LassoHTP: a High-throughput Computational Tool for Lasso Peptide Structure Construction and Modeling

Reecan J. Juarez,[5] Matthew Tremblay,[1] Yaoyukun Jiang,[1] Qianzhen Shao,[1] A. James Link,[6] and Zhongyue J. Yang[1-4,*]

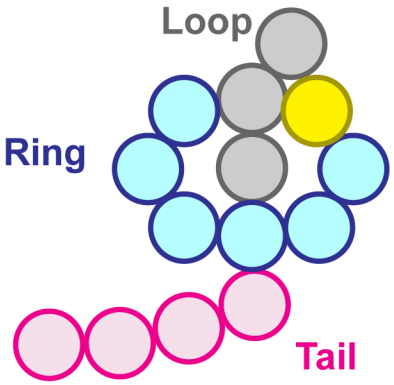
[1]*Department of Chemistry, Vanderbilt University, Nashville, Tennessee 37235, United States* [2]*Center for Structural Biology, Vanderbilt University, Nashville, Tennessee 37235, United States* [3]*Vanderbilt Institute of Chemical Biology, Vanderbilt University, Nashville, Tennessee 37235, United States* [4]*Data Science Institute, Vanderbilt University, Nashville, Tennessee 37235, United States* [5]*Chemical and Physical Biology Program, Vanderbilt University, Nashville, Tennessee 37235, United States* [6]*Department of Chemical and Biological Engineering, Chemistry and Molecular Biology, Princeton University, 207 Hoyt Laboratory, Princeton, New Jersey 08544, United States*

ABSTRACT: Lasso peptides are a sub-class of ribosomally synthesized and post-translationally modified peptides with a slipknot conformation. Often with superior thermal stability, protease resistance, and antimicrobial activity, lasso peptides are promising candidates for bioengineering and pharmseutical applications. To enable high-throughput computational prediction and design of lasso peptides, we developed software, LassoHTP, for automatic lasso peptide structure construction and modeling. LassoHTP consists of three modules, including: scaffold constructor, mutant generator, and molecular dynamics (MD) simulator. Based on a user-provided sequence and conformational annotation, LassoHTP can either generate the structure and conformational ensemble as is or conduct random mutagenesis. We used LassoHTP to construct eight known lasso peptide structures *de novo* and to simulate their conformational ensembles from 100 ns MD simulations. For benchmarking, we calculated the root mean square deviation (RMSD) of these ensembles with reference to their experimental crystal or NMR PDB structures; we also compared these RMSD values against those of the MD ensembles that are initiated from the PDB structures. The results show that the RMSD values of the LassoHTP-initiated ensembles are highly similar to those of the PDB-initiated ensembles with the ΔRMSD ranging from 0.0 to 1.2 Å and averaging at 0.5 Å. LassoHTP offers a computational platform to develop strategies for lasso peptide prediction and design.

Keywords: High-throughput Modeling, RiPPs, Lasso Peptides

1. Introduction

Lasso peptides are a sub-class of ribosomally synthesized and post-translationally modified peptides (RiPPs),[1, 2] some members of which exhibit antibacterial,[3-7] antiviral,[8] and antitumor activities[9]. Their bioactivities and favorable stability make lasso peptides good candidates for drug development campaigns.[6, 7, 10-12] Lasso peptides are a natural example [1]rotaxanes, a class of mechanically-interlocked molecules.[13] The [1]rotaxane slipknot conformation enables lasso peptides to be resilient towards thermal degradation (via the presence of bulky stopper residues) and proteolysis (via burial of the amide backbone).[14-17] The slipknot conformation can be formally represented by the C-terminus peptide tail threading into the macrolactam ring. The ring is connected by an isopeptide bond between the nitrogen of the N-terminus and a glutamate or aspartate carboxylate carbon located at the 7th, 8th, or 9th amino acid position.[18, 19] Two sterically bulky amino acids, called steric locks or upper and lower plugs, situated above and below the ring maintain the strained lariat-knot conformation. By definition, the lasso peptide loop refers to the amino acid sequence between the macrolactam ring and the upper plug; and tail between the lower plug and the C-terminus (Scheme-1).



Scheme-1. Conformational annotation for lasso peptide.

Computational studies have elucidated the physical basis of folding, threading, and thermally-actuated switching for specific lasso peptides.[6, 11, 15, 16] Ferguson et al. showed that uncyclized microcin J25 can spontaneously adopt a left-handed coil configuration using replica-exchange molecular dynamics (MD).[6] Allen et al. applied classical classical MD simulations with umbrella sampling to study the energetic preference of different pulling mechanisms for astexin-3 unthreading.[16] Using a poly-alanine model system, they demonstrated the relationship between the bulkiness of steric plugs and threading barrier. Yang et al. performed multiscale simulations on benenodin-1, a thermally-actuated molecular swtich between two conformers that differ by the position of the steric plug Q15.[15] By combining classical MD and large-scale quantum mechanics calculations, they quantified the entropy and enthalpy that accompany the conformational switching and elucidated the roles of steric plugs in mediating entropy-enthalpy compensation.

Molecular modeling studies deepen the insight into lasso peptide structure and dynamics. They rely on an experimentally-determined lasso peptide structures as a starting point or a reference. Given the large amount of natural lasso peptide sequences whose three-dimensional structure is undetermined,[1, 2, 20] computational structure prediction can greatly accelerate the process of investigating lasso peptide conformational dynamics and thermodynamic properties. However, unlike globular proteins whose structure can be predicted with AlphaFold2,[21] computational tools for lasso peptide structure construction or prediction are yet to be developed. Although lasso peptides share a common threaded scaffold, the task for structure prediction still presents a nontrivial challenge. The challenge lies in the diverse constructs of lasso peptides. Based on the observation of structurally-known lasso peptides, the loop size is three (e.g., xanthomonin-I and xanthomonin-II) for 7-membered ring,[22] but ranges from four (e.g., lariatin A and lariatin B)[23] to eighteen (e.g., ubonodin[4, 5]) for 8-membered ring, and from four (e.g., lihuanodin[24]) to seven

(e.g., caulonodin-V[14]) for 9-membered ring. In addition, the tail size can range from two in microcin J25,[7] citrocin[3] or ubonodin[4] to eighteen in pandonodin[25]. Despite the constructional diversity for known lasso peptides, a significantly greater number of lasso peptides still remain to be discovered and their structures determined.[2] With diverse scaffold constructs, the conformational ensembles are likely to be variable.

Here, we report LassoHTP as a computational tool for automatic, high-throuput lasso peptide structure prediction and conformational ensemble sampling. LassoHTP converts user-input lasso peptide sequences and conformational annotation to structure and conformational ensemble by employing three modules, including: the scaffold constructor, the mutant generator, and the MD simulator. LassoHTP can also be used to perform random mutagenesis with a template sequence or structural model. Finally, we benchmarked LassoHTP in the task of predicting the structure and conformational ensemble for eight known lasso peptides.

2. Design and Implementation

2.1 Architecture of LassoHTP. LassoHTP is designed to translate a user-defined lasso peptide sequence (with annotation of ring, loop, and tail) into a conformational ensemble (Figure 1). LassoHTP involves three modules to automate construction and simulation of lasso peptides, including: a scaffold constructor, a mutant generator, and an MD simulator. Specifically, the scaffold constrctor builds a poly-alanine lasso peptide scaffold; the mutant generator mutates the scaffold into a lasso peptide structure based on either user-defined sequence or random mutagenesis; the MD simulator parameterizes the lasso peptide structure and initiates MD simulations to output a conformational ensemble. The modular architecture ensures flexibility in application, which is similar to EnzyHTP, a software we developed for high-throughput enzyme

modeling.[26] Any single module can be independently executed to build, modify, or model a lasso peptide. Or, the three modules can be sequentially operated in an automatic workflow to convert user-defined lasso peptide sequences to structural ensembles.

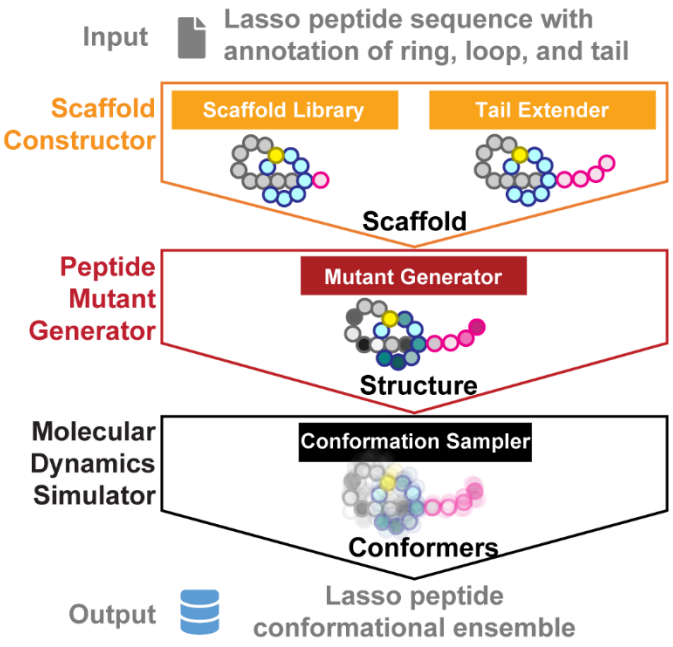


Figure 1. The workflow of LassoHTP to convert user-input sequence into a conformational ensemble via three modules, including: scaffold constructor, peptide mutant generator, and molecular dynamics simulator. To initiate, a user inputs a lasso peptide sequence and indicates the length for the ring, loop, and tail. From the sequence and conformational annotation, the scaffold constructor module generates prototypical poly-alanine lasso peptide scaffold via its native scaffold library and the tail extender function. The peptide mutant generator mutates each residue on the poly-alanine lasso peptide structure to match the input sequence. The molecular dynamics simulator module automatically writes MD input files, constructs a solvent box and initiates each subsequent MD step leading up to production MD.

2.2 Scaffold Construction. With an input of lasso peptide sequence and conformational annotation, LassoHTP's construction module uses a structural library in tandem with a tail-extender function to create a poly-alanine scaffold template (Supporting Information, Figure S1). The template lasso peptide scaffold can be mutated by the mutant generator and is compatible with software AMBER[27] used in MD simulator.

The structural library is a collection of 70 lasso peptide scaffolds consisting of a diverse range of ring and loop size (Supporting Information, Figure S2). Ring sizes range from 7 to 9 amino acids (Figure 2). Loop sizes range from 2 to 10 amino acids for 7- and 9-membered ring scaffolds and from 2 to 18 for 8-membered ring scaffolds. Each scaffold is composed almost entirely of alanine residues apart from the isopeptide moiety, which is the side chain of either an aspartate or glutamate residue (informed in the user-input sequence), of the lasso peptide ring.

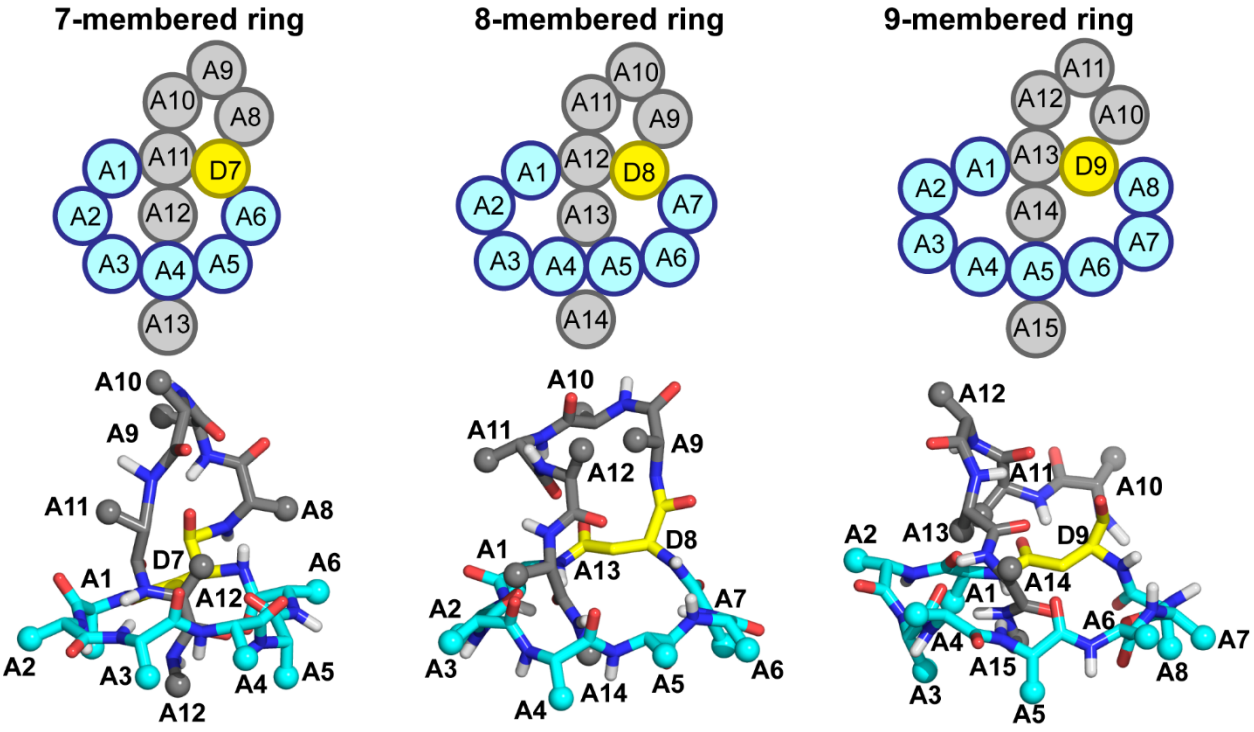


Figure 2. Representative structures in the LassoHTP scaffold library exhibiting various ring sizes with an aspartate isopeptide linker and loop length of 5. Backbone heavy atoms, alanine side-chains, and only polar hydrogens are shown for clarity. The ring, isopeptide bond, and loop moieties are colored cyan, yellow, and gray, respectively.

The poly-alanine scaffolds were mostly constructed through steered MD[28, 29] simulations using AMBER[27] (Figure 3 and detailed in the Computational Methods section); some scaffolds were truncated directly from known lasso peptide PDB structures (noted in Supporting Information, Figure S2). All scaffolds adopt a right-hand wrapping conformation.[30] The isopeptide bond of each scaffold adopts *trans*-isopeptide bond configuration. As a future plan, we will diversify the scaffold library by incorporating scaffold with left-hand wrapping topology and with *cis*-isopeptide bond configuration.

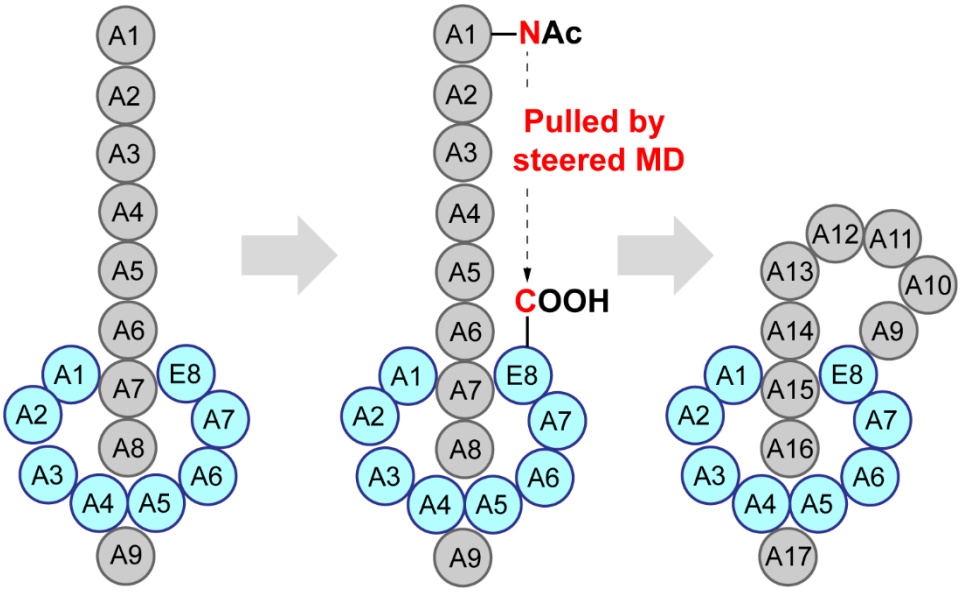


Figure 3. The workflow for constructing a poly-alanine scaffold using steered MD simulation. A linear peptide thread (colored in gray) is docked within the isopeptide ring (colored in cyan).

Through steered MD a pulling force brings the nitrogen of the N-terminus, which is capped with an acetyl group, to 2.0 Å from the carboxylate carbon of the isopeptide linker's C-terminus.

Based on the tail length defined by the user, the tail-extender function complements a selected scaffold by appending additional alanine residues to the C-terminus residue with *trans*-configuration peptide bond via a rotation matrix. The resulting lasso scaffold lays the foundation for the subsequent modules.

2.3 Mutant Generation. LassoHTP's mutant generator takes a lasso peptide scaffold as input, and mutates the scaffold's residues to any of the 20 canonical amino acids. LassoHTP allows users to either build a mutant structure based on input sequence and conformational annotation, or conduct random mutagenesis on a lasso peptide structure (Figure 4). The input scaffold received by mutant generator can be an output of scaffold constructor (i.e., a poly-alanine scaffold) or a stand-alone lasso peptide structural model.

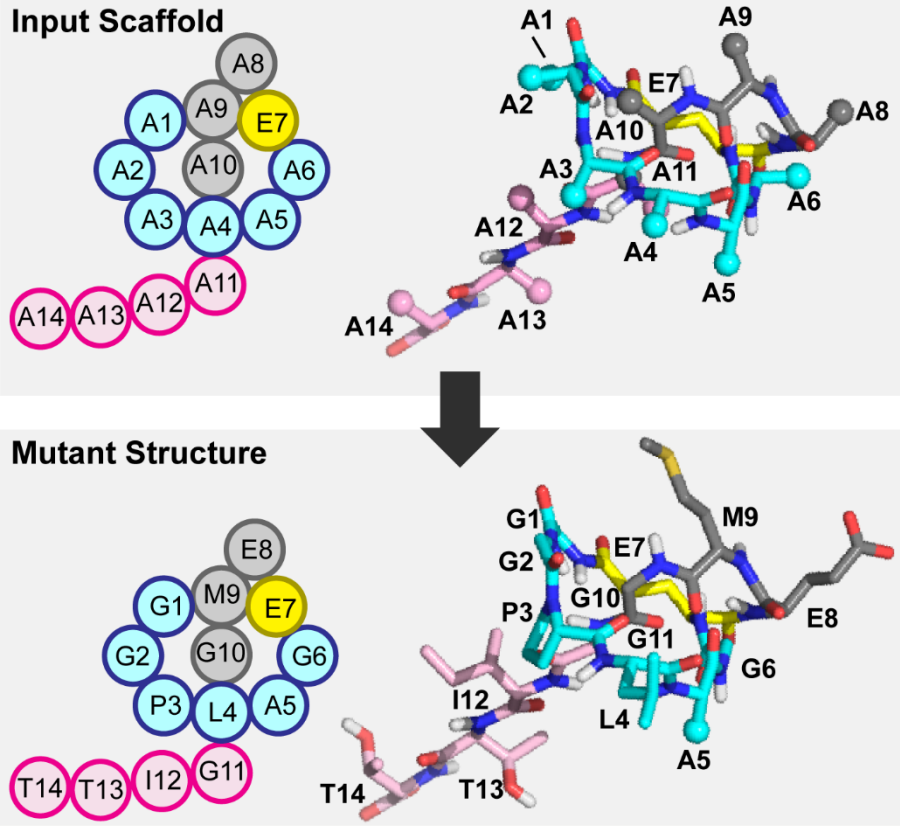


Figure 4. Conversion of the poly-alanine lasso peptide scaffold into the lasso peptide that is consistent with the user-input sequence using the mutant generator module. Using the initial input sequence, the mutant generator module mutates each alanine to match the input amino acid sequence. Sequence shown is xanthomonin-II[22] (PDB ID: 2MFV).

The mutant generator operates on a given lasso peptide scaffold as a Python object. Specifically, the module recognizes the scaffold as a structural object and divides it into subunits such as residues and atoms. Using *seq_parse.py*, the module parses these subunits in accordance with the major sections of a lasso peptide, including the ring, loop, upper and lower plugs, and tail (detailed in the Supporting information, Text S1). Inspired by the framework of EnzyHTP,[26] the parser treats each individual amino acid in the lasso peptide sequence as a "flag" in the form of '[X#Y]' where X is the original residue, # is the positional index, and Y is the mutated residue.

The function uses tLEaP in AMBER to mutate each alanine residue on the scaffold by adding and removing atoms as needed. The module does not mutate the acidic residue that forms the isopeptide moiety to preserve the scaffold's lariat-knot conformation. tLEaP reparameterizes the mutated amino acids and the isopeptide bond (Supporting Information, Table S1).[11]

Regarding random mutation of certain scaffold positions, the mutant generator also allows the user to perform selective mutations on any major sections of a given lasso peptide scaffold (Supporting Information, Figure S3). As such, users can manually input the peptide sequence, choose to randomly generate a sequence, or manually or randomly mutate only certain positions of the scaffold.

2.4 MD Simulation. LassoHTP's MD simulator conducts classical MD simulations of mutant lasso peptides. The input structure can either derive from the mutant generator or provided by the user based on experimentally-determined structures. For any lasso peptide, the module automatically generates input files for minimization, heating, equilibration, and production MD and automatically initiates each stage in the MD process (Supporting Information, input_prmtop_inpcrd.zip file). The production MD yields a final output of MD trajectories for a given lasso peptide, which can then be sequestered into conformational ensembles. The conformational ensembles can be evaluated for structural features such as NMR restraints and thermodynamic properties such as free energy and conformational entropy. Although classical MD is the default setting for the module, enhanced sampling such as umbrella sampling[31] can also be conducted.

3. Computational Methods

3.1 Steered MD. We employed steered MD to construct lasso peptide scaffolds in the library (Supporting Information, Text S2). Each scaffold is built in two stages. The first stage is building an isopeptide ring-linear peptide thread complex (refer to as ring-thread complex below). AMBER's tLEaP module is applied to build a linear poly-alanine peptide thread that is capped at the N- and C-terminus with acetyl and N-methyl groups, respectively. The 9-membered macrolactam isopeptide ring is extracted from astexin-3 PDB structure (PDB ID: 2M8F).[32] The 7- and 8-membered rings were manually constructed by adapting the structure of 9-membered ring using PyMOL. Each residue on the ring is converted to alanine using PyMOL's mutagenesis feature. To parameterize the isopeptide bond, we adopted the force field parameter from the prior work[11] by Lai and Kaznessis and determined atomic charges using the AM1-BCC model[33, 34] (Supporting Information, Table S1). For the canonical amino acids in the peptide, ff14SB force field is employed.[35] Finally, the ring-thread complex is constructed by docking the peptide thread in the center of the ring in an orthogonal position using *com_placement.py*.[36] This is accomplished by docking the center of mass (CoM) of the linear peptide thread's C-terminal residue in the isopeptide ring's CoM and along the ring's z-axis. The final output is a PDB file that geometrically defines the ring-thread complex (*right*, Figure 3).

The second stage is constructing the loop of the lasso peptide scaffold. Using tLEaP, the ring-thread complex is solvated in a TIP3P water[37] octahedral solvent box with a 40 Å buffer. PMEMD.CUDA[38, 39] is used to minimize, heat the system towards 300K, and equilibrate the system. In equilibration, restraints are applied to the N-methyl cap and C-terminal alanine to prevent unthreading (20 kcal/mol·Å). A 30 ps steered MD is conducted using SANDER to direct the peptide thread's N-terminal residue towards the C-terminus of the isopeptide ring using a harmonic restraint of 2000 kcal/mol·Å (i.e., the main chain carboxylate on the glutamate or

asparate). For steered MD, the nitrogen atom of the N-terminal alanine and the carbon atom of the isopeptide ring's C-terminus are defined as target points (*middle*, Figure 3). The target distance for steered MD simulation is set to be 1.5 Å. The resulting structure from the steered MD will be converted to to a lasso peptide scaffold by tLEaP. Speifically, tLEaP is used to bond these the two target atoms (i.e., C and N) and remove the acetyl group and extraneous hydrogen and hydroxyl groups. The final output is a PDB file that defines the lasso peptide's ring and loop structures.

3.2 Classical MD simulation. For each lasso peptide used in the benchmark, we performed classical MD simulations using the pmemd.cuda engine of AMBER with one NVIDIA Pascal GPU.[38, 39] The LassoHTP workflow first generates the force field parameters using ff14SB force field, constructs a 10 Å octahedral solvent box and then automatically composes the input files (Supporting Information, Figure S4), with preset parameters, for 20,000 cycles of minimization, 40 ps heating to 300K, 1 ns NPT equilibration (backbone atoms restrained with a harmonic potential of 2 kcal/mol·K), 1 ns NVT equilibration (unrestrained), and 100 ns production MD. Each simulation uses a time step of 2 fs, Langevin thermostat, and Berendsen barostat. From a 100 ns MD trajectory, 1000 snapshots are evenly extracted to form a conformational ensemble. We characterized the derivation of the conformers from the experimentally-determined structure by calculating the root mean square deviation (i.e., RMSD) of the backbone heavy atoms relative to either the first structure of the NMR ensemble or crystal structure. We used Xmgrace to visualize and produce histograms for each RMSD analysis.

4. Results and Discussion

As the first test, we employed LassoHTP to predict the conformational ensemble (called LHTP-initiated MD, Figure 5) for the wild-type caulosegnin-II[14] (PDB ID: 5D9E) and

benchmarked its consistency against the MD ensemble initiated from the experimentally-determined structure (called PDB-initiated MD, Figure 5). Caulosegnin-II has a glutamate-linked 9-membered ring, a 6 amino acid (aa) loop, and 4 aa tail. We employed wild-type caulosegnin-II as our first test because a well-resolved crystal structure (resolution: 0.86 Å) is known for this lasso peptide (Met17 was oxidize to methionine sulfoxide during crystallization).[14] The comparison between the two MD conformational ensembles normalizes the impact of force field parameters on conformational sampling. Both LHTP-initiated and PDB-initiated MD ensemble were constructed by evenly extracting 1000 snapshots from a 100 ns MD trajectory. The RMSD of the heavy backbone atoms (i.e. C_α, N, C, and O) of each conformer was computed relative to the crystal structure (Supporting information, Text S3).

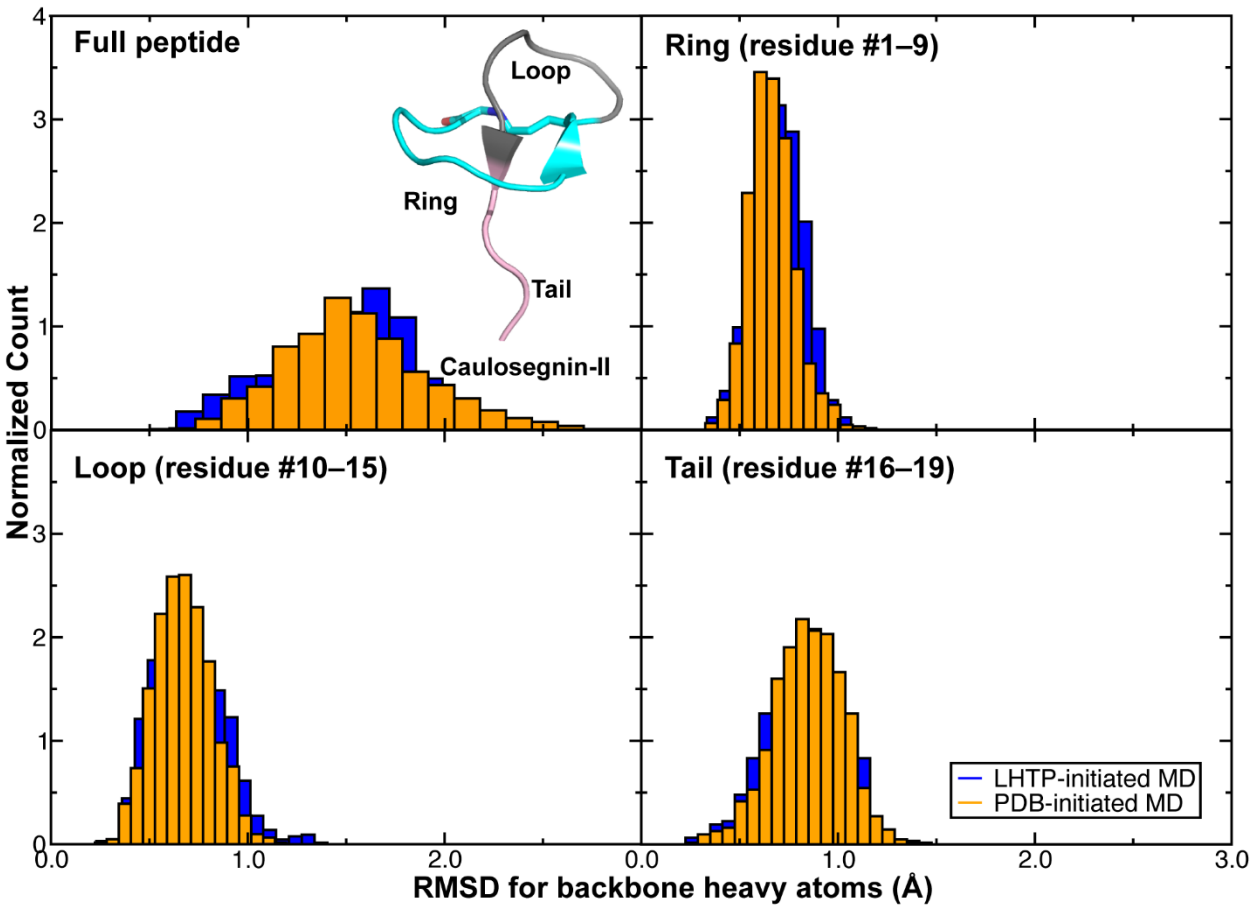


Figure 5. Distribution of RMSD for LassoHTP-initiated and PDB-initiated MD conformational ensemble for caulosegnin-II. Both ensembles are constructed by 1000 snapshots that are evenly extracted from 100 ns MD trajectories. The RMSD distributions for the full peptide as well as sub-structures (i.e., ring, loop, and tail) are shown. A cartoon representation of caulosegnin-II is shown in the inset with the ring, loop, and tail colored in cyan, gray, and pink. For both ensembles, RMSD was calculated using backbone atoms (i.e., C_α, N, C, and O) with reference to the experimentally-determined PDB crystal structure.

The average RMSD for the LHTP-inititated and PDB-inititated MD ensemble are 1.48 Å and 1.55 Å, respectively (*top left*, Figure 5). To further confirm LassoHTP's fidelity, we calculated backbone heavy atom RMSD for the ring, loop, and tail substructures with the reference taken from the corresponding structural moeity in the crystal structure. In contrast, the average RMSD for the LHTP versus PDB ensemble is 0.70 Å vs. 0.66 Å for the ring (*top right*, Figure 5), 0.70 Å vs. 0.67 Å for the loop (*bottom left*, Figure 5), and 0.84 Å vs. 0.85 Å for the tail (*bottom right*, Figure 5). These results show that the RMSD values calculated from the LHTP-initiated ensemble closely align with those from the PDB-initiated ensemble regarding the full peptide and its substructures. The alignment in the tail substructure is especially impressive because the tail is generated in situ by the tail extender function in LassoHTP and is the most flexible part of the peptide. The consistency of the RMSD values in the ring is somewhat expected because of its conformational rigidity. The close match in the loop RMSD values can be attributed to the use of caulosegnin-II loop conformation in the LassoHTP scaffold library (Supporting Information, Figure S2), although the residues are replaced by alanine in the library. The similar RMSD distributions between both ensembles show that LassoHTP yields a reliable lasso peptide molecular model and conformational ensemble.

In addition to caulosegnin-II, we further tested LassoHTP using seven distinct lasso peptide structures that have been determined by NMR (Figure 6 and Supporting Information, Table S2). We followed a similar approach to our testing protocol for caulosegnin-II. The first structural model of each peptide's NMR-resolved structural ensemble was used to inititate the sampling of MD conformational ensemble. We also used the first structural model as a reference structure for RMSD calculations in both LHTP- and PDB-initiated MD ensembles. The seven lasso peptides used in the benchmark are: benenodin-1 conformer 1 (PDB ID: 5TJ1),[17] benenodin-1 conformer 2 (PDB ID: 6B5W),[17] citrocin (PDB ID: 6MW6),[3] the RGD variant (i.e., G12R, I13G, and G14D) of microcin J25 (mccJ25 RGD, PDB ID: 2MMW),[12] streptomonomicin (PDB ID: 2MW3),[40] ubonodin (PDB ID: 6POR),[4, 5] and xanthomonin-II (PDB ID: 2MFV)[22]. They involve a wide range of lasso constructs. Xanthomonin-II has a 7-membered ring; citrocin, mccJ25 RGD, ubonodin, and benenodin-1 have 8-membered rings; and streptomonomicin has a 9-membered ring. The loop size ranges from 4 aa in xanthomonin-II to 18 aa in ubonodin.

Using LassoHTP, the sequence of each lasso peptide, along with the annotation of ring, loop, and tail size, is converted to a lasso peptide structural model. In the benchmark, the poly-alanine scaffolds used to construct the lasso peptide structures were derived from either PDB, including: citrocin (i.e., ring size: 8 aa and loop size: 9 aa), MccJ25 RGD (i.e., ring size: 8 and loop size: 11 aa), and ubonodin (i.e., ring size: 8 aa and loop size: 18 aa), or steered-MD simulations, including: benenodin-1 conformer 1 (i.e., ring size: 8 aa and loop size: 6 aa), benenodin-1 conformer 2 (i.e., ring size: 8 aa and loop size: 7 or 8 aa), streptomonomicin (i.e., ring size: 9 aa and loop size: 4, 5, or 6 aa), and xanthomonin-II (i.e., ring size: 7 aa and loop size: 2, 3, or 4 aa). In addition, only one scaffold construct was used for benenodin-1 conformer 1, citrocin, MccJ25 RGD, and ubonodin due to the adjacency between the steric plugs; multiple

scaffold constructs with various loop/tail sizes (i.e., translational isomers) were employed for benenodin-1 conformer 2, streptomonomicin, and xanthomonin-II because the upper and lower steric plugs in these peptides are gapped by one or a couple of amino acids (Supporting Information, Text S4). In this way, LassoHTP can maximize sampling of the conformational space by automatically modeling multiple translational isomers of the lasso peptide. For these peptides, the RMSD value is averaged over ensembles of all possible constructs.

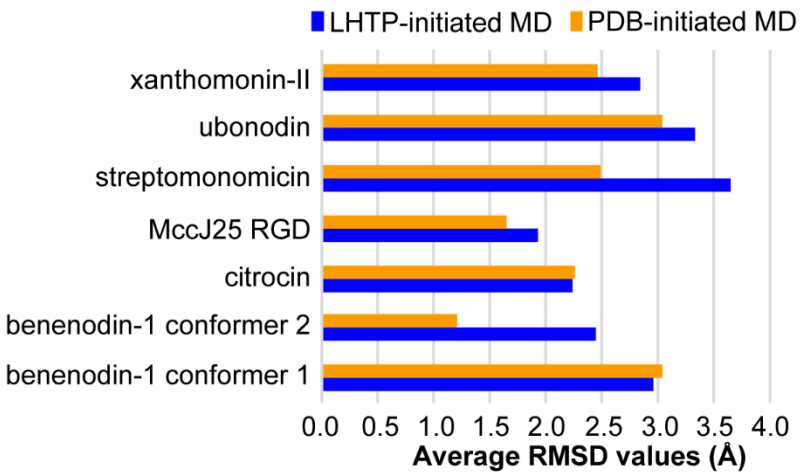


Figure 6. Averge RMSD values of LHTP-initiated (colored in blue) and PDB-initiated (colored in orange) MD ensembles for eight lasso peptides involved in the benchmark. The structures of the lasso peptides were determined mostly by NMR except for caulosegnin-II by X-ray crystallography. Average RMSD calculations include backbone heavy atoms (i.e., C_α, N, C, and O) with reference to the crystal structure or the first structure of the NMR ensemble. For streptomonomicin, xanthomonin-II, and benenodin-1 confromer 2, the average RMSD values for the LHTP-inititated MD ensembles account for multiple LassoHTP constructs with various loop and tail sizes.

In terms of deviation from the reference NMR structure, the average RMSD values of the LHTP-initiated ensembles range from 1.93 Å for mccJ25 RGD to 3.64 Å for streptomonomicin; those of the PDB-initiated ensembles range from 1.21 Å for benenodin-1 conformer 2 to 3.04 Å for ubonodin. In both MD ensembles, the sampled conformers deviate reasonably from the reference structure. Two peptides, ubonodin and streptomonomicin, involve an average RMSD values higher than 3.00 Å in the LHTP-initiated MD ensemble. By inspecting the RMSD values of the substructures, we observed that ubonodin involves a flexible loop (i.e., 2.96 Å), while streptomonomicin involves a flexible tail (i.e., 2.93 Å by averaging over three constructs, Supporting Information, Table S2).

Furthermore, we compared the difference of the average RMSD values between the two ensembles (i.e., ΔRMSD, defined as $RMSD_{LHTP} - RMSD_{PDB}$). The ΔRMSD values range from ~0.0 Å for benenodin-1 conformer 1 and citrocin to ~1.2 Å for streptomonomicin and benenodin-1 conformer 2. Over seven lasso peptides in the benchmark, the average of ΔRMSD values is 0.48 Å, which indicates the conformational similarity between the two ensembles. Benenodin-1 conformer 2 and streptomonomicin involve the largest discrepancy between the two ensembles. Benenodin-1 conformer 2 involves a *cis*-isopeptide bond,[15] which was not considered in the current version of LassoHTP. For streptomonomicin, the NMR structure was determined in methanol. The simulations using water box likely contribute to enlarge the deviation from the reference structure, albeit being more relevant to their biological environment. In the future version of LassoHTP, we will implement functions to construct *cis*-configuration for the peptide or isopeptide bond; we will also implement modules to allow using an organic solvent box for MD sampling of lasso peptides. We should note that the use of 100 ns classical MD might not sufficiently cover the conformational space for lasso peptides with a flexible loop or tail. In

practice, we would recommend using accelerated[41] or replica-exchange[42] MD for conformational sampling like what has been extensively demonstrated in the conformational studies of cyclic peptide.[43-45]

5. **Conclusion**

LassoHTP is a new Python platform that allows automatic structure construction and conformational sampling in a high-throughput fashion. LassoHTP uses a user-input lasso peptide sequence and conformational annotation to construct a lasso peptide structure and generate the structure's conformational ensemble. The LassoHTP workflow accomplishes this via three main modules: the scaffold constructor module, the mutation module, and the high-throughput MD module. The scaffold constructor module constructs the proto-lasso peptide structure (i.e., polyalanine scaffold) by selecting a ring and upper loop structure from the scaffold library followed by appending additional alanine residues to complete the tail length. The mutant generator module changes all alanines on the proto-structure to match the user-given amino acid sequence. The MD simulator module automatically generates input files for the structure, optimizes the structure, and then performs MD sampling to construct a conformational ensemble.

To evaluate LassoHTP's performance, we benchmarked LassoHTP against eight known lasso peptides, including one with a crystal structure (i.e., caulosegnin-II) and seven with NMR structures (i.e., benenodin-1 conformer 1, benenodin-1 conformer 2, citrocin, MccJ25 RGD, xanthomonin-II, streptomonomicin, and ubonodin). We applied LassoHTP to construct *de novo* structures of these lasso peptides and generate their conformational ensembles with MD simulations. Then, we generated conformational ensembles using the experimentally-determined structures as the initial structures. Using backbone RMSD as a metric and the first structure of the

NMR ensemble or, in the case of caulosegnin-II, the crystal structure as reference, we compared the two distinct conformational ensembles for each lasso peptide. All LassoHTP-generated MD ensembles are well consistent with their PDB-initiated MD ensembles, on average deviating no more than 1.2 Å in RMSD. The benchmark shows that LassoHTP can generate valid lasso peptide structures and conformational ensembles *de novo*. As such, lassoHTP provides a foundation for high-throughput computational lasso peptide prediction and design to facilitate experimental discovery.

ASSOCIATED CONTENT

Supporting Information. Sample codes; scaffold structure models; force field parameters; steered MD simulation protocol; sample input files; RMSD values for tested lasso peptides and their substructures (PDF)

input_prmtop_inpcrd.zip and LassoHTP_scaffold_library.zip (ZIP).

Data and Software Availability. The code and sample input for LassoHTP framework is publically available at https://github.com/ChemBioHTP/LassoHTP/. The input files and structures are provided as part of the SI files. AMBER 18 is available from http://ambermd.org/.

AUTHOR INFORMATION

Corresponding Author

*Email: zhongyue.yang@vanderbilt.edu phone: 615-343-9849

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENT

The authors thank Chris Jurich for providing technical support for producing Figure S2. This research was supported by the startup grant from Vanderbilt University. This work was carried out in part using computational resources from the Extreme Science and Engineering Discovery Environment (XSEDE), which is supported by National Science Foundation grant number TG-BIO200057.[46] Support for work on lasso peptides to A.J.L. was provided by NIH grant GM107036. R.J.J. thanks the financial support from the National Institutes of Health Molecular Biophysics Training Grant (MBTP T32 GM008320).

References

1. Montalbán-López, M.; Scott, T. A.; Ramesh, S.; Rahman, I. R.; van Heel, A. J.; Viel, J. H.; Bandarian, V.; Dittmann, E.; Genilloud, O.; Goto, Y.; Grande Burgos, M. J.; Hill, C.; Kim, S.; Koehnke, J.; Latham, J. A.; Link, A. J.; Martínez, B.; Nair, S. K.; Nicolet, Y.; Rebuffat, S.; Sahl, H.-G.; Sareen, D.; Schmidt, E. W.; Schmitt, L.; Severinov, K.; Süssmuth, R. D.; Truman, A. W.; Wang, H.; Weng, J.-K.; van Wezel, G. P.; Zhang, Q.; Zhong, J.; Piel, J.; Mitchell, D. A.; Kuipers, O. P.; van der Donk, W. A., New developments in RiPP discovery, enzymology and engineering. *Natural Product Reports* **2021**, 38, 130-239.
2. Tietz, J. I.; Schwalen, C. J.; Patel, P. S.; Maxson, T.; Blair, P. M.; Tai, H.-C.; Zakai, U. I.; Mitchell, D. A., A new genome-mining tool redefines the lasso peptide biosynthetic landscape. *Nature Chemical Biology* **2017**, 13, 470-478.
3. Cheung-Lee, W. L.; Parry, M. E.; Jaramillo Cartagena, A.; Darst, S. A.; Link, A. J., Discovery and structure of the antimicrobial lasso peptide citrocin. *Journal of Biological Chemistry* **2019**, 294, 6822-6830.
4. Cheung‐Lee, W. L.; Parry, M. E.; Zong, C.; Cartagena, A. J.; Darst, S. A.; Connell, N. D.; Russo, R.; Link, A. J., Discovery of ubonodin, an antimicrobial lasso peptide active against members of the Burkholderia cepacia complex. *ChemBioChem* **2020**, 21, 1335-1340.
5. Do, T.; Thokkadam, A.; Leach, R.; Link, A. J., Phenotype-Guided Comparative Genomics Identifies the Complete Transport Pathway of the Antimicrobial Lasso Peptide Ubonodin in Burkholderia. *ACS Chemical Biology* **2022**.
6. Ferguson, A. L.; Zhang, S.; Dikiy, I.; Panagiotopoulos, A. Z.; Debenedetti, P. G.; James Link, A., An Experimental and Computational Investigation of Spontaneous Lasso Formation in Microcin J25. *Biophysical Journal* **2010**, 99, 3056-3065.
7. Salomón, R. A.; Farías, R. N., Microcin 25, a novel antimicrobial peptide produced by Escherichia coli. *Journal of Bacteriology* **1992**, 174, 7428-7435.
8. Constantine, K. L.; Friedrichs, M. S.; Detlefsen, D.; Nishio, M.; Tsunakawa, M.; Furumai, T.; Ohkuma, H.; Oki, T.; Hill, S.; Bruccoleri, R. E.; Lin, P.-F.; Mueller, L., High-resolution solution structure of siamycin II: Novel amphipathic character of a 21-residue peptide that inhibits HIV fusion. *Journal of Biomolecular NMR* **1995**, 5, 271-286.

9. Son, S.; Jang, M.; Lee, B.; Hong, Y.-S.; Ko, S.-K.; Jang, J.-H.; Ahn, J. S., Ulleungdin, a Lasso Peptide with Cancer Cell Migration Inhibitory Activity Discovered by the Genome Mining Approach. *Journal of Natural Products* **2018**, 81, 2205-2211.
10. Adelman, K.; Yuzenkova, J.; La Porta, A.; Zenkin, N.; Lee, J.; Lis, J. T.; Borukhov, S.; Wang, M. D.; Severinov, K., Molecular Mechanism of Transcription Inhibition by Peptide Antibiotic Microcin J25. *Molecular Cell* **2004**, 14, 753-762.
11. Lai, P.-K.; Kaznessis, Y. N., Free Energy Calculations of Microcin J25 Variants Binding to the FhuA Receptor. *Journal of Chemical Theory and Computation* **2017**, 13, 3413-3423.
12. Knappe, T. A.; Manzenrieder, F.; Mas‐Moruno, C.; Linne, U.; Sasse, F.; Kessler, H.; Xie, X.; Marahiel, M. A., Introducing lasso peptides as molecular scaffolds for drug design: engineering of an integrin antagonist. *Angewandte Chemie International Edition* **2011**, 50, 8714-8717.
13. Sluysmans, D.; Stoddart, J. F., The Burgeoning of Mechanically Interlocked Molecules in Chemistry. *Trends in Chemistry* **2019**, 1, 185-197.
14. Hegemann, J. D.; Fage, C. D.; Zhu, S.; Harms, K.; Di Leva, F. S.; Novellino, E.; Marinelli, L.; Marahiel, M. A., The ring residue proline 8 is crucial for the thermal stability of the lasso peptide caulosegnin II. *Molecular BioSystems* **2016**, 12, 1106-1109.
15. Yang, Z.; Hajlasz, N.; Kulik, H. J., Computational Modeling of Conformer Stability in Benenodin-1, a Thermally Actuated Lasso Peptide Switch. *The Journal of Physical Chemistry B* **2022**, 126, 3398-3406.
16. Allen, C. D.; Chen, M. Y.; Trick, A. Y.; Le, D. T.; Ferguson, A. L.; Link, A. J., Thermal Unthreading of the Lasso Peptides Astexin-2 and Astexin-3. *ACS Chemical Biology* **2016**, 11, 3043-3051.
17. Zong, C.; Wu, M. J.; Qin, J. Z.; Link, A. J., Lasso Peptide Benenodin-1 Is a Thermally Actuated [1]Rotaxane Switch. *Journal of the American Chemical Society* **2017**, 139, 10403-10409.
18. Maksimov, M. O.; Pan, S. J.; James Link, A., Lasso peptides: structure, function, biosynthesis, and engineering. *Natural Product Reports* **2012**, 29, 996-1006.
19. Hegemann, J. D.; Zimmermann, M.; Xie, X.; Marahiel, M. A., Lasso Peptides: An Intriguing Class of Bacterial Natural Products. *Accounts of Chemical Research* **2015**, 48, 1909-1919.
20. Cheung-Lee, W. L.; Link, A. J., Genome mining for lasso peptides: past, present, and future. *Journal of Industrial Microbiology and Biotechnology* **2019**, 46, 1371-1379.
21. Tunyasuvunakool, K.; Adler, J.; Wu, Z.; Green, T.; Zielinski, M.; Žídek, A.; Bridgland, A.; Cowie, A.; Meyer, C.; Laydon, A.; Velankar, S.; Kleywegt, G. J.; Bateman, A.; Evans, R.; Pritzel, A.; Figurnov, M.; Ronneberger, O.; Bates, R.; Kohl, S. A. A.; Potapenko, A.; Ballard, A. J.; Romera-Paredes, B.; Nikolov, S.; Jain, R.; Clancy, E.; Reiman, D.; Petersen, S.; Senior, A. W.; Kavukcuoglu, K.; Birney, E.; Kohli, P.; Jumper, J.; Hassabis, D., Highly accurate protein structure prediction for the human proteome. *Nature* **2021**.
22. Hegemann, J. D.; Zimmermann, M.; Zhu, S.; Steuber, H.; Harms, K.; Xie, X.; Marahiel, M. A., Xanthomonins I–III: A New Class of Lasso Peptides with a Seven‐Residue Macrolactam Ring. *Angewandte Chemie International Edition* **2014**, 53, 2230-2234.
23. Iwatsuki, M.; Tomoda, H.; Uchida, R.; Gouda, H.; Hirono, S.; Ōmura, S., Lariatins, Antimycobacterial Peptides Produced by Rhodococcus sp. K01–B0171, Have a Lasso Structure. *Journal of the American Chemical Society* **2006**, 128, 7486-7491.

24. Cao, L.; Beiser, M.; Koos, J. D.; Orlova, M.; Elashal, H. E.; Schröder, H. V.; Link, A. J., Cellulonodin-2 and Lihuanodin: Lasso Peptides with an Aspartimide Post-Translational Modification. *Journal of the American Chemical Society* **2021**, 143, 11690-11702.
25. Cheung-Lee, W. L.; Cao, L.; Link, A. J., Pandonodin: A Proteobacterial Lasso Peptide with an Exceptionally Long C-Terminal Tail. *ACS Chemical Biology* **2019**, 14, 2783-2792.
26. Shao, Q.; Jiang, Y.; Yang, Z. J., EnzyHTP: A High-Throughput Computational Platform for Enzyme Modeling. *Journal of Chemical Information and Modeling* **2022**, 62, 647-655.
27. D.A. Case, D. S. C., T.E. Cheatham, III, T.A. Darden, R.E. Duke, T.J. Giese, H. Gohlke, A.W. Goetz, D.; Greene, N. H., S. Izadi, A. Kovalenko, T.S. Lee, S. LeGrand, P. Li, C. Lin, J. Liu, T. Luchko, R. Luo,; D. Mermelstein, K. M. M., G. Monard, H. Nguyen, I. Omelyan, A. Onufriev, F. Pan, R. Qi, D.R. Roe, A.; Roitberg, C. S., C.L. Simmerling, W.M. Botello-Smith, J. Swails, R.C. Walker, J. Wang, R.M. Wolf, X.; Wu, L. X., D.M. York and P.A. Kollman *AMBER 2017*, University of California, San Francisco: 2017.
28. Isralewitz, B.; Gao, M.; Schulten, K., Steered molecular dynamics and mechanical functions of proteins. *Current Opinion in Structural Biology* **2001**, 11, 224-230.
29. Lu, H.; Schulten, K., Steered molecular dynamics simulations of force-induced protein domain unfolding.
30. Schröder, H. V.; Zhang, Y.; Link, A. J., Dynamic covalent self-assembly of mechanically interlocked molecules solely made from peptides. *Nature Chemistry* **2021**, 13, 850-857.
31. Kästner, J., Umbrella sampling. *Wiley Interdisciplinary Reviews: Computational Molecular Science* **2011**, 1, 932-942.
32. Maksimov, M. O.; Link, A. J., Discovery and Characterization of an Isopeptidase That Linearizes Lasso Peptides. *Journal of the American Chemical Society* **2013**, 135, 12038-12047.
33. Jakalian, A.; Jack, D. B.; Bayly, C. I., Fast, efficient generation of high ‐ quality atomic charges. AM1 ‐ BCC model: II. Parameterization and validation. *Journal of computational chemistry* **2002**, 23, 1623-1641.
34. Jakalian, A.; Bush, B. L.; Jack, D. B.; Bayly, C. I., Fast, efficient generation of high ‐ quality atomic charges. AM1 ‐ BCC model: I. Method. *Journal of computational chemistry* **2000**, 21, 132-146.
35. Maier, J. A.; Martinez, C.; Kasavajhala, K.; Wickstrom, L.; Hauser, K. E.; Simmerling, C., ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. *Journal of Chemical Theory and Computation* **2015**, 11, 3696-3713.
36. Dickson, C. J. *AMBER-Umbrella COM restraint tutorial*, GitHub, 2020.
37. Mark, P.; Nilsson, L., Structure and Dynamics of the TIP3P, SPC, and SPC/E Water Models at 298 K. *The Journal of Physical Chemistry A* **2001**, 105, 9954-9960.
38. Salomon-Ferrer, R.; Götz, A. W.; Poole, D.; Le Grand, S.; Walker, R. C., Routine Microsecond Molecular Dynamics Simulations with AMBER on GPUs. 2. Explicit Solvent Particle Mesh Ewald. *Journal of Chemical Theory and Computation* **2013**, 9, 3878-3888.
39. Götz, A. W.; Williamson, M. J.; Xu, D.; Poole, D.; Le Grand, S.; Walker, R. C., Routine Microsecond Molecular Dynamics Simulations with AMBER on GPUs. 1. Generalized Born. *Journal of Chemical Theory and Computation* **2012**, 8, 1542-1555.
40. Metelev, M.; Tietz, Jonathan I.; Melby, Joel O.; Blair, Patricia M.; Zhu, L.; Livnat, I.; Severinov, K.; Mitchell, Douglas A., Structure, Bioactivity, and Resistance Mechanism of Streptomonomicin, an Unusual Lasso Peptide from an Understudied Halophilic Actinomycete. *Chemistry & Biology* **2015**, 22, 241-250.

41. Hamelberg, D.; Mongan, J.; McCammon, J. A., Accelerated molecular dynamics: A promising and efficient simulation method for biomolecules. *The Journal of Chemical Physics* **2004**, 120, 11919-11929.
42. Sugita, Y.; Okamoto, Y., Replica-exchange molecular dynamics method for protein folding. *Chemical Physics Letters* **1999**, 314, 141-151.
43. McHugh, S. M.; Rogers, J. R.; Yu, H.; Lin, Y.-S., Insights into How Cyclic Peptides Switch Conformations. *Journal of Chemical Theory and Computation* **2016**, 12, 2480-2488.
44. Miao, J.; Descoteaux, M. L.; Lin, Y.-S., Structure prediction of cyclic peptides by molecular dynamics + machine learning. *Chemical Science* **2021**, 12, 14927-14936.
45. Damjanovic, J.; Miao, J.; Huang, H.; Lin, Y.-S., Elucidating Solution Structures of Cyclic Peptides Using Molecular Dynamics Simulations. *Chemical Reviews* **2021**, 121, 2292-2324.
46. Towns, J.; Cockerill, T.; Dahan, M.; Foster, I.; Gaither, K.; Grimshaw, A.; Hazlewood, V.; Lathrop, S.; Lifka, D.; Peterson, G. D.; Roskies, R.; Scott, J. R.; Wilkins-Diehr, N., XSEDE: Accelerating Scientific Discovery. *Comput. Sci. Eng.* **2014**, 16, 62-74.

Table of Contents Graphic