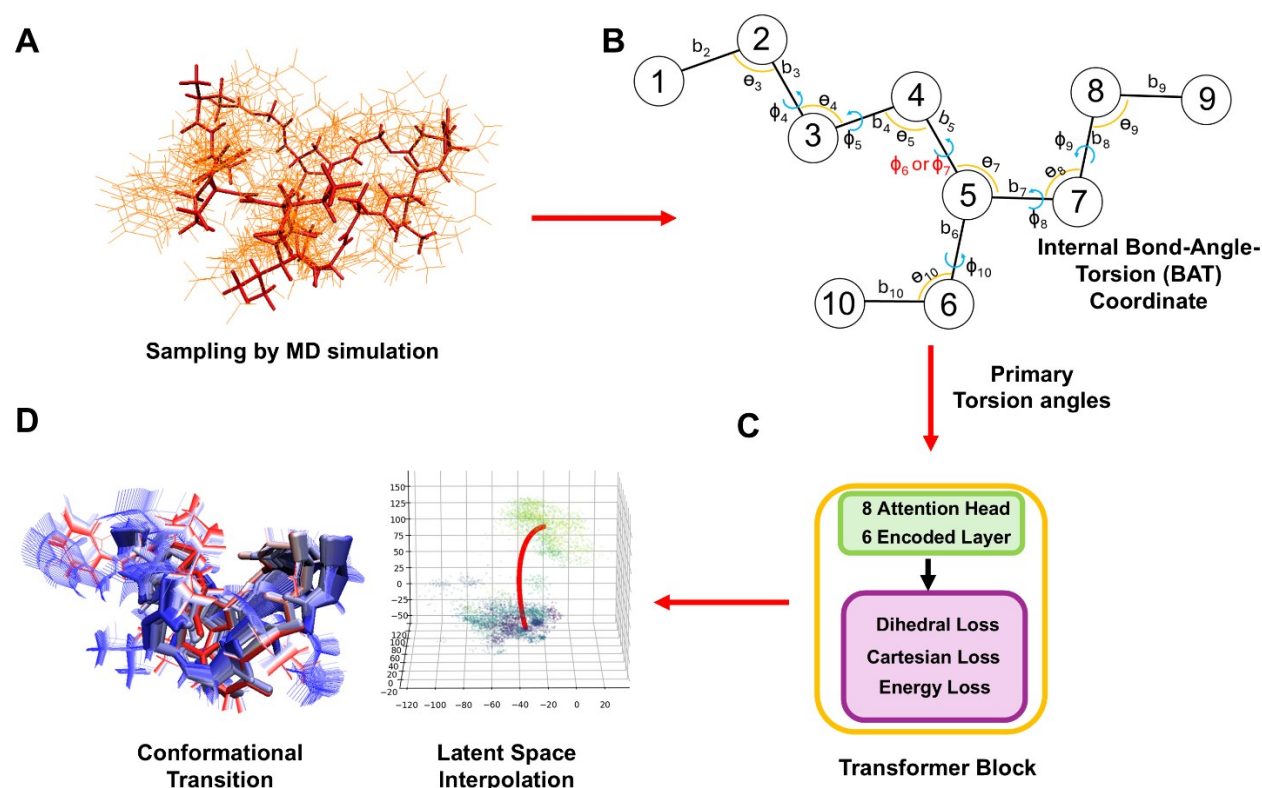
After training, each conformation of the MD simulation was encoded as each unique data point in the 3D latent space (**Figure S3A**). Interpolation between two data points in 3D latent space is nontrivial. Because intermediate data points are lacking, there is no constraint to guide a physically meaningful transition, and interpolation through this empty region could yield infinitely many arbitrary trajectories. A realistic conformational transition, however, should follow the MEP, which represents the most favorable route between the two states. To address this limitation, we applied PCA to project the 3D latent representations onto a 2D principal component space (**Figure S6B**). We then computed the probability density distribution of the 2D data and visualized it as a contour map, which revealed local minima and their corresponding energy gradients (**Figure S6C**). The latent-space free energy landscape is intended only as a qualitative guide for pathway selection and should not be interpreted as a quantitatively accurate thermodynamic free energy surface, since the training dataset is filtered to retain structurally diverse conformations. From the 2D probability contour map, we identified two representative conformations, MD1 and MD2, each located at a distinct local minimum of the 2D contour plot. Then, interpolation was subsequently performed along the energy gradient connecting MD1 and MD2 within the 2D probability landscape, providing an approximation of the MEP (**Figure S3D**). Finally, we can map the 2D MEP back to the closest point in 3D latent space and apply a polynomial fit applied to create a smooth 3D MEP path (**Figure S3E**). The 3D MEP is reconstructed back to primary torsion angles through the decoder followed by BAT to Cartesian coordinate conversion. The smooth conformational transition pathway can then be obtained and analyzed to understand the mechanism in full atomistic detail (**Figure S3F**). For all conformational transitions sampled from the MEP in this study, 10,000 frames were generated to capture detailed torsional rotations and energy changes along the MEP interpolation pathway.

## Results and Discussion

### Overview of ICoN-v1 network architecture and training

To ensure the high quality of the training dataset, we used trajectories generated by all-atom MD simulations with an explicit solvent model to train our ICoN-v1 model. We first converted Cartesian coordinates to internal BAT coordinates. This representation removes 6 external rotational and translation degrees of freedoms (DOFs) to yield  $3N-6$  DOFs (i.e.,  $N-1$  bonds,  $N-2$  angles and  $N-3$  torsions), where  $N$  is number of atoms. Because rotating a few atoms (i.e., methyl rotation) may be described by a single primary torsion, we used “primary” torsions as the input features for model training (see details in Methods). Instead of using all output frames from MD runs, we performed quick minimizations for all frames and removed repeating conformations with a 1.5-Å cutoff root mean square deviation (RMSD) for heavy atoms. This allows the model to efficiently learn from distinct conformations.

ICoN-v1 is built on a Transformer architecture composed of six layers with eight attention heads per layer, enabling the model to efficiently learn correlations among torsional motions. During training, the encoder maps each conformation to a unique point in a three-dimensional (3D) latent space, and the decoder reconstructs the corresponding full conformation from this latent representation. Each molecular conformation was encoded by the neural network into a 3D latent space (**Figure 2**), with each point corresponding to a distinct conformation. Interpolating between two latent points (i.e. conformations with distinct ring puckers) and decoding the intermediate latent representations result in novel synthetic structures along the transition path. This procedure yields smooth and physically interpretable transition pathways between selected conformations (**Figure 2**).



**Figure 2.** Overview of ICoN-v1. **(A)** The dataset comprises conformational ensembles obtained from molecular dynamics (MD) simulations, with the **(B)** Cartesian coordinates converted to Bond-Angle-Torsion (BAT) coordinates to represent cyclic peptide conformations. The primary torsion angles are extracted as input features for training. The bonds joining atoms 5 with 6 and 5 with 7 can be represented by torsions  $\phi_6$  and  $\phi_7$ , respectively.  $\phi_7$  is selected as the primary torsion that leads to conformational changes for atom 7 and  $\phi_6$  is treated as secondary torsion, which is excluded in the input features. **(C)** Transformer model architecture with 8 attention heads and 6 layers (see details in Methods). **(D)** After training, each cyclic peptide conformation is represented as a single point in latent space. Any point in the latent space can be decoded to an atomistic cyclic peptide conformation, including novel synthetic conformations that are not present in the original MD dataset. Interpolation between different local minima, shown as a red curve in the 3D latent space, reveals conformational transition of the cyclic peptide.

### Exploring transient conformations with latent sampling along minimum energy path

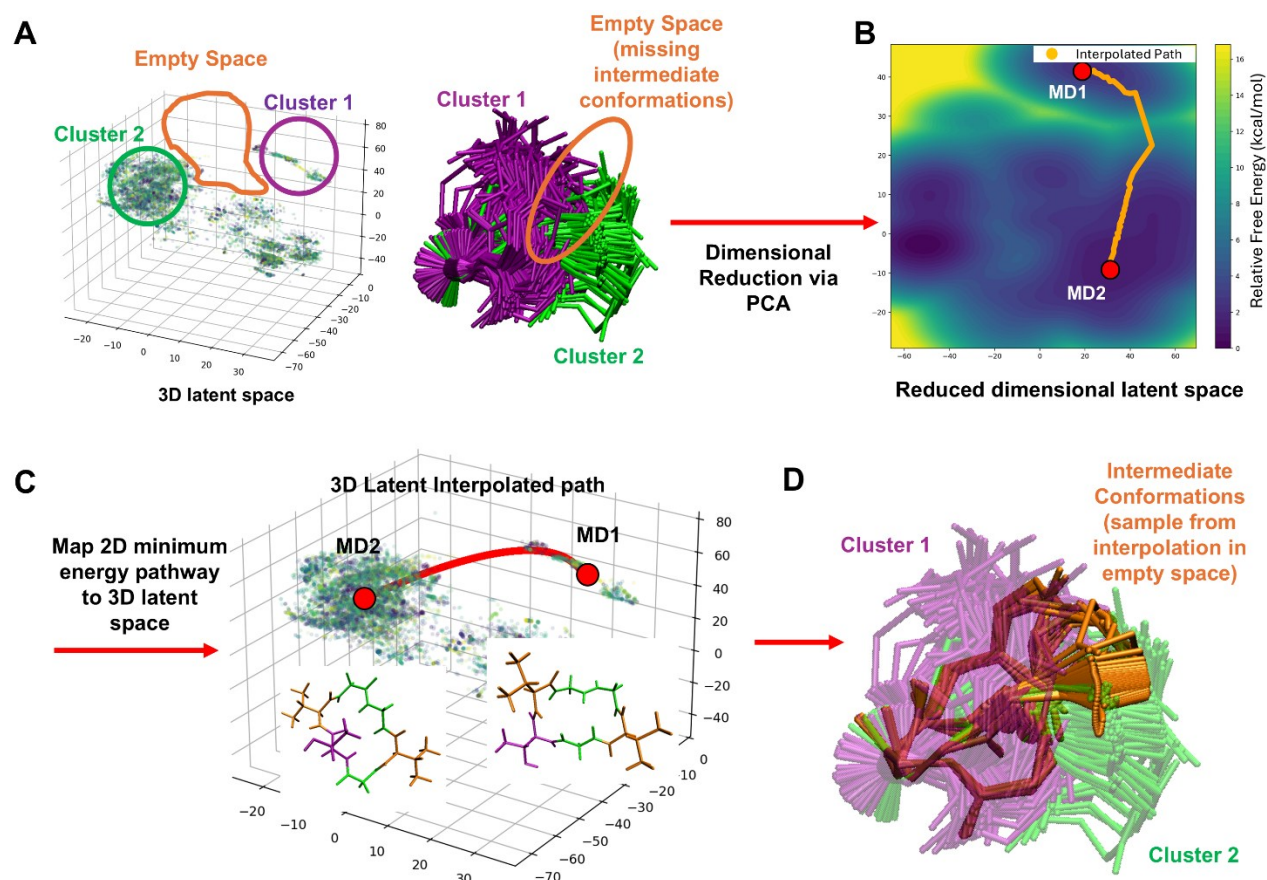
We first investigated cyclo-(TVGGVG) and cyclo-(VVGGVG) (**Figure 1A**) to demonstrate how to use ICoN-v1 to generate smooth conformational transitions and explain how mutating THR1 to VAL1 disrupted the highly packed intramolecular H-bond observed in THR1. The training data

was obtained from NMR and MD simulations by the Lin group <sup>22</sup>. After encoding each conformation into the latent space, the dataset can be grouped by two well-defined clusters for cyclo-(TVGGVG): Cluster 1 (purple circle) represents a state sampled only in MD simulations, and Cluster 2 (green circle) the dominant conformation observed in both MD and NMR (**Figure 3A**). The sparse region between these two clusters contains few or no datapoints, so the intermediates are absent in the reported conformations.

To sample conformational transition, one can interpolate points in the latent space constructed by ICoN-v1, as demonstrated in **Figure 2D**. However, arbitrarily interpolating the two clusters does not guarantee a thermodynamically favorable transition pathway. We instead determined a minimum-energy pathway (MEP) in the latent space that follows the underlying energy gradient and connects two local energy minima through the most probable route. Constructing an MEP directly in the 3D latent space is numerically challenging because of the difficulty of building a well-resolved probability contour map in 3D. We projected the latent space onto two principal components, reducing it to 2 dimensions (2D), which allows for easier construction of probability contour maps (see Methods and **Figure S3** for details). After selecting two local minima, MD1 and MD2, which are also conformations in Cluster 1 and Cluster 2 sampled by MD runs, a 2D MEP path was generated following the energy gradient, which can be easily visualized in the 2D space (**Figure 3B**). The 2D MEP path was then mapped back to the 3D latent space and polynomial fitting was used to reconstruct the 3D MEP for performing a mathematically continuous interpolation. Notably, except for MD1 and MD2, other points in the 3D MEP were not present in the MD simulations (**Figure 3C**). Novel synthetic conformations generated by interpolating points along this 3D MEP path in latent space produced a smooth and continuous conformational transition from MD1 to MD2, as shown by the orange conformations in **Figure 3D**.

Of note, the proposed MEP represents only one possible transition pathway. In principle, cyclic peptides may use substantially different conformational pathways, especially when they rearrange conformations to adapt to distinct environments. For example, during membrane permeation, a cyclic peptide can undergo a transition route that differs substantially from conformational transition pathways observed in an aqueous environment, driven by the shift from water to the

nonpolar membrane interior. To further explore transitions driven by environmental changes, one could intentionally interpolate energy paths that pass through high-energy regions to probe rare or transient conformational transitions, such as those occurring during membrane permeation. Moreover, thorough conformational sampling may be achieved by systematically, or even exhaustively, exploring transition pathways between datapoints in latent space. Analyzing these paths may uncover distinct sets of concerted torsional rotations and their energy landscapes, which help to identify their mechanisms, quantify energy barriers and estimate kinetic properties.



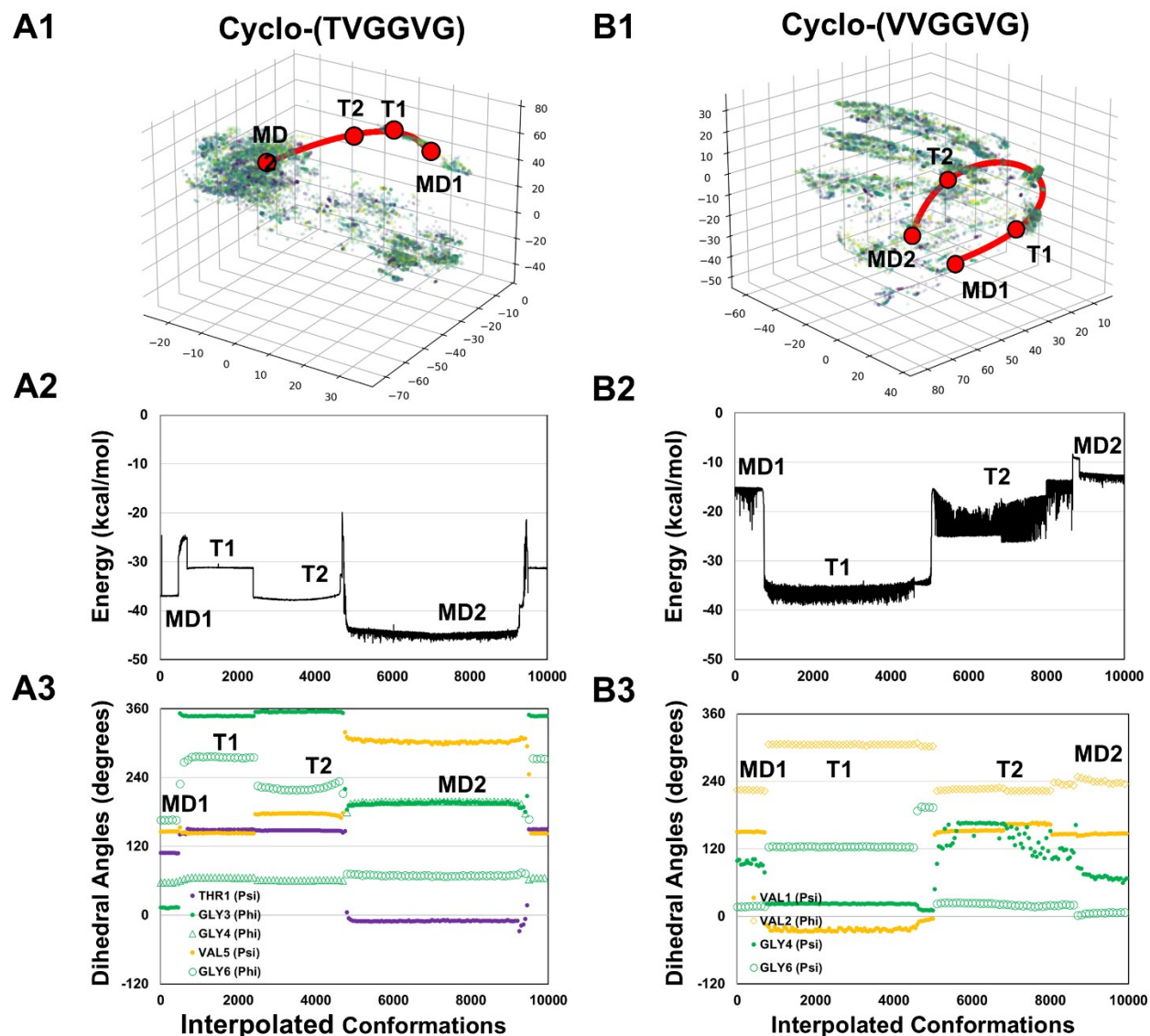
**Figure 3.** Minimum-energy pathway (MEP)-guided conformational transition of cyclo-[IVGGVG]. **(A)** (Left panel) 3D latent space of cyclo-[IVGGVG]. Each dot represents different frames of the molecular dynamics (MD) conformation used as training data encoded in the latent space. The green circle denotes Cluster 1, the purple circle denotes Cluster 2, and the orange outline marks the gap where intermediate conformations are missing. (Right panel) Overlaying 100 conformations of Cluster 1 (purple) and Cluster 2 (green). **(B)** A 2D probability contour map generated by projecting latent space datapoints using principal component analysis, with two local minima, MD1 and MD2, selected as interpolation endpoints. The orange path indicates the MEP connecting MD1 and MD2. **(C)** The 3D MEP (orange path in B) is mapped back onto the 3D latent space, producing the red curve that guides interpolation. **(D)** Overlaying 100 synthetic conformations (orange) generated along the red curve in C reveals the missing intermediate conformations between Cluster 1 (purple) and Cluster 2 (green).

### Revealing concerted torsional rotations during conformational transition of cyclic hexapeptides

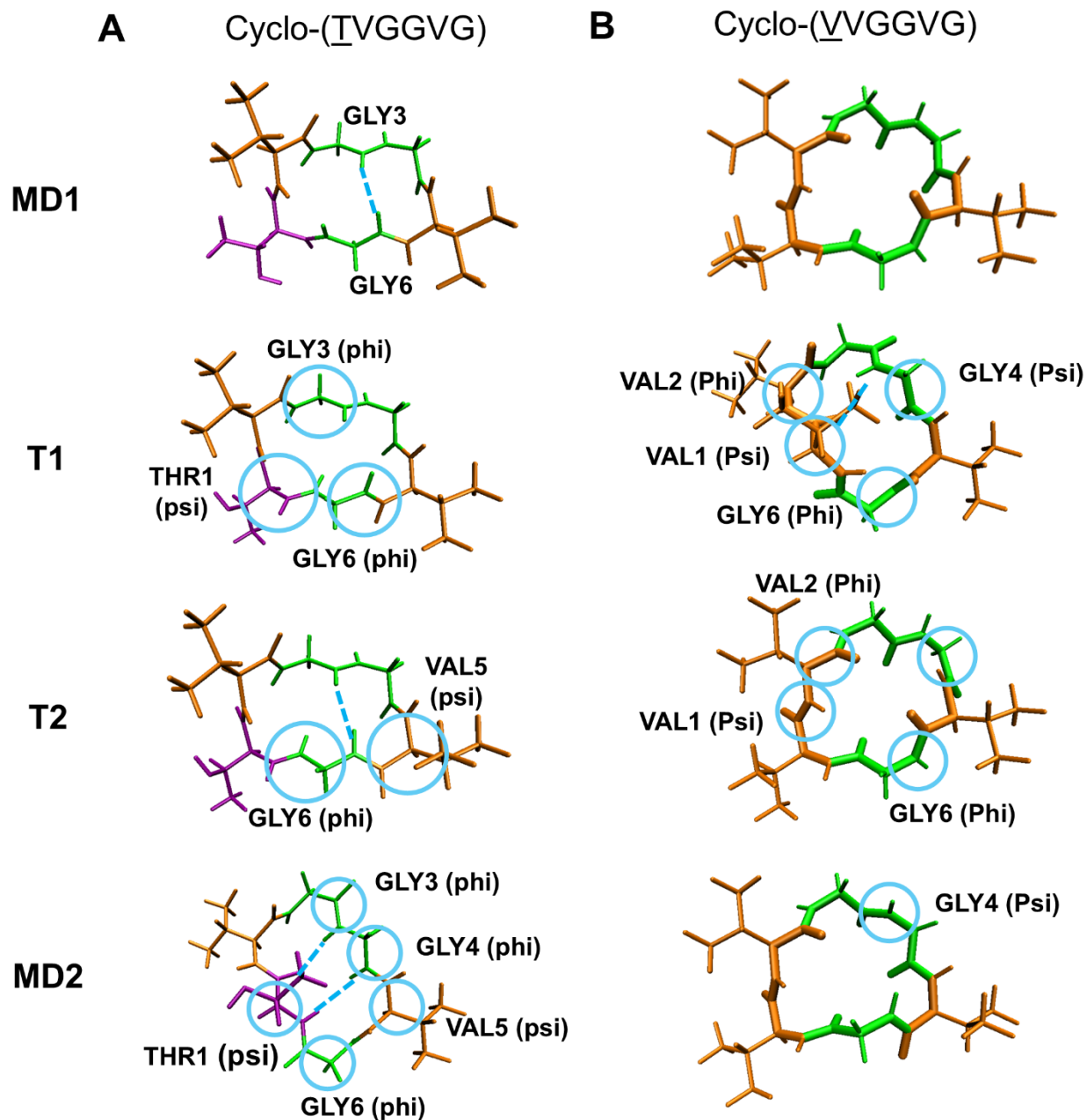
The MEP-guided interpolation revealed a stepwise conformational transition between two local minima MD1 and MD2, for both cyclic hexapeptides cyclo-(TVGGVG) and cyclo-(VVGGVG) (**Figure 4-A1 and -B1**). The potential energies computed along the interpolated curve indicate that after crossing an energy barrier, cyclo-(TVGGVG) transitions from the minimum MD1 to intermediate state T1, then proceeds to another intermediate state T2 before reaching the global minimum MD2 sampled by both NMR and MD (**Figure 4-A2**). Because no original MD conformations directly corresponded to the T1 or T2 states, we analyzed the generated intermediate conformations along the MEP pathway. A total of 10,000 intermediate conformations were generated, and the conformations at interpolation frames 1800 and 4000 were selected as representative T1 and T2 states, respectively, based on their correspondence to local minima in the MM potential energy profile (Figures 4-A2, 4-B2, and 5). Analyzing their torsional rotations revealed that concerted rotation of GLY3/6 (Phi) and THR1 (Psi) drove the transition from MD1 to T1, which slightly altered the ring pucker and disrupted the intramolecular H-bond. Subsequently, GLY6 (Phi) switched again, accompanied by rotation of VAL5 (Psi) to move T1 to T2 to strengthen the polar attraction and allow the H-bond to reform. A new set of concerted torsion rotations, THR1 (Psi), GLY3 (Phi), GLY4 (Phi), VAL5 (Psi), and GLY6 (Phi) shown in **Figure 5-MD2** and detailed in **Figure S4**, changed the ring pucker from T2 to more folded MD2.

Among the multiple local minima sampled by MD, two conformations were selected as the interpolation endpoints (MD1 and MD2), as shown in **Figure 4-B1**. Starting from MD1, the MEP-guided interpolation pathway led cyclo-(VVGGVG) through the global energy minimum at T1, followed by crossing the energy barrier to T2, and finally reaching the second local minimum at MD2. Of note, the conformation of MD1 and T2 were similar, and our model showed that similar sets of dihedrals are used when transitioning from MD1 to T1 and from T1 to T2. The only difference was that GLY4 (psi) rotated from 100 to 0 degrees from MD1 to T1 (**Figure 4-B2 and 5B**) but from 0 to 140 degrees from T1 to T2 (**Figure S5**). This finding highlights the importance of GLY4 (psi) to pre-organize the conformations to MD2. Assuming all dihedral rotations remain the same, we can expect that T2 will likely return to MD1. This observation also highlights the ability of our model to capture subtle differences in the latent space.

In cyclo-(TVGGVG), the polar sidechain shifts from an outward orientation in T2 to an inward position in MD2, forming intramolecular hydrogen bonds with GLY3 and GLY4. This transition produces a compact ring conformation driven by THR1/THR5 (psi) rotations, with GLY3/GLY6 (phi) angles facilitating the stabilization of the hydrogen bonds in the global minimum (**Figure 5A**). In contrast, cyclo-(VVGGVG) reorients the backbone carbonyl to form an H-bond with the GLY4 –NH group, changing the ring from a wide-open MD1 conformation to the more compact global-minimum T1 structure by rotating VAL1/2 (phi) (**Figure 5B-T1**), which is adjacent to the mutation site and serves as a key contributor. Its contribution is despite not directly participating in hydrogen bonding because the steric hindrance from the VAL1 sidechain of cyclo-(VVGGVG) prevents the highly packed H-bond contact observed in cyclo-(TVGGVG) (**Figure 5A-MD2**). Such transition pathways are difficult to resolve directly from conventional MD simulations in which transient conformations are short-lived. In contrast, the latent representation learned by ICoN-v1 captured the concerted torsional rotations in different transient states, which emerged naturally from the learned energy landscape, thus providing mechanistic insight beyond what can be inferred from typical MD analysis alone.



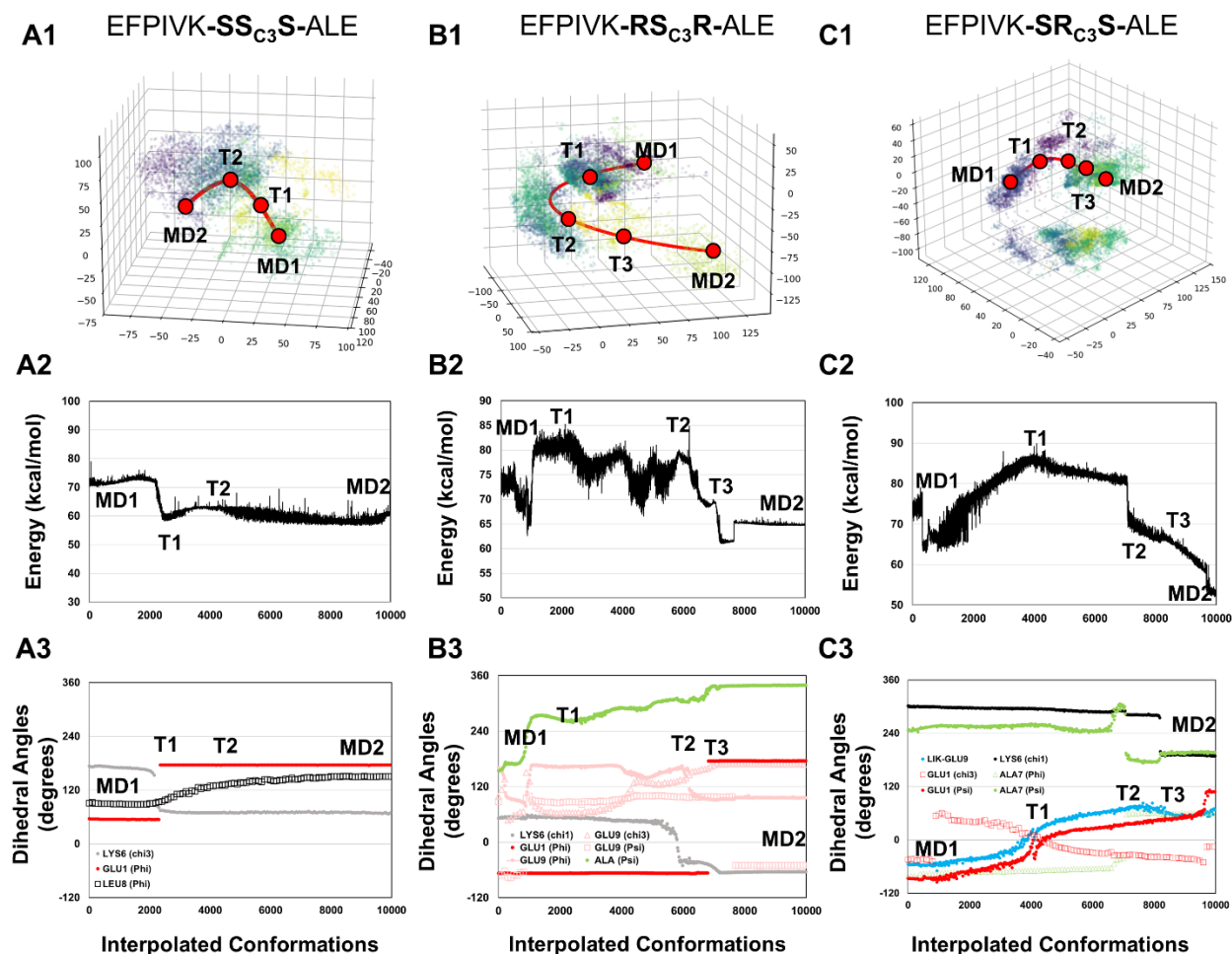
**Figure 4.** Conformational transition pathway of cyclo-(TVGGVG) and cyclo-(VVGGVG). (**A1 and B1**) Training datapoints in the 3D latent space. Each point in the latent space represents a single MD frame and is colored according to its position in the trajectory using the Viridis colormap. The minimum-energy pathway (MEP)-guided interpolation curve (red curve), with two end datapoints, MD1 and MD2, and local minima, T1 and T2, are the key transition steps. (**A2 and B2**) MM potential energy for conformations along the interpolation curve (red line in **A1 and B1**). (**A3 and B3**) Changes in the major torsion rotations of the interpolation curve (red line in **A1 and B1**) contribute to the conformational transition.



**Figure 5.** Representative conformations of minimum-energy pathway (MEP)-guided interpolation curve (red line of **Figures 4A1** and **4B1**). Four conformations, MD1, T1, T2, and MD2, are shown in (A) cyclo-(TVGGVG) and (B) cyclo-(VVGGVG). The hydrogen bonds that exist in the conformations of the low-energy minima, MD1 and MD2 of cyclo-(TVGGVG) and T1 of cyclo-(VVGGVG), are labeled in blue dashed lines. The important torsion rotations are circled in blue.

### Conformational transitions of stereo-diversified bicyclic peptides

We investigated five bicyclic peptides with the cyclization step connecting residues GLU1, LYS6 and GLU9 to the central linker. The EFPIVK-linker-ALE sequence covalently connects the residues to the C1, C3 and C4 carbons, whereas the EWPHRK-linker-RLE and EWPHRK-linker-RIE sequences connect them to C1, C2 and C4. As a result, we explicitly denote only the C3 or C2 position (i.e., SS<sub>C3</sub>S or RR<sub>C2</sub>R) when describing their stereo structures. Using the MEP-guided interpolation strategy in the 3D latent space, we investigated how linker chirality reshapes conformational transitions and how a single residue LEU/ILE substitution contributes to the conformational transition mechanisms of the bicyclic peptides (**Figure 1B**). Our previous study sampled diverse conformations of stereo-diversified MYC-targeting bicyclic peptides using MD simulations; however, how these bicyclic peptides lead to different conformations was not understood<sup>23</sup>. All bicyclic peptides studied here have various local minima, including those with open and folded conformations. Unsurprisingly, folded conformations are more stable in an aqueous environment because they can form more intramolecular contacts. To investigate high- to low-energy conformational transition paths for each system, we constructed MEP (**Figure S6 and 7**) and selected two endpoints, MD1 (high-energy state) and MD2 (low-energy state), for sampling their structural transitions using interpolation of MEP-guided curve in the latent space (**Figures 6A**). Based on the computed potential energy (**Figure 6B**) and corresponding torsional rotations along the interpolation curve (**Figure 6C**), we selected intermediate conformations (i.e., T1-T3) to illustrate the structural rearrangements and the mechanisms driving the transitions.



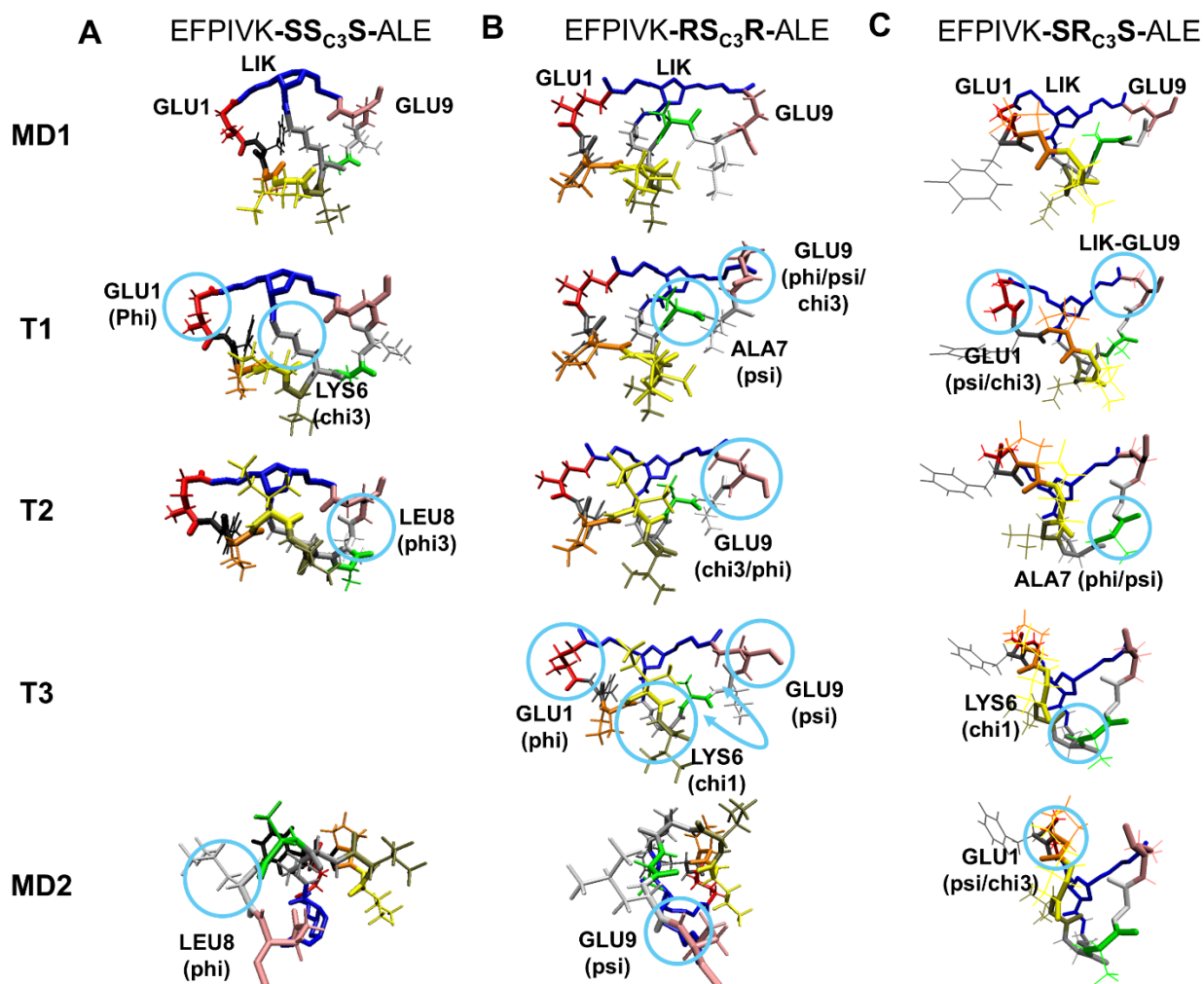
**Figure 6.** Conformational transition pathways of EFPIVK-SS<sub>C3</sub>S-ALE, EFPIVK-RS<sub>C3</sub>R-ALE, and EFPIVK-SR<sub>C3</sub>S-ALE. **(A1, B1, and C1)** Training datapoints in the 3D latent space. Each point in the latent space represents a single MD frame and is colored according to its position in the trajectory using the Viridis colormap. The minimum-energy pathway (MEP)-guided interpolation curve (red line), with two end datapoints, MD1 and MD2, and local minima, T1, T2 and T3, are the key transition steps. **(A2, B2 and C2)** MM potential energy for conformations along the interpolation curve (red line in **A1, B1 and C1**). **(A3, B3, and C3)** Changes in the major torsion rotations of the interpolation curve (red line in **A1, B1, and C1**) contribute to the conformational transition.

How do EFPIVK-SS<sub>C3</sub>S-ALE, EFPIVK-RS<sub>C3</sub>R-ALE, and EFPIVK-SR<sub>C3</sub>S-ALE, with the same sequence but different stereo-structures, vary their high-energy (open) to low-energy (folded) transition?

With guidance by the probability contour map (**Figure S6**), we selected an open conformation (high-energy state) for each bicyclic peptide. Unsurprisingly, folded conformations are more stable (low-energy state) in an aqueous environment because they can bury non-polar sidechains to form more intramolecular contacts while extending polar sidechains to interact with water molecules. To investigate open (high energy) to folded (low energy) conformational transition paths for each system, we constructed an MEP and selected two endpoints, MD1 (open) and MD2 (folded) (**Figure 6-A1 and Figure S6**). Although the side chains of GLU1 and GLU9 were spatially separated for the three peptides with the same sequence, the chirality yielded different open conformations (**Figure 7-MD1**). Different open conformations are anticipated because experiments showed that these bicyclic peptides had various binding affinity to the MYC epitope<sup>23</sup>. Of note, the S configuration in the chiral center C3, EFPIVK-SS<sub>C3</sub>S-ALE and EFPIVK-RS<sub>C3</sub>R-ALE, share highly similar folded conformations to MD2, where PRO3 and ILE4 are in close contact with LYS6 (**Figure S8**). However, their transition pathways were remarkably different. EFPIVK-SS<sub>C3</sub>S-ALE follows a straightforward transition pathway, with the potential energy landscape featuring only a single intermediate conformation T1. The transition from MD1 to T1 is initiated by rotations of GLU1 ( $\phi$ ) and LYS6 ( $\chi_3$ ) to drive PRO3 and ILE4 toward the linker region and alter the ring pucker, followed by rotation of LEU8 ( $\phi$ ) to pass a small energy barrier (T2) (**Figure 6-A2 and -A3**) to optimize the intramolecular attraction of LEU8 with surrounding residues and reach the global energy minimum, the folded conformation MD2.

The interpolation curve reveals that EFPIVK-RS<sub>C3</sub>R-ALE undergoes an open-to-folded transition via a series of concerted torsional rotations that modulate the ring pucker (**Figure 6-B1**). To characterize this process, we selected three representative intermediate conformations (T1–T3) that capture the key switching events in these concerted torsion rotations. **Figure 6-B2 and -B3** demonstrate that EFPIVK-RS<sub>C3</sub>R-ALE follows a distinct set of torsional pathways and traverses a more rugged energy landscape yet ultimately converges to the same folded minimum MD2 conformation observed for EFPIVK-SS<sub>C3</sub>S-ALE. ALA7 ( $\Psi$ ) and GLU9 ( $\Phi/\Psi/\chi_3$ ) initiate the

conformational transition to overcome ~20 kcal/mol energy change and bring EFPIVK-RS<sub>C3</sub>R-ALE to a local energy barrier (T1). The two torsions continue rotating non-linearly to guide the molecule while visiting a few local minima before reaching another energy barrier (T2). To open space for PRO3/ILE4/VAL5 to move inward and interact with the linker, GLU1 (phi), LYS6 (chi3) and GLU9 (phi) on opposite sides of the ring rotate in a coordinated fashion to reach the local minimum T3, followed by adjusting GLU9 (Psi) to complete the transition to MD2 (**Figure 7B**).



**Figure 7.** Representative conformations of minimum-energy pathway (MEP)-guided interpolation curve (red lines of **Figure 6-A1, -B1 and -C1**), MD1, T1, T2, T3, and MD2 of **(A)** EFPIVK-SS<sub>C3</sub>S-ALE, **(B)** EFPIVK-RS<sub>C3</sub>R-ALE, and **(C)** EFPIVK-SR<sub>C3</sub>S-ALE. The important torsion rotations shown in **Figure 6-A3, -B3 and -C3** are circled in blue.

With an R configuration in the chiral center C3, EFPIVK-**SR**<sub>C3</sub>**S**-ALE has a distinct folding state of MD2, with GLU1 and GLU9 forming close contacts (**Figure 7C-MD2**). In EFPIVK-**SS**<sub>C3</sub>**S**-ALE, GLU1 ( $\phi$ ) initiates the conformational transition. In contrast, although EFPIVK-**SR**<sub>C3</sub>**S**-ALE also initiates the transition through GLU1, it employs a different set of torsional degrees of freedom. As shown in **Figure 6-C2 and -C3**, coordinated rotation of GLU1 ( $\psi/\chi_3$ ) together with linker-GLU9, enable the system to overcome the  $\sim 20$  kcal/mol MM energy barrier and reach a local intermediate T1. The additional rotation of ALA7 ( $\Phi/\Psi$ ) then reorganizes the backbone to form T2, thus enabling the subsequent rotation of LYS6 ( $\chi_1$ ) for the T2 to T3 transition. A final adjustment of GLU1 ( $\psi$  and  $\chi_3$ ) completes the folding process and stabilizes GLU1–GLU9 contacts (MD2) (**Figure 7C**). To verify whether the interpolated pathway captures a transition similar to what is observed in MD, we overlaid 100 conformations sampled from both the MD trajectory and the interpolated trajectory for EFPIVK-SRC3S-ALE (Figure S9). The interpolated conformations exhibit a smooth structural transition that closely follows the conformational transition observed in the MD trajectory, supporting the ability of the latent-space interpolation to generate physically reasonable intermediate conformations.

Residues GLU1, LYS6, and GLU9 directly connect to the linker region, so rotating backbone  $\phi$  or  $\psi$  angles of these residues contribute substantially to changing the ring conformations. However, different stereo-structures use distinct pathways, each identified by specific combinations of torsions in these three residues, coupled with coordinated rotations of torsions in ALA7 or LEU8 lead to different intermediate conformations and finally achieve an open to folded conformational transition. Although ALA7 does not directly attach to the linker, its backbone rotation can notably fluctuate the ring pucker and largely change the energy by  $\sim 20$  kcal/mol (**Figure 6-B2 and -C2**). Of note, torsional rotations in PRO3 and ILE4 were not observed for all three bicyclic peptides with the EFPIVK-linker-ALE sequence. Instead, motions from nearby residues resulted in strengthening interactions between PRO3/ILE4 with other residues (i.e., PRO3-LEU8 or ILE4-linker shown in **Figure S10**). These interactions lowered the potential energy by  $\sim 10$  kcal/mol, which suggests their mechanistic importance in stabilizing the conformations and driving the system toward the folded global minimum. Mutating PRO3 or ILE4 to a bulkier residue such as TYR or TRP introduced substantial steric hindrance, which

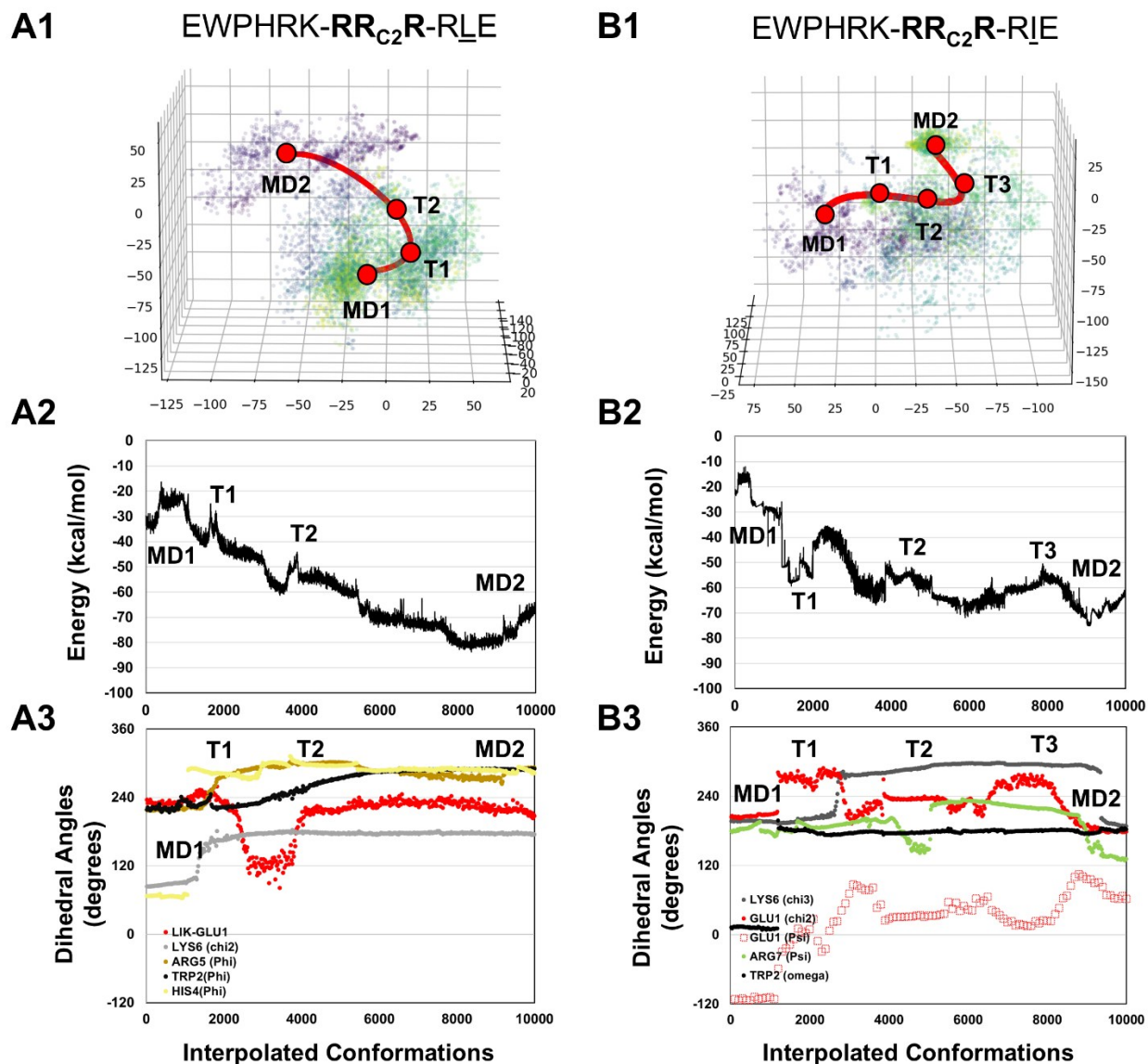
could disrupt the formation of folded conformations and shift the energy landscape toward more open conformations. Depending on the application, the open conformations may improve binding to their protein target but potentially impair the membrane permeability of the bicyclic peptide.

How does a single point mutation, from EWPHRK-RR<sub>C2</sub>R-RL<sub>E</sub> to EWPHRK-RR<sub>C2</sub>R-RI<sub>E</sub>, alter the conformational transitions from high to low energy states?

Unlike proteins, in which the highly similar residues LEU and ILE usually yield nearly identical structures, mutating LEU to ILE in EWPHRK-RR<sub>C2</sub>R-RL<sub>E</sub>/IE can completely reshape the conformational ensemble. In contrast to the EFPIVK-linker-ALE bicyclic peptides, which exhibit diverse potential energy landscapes during transitions from high to low energy states, the EWPHRK-RR<sub>C2</sub>R-RL<sub>E</sub>/IE bicyclic peptides follow a smoother downhill energy landscape with no major energy barriers (**Figure 8-A2 and -B2**). Moreover, both bicyclic peptides share the same high-energy folded conformation (MD1) and a similar open intermediate (T2). In MD1, GLU1 and GLU9 are far apart and sidechains of ARG5 and ARG7 rotate inward to interact with TRP2 and HIS4 to form cation- $\pi$  interactions to stabilize the folded conformation. However, positioning both positively charged ARG5 and ARG7 in close proximity introduces unfavorable electrostatic repulsion, thus resulting in a high-energy folded conformation. As a result, a natural first step in moving the energy landscape downhill is to relieve the ARG5-ARG7 repulsion by changing the ring pucker. Of note, bicyclic peptides with LEU8 or ILE8 substitutions use different strategies to alter the ring, which gradually leads to two distinct low-energy folded conformations (MD2).

Beginning from the folded MD1 conformation, outward rotations of TRP2 ( $\phi$ ), HIS4 ( $\phi$ ), and ARG5 ( $\phi$ ) in EWPHRK-RR<sub>C2</sub>R-RL<sub>E</sub> (**Figure 8-A3**) drive the system into the T1 intermediate state (**Figure 9A-T1**). The concerted rotations then shift torsions HIS4 ( $\phi$ ) and LYS6 ( $\chi_2$ ) to guide ARG5 and ARG7 to move outward to produce an open conformation T2. Subsequent sequential rotations of TRP2 ( $\phi$ ), ARG5 ( $\phi$ ) and GLU1-linker torsions further refold the ring to another low-energy folded conformation (MD2) (**Figure 9A-MD2 and Movie S1**). In

EWPHRK-RR<sub>C2</sub>R-R<sub>I</sub>E, the peptide uniquely rotates the TRP2 omega bond and GLU1 (psi/chi2) to disrupt the cation- $\pi$  interactions (**Figure 8-B3**) between TRP2/HIS4 and ARG5/ARG7 appearing in MD1, thereby slightly altering the ring pucker (T1) for subsequent rotation of ARG7 to open the ring (T2). The transition from T2 to T3 involves rotations in GLU1 (psi/chi2), with an open conformation, followed by varying degrees of coordinated LYS6 (chi3) and GLU1 (psi and chi2) and ARG7 (psi) rotations to finally form contact between PRO3-GLU9 and HIS4-ILE8, thereby leading to another final folded, low-energy conformation (MD2) (**Figure 9B and Movie S2**).



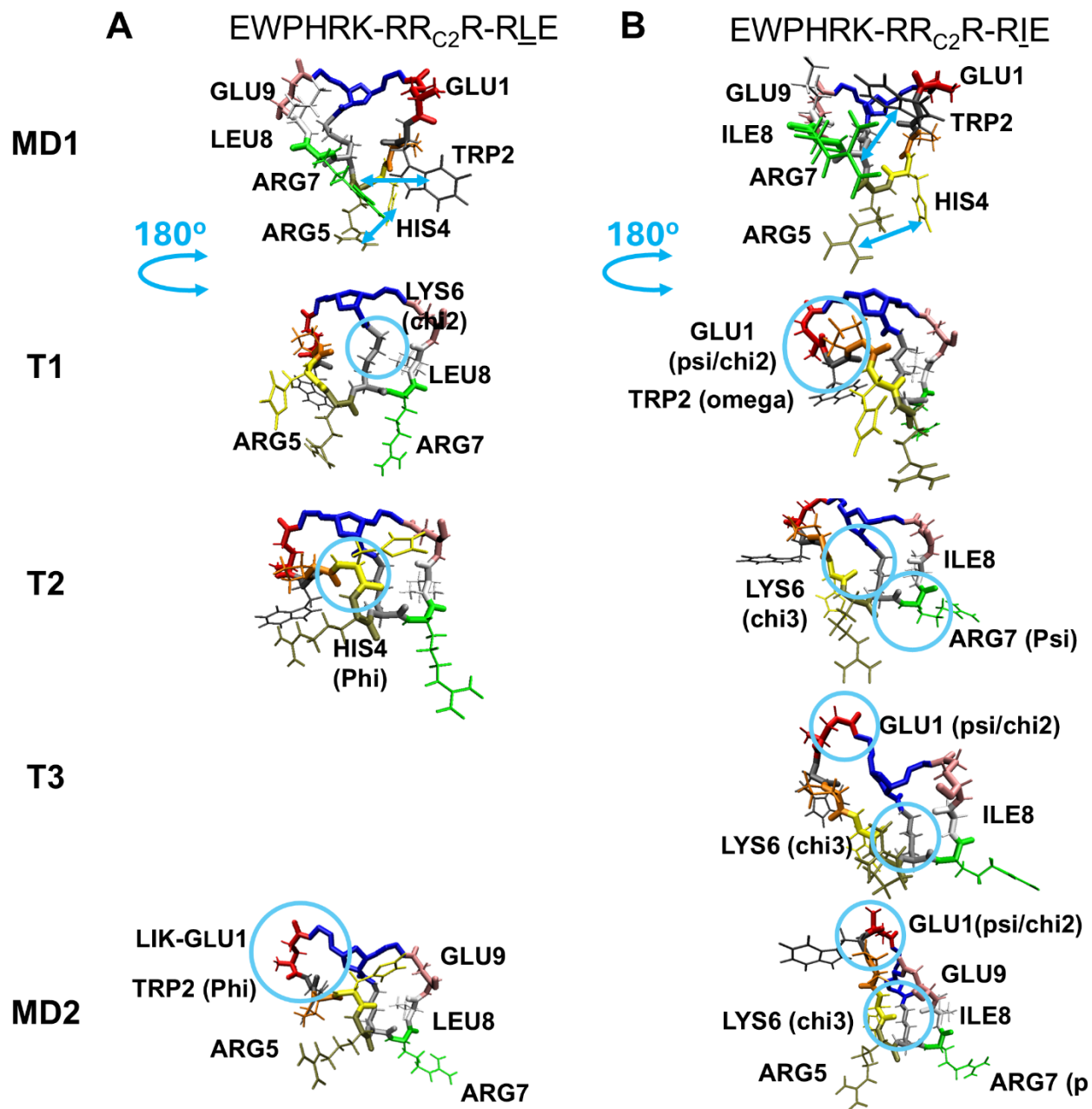
**Figure 8.** Conformational transition pathways of EWPHRK-RR<sub>C2</sub>R-RLE and EWPHRK-RR<sub>C2</sub>R-RLE. **(A1 and B1)** Training datapoints in the 3D latent space. Each point in the latent space represents a single MD frame and is colored according to its position in the trajectory using the Viridis colormap. The minimum-energy pathway (MEP)-guided interpolation curve (red line), with two end datapoints, MD1 and MD2, and local minimum, T1, T2 and T3, which highlight the key transition steps. **(A2 and B2)** MM potential energy for conformations along the interpolation curve (red line in **A1 and B1**). **(A3 and B3)** Changes in the major torsion rotations of the interpolation curve (red line in **A1 and B1**) contribute to the conformational transition.

Investigating the transition pathways for EWPHRK-RR<sub>C2</sub>R-RL/IE explains how LEU8 or ILE8 substitution affects conformational ensembles. Even small sequence changes in bicyclic peptides are well known to produce often unpredictable and large structural variations<sup>23</sup>. These systems exhibit highly concerted torsional rotations to lead to different conformations. Even when transitioning from one folded conformation to another, a bicyclic peptide may transiently open the ring and then re-fold again into different folded conformations. As demonstrated in EWPHRK-RR<sub>C2</sub>R-RL/IE, even when our latent space interpolation begins from the same high-energy folded MD1 conformation, the two molecules use a completely different set of torsion rotations (i.e., TRP2(phi)/HIS4(phi)/ARG7(phi) in LEU8 substitution and TRP2(omega)/GLU1(psi/chi2) in ILE8 substitution) to disrupt interactions between ARG5/ARG7 and TRP2/HIS4 and change the ring structure. Of note, EWPHRK-RR<sub>C2</sub>R-R<sub>I</sub>E uses the peptide bond of TRP, TRP2 (omega), to change the molecular conformation. In structural proteins, cis-trans isomerization of an omega bond typically involves a 15~20 kcal/mol free energy barrier<sup>27</sup> and thus often requires a cis-trans isomerase to lower the free energy barrier for the transition. However, in cyclic peptides, the conformational constraints imposed by cyclization lower the energetic barrier for cis-trans omega rotations, thus enabling access to a broader conformational ensemble than typical proteins. For example, the free energy barrier for omega cis–trans isomerization decreases from approximately 18.7 (kcal/mol) in the linear triglycine peptide (Ace-GGG-Nme) to 14.9 (kcal/mol) in the cyclic tetrapeptide cyclo-(GGGG)<sup>28</sup>. This enhanced flexibility allows cyclic peptides to adopt conformations capable of engaging transients or “undruggable” binding sites on target proteins, which highlights a key advantage of cyclic scaffolds in molecular recognition and design.

After the bicyclic ring opens to reach a similar intermediate T2, the bicyclic peptides with LEU8 or ILE8 again use distinct sets of torsional rotations to guide the bicyclic peptides into different low-energy folded conformations, each corresponding to its own MD2 endpoint. Moving LYS6(chi3) refolds EWPHRK-RR<sub>C2</sub>R-R<sub>I</sub>E (**Figure 8-B3**), where the linear sidechain of ILE8 can be more flexible to adjust its conformation, thus allowing the molecule to adopt a new low-energy folded conformation. However, the branched sidechain of LEU8 is bulkier spatially to prevent rotations of LYS6 from rotating further for conformational transition. After initial changes, the torsions of LYS6 remain stable throughout the subsequent transitions. (**Figure 8-A3**). Instead,

although far from LEU8, the torsional rotation of TRP2 ( $\phi$ ) plays a pivotal role in bringing HIS4 to contact with the linker, refold the cyclic peptide, and bring the molecule to a low-energy folded form. By examining their transition pathways, we propose that TRP2 functions as a leading residue that differentiates the conformational transition routes.

By illustrating the diverse conformational transitions governed by varying concerted torsional rotations along the pathways, we gain deeper insight into how bicyclic peptides adapt to different environments. Importantly, such an investigation is only possible with a high-quality DL model that learns the underlying physics and captures these dynamically switching sets of torsional rotations. Methods based on linear combinations of torsions using fixed rotational magnitude are inherently unable to present these highly complex and cooperated motions shown here. The folded conformations of bicyclic peptides can diverge enough to have different peptide–water interactions, which may influence membrane permeability. With EWPHRK-RR<sub>C2</sub>R-R<sub>IE</sub> used as an example, the close packing of ARG5/ARG7 and TRP2/HIS4 in MD1 buries their polar function groups (**Figure S11-MD1**), thus resulting in a peptide–water electrostatic attraction of -490.4 kcal/mol. This situation is less favorable than in MD2, where both ARG5/ARG7 extend outward to interact with water (**Figure S11-MD2**), thereby increasing the solvent–solute electrostatic attraction to -555.8 kcal/mol. Nevertheless, MD1 with less favorable interactions with polar environments may become the dominant conformation in nonpolar environments, such as within a lipid membrane or in nonpolar solvent. Here, we demonstrate the chameleonic behavior<sup>29</sup> of these bicyclic peptides by studying their conformational transition pathways involving concerted torsional rotations.



**Figure 9.** Representative conformations of minimum-energy pathway (MEP)-guided interpolation curve (red lines of **Figure 8-A1 and -B1**), MD1, T1, T2, T3, and MD2 of **(A)** EWPHRK-RR<sub>C2</sub>R-R<sub>L</sub>E and **(B)** EWPHRK-RR<sub>C2</sub>R-R<sub>I</sub>E. The conformations are rotated 180 degrees (shown in blue arrows) from MD1 to T1 for easier visualization of important torsion rotation. The important torsion rotations shown in **Figure 8-A3 and -B3** are circled in blue.

## Conclusions

This study uses a DL model, ICoN-v1, to reveal the underlying physics governing conformational transitions in bicyclic peptides. By using primary torsion angles to represent atomic motions and incorporating an energy-guided loss function during training, the model learns physically meaningful conformational landscapes and constructs a 3D latent space that encodes molecular conformational information. The latent space can be viewed as a synthetic configuration space that captures the conformational free-energy profile. Interpolating points between local energy minima enables the discovery of conformational transition paths. Analyses of MEP-guided interpolation further offer mechanistic insights into these transitions.

We show how a single mutation with the same residue property substitution LEU8 to ILE8 can introduce distinct transient states and energy barriers via different TRP2 backbone rotations. Our analysis further reveals that chirality alone does not determine the balance between open and folded conformations. Distal residues, such as ALA7 and LEU8, can contribute up to  $\sim 20$  kcal/mol to the potential energy landscape, whereas PRO3 and ILE4 stabilize folded conformations. With cyclic hexapeptides, we also demonstrate how a THR-to-VAL mutation disrupts the H-bond network and reshapes the free-energy profile. Our model goes beyond end-state comparisons by reconstructing complete transition pathways and exposing intermediate states, which are unreported in existing MD trajectories yet critically govern the conformational ensemble sampled by MD runs. These findings provide mechanistic insights into cyclic peptide conformational landscapes and identify key residues and dihedral angles that control conformational ensembles and energy barriers. Such information can guide rational cyclic peptide design for various applications.

Because detailed, system-specific understanding is essential for molecular design, ICoN-v1 is not intended to construct a universal model, and a system-dependent 3D latent space should be constructed for each molecular system. However, the training dataset is not limited to MD trajectories and can be trained on conformational ensembles generated by any sampling method. However, like many DL approaches, the model can suffer from domain limitations when generating synthetic conformations far from the training distribution. Incorporating physics-based noise, such as Langevin Dynamic style perturbations, may help expand the accessible

conformational space and improve the model’s ability to explore previously unseen regions of the energy landscape.

### **Outlook: Current Challenges and Future Directions**

Several challenges remain for latent-space interpolation methods, including the BAT internal-coordinate representation, the lack of rigorous kinetic information and the protein-solvent interaction along interpolated transition pathways.

The BAT representation presents three primary limitations. First, because BAT coordinates are defined sequentially, errors in torsional angles near the root atom can propagate along the molecular chain, leading to unphysical global conformations. In the current implementation, this issue is partially mitigated by combining structural alignment with a Cartesian reconstruction loss during training, thereby improving overall reconstruction accuracy. Second, reconstruction from BAT coordinates is inherently sequential, with atoms generated one at a time, making the procedure computationally inefficient for large biomolecular systems. Third, the sequential dependency limits scalability as system size increases. These challenges can be alleviated by partitioning a protein into multiple BAT fragments, typically based on secondary structural elements. Each fragment is assigned to an independent root atom, reducing error propagation while enabling parallel reconstruction. We are currently developing an efficient parallel BAT reconstruction algorithm together with a multi-fragment BAT representation to improve the scalability of ICoN-v1 for large and complex protein systems.

It is also important to distinguish ICoN-v1 from enhanced sampling methods. Enhanced sampling techniques are designed to generate conformational ensembles under specific thermodynamic conditions by explicitly exploring the underlying free-energy landscape. In contrast, ICoN-v1 is designed to learn a underlying physics of the conformational data, including MD simulations, enhanced sampling trajectories, and experimental structures such as NMR ensembles, and to generate smooth conformational transitions between metastable states. Rather than replacing enhanced sampling methods, ICoN-v1 is intended to complement them.

The interpolated pathways generated by ICoN-v1 should not be interpreted as physical dynamical trajectories and therefore do not directly provide kinetic information. Nevertheless, the corresponding energy barrier may provide an initial approximation of the underlying transition landscape and identify physically plausible conformations for subsequent kinetic refinement. For example, short MD simulations with an explicit solvent model initiated from

structures sampled along the interpolated pathway could be used to reconstruct a more accurate free-energy profile and estimate kinetic properties. Furthermore, these MD simulations explicitly capture protein–solvent interactions that are absent from the implicit solvent model used during ICoN-v1 training. The resulting trajectories can then be incorporated into subsequent training to further refine the latent-space representation, yielding a more physically faithful MEP while improving the handling of solvent-mediated conformational transitions. Such an iterative framework enables machine learning and molecular simulations to mutually enhance one another, progressively improving both the accuracy of the latent representation and the quality of the predicted transition pathways.

### **Data and Software Availability**

The code, pre-trained ICoN-v1 model parameters, and data sets are available on <https://github.com/chang-group>

### **Supporting Information**

Additional training detail, model architecture, post-training analysis, and figures with detailed stereochemistry, torsional rotations and energy components are included. (PDF)

Additional movies of bi-cyclic peptides interpolated trajectories are provided. (MP4)

### **Acknowledgements.**

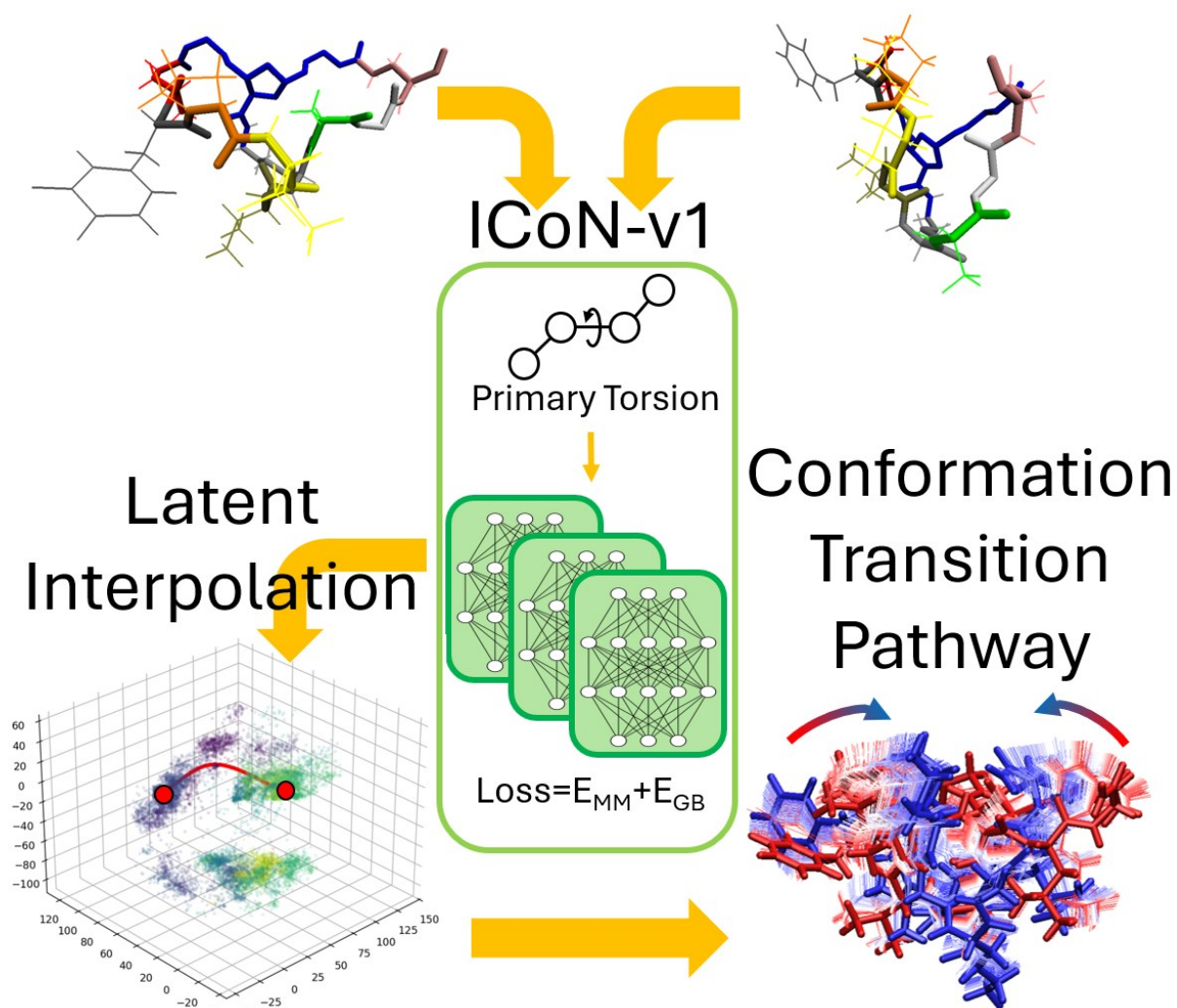
We thank Dr. Min Xue for valuable discussion regarding the bicyclic peptide sequences and their associated experimental results. and. We thank Dr. Yu-Shan Lin for providing both MD and NMR conformations of cyclic hexapeptides peptides. This study was supported by the US National Institutes of Health (R01GM-109045 and R35GM-164210 to C.C.), the US National Science Foundation (MCB-2437134 to C.C.), a UCR RAISE fellowship to T.H., and UCR Academic Senate fellowship to C. C.

## Reference

- (1) Chan, L.; Hutchison, G. R.; Morris, G. M. Understanding Ring Puckering in Small Molecules and Cyclic Peptides. *J. Chem. Inf. Model.* **2021**, *61* (2), 743–755. <https://doi.org/10.1021/acs.jcim.0c01144>.
- (2) Ji, X.; Nielsen, A. L.; Heinis, C. Cyclic Peptides for Drug Development. *Angew. Chem. Int. Ed.* **2024**, *63* (3). <https://doi.org/10.1002/anie.202308251>.
- (3) Corrigendum: Cyclic Peptides for Drug Development. *Angew. Chem. Int. Ed.* **2024**, *63* (10). <https://doi.org/10.1002/anie.202319807>.
- (4) Malde, A. K.; Hill, T. A.; Iyer, A.; Fairlie, D. P. Crystal Structures of Protein-Bound Cyclic Peptides. *Chem. Rev.* **2019**, *119* (17), 9861–9914. <https://doi.org/10.1021/acs.chemrev.8b00807>.
- (5) Ramelot, T. A.; Palmer, J.; Montelione, G. T.; Bhardwaj, G. Cell-Permeable Chameleonic Peptides: Exploiting Conformational Dynamics in de Novo Cyclic Peptide Design. *Curr. Opin. Struct. Biol.* **2023**, *80*, 102603. <https://doi.org/10.1016/j.sbi.2023.102603>.
- (6) Over, B.; Matsson, P.; Tyrchan, C.; Artursson, P.; Doak, B. C.; Foley, M. A.; Hilgendorf, C.; Johnston, S. E.; Lee, M. D.; Lewis, R. J.; McCarren, P.; Muncipinto, G.; Norinder, U.; Perry, M. W. D.; Duvall, J. R.; Kihlberg, J. Structural and Conformational Determinants of Macrocyclic Cell Permeability. *Nat. Chem. Biol.* **2016**, *12* (12), 1065–1074. <https://doi.org/10.1038/nchembio.2203>.
- (7) Wang, S.; König, G.; Roth, H.-J.; Fouché, M.; Rodde, S.; Riniker, S. Effect of Flexibility, Lipophilicity, and the Location of Polar Residues on the Passive Membrane Permeability of a Series of Cyclic Decapeptides. *J. Med. Chem.* **2021**, *64* (17), 12761–12773. <https://doi.org/10.1021/acs.jmedchem.1c00775>.
- (8) Zhang, H.; Chen, S. Cyclic Peptide Drugs Approved in the Last Two Decades (2001–2021). *RSC Chem. Biol.* **2022**, *3* (1), 18–31. <https://doi.org/10.1039/D1CB00154J>.
- (9) Zorzi, A.; Deyle, K.; Heinis, C. Cyclic Peptide Therapeutics: Past, Present and Future. *Curr. Opin. Chem. Biol.* **2017**, *38*, 24–29. <https://doi.org/10.1016/j.cbpa.2017.02.006>.
- (10) Buckton, L. K.; Rahimi, M. N.; McAlpine, S. R. Cyclic Peptides as Drugs for Intracellular Targets: The Next Frontier in Peptide Therapeutic Development. *Chemistry – A European Journal* **2021**, *27* (5), 1487–1513. <https://doi.org/10.1002/chem.201905385>.
- (11) Buyanova, M.; Pei, D. Targeting Intracellular Protein–Protein Interactions with Macrocyclic Peptides. *Trends Pharmacol. Sci.* **2022**, *43* (3), 234–248. <https://doi.org/10.1016/j.tips.2021.11.008>.

- (12) Limbach, M. N.; Antevska, A.; Oluwatoba, D. S.; Gray, A. L. H.; Carroll, X. B.; Hoffmann, C. M.; Wang, X.; Voehler, M. W.; Steren, C. A.; Do, T. D. Atomic View of Aqueous Cyclosporine A: Unpacking a Decades-Old Mystery. *J. Am. Chem. Soc.* **2022**, *144* (28), 12602–12607. <https://doi.org/10.1021/jacs.2c01743>.
- (13) Jwad, R.; Weissberger, D.; Hunter, L. Strategies for Fine-Tuning the Conformations of Cyclic Peptides. *Chem. Rev.* **2020**, *120* (17), 9743–9789. <https://doi.org/10.1021/acs.chemrev.0c00013>.
- (14) Rüdiger, S. H.; Matabaro, E.; Sonderegger, L.; Güntert, P.; Künzler, M.; Gossert, A. D. Conformations of Macrocyclic Peptides Sampled by Nuclear Magnetic Resonance: Models for Cell-Permeability. *J. Am. Chem. Soc.* **2023**, *145* (50), 27601–27615. <https://doi.org/10.1021/jacs.3c09367>.
- (15) Zhu, K.; Trizio, E.; Zhang, J.; Hu, R.; Jiang, L.; Hou, T.; Bonati, L. Enhanced Sampling in the Age of Machine Learning: Algorithms and Applications. *Chem. Rev.* **2026**, *126* (1), 671–713. <https://doi.org/10.1021/acs.chemrev.5c00700>.
- (16) Hénin, J.; Lelièvre, T.; Shirts, M. R.; Valsson, O.; Delemotte, L. Enhanced Sampling Methods for Molecular Dynamics Simulations [Article v1.0]. *Living J. Comput. Mol. Sci.* **2022**, *4* (1), 1583. <https://doi.org/10.33011/livecoms.4.1.1583>.
- (17) Janson, G.; Valdes-Garcia, G.; Heo, L.; Feig, M. Direct Generation of Protein Conformational Ensembles via Machine Learning. *Nat. Commun.* **2023**, *14* (1), 774. <https://doi.org/10.1038/s41467-023-36443-x>.
- (18) Zhu, J.; Li, Z.; Tong, H.; Lu, Z.; Zhang, N.; Wei, T.; Chen, H.-F. Phanto-IDP: Compact Model for Precise Intrinsically Disordered Protein Backbone Generation and Enhanced Sampling. *Brief. Bioinform.* **2023**, *25* (1). <https://doi.org/10.1093/bib/bbad429>.
- (19) Rettie, S. A.; Campbell, K. V.; Bera, A. K.; Kang, A.; Kozlov, S.; Bueso, Y. F.; De La Cruz, J.; Ahlrichs, M.; Cheng, S.; Gerben, S. R.; Lamb, M.; Murray, A.; Adebomi, V.; Zhou, G.; DiMaio, F.; Ovchinnikov, S.; Bhardwaj, G. Cyclic Peptide Structure Prediction and Design Using AlphaFold2. *Nat. Commun.* **2025**, *16* (1), 4730. <https://doi.org/10.1038/s41467-025-59940-7>.
- (20) Grambow, C. A.; Weir, H.; Cunningham, C. N.; Biancalani, T.; Chuang, K. V. CREMP: Conformer-Rotamer Ensembles of Macrocyclic Peptides for Machine Learning. *Sci. Data* **2024**, *11* (1), 859. <https://doi.org/10.1038/s41597-024-03698-y>.
- (21) Ruzmetov, T.; Hung, T. I.; Jonnalagedda, S. P.; Chen, S.; Fasihianifard, P.; Guo, Z.; Bhanu, B.; Chang, C. A. Sampling Conformational Ensembles of Highly Dynamic Proteins via Generative Deep Learning. *J. Chem. Inf. Model.* **2025**, *65* (5), 2487–2502. <https://doi.org/10.1021/acs.jcim.4c01838>.

- (22) Miao, J.; Ghosh, A. P.; Ho, M. N.; Li, C.; Huang, X.; Pentelute, B. L.; Baleja, J. D.; Lin, Y.-S. Assessing the Performance of Peptide Force Fields for Modeling the Solution Structural Ensembles of Cyclic Peptides. *J. Phys. Chem. B* **2024**, *128* (22), 5281–5292. <https://doi.org/10.1021/acs.jpcc.4c00157>.
- (23) Li, Z.; Huang, Y.; Hung, T. I.; Sun, J.; Aispuro, D.; Chen, B.; Guevara, N.; Ji, F.; Cong, X.; Zhu, L.; Wang, S.; Guo, Z.; Chang, C.; Xue, M. MYC-Targeting Inhibitors Generated from a Stereodiversified Bicyclic Peptide Library. *J. Am. Chem. Soc.* **2024**. <https://doi.org/10.1021/jacs.3c09615>.
- (24) Chang, C.-E.; Potter, M. J.; Gilson, M. K. Calculation of Molecular Configuration Integrals. *J. Phys. Chem. B* **2003**, *107* (4), 1048–1055. <https://doi.org/10.1021/jp027149c>.
- (25) Altis, A.; Nguyen, P. H.; Hegger, R.; Stock, G. Dihedral Angle Principal Component Analysis of Molecular Dynamics Simulations. *J. Chem. Phys.* **2007**, *126* (24). <https://doi.org/10.1063/1.2746330>.
- (26) Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A. N.; Kaiser, L.; Polosukhin, I. Attention Is All You Need. *ArXiv* **2023**. <https://doi.org/https://doi.org/10.48550/arXiv.1706.03762>.
- (27) Scherer, G.; Kramer, M. L.; Schutkowski, M.; Reimer, U.; Fischer, G. Barriers to Rotation of Secondary Amide Peptide Bonds. *J. Am. Chem. Soc.* **1998**, *120* (22), 5568–5574. <https://doi.org/10.1021/ja980181t>.
- (28) Oakley, M. T.; Oheix, E.; Peacock, A. F. A.; Johnston, R. L. Computational and Experimental Investigations into the Conformations of Cyclic Tetra- $\alpha/\beta$ -Peptides. *J. Phys. Chem. B* **2013**, *117* (27), 8122–8134. <https://doi.org/10.1021/jp4043039>.
- (29) Whitty, A.; Zhong, M.; Viarengo, L.; Beglov, D.; Hall, D. R.; Vajda, S. Quantifying the Chameleonic Properties of Macrocycles and Other High-Molecular-Weight Drugs. *Drug Discov. Today* **2016**, *21* (5), 712–717. <https://doi.org/10.1016/j.drudis.2016.02.005>.



TOC Graphic