Computational Modeling of Conformer Stability in Benenodin-1, A Thermally-Actuated Lasso Peptide Switch

Zhongyue Yang[1], Natalia Hajlasz[1], and Heather J. Kulik[1,*]

[1]*Department of Chemical Engineering, Massachusetts Institute of Technology, Cambridge, MA 02139*

ABSTRACT: Benenodin-1 is a thermally-actuated lasso peptide rotaxane switch with two primary translational isomers that differ in the relative position of residue Gln15. The conversion from one conformer to the other involves substantial enthalpy-entropy compensation: one conformer is energetically-favored and the other is entropically-favored. Here, we take a multi-scale quantum mechanical (QM) and classical molecular dynamics (MD) approach to reveal residue-specific sources of these differences in stability. QM reveals that the two benenodin-1 conformers involve distinct hydrogen bonding networks, with the enthalpically-favored conformer having more intra-peptide hydrogen bonds between Gln15 sidechain and nearby residues. Evaluation of configurational entropy over the MD-sampled geometries reveals that the entropically-favored conformer has enhanced conformational flexibility. By computing the by-residue-sum entropies, we identify the role of Gln15 and neighboring Glu14 in mediating the entropic variation during the switching process. These computational insights help explain the effects of Glu14Ala and Gln15Ala mutations on the conformational population of benenodin-1 observed experimentally.

1. Introduction.

The macromolecular structure of proteins imparts unique properties that allow them to play diverse roles in the cell. While most proteins are large and globular in nature, consisting of α-helical and β-sheet domains connected by loops, there are notable exceptions. Small peptides can take on more diverse structures, e.g., as intrinsically disordered proteins[1-3] and cyclic peptides[4-6]. In comparison to their large, globular counterparts, these peptides remain poorly understood, motivating further experimental and computational investigation. Exemplary of this class of peptides, are the lasso peptides, which are a type of ribosomally-derived natural products.[7-9] Ever since their first discovery in 1991,[10] lasso peptides have been increasingly reported as candidates for new antibiotics,[11-14] enzyme inhibitors,[12,15] and receptor antagonists,[10] (e.g., microcin J25[15,16]). Lasso peptides involve a 1-rotaxane topology with a macrolactam ring threaded by the C-terminal tail.[8,17] The ring is held in position by sterically bulky residues above and below the ring, which act as plugs. The ring in the lasso peptide is connected by an isopeptide bond between the N-terminal α-amino group and the carboxylic acid side chain of an aspartate or glutamate, which can be located at positions 7, 8, or 9 of the amino acid sequence depending on the specific Lasso peptide. Lasso peptides exhibit unique properties, including protease stability, thermostability, and thermal switchability.[17]

Computational modeling can provide valuable insight into the dynamics[18-24] and non-covalent interactions[25-28] of proteins that govern structure-function relationships. While molecular modeling with classical molecular mechanics (MM) force fields is often the method of choice for protein modeling[29,30], challenges remain for their application in non-globular, small peptides[31,32]. Thanks to key advances in quantum mechanical (QM) modeling methodology[33-39] combined with the modest size of small peptides has made peptides increasingly tractable for

fully QM study[33,40-42]. While the levels of approximate density functional theory (DFT) with modest basis sets typically used in QM modeling of peptides also can suffer from inaccuracies due to self-interaction error[43] or basis set incompleteness[33,42,44,45], QM modeling has been shown to be superior for de novo prediction of properties of peptides with non-folded structures[33]. At the same time, the sampling required for understanding conformational ensembles is typically prohibitive, motivating a multi-scale approach[46-49].
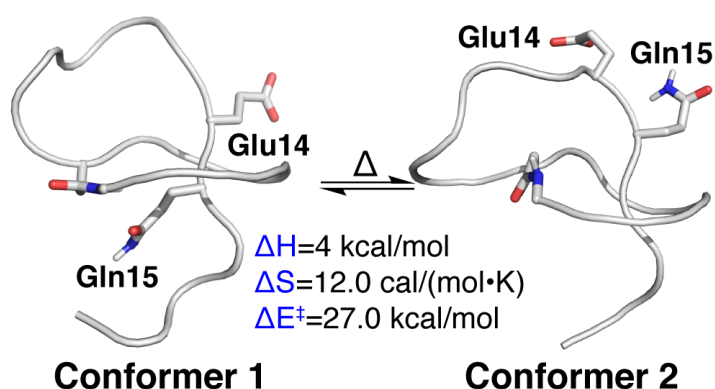
In this work, we introduce a multi-scale approach aimed at bridging potential limitations of MM and QM modeling on benenodin-1, a representative lasso peptide. Derived from *A. benevestitus*, lasso peptide benenodin-1 was discovered as a natural example of a switchable mechanically interlocked molecule.[50,51] In benenodin-1, the residues Glu14 and Gln15 act as steric plugs that bracket the lasso ring. This contrasts with most other lasso peptides which have bulkier aromatic or large positively charged amino acids that act as steric lock residues. Upon heating, benenodin-1 ΔC5 (i.e., the natural product with the last five residues removed) has been observed to achieve equilibria among four translational isomers. The two primary isomers in the ensemble are: Conformer 1, where the lasso peptidic wheel is between Glu14 and Gln15, and Conformer 2, where the lasso peptidic wheel is between Glu15 and Lys17.[51] Although the structural ensembles of these conformers have been characterized using NMR, mass spectroscopy, and HPLC experiments, computational simulations are needed to elucidate the ways the electronic structure and dynamics of individual amino acids that contribute to the overall thermal switchability of benenodin-1.

While this study represents the first detailed computational study of benenodin-1 with large-scale QM, other studies have been carried out on the broader family of lasso peptides. Molecular mechanics has been employed to elucidate the molecular details of lasso peptide

formation and un-threading. Ferguson et. al performed replica exchange MD[52] to study the wrapping mechanism of uncyclized analogue of microcin J-25 (i.e., pro-microcin J25).[53] The studies reveal dynamic pathways of pro-microcin J25 to generate either a loop-tail peptide or a coil-conformation that can be potentially converted to lasso topology. The study implicated important roles of enzyme machinery to wrap the uncyclized peptide to form a lasso geometry. Caitlin et. al conducted biased MD simulations to understand the thermal unthreading mechanism (i.e., loop-pulling or tail-pulling) of lasso peptide astexin-2 and astexin-3.[54] They computed the free energy barriers associated with different amino acids passing through a macrolactam ring. Applying these modeling data, they established a quantitative model of the strength of amino acid steric plugs to estimate the thermostability of a lasso peptide, albeit with a significant energetic uncertainty around $8k_BT$ (i.e., 5 kcal/mol at room temperature).

In experiments, benenodin-1 was expressed in a truncated form that lacks 5 C-terminal amino acids (benenodin-1 ΔC5).[50,51] Based upon characterization by NMR, MALDI-MS and HPLC, benenodin-1 ΔC5 can switch among four geometric isomers, with two primary isomers being Conformer 1 and Conformer 2 (Figure 1). One major difference between the two conformers lies in the position of the Gln15 residue. Gln15 locates beneath the lasso ring in Conformer 1 but flips above the ring in Conformer 2. Consequently, Conformer 1 has 6 amino acids (i.e., aa) in the threaded loop (i.e., from Ser9 to Glu14) and 5 aa in the tail (i.e., from Gln15 to Met19), while Conformer 2 has a loop of 8 aa (i.e., from Ser9 to Ala16) and a tail of only 3 aa (i.e., from Lys17 to Met19). For benenodin-1 ΔC5, the conversion from Conformer 1 to Conformer 2 was characterized to have a 27 kcal/mol activation energy barrier.[50] Although the switching process does not involve direct bond rearrangement, the activation energy observed for this process is comparable to (if not higher than) typical chemical reactions that involve bond

rearrangement. This highlights the strength of the "mechanical bond" that exists in rotaxane or other mechanically interlocked molecules.[55,56] It has been proposed that the activation barrier for the lasso peptide switching or threading depends on the size of the "steric plug" residues (e.g., Glu14 and Gln15 in Conformer 1).[54] This model is chemically intuitive and is consistent with the observation that lasso peptides with bulkier steric plugs (such as phenylalanine) do not easily unthread.[57] Nevertheless, benenodin-1 plug residues are less bulky than other well studied lasso peptides, motivating investigation of their roles in hindering rearrangement.

Figure 1. Thermally switchable conformers for benenodin-1 ΔC5, Conformer 1 and Conformer 2, and their associated, experimentally measured thermodynamic parameters, including enthalpy (ΔH), entropy (ΔS), and activation energy (ΔE‡). The lasso peptide backbone is colored in grey. The sidechains of Glu14 and Gln15 and the isopeptide bond (between Gly1 and Asp8) are shown in sticks, where carbon, oxygen, nitrogen atoms are colored in grey, red, and blue.

Despite the high activation barrier, Conformer 1 and Conformer 2 were observed to co-exist in experiment and the composition in the structural ensemble varies in different temperatures (i.e., Conformer 1 to Conformer 2 is 2:1 at 286 K and 1:2 at 368 K).[50] Experimental kinetic and thermodynamic studies indicate that the conversion from Conformer 1 to Conformer 2 is energetically uphill by 4 kcal/mol but entropically favored by 12 cal/mol/K. The mechanism behind this thermodynamic trend is not easily comprehensible due to the structural and dynamic complexity of lasso peptide. Furthermore, the impact of mutation on the

conformational population appears puzzling. For example, mutation of Gln15 to alanine, which presumably eases the up-threading of lasso tail to generate Conformer 2, in fact leads to exclusive observation of Conformer 1 in the structural ensemble. However, mutation of Glu14 to alanine leads to exclusive observation of Conformer 2. These phenomena imply the important roles played by Glu14 and Gln15 during the switching process: besides serving as a steric plug to hold the peptide chain in place to conform a lasso shape, these two residues might also facilitate the stabilization of the energy state of Conformer 1 and the enhancement the conformational flexibility of Conformer 2. Therefore, we combined GPU-accelerated quantum mechanics (QM)[34,58,59] and classical molecular dynamics (MD) simulations to unravel the atomic-level mechanism behind the lasso peptide switch.

Notably, for peptides such as benenodin-1 that interconvert near room temperature, a quantitative description of this thermal switching is particularly difficult because the free energy difference between the two conformational states are likely within 3 kcal/mol.[51] In addition, thermal switchability involves a subtle balance between the change of enthalpy and entropy. Evaluating these two parameters separately is essential to elucidate the driving force behind the conformational switching behavior. Thus, we carry out a combined study that employs QM in the form of dispersion-corrected, range-separated hybrid DFT along with MM sampling of the thermal switching of benenodin-1. The study reveals quantitatively how rearrangement of hydrogen bonding networks and changes of residue dynamics contribute to the formation of the two primary benenodin-1 translational isomers, Conformer 1 and Conformer 2. These results are consistent with both experimental observations and hypotheses.[50,51] We expect that the computational studies can potentially facilitate the rational design and engineering of lasso peptide mutants as thermally responsive biomolecules.

2. Computational Details.

Lasso Peptide Structure and Preparation. Solution NMR structures for the lasso peptide benenodin-1 were obtained from the Protein Data Bank (PDB IDs: 5TJ1 and 6B5W), and all crystallizing agents were removed.[50] These structures correspond to two distinct states (i.e., Conformer 1 PDB ID: 5TJ1 and Conformer 2 PDB ID: 6B5W) that have been resolved experimentally and differ by the position of the lasso ring.[50] For each conformer, 20 structures were available from the NMR ensemble deposited in the PDB.[50] Residue protonation states were assigned using the H++ webserver[60-62] assuming a pH of 7.0 and a dielectric constant of 78.4 to mimic aqueous solvent with all other defaults applied (Supporting Information Table S1). The resulting protonated structure consists of 282 atoms with zero net charge.

Structures were prepared with the AMBER[63] tleap utility for classical molecular dynamics (MD) with the AMBER ff14SB force field[64] for the peptide. Benenodin-1 contains an isopeptide bond between Gly1–NH_3^+ and Asp8–COO^-, which was parameterized using AMBER99 force field parameters[65] (Supporting Information Table S2). For each of the 20 benenodin-1 NMR conformers in the Conformer 1 and Conformer 2 states, we constructed the model by solvating the conformer in a periodic rectangular box with at least a 12.0 Å buffer of TIP4P-Ew[30,66,67] water (Supporting Information Tables S3-S4). The TIP4P-Ew water model was chosen for its good performance in NMR structure prediction in prior studies.[30] The AMBER MD input files (i.e. topology and coordinate files) for the models are provided in the Supporting Information .zip file.

Classical MD Simulations. The peptide models were equilibrated with MD using the GPU-accelerated pmemd form of AMBER.[68,69] The equilibration steps were as follows: i) 2000 minimization steps, ii) 10-ps NVT heating to 300 K with a Langevin thermostat with collision

frequency of 1.0 ps^{-1} using a random seed, and iii) 10-ps NpT equilibration using the Berendsen barostat with a pressure relaxation time of 2 ps. These steps were followed by a 100 ns NpT production run. Throughout the equilibration and production runs, the peptide backbone atoms (i.e., Cα, C, O, and N) were restrained using harmonic potentials with 20 kcal/mol$^{.}$Å^2 force constants. The positional restraints were employed here to ensure the backbone in each NMR conformer was preserved while sidechain motions were sampled (Supporting Information Figure S1). The restraining potential on the backbone is also used to ensure the high percentage of the *cis*-configurations (55% of NMR configurations) of the isopeptide bond in benenodin-1 in comparison to typical peptides (<0.1%)[70] (Supporting Information Figure S2). The SHAKE algorithm[71] was applied to all bonds with hydrogen atoms in combination with a 2-fs timestep for all MD simulations. For the long-range electrostatics, the particle mesh Ewald method was used with a 10-Å real space electrostatic cutoff. For each of the Conformer 1 and Conformer 2 states, the production runs result in a total of 2.0 µs (i.e., 20 models x 100 ns/model) trajectories, from which we removed all water molecules. We then extracted snapshots every 100 ps for a total of 200,000 peptide structures for the conformational entropy calculations. We also extracted 100 peptide conformers for QM geometry optimizations spaced 20 ns apart (see next).

Conformational Entropy Calculations. The PARENT program[72] was employed to compute the conformational entropy. With the protein structural ensemble as input, the program first transforms the peptide Cartesian coordinates to internal bond, angle, and torsion angle coordinates. With a default setting of 50 bins, a histogram-based approach was then employed to estimate the probability density function for each degree of freedom (DOF) and the joint-probability density function between DOFs. These probability density functions are employed to compute the one-dimensional entropy terms and the higher-order entropy terms using mutual
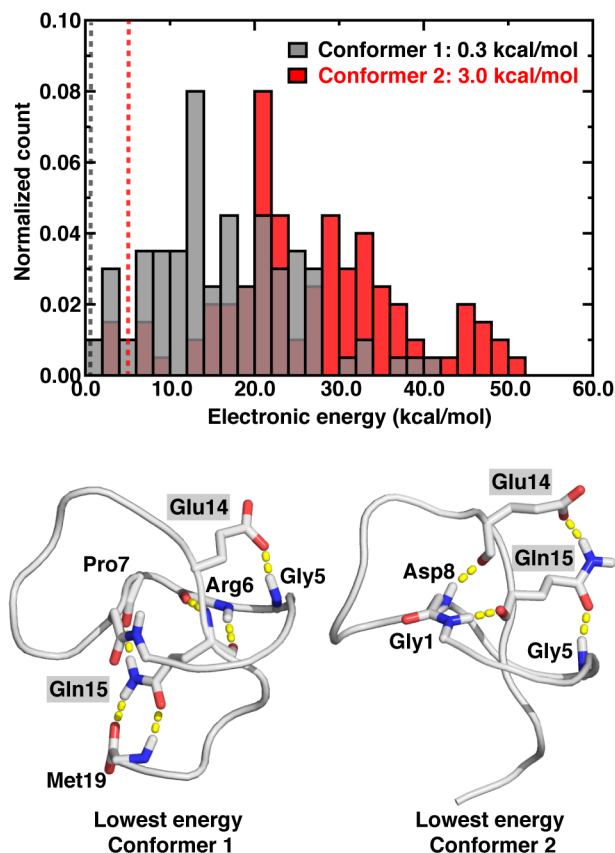
information spanning tree approximations[73]. To ensure statistical convergence, we benchmarked the dependence of total conformational entropy on structural ensemble sizes for Conformer 1 and Conformer 2 (i.e., the number of conformers) and determined the analysis was converged with an ensemble size of 100,000 structures (Supporting Information Figure S3).

Quantum Mechanical Modeling. We performed geometry optimizations and single point energy calculations corrections on 100 conformers for each of the Conformer 1 and Conformer 2 states with range-separated hybrid density functional theory (DFT). All DFT calculations were carried out using a developer version of the GPU-accelerated code TeraChem[34,58,59]. The long-range-corrected, range-separated ωPBEh[74] (ω=0.2 bohr^{-1}) exchange-correlation functional was employed in conjunction with the 6-31G[75] basis set for geometry optimizations. Additionally, empirical dispersion (i.e., D3) corrections were included[76,77] with Becke-Johnson damping. The full 282-atom benenodin-1 peptide was modeled as a closed shell singlet with a net charge of 0, and the solvent environment was modeled implicitly. The C-PCM implicit solvent model[78,79] implemented in TeraChem[40,80] was employed with ε=78.4 and default 1.2x Bondi radii[81] to build the cavity. Basis set and functional sensitivities were found to be modest in ranking relative conformer energies (Supporting Information Figures S4-S6). On each of these conformers, single point energies were obtained with ωPBEh and B3LYP functionals in conjunction with a larger, polarized 6-31G(d) and 6-311G(d,p) basis set. Geometry optimizations were carried out in translation-rotation internal coordinates (TRIC)[82] with L-BFGS. The convergence criteria were loosened from their defaults by an order of magnitude due to the large size of the model, with a maximum for the gradient of 4.5×10^{-3} hartree/bohr and maximum change in energy between steps of 1.0×10^{-5} hartree. All final QM-optimized geometries and single point energies are provided in the Supporting Information .zip file.

3. Results and Discussion.

3a. Loss of Hydrogen Bonding Interactions Elevates the Energy State of Conformer 2.

We first performed QM calculations to study the electronic energy distribution of the two conformational states (shown in gray and red for Conformer 1 and Conformer 2, respectively, top, Figure 2). We conducted geometry optimizations with range separated hybrid DFT along with a modest double-ζ basis set and refined these energies with single points using polarization functions (see Computational Details). Benchmarks with other methods and basis sets yielded comparable results and are reported in the Supporting Information (Supporting Information Figures S4-S5). Calculations were carried out on representative geometries of Conformer 1 and Conformer 2 obtained from MD simulation (Figure 2). The distribution in energies for structures sampled in each state corresponds roughly to a right-skewed Gaussian distribution with a longer tail extending to higher energy conformational states. Despite having a wide range of energy distribution, Conformer 1 is statistically more favored than Conformer 2 by 2.7 kcal/mol using Boltzmann averaging. We adopted Boltzmann averaging due to its relevance to characterize energy state of canonical ensemble in thermodynamic equilibrium.[83,84] The computed energy gap of 2.7 kcal/mol is consistent with the experimental trend for ΔH (i.e., 4 kcal/mol). The agreement is particularly good considering that we neglected vibrational zero-point energy corrections due to the high cost of computing Hessians for such large systems. We also neglected the PΔV term, which is known to contribute a trivial part to chemical processes in solution under ambient conditions. The current model also employs an implicit solvent model to describe the aqueous environment in the QM energy calculation, and explicit treatment of water molecules in future work could also lead to shifts in hydrogen bonding between residues (see next).

Figure 2. (Top) Distribution of electronic energies for 100 optimized geometries of Conformer 1 (grey) and Conformer 2 (red). The geometries were sampled by using NMR-guided MD simulations. Each geometry was optimized using ωPBEh/6-31G method followed by single point energy corrections using ωPBEh-D3/6-311g(d,p). Both calculations were conducted in implicit water model. The relative energy values of the geometries were computed with reference to the energy of the lowest-energy geometry of Conformer 1. The gray and red dashed line represent the averaged energy values for Conformer 1 and Conformer 2 using Boltzmann averaging. (Bottom) The lowest-energy geometries in the Conformer 1 and Conformer 2 ensemble. The lasso peptide backbone is colored in gray. The sidechains of residues (i.e., Gly5, Arg6, Pro7, Glu14, Gln15, and Met19) and the isopeptide bond (between Gly1 and Asp8) are shown in sticks, where carbon, oxygen, nitrogen atoms are colored in grey, red, and blue. The hydrogen bonds between Glu14/Gln15 and nearby residues are colored in yellow dashes.

To understand why Conformer 1 is more stable than Conformer 2, we studied the lowest-energy geometries for each conformational state (Figure 2). For these two cases, there are notable differences in the hydrogen bonds (H-bonds) formed between Glu14/Gln15 and nearby

residues (Figure 2). Changes in the orientation of these two residues induce significant rearrangement of hydrogen bonding network during the conversion from Conformer 1 to Conformer 2. [50,51] Specifically, Gln15 forms H-bonds to the charged C-terminus of Met19 and polar Pro7 as well as positively charged Arg6 in Conformer 1 (Figure 2). In conformer 2, the hydrogen bonding pattern of Gln15 is different (Figure 2). The backbone of Gln15 forms hydrogen bonds to the backbone amide nitrogen of nonpolar Gly1, its sidechain oxygen acts as a hydrogen bond to the backbone amide nitrogen of Gly5, and the sidechain nitrogen acts as a hydrogen bond donor to Glu14 in Conformer 2 (Figure 2). Although the number of H-bonding residues remains the same for Gln15, this process involves a replacement of two charge-assisted, double-hydrogen bonds[27,85] in Conformer 1 (i.e., Gln15-Met19 and Gln15-Arg6) by two single-hydrogen bonds, leading to a reduction of the total number of hydrogen bonds. At the same time, Glu14 shifts its H-bonding partner from nonpolar Gly5 in Conformer 1 to both a sidechain-sidechain hydrogen bond polar Gln15 and a backbone-backbone hydrogen bond to the amide nitrogen of Asp8 in Conformer 2. Overall, the conversion from Conformer 1 to Conformer 2 involves a net loss of one intra-peptide hydrogen bond, which contributes to destabilize Conformer 2. Overall, the relatively similar number of hydrogen bonds is consistent with the only slight (ca. 2-3 kcal/mol) destabilization of Conformer 2 over Conformer 1.

3b. Configurational Entropy Analysis Reveals the Important Roles of Glu14 and Gln15.

Next, we investigated the change of configurational entropy[72,73,86] during the conformational conversion. The total configurational entropy of a molecule can be expressed as:

$$S(x_1, \ldots, x_m) = -R \int J(x_1, \ldots, x_m) \, \rho(x_1, \ldots, x_m) ln\rho(x_1, \ldots, x_m) dx_1 \ldots dx_m \qquad (1)$$

where m is the total number of degrees of freedom (i.e., $3N$–6 for nonlinear molecules with N being the number of atoms). In eqn. (1), J represents the Jacobian determinant associated with the coordinate variables x_i (i = 1, …, m), where x_i are the internal (i.e., bond, angle, torsion) coordinates. Internal coordinates are preferred to avoid spurious correlations that may be present the Cartesian coordinate representation. Using the second-order mutual information spanning tree (MIST) method,[73] the full dimensional configurational entropy can be truncated to:
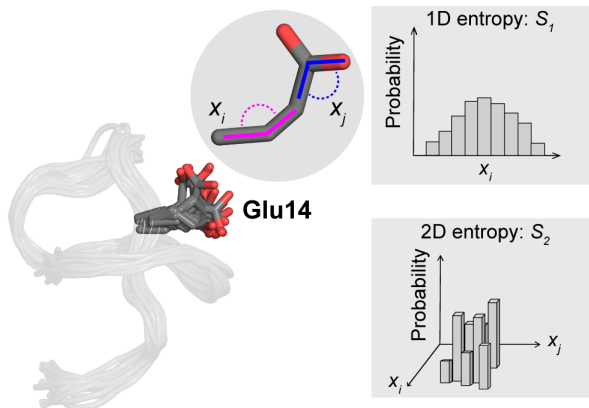
$$S_m^{MIST_2}(x_1, \dots, x_m) = \sum_{i=1}^{m}[S_1(x_i) - \underset{j \in \{1, \dots, i-1\}}{max} I_2(x_i; x_j)] \tag{2}$$

where S_1 represents the one-dimensional (1D) entropy calculated using the marginal probability density function of x_i. The I_2 term represents the mutual information between two coordinate variables x_i and x_j, which can be expressed as:

$$I_2(x_i; x_j) \equiv S_1(x_i) + S_1(x_j) - S_2(x_i, x_j) \tag{3}$$

where S_2 represents the entropy calculated using the two-dimensional (2D) joint probability density function of x_i and x_j.

By combining eqns. (2) and (3), we can calculate the configurational entropy value of a conformational ensemble using the 1D- and 2D-entropy terms derived from the probability density functions of the coordinate variables x_i and x_j. The distribution of the 1D and 2D probability density functions can be estimated using a histogram-based approach (see Sec. 2). An example of the outcomes of this analysis can be observed for how representative angle variables are coupled for the Glu14 residue in ensemble of Conformer 1 (Figure 3).
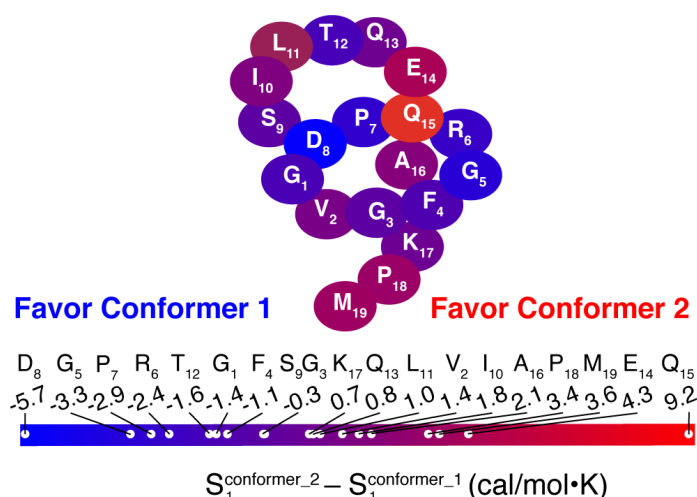
Figure 3. Schematic for computing the conformational entropy. $S_1(x_i)$ represent the one-dimensional entropy for the variable x_i. The coordinate variables, x_i and x_j, are shown for the example of two adjacent angles of Glu14. $S_2(x_i, x_j)$ represents the two-dimensional entropy for the joint probability distribution of x_i and x_j. The conformational ensemble of Conformer 1 is represented by an overlap of 10 geometries at left. The lasso peptide backbone is colored in gray. The sidechain of Glu14 is shown in sticks, where the carbon and oxygen atoms are colored in grey and red, respectively.

We extracted geometries from 2.0 μs of MD trajectories to compute the converged configurational entropy values for Conformer 1 and Conformer 2 (Supporting Information Figure S3). From this analysis, we conclude that the entropy of Conformer 2 is higher than that of Conformer 1 by 14.4 cal/(mol·K). This difference is consistent with the experimentally measured ΔS value between these two conformers (i.e., 12.0 cal/(mol·K)). Using eqn. (2), we decomposed the 14.4 cal/(mol·K) entropy difference to a sum of the 1D entropy (i.e., 9.2 cal/(mol·K)) and negative 2D mutual information (i.e., 5.2 cal/(mol·K)). This indicates that the ensemble of Conformer 2 exhibits a broader conformational distribution (i.e., 1D entropy is larger) and more de-coupled intra-peptide residue motion (i.e., the 2D mutual information is smaller).

To understand the origin of entropic favorability of Conformer 2 over Conformer 1, we computed the variation of by-residue-sum 1D entropy values and negative 2D mutual

information values for each benenodin-1 ΔC5 residue. We sorted the entropic variation values of all residues and ranked them based on their favorability in Conformer 1 or Conformer 2. For the 1D entropy change from Conformer 1 to 2, the entropy values for 9 residues exhibit negative changes, indicating the shrinkage of their conformational space, whereas 10 residues have positive changes and thus expansion of their conformational variation (Figure 4). Thus, the changes of residue entropy values are compensating. This phenomenon may be expected because the enhanced flexibility of a residue may restrict the motion of other spatially proximal residues.[87] Gln15 has the most positive change of 1D entropy (i.e., 9.2 cal/(mol·K)) from Conformer 1 to Conformer 2, which is more than twice the entropy change of the following residue on the rank (i.e., Glu14 with 4.3 cal/(mol·K)). This indicates that Gln15's conformational space is significantly enlarged in Conformer 2. Two reasons may contribute to this phenomenon. First, unlike the Conformer 1 in which Gln15 and Glu14 are constrained to bracket the lasso ring, Conformer 2 leaves sufficiently large space for Gln15 and Glu14's side chain to adopt a variety of vibrational modes and states. The lengthening of the threaded loop (i.e., 8 aa in Conformer 2, compared to 6 aa in Conformer 1) also mobilizes the backbone atoms of Gln15 and Glu14, which, combined with their more flexible side chain, further enhances the conformational space of Gln15 and Glu14.
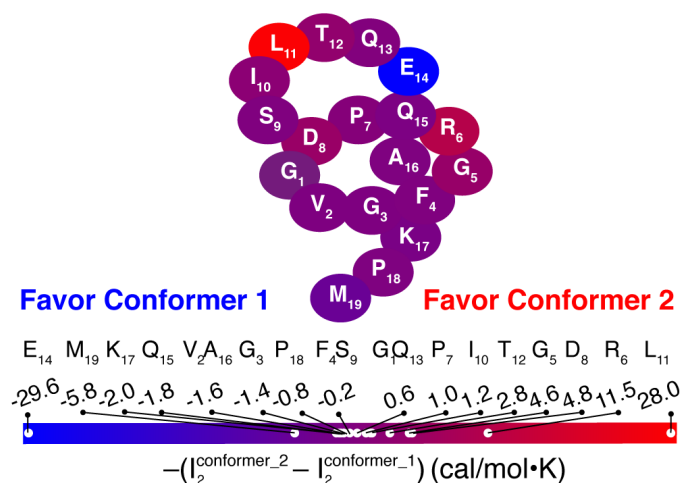
Favor Conformer 1   **Favor Conformer 2**

D$_8$ G$_5$ P$_7$ R$_6$ T$_{12}$ G$_1$ F$_4$ S$_9$G$_3$ K$_{17}$Q$_{13}$ L$_{11}$ V$_2$ I$_{10}$ A$_{16}$ P$_{18}$M$_{19}$ E$_{14}$ Q$_{15}$
-5.7 -3.3 -2.9 -2.4 -1.6 -1.4 -1.1 -0.3 0.7 0.8 1.0 1.4 1.8 2.1 3.4 3.6 4.3 9.2

$$S_1^{conformer\_2} - S_1^{conformer\_1} \text{ (cal/mol·K)}$$

Figure 4. Variation of by-residue-sum 1D entropy values for each benenodin-1 residue upon converting from Conformer 1 to Conformer 2. All residues are sorted and populated on a continuous gradient bar based on their entropic variation, in which a negative or positive value indicates their favorability towards Conformer 1 (blue) or 2 (red), respectively.

Consistent with the hypothesis that compensation can occur due to changes in the lasso ring, the residues on the lasso tail from Ala16 to Met19 have slightly increased 1D entropy, albeit with a lower value than Gln14 or Glu15, whereas the residues on the lasso ring from Gly1 to Asp8 involve a decreased 1D entropy value (Figure 4). The loss of charge-assisted, double-hydrogen bonds between the Gln15 sidechain and Met19 also allows Gln15 to adopt more rotamers. Notably, Met19 has the third most-enhanced entropy values by rank from Conformer 1 to 2 (i.e., 3.6 cal/(mol·K)), which is likely a consequence of the C-terminal carboxylate not forming any strong hydrogen bonds, leading to enhanced backbone and side-chain motion. These differences in Conformer 1 and 2 depend on the size and polar nature of Gln15's side chain. Thus, mutating Gln15 to alanine should reduce the favorable shift in entropy of Conformer 2 because of a reduced bulkiness and abolished hydrogen-bonding capability of the side chain, helping to explain the exclusive observation of Conformer 1 in the Gln15Ala mutant in experiment.[50]

As in the case of 1D mutual information, the changes of residue 2D mutual information values also appear to involve compensation. From Conformer 1 to Conformer 2, the mutual information becomes larger for 10 residues and smaller for 9 residues (Figure 5). An increased mutual information value indicates an enhanced correlated motion among residues. The same group of residues that involved an enhanced 1D entropy (i.e., from Glu14 to Met19) also exhibit enhanced mutual information that could be a consequence of enhanced collective motion of threading and stretching. A large portion of residues on the lasso ring that have diminished 1D entropy also show decreased mutual information (e.g., from Gly5 to Asp8, now colored in red, Figure 5). Conversely, Glu14 exhibits the most significant increase in mutual information (Figure 5). As the magnitude of mutual information changes proportionally with the size of the residue, replacing Glu14 with alanine to reduce the size would reduce the magnitude of coupling and eliminate the enhanced mutual information by a significant amount. As such, compared to the wild-type benenodin-1, the Glu14Ala is expected to have an even more favorable the entropic increase in Conformer 2. This computational analysis thus explains the exclusive observation of Conformer 2 in the Glu14Ala mutant.[50]

Figure 5. Variation of by-residue-sum two-dimensional mutual information values for each benenodin-1 residue upon converting from Conformer 1 to Conformer 2. The negative mutual

information values are adopted here to ensure a consistent sign notation to that used in Figure 4. All residues are sorted and populated on a continuous gradient bar based on their mutual information variation, in which a negative or positive value indicates their favorability towards Conformer 1 (blue) or 2 (red), respectively.

4. Conclusions.

The protein benenodin-1 is a naturally occurring mechanically-interlocked lasso peptide rotaxane switch that can exist in either Conformer 1 or Conformer 2. Experimentally, the conversion from Conformer 1 to Conformer 2 is enthalpically uphill but entropically favorable, but the origin of these differences had been unknown. Through a combination of QM and MM MD calculations, we aimed to decompose by-residue contributions to these differences in the conformer. Based on the lowest energy geometries computed by QM simulations, we showed that Conformer 1 and Conformer 2 involve different hydrogen bonding interactions as a result of the differences in placement of Gln15. Compared to Conformer 1, Conformer 2 involves a net loss of hydrogen bonds, which contributes to the diminished (i.e., by around 3 kcal/mol) stability.

Based on the ensemble of geometries sampled using MD, we computed the configurational entropies for Conformer 1 and Conformer 2 within the second-order mutual information spanning tree approximation. In contrast to Conformer 1, Conformer 2 involves a substantially enhanced conformational flexibility (i.e., enhanced 1D entropy values) and reduced intra-peptide correlated motions (i.e., reduced 2D mutual information values). From the by-residue-sum entropies, Gln15 involves the largest increase in 1D entropy values, indicating its significantly increased conformational flexibility in Conformer 2. In addition, Glu14 had the largest increase in 2D mutual information values from Conformer 1 to 2, indicating its significantly enhanced correlated motions with the rest of the residues in the peptide. Mutation of

Glu14 to Ala could thus be expected to favor Conformer 2, whereas Gln15 mutation to Ala could be expected to favor Ala, consistent with experimental observations. Further calculations on the Glu14Ala and Gln15Ala mutants are needed to elucidate how other weak interactions and protein dynamics collectively affect the equilibrium distribution. Nonetheless, we expect that these quantitative studies inform a strategy to evaluate and even predict the molecular switching behaviors of lasso peptide, which can be used to rationally design and engineer other thermally-responsive biomolecules.

ASSOCIATED CONTENT

Supporting Information. Protonation state of amino acids in Benenodin-1; Parameters for the isopeptide bond in Benenodin-1; System size and box dimension for the solvated Conformer 1 conformers; System size and box dimension for the solvated Conformer 2; Structural ensemble for typical Conformer 1 and Conformer 2 conformers; Benchmark of conformational entropy against structural ensemble sizes; Benchmark of electronic energy against method and basis set for Conformer 1; Benchmark of electronic energy against method and basis set for Conformer 2; Benchmark of electronic energy for lower-level method for Conformer 1 and Conformer 2; Distribution of configurations for the isopeptides of benenodin-1 (PDF)

QM-optimized xyz structures; AMBER prmtop and inpcrd files for MD simulations (ZIP)

This material is available free of charge via the Internet at http://pubs.acs.org.

AUTHOR INFORMATION

Corresponding Author

*email:hjkulik@mit.edu

Notes

The authors declare no competing financial interest.

REFERENCES

(1)   Tompa, P. Intrinsically Disordered Proteins: A 10-Year Recap. *Trends in biochemical sciences* **2012,** *37*, 509-516.
(2)   Oldfield, C. J.; Dunker, A. K. Intrinsically Disordered Proteins and Intrinsically Disordered Protein Regions. *Annual review of biochemistry* **2014,** *83*, 553-584.
(3)   Wright, P. E.; Dyson, H. J. Intrinsically Disordered Proteins in Cellular Signalling and Regulation. *Nature reviews Molecular cell biology* **2015,** *16*, 18-29.
(4)   Kessler, H. Conformation and Biological Activity of Cyclic Peptides. *Angewandte Chemie International Edition in English* **1982,** *21*, 512-523.
(5)   Tapeinou, A.; Matsoukas, M. T.; Simal, C.; Tselios, T. Review Cyclic Peptides on a Merry-Go-Round; Towards Drug Design. *Peptide Science* **2015,** *104*, 453-461.
(6)   Elashal, H. E.; Cohen, R. D.; Elashal, H. E.; Zong, C.; Link, A. J.; Raj, M. Cyclic and Lasso Peptides: Sequence Determination, Topology Analysis, and Rotaxane Formation. *Angewandte Chemie* **2018,** *130*, 6258-6262.
(7)   Arnison, P. G.; Bibb, M. J.; Bierbaum, G.; Bowers, A. A.; Bugni, T. S.; Bulaj, G.; Camarero, J. A.; Campopiano, D. J.; Challis, G. L.; Clardy, J. Ribosomally Synthesized and Post-Translationally Modified Peptide Natural Products: Overview and Recommendations for a Universal Nomenclature. *Natural product reports* **2013,** *30*, 108-160.
(8)   Do, T.; Link, A. J. Protein Engineering in Ribosomally Synthesized and Post-Translationally Modified Peptides (Ripps). *Biochemistry* **2022,** DOI:10.1021/acs.biochem.1c00714 10.1021/acs.biochem.1c00714.
(9)   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.et al. New Developments in Ripp Discovery, Enzymology and Engineering. *Natural Product Reports* **2021,** *38*, 130-239.

(10)   Weber, W.; Fischli, W.; Hochuli, E.; Kupfer, E.; Weibel, E. K. Anantin-a Peptide Antagonist of the Atrial Natriuretic Factor (Anf). *The Journal of antibiotics* **1991,** *44*, 164-171.

(11)   Salomon, R. A.; Farías, R. N. Microcin 25, a Novel Antimicrobial Peptide Produced by Escherichia Coli. *Journal of bacteriology* **1992,** *174*, 7428-7435.

(12)   Constantine, K. L.; Friedrichs, M. S.; Detlefsen, D.; Nishio, M.; Tsunakawa, M.; Furumai, T.; Ohkuma, H.; Oki, T.; Hill, S.; Bruccoleri, R. E. 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.

(13)   Tsunakawa, M.; Hu, S.-L.; Hoshino, Y.; Detlefson, D. J.; Hill, S. E.; Furumai, T.; White, R. J.; Nishio, M.; Kawano, K.; Yamamoto, S. Siamycins I and Ii, New Anti-Hiv Peptides: I. Fermentation, Isolation, Biological Activity and Initial Characterization. *The Journal of antibiotics* **1995,** *48*, 433-434.

(14)   Pan, S. J.; Cheung, W. L.; Fung, H. K.; Floudas, C. A.; Link, A. J. Computational Design of the Lasso Peptide Antibiotic Microcin J25. *Protein Engineering, Design and Selection* **2011,** *24*, 275-282.

(15)   Braffman, N. R.; Piscotta, F. J.; Hauver, J.; Campbell, E. A.; Link, A. J.; Darst, S. A. Structural Mechanism of Transcription Inhibition by Lasso Peptides Microcin J25 and Capistruin. *Proceedings of the National Academy of Sciences* **2019,** *116*, 1273.

(16)   Solbiati, J. O.; Ciaccio, M.; Farías, R. N.; González-Pastor, J. E.; Moreno, F.; Salomón, R. A. Sequence Analysis of the Four Plasmid Genes Required to Produce the Circular Peptide Antibiotic Microcin J25. *Journal of bacteriology* **1999,** *181*, 2659-2662.

(17)   Cheung-Lee, W. L.; Link, A. J. Genome Mining for Lasso Peptides: Past, Present, and Future. *Journal of industrial microbiology & biotechnology* **2019**, 1-9.

(18)   Blaha-Nelson, D.; Krüger, D. M.; Szeler, K.; Ben-David, M.; Kamerlin, S. C. L. Active Site Hydrophobicity and the Convergent Evolution of Paraoxonase Activity in Structurally Divergent Enzymes: The Case of Serum Paraoxonase 1. *Journal of the American Chemical Society* **2017,** *139*, 1155-1167.

(19)   Crean, R. M.; Gardner, J. M.; Kamerlin, S. C. L. Harnessing Conformational Plasticity to Generate Designer Enzymes. *Journal of the American Chemical Society* **2020**.

(20)   Lau, E. Y.; Bruice, T. C. Importance of Correlated Motions in Forming Highly Reactive near Attack Conformations in Catechol O-Methyltransferase. *Journal of the American Chemical Society* **1998,** *120*, 12387-12394.

(21)   Horowitz, S.; Dirk, L. M. A.; Yesselman, J. D.; Nimtz, J. S.; Adhikari, U.; Mehl, R. A.; Scheiner, S.; Houtz, R. L.; Al-Hashimi, H. M.; Trievel, R. C. Conservation and Functional Importance of Carbon–Oxygen Hydrogen Bonding in Adomet-Dependent Methyltransferases. *Journal of the American Chemical Society* **2013,** *135*, 15536-15548.

(22)   Phatak, P.; Sumner, I.; Iyengar, S. S. Gauging the Flexibility of the Active Site in Soybean Lipoxygenase-1 (Slo-1) through an Atom-Centered Density Matrix Propagation (Admp) Treatment That Facilitates the Sampling of Rare Events. *The Journal of Physical Chemistry B* **2012,** *116*, 10145-10164.

(23)   Lu, X.; Ovchinnikov, V.; Demapan, D.; Roston, D.; Cui, Q. Regulation and Plasticity of Catalysis in Enzymes: Insights from Analysis of Mechanochemical Coupling in Myosin. *Biochemistry* **2017,** *56*, 1482-1497.

(24)   Patra, N.; Ioannidis, E. I.; Kulik, H. J. Computational Investigation of the Interplay of Substrate Positioning and Reactivity in Catechol O-Methyltransferase. *PLoS ONE* **2016, *11***, e0161868.

(25)   Parrish, R. M.; Thompson, K. C.; Martínez, T. J. Large-Scale Functional Group Symmetry-Adapted Perturbation Theory on Graphical Processing Units. *Journal of chemical theory and computation* **2018, *14***, 1737-1753.

(26)   Wang, L.; Fried, S. D.; Boxer, S. G.; Markland, T. E. Quantum Delocalization of Protons in the Hydrogen-Bond Network of an Enzyme Active Site. *Proc. Natl. Acad. Sci. U. S. A.* **2014, *111***, 18454-18459.

(27)   Qi, H. W.; Kulik, H. J. Evaluating Unexpectedly Short Non-Covalent Distances in X-Ray Crystal Structures of Proteins with Electronic Structure Analysis. *Journal of Chemical Information and Modeling* **2019, *59***, 2199-2211.

(28)   Vennelakanti, V.; Qi, H. W.; Mehmood, R.; Kulik, H. J. When Are Two Hydrogen Bonds Better Than One? Accurate First-Principles Models Explain the Balance of Hydrogen Bond Donors and Acceptors Found in Proteins. *Chem. Sci.* **2021**, *Accepted manuscript*.

(29)   Riniker, S. Fixed-Charge Atomistic Force Fields for Molecular Dynamics Simulations in the Condensed Phase: An Overview. *Journal of chemical information and modeling* **2018, *58***, 565-578.

(30)   Beauchamp, K. A.; Lin, Y.-S.; Das, R.; Pande, V. S. Are Protein Force Fields Getting Better? A Systematic Benchmark on 524 Diverse Nmr Measurements. *Journal of chemical theory and computation* **2012, *8***, 1409-1414.

(31)   Kulik, H. J.; Luehr, N.; Ufimtsev, I. S.; Martinez, T. J. Ab Initio Quantum Chemistry for Protein Structure. *The Journal of Physical Chemistry B* **2012, *116***, 12501-12509.

(32)   Rauscher, S.; Gapsys, V.; Gajda, M. J.; Zweckstetter, M.; de Groot, B. L.; Grubmüller, H. Structural Ensembles of Intrinsically Disordered Proteins Depend Strongly on Force Field: A Comparison to Experiment. *Journal of chemical theory and computation* **2015, *11***, 5513-5524.

(33)   Kulik, H. J.; Luehr, N.; Ufimtsev, I. S.; Martinez, T. J. Ab Initio Quantum Chemistry for Protein Structures. *The Journal of Physical Chemistry B* **2012, *116***, 12501-12509.

(34)   Ufimtsev, I. S.; Martínez, T. J. Quantum Chemistry on Graphical Processing Units. 3. Analytical Energy Gradients, Geometry Optimization, and First Principles Molecular Dynamics. *Journal Of Chemical Theory And Computation* **2009, *5***, 2619-2628.

(35)   Isborn, C. M.; Luehr, N.; Ufimtsev, I. S.; Martinez, T. J. Excited-State Electronic Structure with Configuration Interaction Singles and Tamm-Dancoff Time-Dependent Density Functional Theory on Graphical Processing Units. *Journal of Chemical Theory and Computation* **2011, *7***, 1814-1823.

(36)   Ufimtsev, I. S.; Luehr, N.; Martínez, T. J. Charge Transfer and Polarization in Solvated Proteins from Ab Initio Molecular Dynamics. *The Journal of Physical Chemistry Letters* **2011, *2***, 1789-1793.

(37)   Ochsenfeld, C.; Kussmann, J.; Lambrecht, D. S. Linear-Scaling Methods in Quantum Chemistry. *Reviews in computational chemistry* **2007, *23***, 1.

(38)   Eichkorn, K.; Weigend, F.; Treutler, O.; Ahlrichs, R. Auxiliary Basis Sets for Main Row Atoms and Transition Metals and Their Use to Approximate Coulomb Potentials. *Theoretical Chemistry Accounts* **1997, *97***, 119-124.

(39)   Eichkorn, K.; Treutler, O.; Öhm, H.; Häser, M.; Ahlrichs, R. Auxiliary Basis Sets to Approximate Coulomb Potentials. *Chemical physics letters* **1995, *240***, 283-290.

(40)   Liu, F.; Luehr, N.; Kulik, H. J.; Martínez, T. J. Quantum Chemistry for Solvated Molecules on Graphical Processing Units Using Polarizable Continuum Models. *Journal of chemical theory and computation* **2015,** *11*, 3131-3144.

(41)   Schmitz, S.; Seibert, J.; Ostermeir, K.; Hansen, A.; Göller, A. H.; Grimme, S. Quantum Chemical Calculation of Molecular and Periodic Peptide and Protein Structures. *The Journal of Physical Chemistry B* **2020,** *124*, 3636-3646.

(42)   Sure, R.; Grimme, S. Corrected Small Basis Set Hartree-Fock Method for Large Systems. *Journal of computational chemistry* **2013,** *34*, 1672-1685.

(43)   Becke, A. D. Perspective: Fifty Years of Density-Functional Theory in Chemical Physics. *The Journal of chemical physics* **2014,** *140*, 18A301.

(44)   Kulik, H. J.; Seelam, N.; Mar, B.; Martinez, T. J. Adapting Dft+U for the Chemically-Motivated Correction of Minimal Basis Set Incompleteness. *The Journal of Physical Chemistry A* **2016,** *120*, 5939-5949.

(45)   Otero-de-la-Roza, A.; DiLabio, G. A. Improved Basis-Set Incompleteness Potentials for Accurate Density-Functional Theory Calculations in Large Systems. *Journal of Chemical Theory and Computation* **2020,** *16*, 4176-4191.

(46)   Vennelakanti, V.; Nazemi, A.; Mehmood, R.; Steeves, A. H.; Kulik, H. J. Harder, Better, Faster, Stronger: Large-Scale Qm and Qm/Mm for Predictive Modeling in Enzymes and Proteins. *Current opinion in structural biology* **2022,** *72*, 9-17.

(47)   Cui, Q.; Pal, T.; Xie, L. Biomolecular Qm/Mm Simulations: What Are Some of the "Burning Issues"? *J. Phys. Chem. B* **2021,** *125*, 689-702.

(48)   Lin, H.; Truhlar, D. Qm/Mm: What Have We Learned, Where Are We, and Where Do We Go from Here? *Theor. Chem. Acc.* **2007,** *117*, 185-199.

(49)   Senn, H. M.; Thiel, W. Qm/Mm Methods for Biomolecular Systems. *Angew. Chem., Int. Ed.* **2009,** *48*, 1198-1229.

(50)   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.

(51)   Schröder, H. V.; Stadlmeier, M.; Wühr, M.; Link, A. J. The Shuttling Cascade in Lasso Peptide Benenodin-1 Is Controlled by Non-Covalent Interactions. *Chemistry–A European Journal* **2021**.

(52)   Sugita, Y.; Okamoto, Y. Replica-Exchange Molecular Dynamics Method for Protein Folding. *Chemical Physics Letters* **1999,** *314*, 141-151.

(53)   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.

(54)   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.

(55)   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.

(56)   Stoddart, J. F. Mechanically Interlocked Molecules (Mims)—Molecular Shuttles, Switches, and Machines (Nobel Lecture). *Angewandte Chemie International Edition* **2017,** *56*, 11094-11125.

(57) Zimmermann, M.; Hegemann, J. D.; Xie, X.; Marahiel, M. A. The Astexin-1 Lasso Peptides: Biosynthesis, Stability, and Structural Studies. *Chemistry & biology* **2013,** *20,* 558-569.

(58) Ufimtsev, I. S.; Martínez, T. J. Quantum Chemistry on Graphical Processing Units. 1. Strategies for Two-Electron Integral Evaluation. *Journal Of Chemical Theory And Computation* **2008,** *4,* 222-231.

(59) Ufimtsev, I. S.; Martinez, T. J. Quantum Chemistry on Graphical Processing Units. 2. Direct Self-Consistent-Field Implementation. *Journal of Chemical Theory and Computation* **2009,** *5,* 1004-1015.

(60) Anandakrishnan, R.; Aguilar, B.; Onufriev, A. V. H++ 3.0: Automating Pk Prediction and the Preparation of Biomolecular Structures for Atomistic Molecular Modeling and Simulations. *Nucleic Acids Research* **2012,** *40,* W537-W541.

(61) Gordon, J. C.; Myers, J. B.; Folta, T.; Shoja, V.; Heath, L. S.; Onufriev, A. H++: A Server for Estimating Pkas and Adding Missing Hydrogens to Macromolecules. *Nucleic Acids Research* **2005,** *33,* W368-W371.

(62) Myers, J.; Grothaus, G.; Narayanan, S.; Onufriev, A. A Simple Clustering Algorithm Can Be Accurate Enough for Use in Calculations of Pks in Macromolecules. *Proteins: Structure, Function, and Bioinformatics* **2006,** *63,* 928-938.

(63) D.A. Case, J. T. B., R.M. Betz, D.S. Cerutti, T.E. Cheatham, III, T.A. Darden, R.E. Duke, T.J. Giese, H. Gohlke, A.W. Goetz, N. Homeyer, S. Izadi, P. Janowski, J. Kaus, A. Kovalenko, T.S. Lee, S. LeGrand, P. Li, T. Luchko, R. Luo, B. Madej, K.M. Merz, G. Monard, P. Needham, H. Nguyen, H.T. Nguyen, I. Omelyan, A. Onufriev, D.R. Roe, A. Roitberg, R. Salomon-Ferrer, C.L. Simmerling, W. Smith, J. Swails, R.C. Walker, J. Wang, R.M. Wolf, X. Wu, D.M. York and P.A. Kollman 2015.

(64) 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.

(65) Wang, J.; Cieplak, P.; Kollman, P. A. How Well Does a Restrained Electrostatic Potential (Resp) Model Perform in Calculating Conformational Energies of Organic and Biological Molecules? *Journal of computational chemistry* **2000,** *21,* 1049-1074.

(66) Horn, H. W.; Swope, W. C.; Pitera, J. W.; Madura, J. D.; Dick, T. J.; Hura, G. L.; Head-Gordon, T. Development of an Improved Four-Site Water Model for Biomolecular Simulations: Tip4p-Ew. *The Journal of chemical physics* **2004,** *120,* 9665-9678.

(67) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of Simple Potential Functions for Simulating Liquid Water. *The Journal of chemical physics* **1983,** *79,* 926-935.

(68) 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.

(69) 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.

(70) Stewart, D. E.; Sarkar, A.; Wampler, J. E. Occurrence and Role Ofcis Peptide Bonds in Protein Structures. *Journal of molecular biology* **1990,** *214,* 253-260.

(71)   Ryckaert, J.-P.; Ciccotti, G.; Berendsen, H. J. C. Numerical Integration of the Cartesian Equations of Motion of a System with Constraints: Molecular Dynamics of N-Alkanes. *Journal of Computational Physics* **1977,** *23*, 327-341.

(72)   Fleck, M.; Polyansky, A. A.; Zagrovic, B. Parent: A Parallel Software Suite for the Calculation of Configurational Entropy in Biomolecular Systems. *Journal of chemical theory and computation* **2016,** *12*, 2055-2065.

(73)   King, B. M.; Tidor, B. Mist: Maximum Information Spanning Trees for Dimension Reduction of Biological Data Sets. *Bioinformatics* **2009,** *25*, 1165-1172.

(74)   Rohrdanz, M. A.; Martins, K. M.; Herbert, J. M. A Long-Range-Corrected Density Functional That Performs Well for Both Ground-State Properties and Time-Dependent Density Functional Theory Excitation Energies, Including Charge-Transfer Excited States. *The Journal of chemical physics* **2009,** *130*, 054112.

(75)   Harihara, P. C.; Pople, J. A. Influence of Polarization Functions on Molecular-Orbital Hydrogenation Energies. *Theor Chim Acta* **1973,** *28*, 213-222.

(76)   Grimme, S.; Ehrlich, S.; Goerigk, L. Effect of the Damping Function in Dispersion Corrected Density Functional Theory. *Journal of computational chemistry* **2011,** *32*, 1456-1465.

(77)   Grimme, S. Density Functional Theory with London Dispersion Corrections. *Wiley Interdisciplinary Reviews: Computational Molecular Science* **2011,** *1*, 211-228.

(78)   Lange, A. W.; Herbert, J. M. A Smooth, Nonsingular, and Faithful Discretization Scheme for Polarizable Continuum Models: The Switching/Gaussian Approach. *The Journal of Chemical Physics* **2010,** *133*, 244111.

(79)   York, D. M.; Karplus, M. A Smooth Solvation Potential Based on the Conductor-Like Screening Model. *The Journal of Physical Chemistry A* **1999,** *103*, 11060-11079.

(80)   Liu, F.; Sanchez, D. M.; Kulik, H. J.; Martinez, T. J. Exploiting Graphical Processing Units to Enable Quantum Chemistry Calculation of Large Solvated Molecules with Conductor-Like Polarizable Continuum Models. *International Journal of Quantum Chemistry* **2019,** *119*, e25760.

(81)   Bondi, A. v. Van Der Waals Volumes and Radii. *J. Phys. Chem.* **1964,** *68*, 441-451.

(82)   Wang, L.-P.; Song, C. Geometry Optimization Made Simple with Translation and Rotation Coordinates. *The Journal of chemical physics* **2016,** *144*, 214108.

(83)   Ryde, U. How Many Conformations Need to Be Sampled to Obtain Converged Qm/Mm Energies? The Curse of Exponential Averaging. *Journal of Chemical Theory and Computation* **2017,** *13*, 5745-5752.

(84)   Cao, L.; Do, T.; Link, A. J. Mechanisms of Action of Ribosomally Synthesized and Posttranslationally Modified Peptides (Ripps). *Journal of Industrial Microbiology and Biotechnology* **2021,** *48*, kuab005.

(85)   Vennelakanti, V.; Qi, H. W.; Mehmood, R.; Kulik, H. J. When Are Two Hydrogen Bonds Better Than One? Accurate First-Principles Models Explain the Balance of Hydrogen Bond Donors and Acceptors Found in Proteins. *Chemical Science* **2021,** *12*, 1147-1162.

(86)   Killian, B. J.; Yundenfreund Kravitz, J.; Gilson, M. K. Extraction of Configurational Entropy from Molecular Simulations Via an Expansion Approximation. *The Journal of Chemical Physics* **2007,** *127*, 024107.

(87)   Polyansky, A. A.; Kuzmanic, A.; Hlevnjak, M.; Zagrovic, B. On the Contribution of Linear Correlations to Quasi-Harmonic Conformational Entropy in Proteins. *Journal of Chemical Theory and Computation* **2012,** *8*, 3820-3829.