Title

LassoHTP: A High-Throughput Computational Tool for Lasso Peptide Structure Construction and Modeling.

Journal

Journal of chemical information and computer sciences, 63(2)

Authors

Juarez, Reecan

Jiang, Yaoyukun

Tremblay, Matthew

et al.

Publication Date

2023-01-23

DOI

10.1021/acs.jcim.2c00945

Peer reviewed

HHS Public Access

Author manuscript

J Chem Inf Model. Author manuscript; available in PMC 2023 April 20.

Published in final edited form as:

J Chem Inf Model. 2023 January 23; 63(2): 522–530. doi:10.1021/acs.jcim.2c00945.

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

Reecan J. Juarez[5], **Yaoyukun Jiang**[1], **Matthew Tremblay**[1], **Qianzhen Shao**[1], **A. James Link**[7], **Zhongyue J. Yang**[1,2,3,4,6,*]

[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 Biomolecular Engineering, Vanderbilt University, Nashville, Tennessee 37235, United States

[7]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. With superior thermal stability, protease resistance, and antimicrobial activity, lasso peptides are promising candidates for bioengineering and pharmaceutical applications. To enable high-throughput computational prediction and design of lasso peptides, we developed a 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. Provided 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 for 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

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

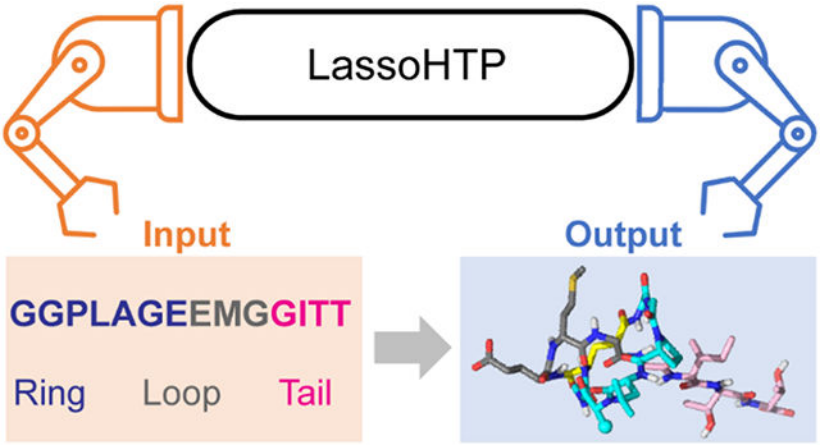
Supporting Information. Sample codes; scaffold structure models; force field parameters; steered MD simulation protocol; sample input files; RMSD values for tested lasso peptides and their substructures; NOE violation analyses; detail of dihedral PCA of lasso peptide conformational ensembles; scaffold comparison (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/.

The authors declare no competing financial interest.

compared these RMSD values against those of the MD ensembles that are initiated from the PDB structures. Dihedral principal component analysis was also conducted. The results show that the LassoHTP-initiated ensembles are similar to those of the PDB-initiated ensembles. LassoHTP offers a computational platform to develop strategies for lasso peptide prediction and design.

Graphical Abstract



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 potential candidates for drug development campaigns.[6, 7, 10-12] Lasso peptides are a natural example of [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] 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. We define the lasso peptide loop as the amino acid sequence between the macrolactam ring and the upper plug; and tail between the lower plug and the C-terminus (Scheme-1).

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 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 switch 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 structure as a starting point or a reference. Given the large amount of natural lasso peptide sequences whose three-dimensional structure is undetermined,[1, 2, 19] 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,[20] 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 rings,[21] but ranges from four (e.g., lariatin A and lariatin B)[22] to eighteen (e.g., ubonodin[4, 5]) for 8-membered rings, and from four (e.g., lihuanodin[23]) to seven (e.g., caulonodin-V[14]) for 9-membered rings. In addition, the tail size can range from two in microcin J25,[7] citrocin[3] or ubonodin[4] to eighteen in pandonodin[24]. Despite the structural 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-throughput 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 constructor 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.[25] Any single module can be independently executed to build, modify, or model a lasso peptide. The three modules can be sequentially operated in an automatic workflow to convert user-defined lasso peptide sequences to structural ensembles.

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 (Figure S1). The template lasso peptide scaffold can be mutated by the mutant generator and is compatible with software AMBER[26] used in the MD simulator module.

The structural library is a collection of 70 lasso peptide scaffolds with one scaffold per topology. Each scaffold is represented by a unique combination of ring size, loop size, and isopeptide linker type. Specifically, 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. The molecular geometry for each scaffold is shown in Figure S2.

The majority of the poly-alanine scaffolds (i.e., 56 out of 70) were constructed through steered MD[27] simulations using AMBER[26] (Figure 3 and detailed in the Computational Methods section). The rest of the scaffolds were truncated directly from known lasso peptide PDB structures (annotated in Figure S2). All scaffolds adopt a right-hand wrapping conformation.[28] The isopeptide bond of each scaffold adopts a *trans*-isopeptide bond configuration. To further develop LassoHTP, we will expand the scaffold library by incorporating scaffold with left-hand wrapping topology and with *cis*-isopeptide bond configuration.

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 the mutant generator can be an output of scaffold constructor (i.e., a poly-alanine scaffold) or a stand-alone lasso peptide structural model.

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 Text S1). Inspired by the framework of EnzyHTP,[25] 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 reparametrizes the mutated amino acids and the isopeptide bond (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 (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 (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, more advanced sampling techniques, such as umbrella sampling,[29] can also be conducted.

3. Computational Methods

3.1 Steered MD.

We employed steered MD to construct lasso peptide scaffolds in the library (Text S2). Each scaffold is built in two stages. The first stage is building an isopeptide ring-linear peptide thread complex (referred 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's PDB structure (PDB ID: 2M8F).[30] 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 benchmarked AM1-BCC and RESP charge models in the task of RMSD calculations using 10 ns MD trajectories (Table S2). The results show that the averaged RMSD value using AM1-BCC is 2.75 Å and that using RESP is 2.74 Å – the performance appears similar (Table S2). We adopted the AM1-BCC model in our simulations. Notably, this charge model was also used in the prior work[11] by Lai and Kaznessis[31] (Table S1). For the canonical amino acids in the peptide, ff14SB force field is employed.[32] 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*.[33] 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[34] octahedral solvent box with a 40 Å buffer. PMEMD.CUDA[35] 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^2). 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^2 (i.e., the main chain carboxylate on the glutamate or aspartate). 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 a lasso peptide scaffold by tLEaP. Specifically, 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.[35] Notably, the code (Class_Conf.py) also allows users to select sander.MPI as an MD engine. The LassoHTP workflow first generates the force field parameters using ff14SB force field, constructs an octahedral solvent box of size 10 Å and then automatically composes the input files (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 a 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] For both LHTP-initiated and PDB-initiated MD ensembles, the trajectories were calculated using the same types of force field parameters. The ensembles 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 (Text S3).

The average RMSD for the LHTP-initiated and PDB-initiated 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 moiety 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 expected because of its conformational rigidity. The close match in the loop RMSD values can be attributed to the use of caulosegnin-II's loop conformation in the LassoHTP scaffold library (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 Table S3). 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 initiate 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 (Table S4) are: benenodin-1 conformer 1,[17] benenodin-1 conformer 2,[17] citrocin,[3] the RGD variant (i.e., G12R, I13G, and G14D) of microcin J25,[12] streptomonomicin,[36] ubonodin,[4, 5] and xanthomonin-II[21]. 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 size (RS), loop size (LS), 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., RS: 8 aa; LS: 9 aa), MccJ25 RGD (i.e., RS: 8; LS: 11 aa), and ubonodin (i.e., RS: 8 aa; LS: 18 aa), or steered-MD simulations, including: benenodin-1 conformer 1 (i.e., RS: 8 aa; LS: 6 aa), benenodin-1 conformer 2 (i.e., RS: 8 aa; LS: 7 or 8 aa), streptomonomicin (i.e., RS: 9 aa; LS: 4, 5, or 6 aa), and xanthomonin-II (i.e., RS: 7 aa; LS: 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 (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.

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. They are also reasonably consistent with the experimentally characterized NMR ensembles based on the calculation of NOE violation (i.e., average RMSE from LHTP: 2.10 Å and NMR: 1.42 Å, Table S5). 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, Table S3).

To understand the impact of MD simulations on the initial scaffold, we compared the average RMSD for lasso peptide constructs before and after MD (Table S6). Across the 14 constructs, the average RMSD before and after MD is 3.08 Å and 2.70 Å, respectively, suggesting that MD simulations have a net effect to optimize the conformational ensemble derived from the initial scaffold. 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. The average ΔRMSD is 0.48 Å. To further elucidate the difference of LHTP- and NMR PDB-initiated ensembles, we performed principal component analysis (PCA) of peptide backbone dihedral angles derived from sampled conformers, in which the dihedral angle values are normalized to address the issue with circular statistics (Text S5). The backbone dihedral angle arrays of the conformers are projected onto the plane formed by the top two principal components, noted as PC1 and PC2 (Figure 7). All tested lasso peptides involve overlap of distributions of PC-projected points, which indicates the conformational similarity between the LHTP- and NMR-initiated ensembles. For ubonodin and microcin J25 (RGD mutant), the distributions of PC-projected points are very well aligned. For benenodin-1 conformer 2 and streptomonomicin, the distributions of PC-projected points are moderately aligned on the PC plane, which is consistent with their greater ΔRMSD values. 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. For xanthomonin-II, citrocin, and benenodin conformer 1, the distributions of PC-projected are moderately aligned albeit a small ΔRMSD value. By inspecting the structural superposition of the initial LHTP and NMR scaffolds, we observed that across the three scaffolds, the ring and loop are geometrically similar, but the tail, the most flexible moiety in lasso peptide, adopts

different conformation (Figure S5). This is caused by the linear peptide growth algorithm implemented in the tail extender of the scaffold constructor (Figure 1). As an immediate next step, we will implement other types of peptide growth algorithm that helps generate scaffolds with more diverse tail conformations.

For the next steps, we will further develop LassoHTP to overcome limitations in the current version. First, the *cis*-configuration of isopeptide bond, which was indicated in our previous computational modeling of benenodin-1,[15] was not considered in the scaffold library of LassoHTP. Incorporating the *cis*-isopeptide bond in the scaffold library is important for conformational sampling because the conversion from *trans*- to *cis*-isopeptide bond is difficult to observe within a limited MD sampling time scale. To address this, we will implement functions in LassoHTP to construct *cis*-configuration for the peptide or isopeptide bond. Second, the conformational sampling of lasso peptides should be further enhanced. Since the current work emphasizes the proof of concept of LassoHTP, we only applied 100 ns classical MD for sampling. This time scale might not sufficiently cover the conformational space for lasso peptides with a flexible loop or tail. As such, we will implement functions to employ accelerated[37] or replica-exchange[38] MD for conformational sampling. These simulation strategies have been shown to be essential for investigating constrained biomolecules (e.g., cyclic peptide[39]) and are likely very useful for studying lasso peptides with unknown structural scaffolds.

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 alanine residues 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 computational platform for high-throughput lasso peptide prediction and discovery.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGMENT

The authors thank Chris Jurich for providing technical support for producing Figure S2. The authors thank Xinchun Ran for performing principal component analysis. 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.[40] 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). M.T. acknowledges financial support from the Vanderbilt Undergraduate Summer Research Program and the Department of Chemistry.

References

(1). Montalbán-López M; Scott TA; Ramesh S; Rahman IR; van Heel AJ; Viel JH; Bandarian V; Dittmann E; Genilloud O; Goto Y; Grande Burgos MJ; Hill C; Kim S; Koehnke J; Latham JA; Link AJ; Martínez B; Nair SK; Nicolet Y; Rebuffat S; Sahl HG; Sareen D; Schmidt EW; Schmitt L; Severinov K; Süssmuth RD; Truman AW; Wang H; Weng JK; van Wezel GP; Zhang Q; Zhong J; Piel J; Mitchell DA; Kuipers OP; van der Donk WA New developments in RiPP discovery, enzymology and engineering. Nat. Prod. Rep 2021, 38 (1), 130–239, 10.1039/D0NP00027B. DOI: 10.1039/D0NP00027B. [PubMed: 32935693]

(2). Tietz JI; Schwalen CJ; Patel PS; Maxson T; Blair PM; Tai H-C; Zakai UI; Mitchell DA A new genome-mining tool redefines the lasso peptide biosynthetic landscape. Nat. Chem. Bio 2017, 13 (5), 470–478. DOI: 10.1038/nchembio.2319. [PubMed: 28244986]

(3). Cheung-Lee WL; Parry ME; Jaramillo Cartagena A; Darst SA; Link AJ Discovery and structure of the antimicrobial lasso peptide citrocin. J. Biol. Chem 2019, 294 (17), 6822–6830. DOI: 10.1074/jbc.RA118.006494. [PubMed: 30846564]

(4). Cheung-Lee WL; Parry ME; Zong C; Cartagena AJ; Darst SA; Connell ND; Russo R; Link AJ Discovery of ubonodin, an antimicrobial lasso peptide active against members of the Burkholderia cepacia complex. ChemBioChem 2020, 21 (9), 1335–1340. [PubMed: 31765515]

(5). Do T; Thokkadam A; Leach R; Link AJ Phenotype-Guided Comparative Genomics Identifies the Complete Transport Pathway of the Antimicrobial Lasso Peptide Ubonodin in Burkholderia. ACS Chem. Biol 2022. DOI: 10.1021/acschembio.2c00420.

(6). Ferguson AL; Zhang S; Dikiy I; Panagiotopoulos AZ; Debenedetti PG; James Link A An Experimental and Computational Investigation of Spontaneous Lasso Formation in Microcin J25. Biophys. J 2010, 99 (9), 3056–3065. DOI: 10.1016/j.bpj.2010.08.073. [PubMed: 21044604]

(7). Salomón RA; Farías RN Microcin 25, a novel antimicrobial peptide produced by Escherichia coli. J. Bacteriol 1992, 174 (22), 7428–7435. DOI: 10.1128/jb.174.22.7428-7435.1992 (acccessed 2022/07/12). [PubMed: 1429464]

(8). Constantine KL; Friedrichs MS; Detlefsen D; Nishio M; Tsunakawa M; Furumai T; Ohkuma H; Oki T; Hill S; Bruccoleri RE; Lin PF; Mueller L High-resolution solution structure of siamycin II: Novel amphipathic character of a 21-residue peptide that inhibits HIV fusion. J. Biomol. NMR 1995, 5 (3), 271–286. DOI: 10.1007/BF00211754. [PubMed: 7787424]

(9). Son S; Jang M; Lee B; Hong Y-S; Ko S-K; Jang J-H; Ahn JS Ulleungdin, a Lasso Peptide with Cancer Cell Migration Inhibitory Activity Discovered by the Genome Mining Approach. J. Nat. Prod. 2018, 81 (10), 2205–2211. DOI: 10.1021/acs.jnatprod.8b00449. [PubMed: 30251851]

(10). Adelman K; Yuzenkova J; La Porta A; Zenkin N; Lee J; Lis JT; Borukhov S; Wang MD; Severinov K Molecular Mechanism of Transcription Inhibition by Peptide Antibiotic Microcin J25. Mol. Cell 2004, 14 (6), 753–762. DOI: 10.1016/j.molcel.2004.05.017 (acccessed 2022/07/12). [PubMed: 15200953]

(11). Lai P-K; Kaznessis YN Free Energy Calculations of Microcin J25 Variants Binding to the FhuA Receptor. J. Chem. Theory Comput 2017, 13 (7), 3413–3423. DOI: 10.1021/acs.jctc.7b00417. [PubMed: 28622469]

(12). Knappe TA; Manzenrieder F; Mas-Moruno C; Linne U; Sasse F; Kessler H; Xie X; Marahiel MA Introducing lasso peptides as molecular scaffolds for drug design: engineering of an integrin antagonist. Angew. Chem. Int. Ed 2011, 50 (37), 8714–8717.

(13). Sluysmans D; Stoddart JF The Burgeoning of Mechanically Interlocked Molecules in Chemistry. Trends Chem. 2019, 1 (2), 185–197. DOI: 10.1016/j.trechm.2019.02.013.

(14). Hegemann JD; Fage CD; Zhu S; Harms K; Di Leva FS; Novellino E; Marinelli L; Marahiel MA The ring residue proline 8 is crucial for the thermal stability of the lasso peptide caulosegnin II. Mol. BioSyst 2016, 12 (4), 1106–1109, 10.1039/C6MB00081A. DOI: 10.1039/C6MB00081A. [PubMed: 26863937]

(15). Yang Z; Hajlasz N; Kulik HJ Computational Modeling of Conformer Stability in Benenodin-1, a Thermally Actuated Lasso Peptide Switch. J. Phys. Chem. B 2022, 126 (18), 3398–3406. DOI: 10.1021/acs.jpcb.2c00762. [PubMed: 35481742]

(16). Allen CD; Chen MY; Trick AY; Le DT; Ferguson AL; Link AJ Thermal Unthreading of the Lasso Peptides Astexin-2 and Astexin-3. ACS Chem. Biol 2016, 11 (11), 3043–3051. DOI: 10.1021/acschembio.6b00588. [PubMed: 27588549]

(17). Zong C; Wu MJ; Qin JZ; Link AJ Lasso Peptide Benenodin-1 Is a Thermally Actuated [1]Rotaxane Switch. J. Am. Chem. Soc 2017, 139 (30), 10403–10409. DOI: 10.1021/jacs.7b04830. [PubMed: 28696674]

(18). Maksimov MO; Pan SJ; James Link A Lasso peptides: structure, function, biosynthesis, and engineering. Nat. Prod. Rep 2012, 29 (9), 996–1006, 10.1039/C2NP20070H. DOI: 10.1039/C2NP20070H. [PubMed: 22833149] Hegemann JD; Zimmermann M; Xie X; Marahiel MA Lasso Peptides: An Intriguing Class of Bacterial Natural Products. Acc. Chem. Res. Acc. Chem. Res 2015, 48 (7), 1909–1919. DOI: 10.1021/acs.accounts.5b00156. [PubMed: 26079760]

(19). Cheung-Lee WL; Link AJ Genome mining for lasso peptides: past, present, and future. J. Ind. Microbiol. Biotechnol 2019, 46 (9-10), 1371–1379. DOI: 10.1007/s10295-019-02197-z (acccessed 7/12/2022). [PubMed: 31165971]

(20). Tunyasuvunakool K; Adler J; Wu Z; Green T; Zielinski M; Žídek A; Bridgland A; Cowie A; Meyer C; Laydon A; Velankar S; Kleywegt GJ; Bateman A; Evans R; Pritzel A; Figurnov M; Ronneberger O; Bates R; Kohl SAA; Potapenko A; Ballard AJ; Romera-Paredes B; Nikolov S; Jain R; Clancy E; Reiman D; Petersen S; Senior AW; Kavukcuoglu K; Birney E; Kohli P; Jumper J; Hassabis D Highly accurate protein structure prediction for the human proteome. Nature 2021. DOI: 10.1038/s41586-021-03828-1.

(21). Hegemann JD; Zimmermann M; Zhu S; Steuber H; Harms K; Xie X; Marahiel MA Xanthomonins I–III: A New Class of Lasso Peptides with a Seven-Residue Macrolactam Ring. Angew. Chem. Int. Ed 2014, 53 (8), 2230–2234.

(22). 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. J. Am. Chem. Soc 2006, 128 (23), 7486–7491. DOI: 10.1021/ja056780z. [PubMed: 16756302]

(23). Cao L; Beiser M; Koos JD; Orlova M; Elashal HE; Schröder HV; Link AJ Cellulonodin-2 and Lihuanodin: Lasso Peptides with an Aspartimide Post-Translational Modification. J. Am. Chem. Soc 2021, 143 (30), 11690–11702. DOI: 10.1021/jacs.1c05017. [PubMed: 34283601]

(24). Cheung-Lee WL; Cao L; Link AJ Pandonodin: A Proteobacterial Lasso Peptide with an Exceptionally Long C-Terminal Tail. ACS Chem. Biol 2019, 14 (12), 2783–2792. DOI: 10.1021/acschembio.9b00676. [PubMed: 31742991]

(25). Shao Q; Jiang Y; Yang ZJ EnzyHTP: A High-Throughput Computational Platform for Enzyme Modeling. J. Chem. Inf. Model 2022, 62 (3), 647–655. DOI: 10.1021/acs.jcim.1c01424. [PubMed: 35073075]

(26). AMBER 2017; University of California, San Francisco: 2017. (accessed.

(27). Isralewitz B; Gao M; Schulten K Steered molecular dynamics and mechanical functions of proteins. Curr. Opin. Struct. Biol 2001, 11 (2), 224–230. DOI: 10.1016/S0959-440X(00)00194-9. [PubMed: 11297932] Lu H; Schulten K Steered molecular dynamics simulations of force-induced protein domain unfolding. (0887-3585 (Print)). From 1999 Jun 1.

(28). Schröder HV; Zhang Y; Link AJ Dynamic covalent self-assembly of mechanically interlocked molecules solely made from peptides. Nat. Chem 2021, 13 (9), 850–857. DOI: 10.1038/s41557-021-00770-7. [PubMed: 34426684]

(29). Kästner J Umbrella sampling. Wiley Interdiscip. Rev.: Comput. Mol. Sci 2011, 1 (6), 932–942.

(30). Maksimov MO; Link AJ Discovery and Characterization of an Isopeptidase That Linearizes Lasso Peptides. J. Am. Chem. Soc 2013, 135 (32), 12038–12047. DOI: 10.1021/ja4054256. [PubMed: 23862624]

(31). Jakalian A; Jack DB; Bayly CI Fast, efficient generation of high-quality atomic charges. AM1-BCC model: II. Parameterization and validation. J. Comput.Chem 2002, 23 (16), 1623–1641. [PubMed: 12395429] Jakalian A; Bush BL; Jack DB; Bayly CI Fast, efficient generation of high-quality atomic charges. AM1-BCC model: I. Method. J. Comput. Chem 2000, 21 (2), 132–146.

(32). Maier JA; Martinez C; Kasavajhala K; Wickstrom L; Hauser KE; Simmerling C ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. J. Chem. Theory Comput 2015, 11 (8), 3696–3713. DOI: 10.1021/acs.jctc.5b00255. [PubMed: 26574453]

(33). AMBER-Umbrella COM restraint tutorial; GitHub, 2020. https://github.com/callumjd/AMBER-Umbrella_COM_restraint_tutorial/ (accessed.

(34). Mark P; Nilsson L Structure and Dynamics of the TIP3P, SPC, and SPC/E Water Models at 298 K. J. Phys. Chem. A 2001, 105 (43), 9954–9960. DOI: 10.1021/jp003020w.

(35). Salomon-Ferrer R; Götz AW; Poole D; Le Grand S; Walker RC Routine Microsecond Molecular Dynamics Simulations with AMBER on GPUs. 2. Explicit Solvent Particle Mesh Ewald. J. Chem. Theory Comput 2013, 9 (9), 3878–3888. DOI: 10.1021/ct400314y. [PubMed: 26592383] Götz AW; Williamson MJ; Xu D; Poole D; Le Grand S; Walker RC Routine Microsecond Molecular Dynamics Simulations with AMBER on GPUs. 1. Generalized Born. J. Chem. Theory Comput 2012, 8 (5), 1542–1555. DOI: 10.1021/ct200909j. [PubMed: 22582031]

(36). 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. Chem. Biol 2015, 22 (2), 241–250. DOI: 10.1016/j.chembiol.2014.11.017. [PubMed: 25601074]

(37). Hamelberg D; Mongan J; McCammon JA Accelerated molecular dynamics: A promising and efficient simulation method for biomolecules. J. Chem. Phys 2004, 120 (24), 11919–11929. DOI: 10.1063/1.1755656 (acccessed 2022/07/12). [PubMed: 15268227]

(38). Sugita Y; Okamoto Y Replica-exchange molecular dynamics method for protein folding. Chem. Phys. Lett 1999, 314 (1), 141–151. DOI: 10.1016/S0009-2614(99)01123-9.

(39). McHugh SM; Rogers JR; Yu H; Lin Y-S Insights into How Cyclic Peptides Switch Conformations. J. Chem. Theory Comput 2016, 12 (5), 2480–2488. DOI: 10.1021/acs.jctc.6b00193. [PubMed: 27031286] Miao J; Descoteaux ML; Lin Y-S Structure prediction of cyclic peptides by molecular dynamics + machine learning. Chem. Sci 2021, 12 (44), 14927–14936, 10.1039/D1SC05562C. DOI: 10.1039/D1SC05562C. [PubMed: 34820109] Damjanovic J; Miao J; Huang H; Lin Y-S Elucidating Solution Structures of Cyclic Peptides Using Molecular Dynamics Simulations. Chem. Rev 2021, 121 (4), 2292–2324. DOI: 10.1021/acs.chemrev.0c01087. [PubMed: 33426882]

(40). Towns J; Cockerill T; Dahan M; Foster I; Gaither K; Grimshaw A; Hazlewood V; Lathrop S; Lifka D; Peterson GD; Roskies R; Scott JR; Wilkins-Diehr N XSEDE: Accelerating Scientific Discovery. Comput. Sci. Eng 2014, 16 (5), 62–74. DOI: 10.1109/MCSE.2014.80.

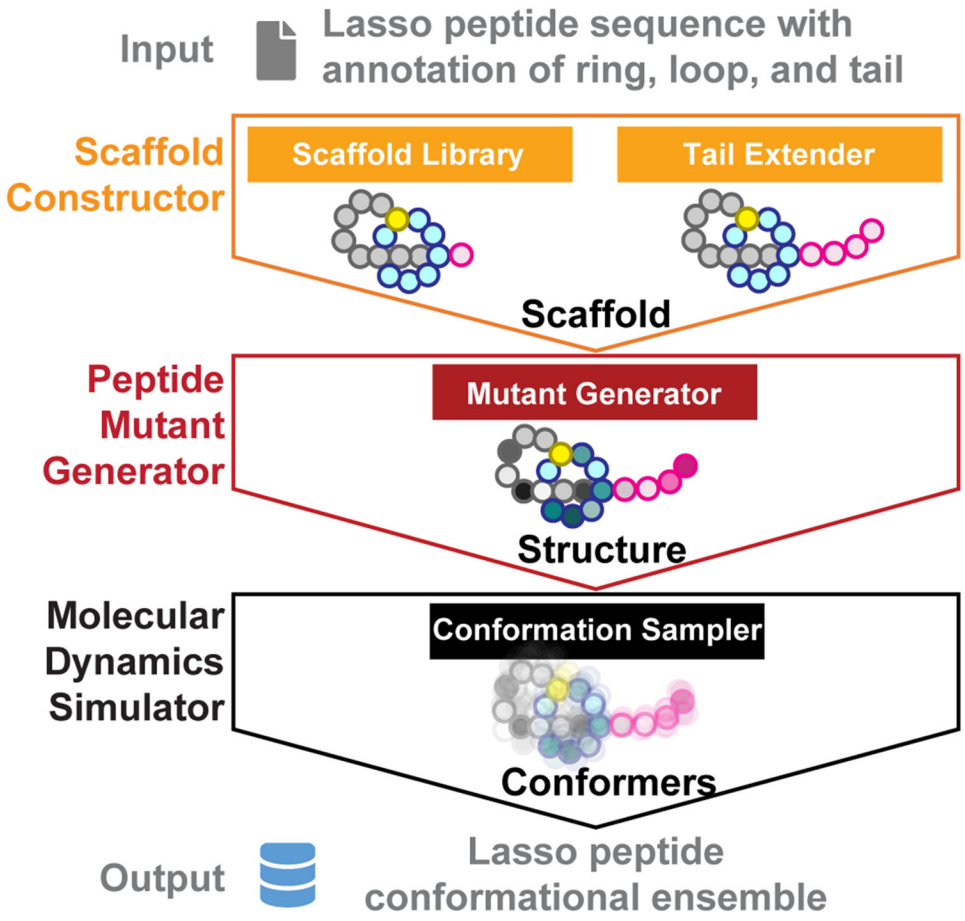


Figure 1.

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 a 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.

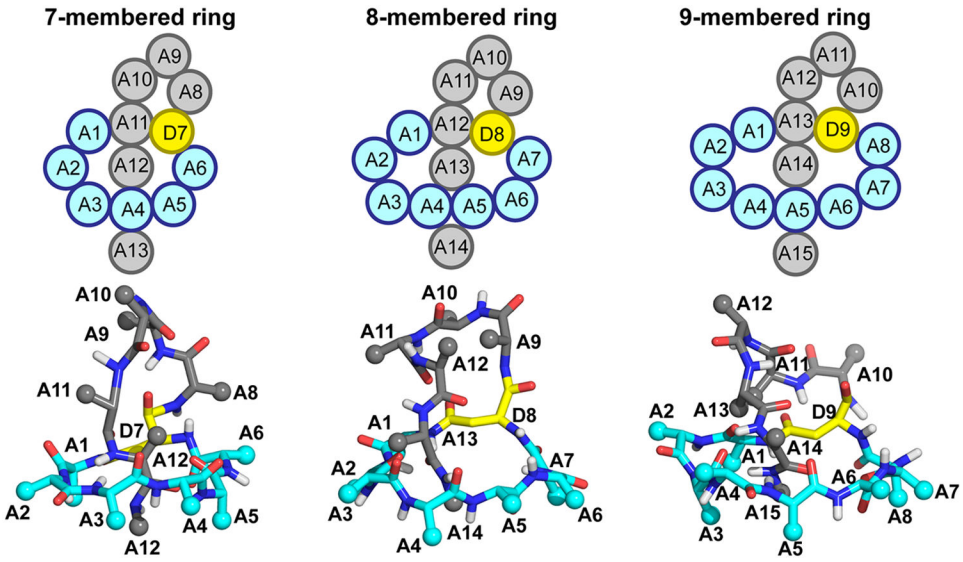


Figure 2.

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.

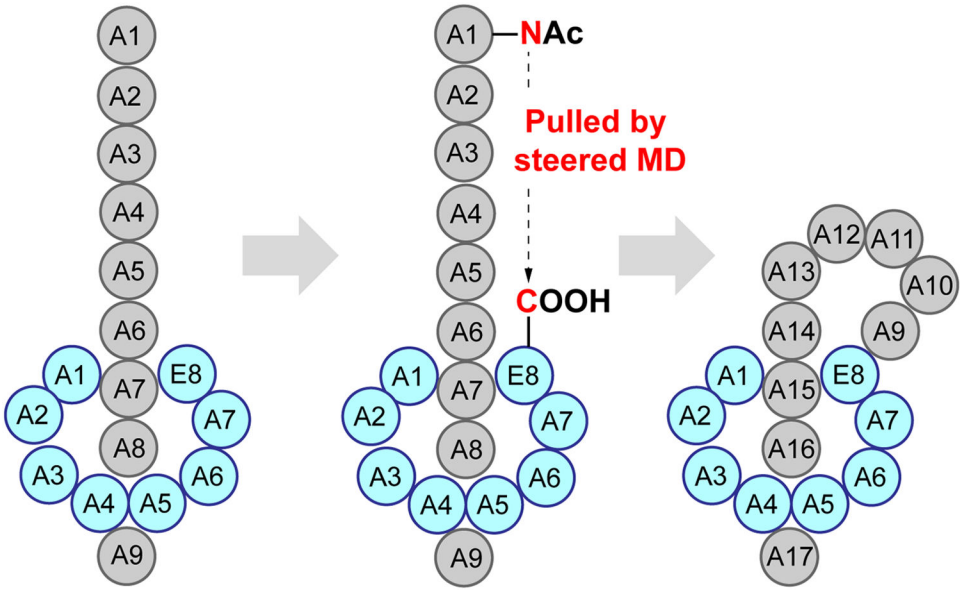


Figure 3.

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.

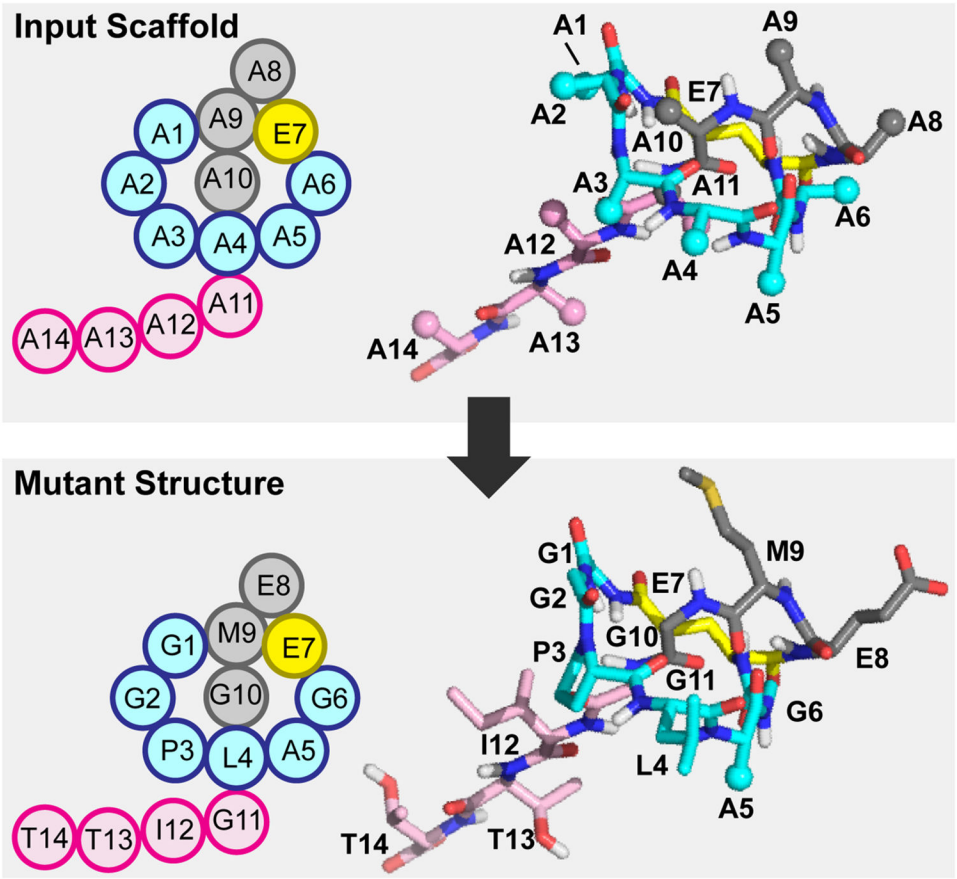


Figure 4.

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[21] (PDB ID: 2MFV).

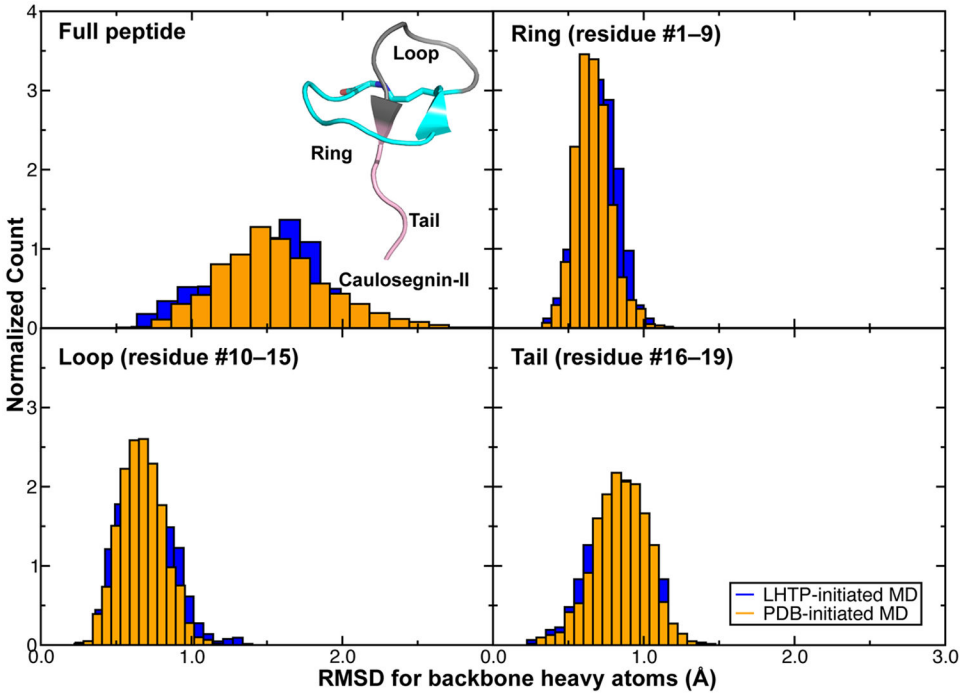


Figure 5.

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.

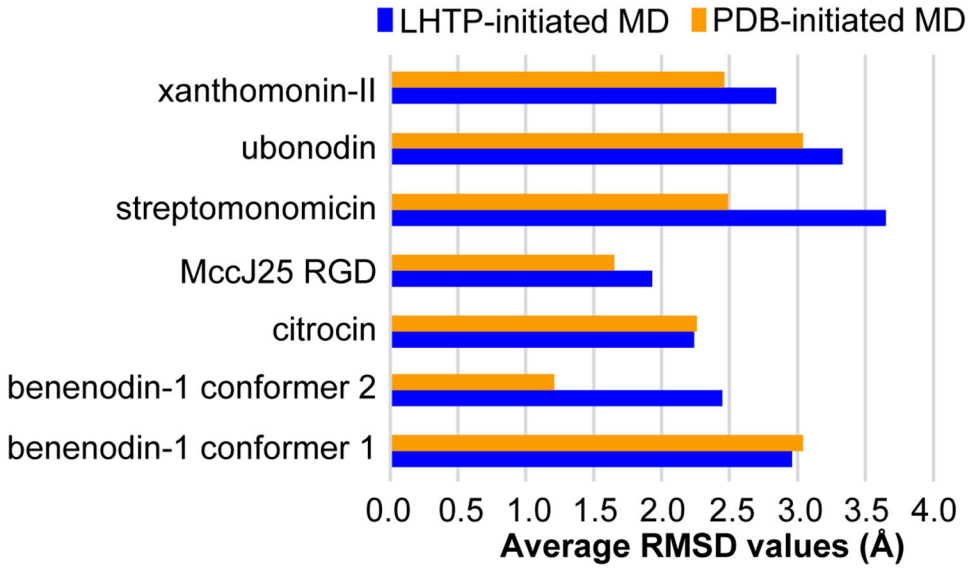


Figure 6.

Average 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 conformer 2, the average RMSD values for the LHTP-initiated MD ensembles account for multiple LassoHTP constructs with various loop and tail sizes.

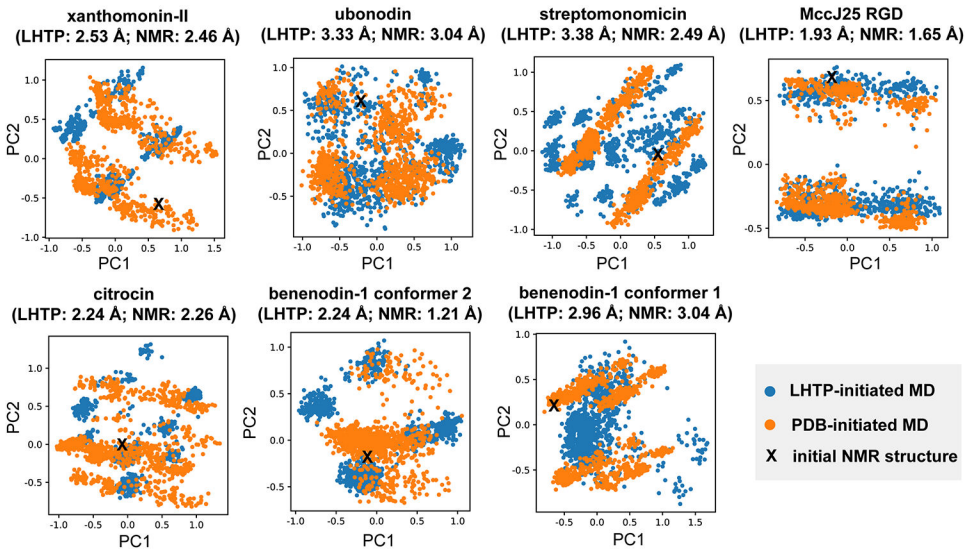
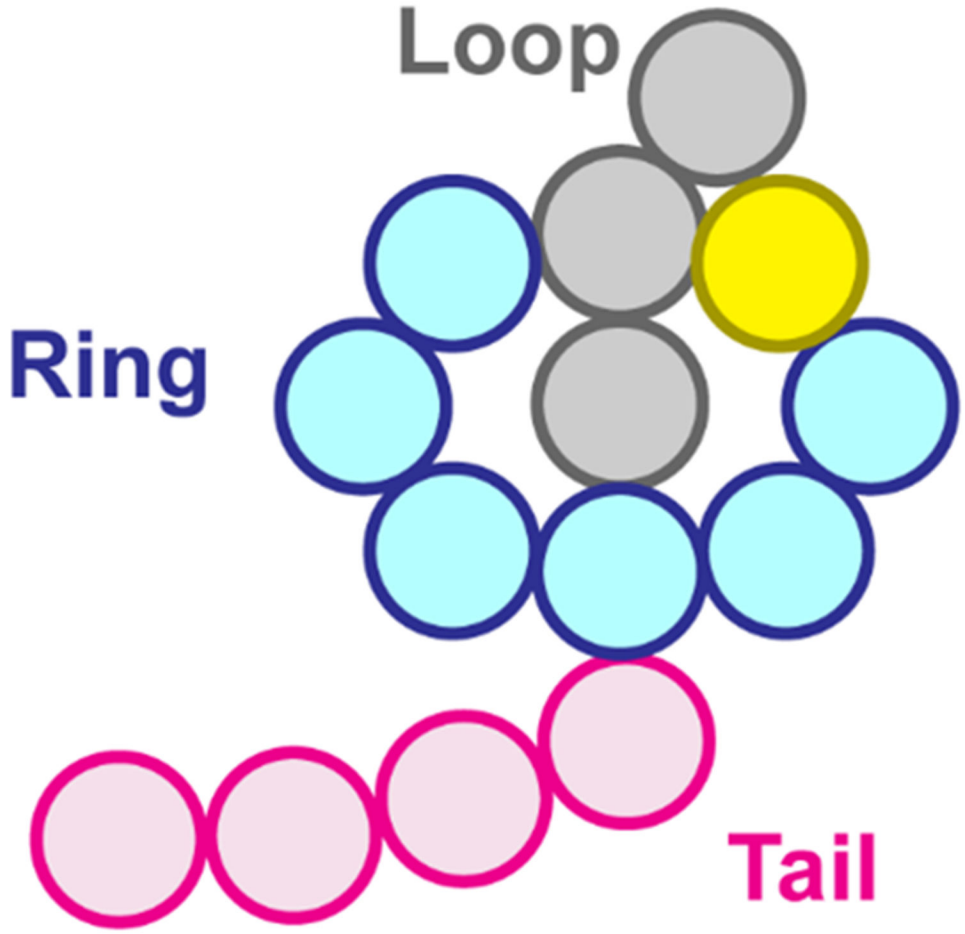


Figure 7.

Dihedral principal component analysis of lasso peptide conformational ensembles initiated from LHTP-constructed scaffold (blue) and NMR PDB (orange). PC1 and PC2, the top two principal components with greatest variance values, are used to project the dihedral angle array of each conformer on a two-dimensional plane. The NMR PDB structure that was applied to initiate the MD ensemble is annotated on the PC plane with the label “X”. The values in the parentheses are average RMSD values derived from the top performing structural construct of each lasso peptide.



Scheme-1.
Conformational annotation for lasso peptide.