Thermal Unthreading of the Lasso Peptides Astexin-2 and Astexin-3

Caitlin D. Allen[1], **Maria Y. Chen**[1], **Alexander Y. Trick**[3], **Dan Thanh Le**[1], **Andrew L. Ferguson**[3,4,5], and **A. James Link**[1,2,5]

[1]Department of Chemical and Biological Engineering, Princeton University, Princeton, NJ 08544

[2]Department of Molecular Biology, Princeton University, Princeton, NJ 08544

[3]Department of Materials Science and Engineering, University of Illinois at Urbana-Champaign, Urbana, IL 61801

[4]Department of Chemical and Biomolecular Engineering, University of Illinois at Urbana-Champaign, Urbana, IL 61801

Abstract

Lasso peptides are a class of knot-like polypeptides in which the C-terminal tail of the peptide threads through a ring formed by an isopeptide bond between the N-terminal amine group and a sidechain carboxylic acid. The small size (~20 amino acids) and simple topology of lasso peptides make them a good model system for studying the unthreading of entangled polypeptides, both with experiments and atomistic simulation. Here we present an in-depth study of the thermal unthreading behavior of two lasso peptides astexin-2 and astexin-3. Quantitative kinetics and energetics of the unthreading process were determined for variants of these peptides using a series of chromatography and mass spectrometry experiments and biased molecular dynamics (MD) simulations. In addition, we show that the Tyr15Phe variant of astexin-3 unthreads via an unprecedented "tail pulling" mechanism. MD simulations on a model ring-thread system coupled with machine learning approaches also led to the discovery of physicochemical descriptors most important for peptide unthreading.

Graphical abstract

[5]To whom correspondence should be addressed: alf@illinois.edu or ajlink@princeton.edu.
[1,2]207 Hoyt Laboratory, Princeton University, Princeton, NJ 08544
[3,4]204 MSEB, University of Illinois at Urbana-Champaign, 1304 W. Green St., Urbana, IL 61801

SUPPORTING INFORMATION

The Supporting Information is available free of charge on the ACS Publications website at DOI: TBD

- Supporting methods; fifteen figures (Figures S1–S15); five tables (Tables S1–S5); four movie captions (Movies S1–S4)

reaction coordinate
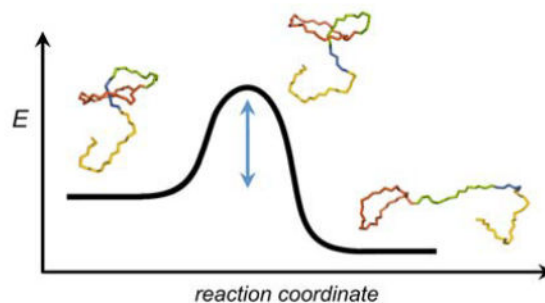
Keywords

lasso peptide; protein folding; RiPPs; natural products; bioconjugation; molecular dynamics

INTRODUCTION

Proteins with knots or slipknots represent ~1% of published structures in the protein data bank.[1, 2] These entangled structures are hypothesized to play several different roles in proteins including ligand or coenzyme binding[3, 4] and enhanced structural stability.[5] Insights into the energetics and kinetics of both folding and unfolding processes of these proteins and peptides have been gained using a variety of experimental and computational techniques. For example, atomic force microscopy (AFM) has been employed to study the unfolding of the slipknotted protein AFV3-109 in conjunction with molecular dynamics simulations.[6] The folding of trefoil knotted enzymes YbeA and YibK have been studied *in vitro* using chemical denaturation and proteins synthesized in a cell-free system, leading to the insights that the knotted segments form spontaneously and that their formation can be accelerated by chaperonins.[7, 8] In addition to these knotted systems in which the entanglements are formed solely by the protein backbone, other knot-like proteins and peptides are established via covalent sidechain-sidechain or sidechain-backbone crosslinks. A protein example of this type of entangled structure is the cytokine leptin[9] which exhibits a "pierced lasso" architecture[10, 11] formed by a disulfide bond. A well-studied example of knotting caused by disulfide bonding is the cystine knot motif, which occurs both in peptides[12] and in larger proteins.[13] Another example of this knot-like structure is the lasso peptide.[14, 15] The C-terminus of a lasso peptide is threaded through an isopeptide-bonded ring of between 7 and 9 amino acids to generate a structure that resembles a lasso (Figure 1). Lasso peptides are arguably the simplest knot-like peptides since they contain only a single entanglement. Their small size also allows for atomistic simulation of lasso peptide unthreading by molecular dynamics. We propose that lasso peptides can serve as a model system for understanding the energetics and kinetics of unthreading of more topologically complicated molecules.

The thermal unthreading of lasso peptides has been previously studied using several techniques.[16–23] The most common method involves HPLC analysis of the lasso peptide before and after heating since the threaded and unthreaded peptides usually exhibit different retention times.[17, 23] Another commonly employed assay to assess unthreading is carboxypeptidase treatment of the peptides. Carboxypeptidase removes amino acids from

the C-terminal tail of the peptide, and when a lasso peptide is in its threaded state, carboxypeptidase cleavage is hindered by the ring portion of the lasso peptide (Figure 1). In contrast, when a lasso peptide is unthreaded, the carboxypeptidase can degrade the C-terminus all the way back to the isopeptide bond. Mass spectrometry has also been employed to examine whether a lasso peptide is threaded. Work on the antimicrobial peptide microcin J25 demonstrated that the threaded species was less prone to fragmentation in MS[2] experiments than the unthreaded form,[24] though these observations are not necessarily generalizable to all lasso peptides.[25] Careful selection of the fragmentation method can lead to insights about the topology of the lasso peptide.[26] It has also been demonstrated that ion mobility mass spectrometry can differentiate between the threaded and unthreaded states of lasso peptides in high charge states.[27] Despite this large body of work, there is still no detailed analysis of the energetics and kinetics of lasso peptide unthreading.

In previous genome mining efforts in our laboratory, we discovered three lasso peptides produced by the α-proteobacterium *Asticcacaulis excentricus.*[23, 28] We have named these peptides astexins-1, -2, and -3. The genes for the precursor proteins to astexins-2 and -3 are found within the same gene cluster and share greater than 50% sequence identity (Figure 1). The NMR structures of astexins-2 and -3[23, 29] are also highly similar (Figure 1), but these peptides exhibit different thermal unthreading behavior. Whereas astexin-2 is thermolabile and unthreads completely within 2 hours at 98 °C, astexin-3 is thermostable and does not unthread at all after 3 hours of heating at 95 °C.[23] Here we use mutagenesis to probe the differences in unthreading between these two peptides. We also carry out a detailed analysis of the kinetics and energetics of the thermal unthreading of these peptides with a combination of experiments and molecular dynamics simulations.

RESULTS AND DISCUSSION

Thermostability of astexin-2 and -3 variants

The astexin-2 and -3 peptides are both 24 aa long with 9 aa isopeptide-bonded rings. These two peptides have similar sequences, with 13/24 positions identical. However, as discussed above, these peptides exhibit different thermostability. We ran HPLC unthreading assays on astexin-2 and astexin-3 at 95 °C. For all of these assays, the full-length astexin-3 peptide was used while a truncation variant of astexin-2 lacking its three C-terminal amino acids (astexin-2 C3)[23] was used for all variants of astexin-2. This truncation variant is the one produced in the highest yield from heterologous expression in *E. coli*. We found that astexin-3 does not unthread after even 8 hours of heat treatment. In contrast, astexin-2 C3 was fully unthreaded after 3 hours at 95 °C (Figure S1). Most lasso peptides include large residues on either side of the ring, steric lock residues, which contribute to the mechanical and thermal stability of the peptide. In astexin-2, these residues are Phe-15 and Arg-16 while astexin-3 contains tyrosine at position 15 and tryptophan at position 16 acting as steric locks. We first probed the role of these steric lock residues using mutagenesis. Our group has previously carried out alanine scanning mutagenesis throughout the ring and loop portions of astexins-2 and -3, finding that single alanine substitutions are tolerated at nearly every position of these peptides.[29] However, when we expressed either the Y15A or W16A

variants of astexin-3 in *E. coli*, no cyclized lasso peptide products were observed. The Y15A/W16A double variant was also not produced.

As a more conservative substitution, we next swapped individual steric lock residues between the two peptides. The Y15F and W16R variants of astexin-3 were successfully produced, and both of these variants were unthreaded upon heating (Figure S1 and Figure S2). In fact, the W16R variant of astexin-3 exhibits unthreading even at room temperature (Figure S2). The unthreaded state of the Y15F and W16R variants were confirmed by a carboxypeptidase assay (Figure S2, Figure S3). These results demonstrate that both of the steric lock residues contribute to the thermostability of astexin-3, though W16 seems to be the most critical. The corresponding F15Y and R16W variants of astexin-2 C3 were also successfully produced. Both of these variants also still unthreaded at 95 °C (Figure S1). Next we tested swaps of both steric lock residues at the same time, generating a Y15F/W16R variant of astexin-3 and a F15Y/R16W variant of astexin-2 C3. The Y15F/W16R variant of astexin-3 was as thermolabile as the W16R variant, exhibiting unthreading even at room temperature (Figure S2). In contrast, the F15Y/R16W variant of astexin-2 C3 did not unthread appreciably after 8 h of heating at 95 °C. These results show that simply grafting the steric lock residues of a thermostable lasso peptide onto a thermolabile peptide can transfer the thermostability properties.

Kinetics and energetics of astexin-2 C3 and -3 variant unthreading

To further probe the unthreading of the astexin-2 C3 and astexin-3 variants, we carried out time courses of unthreading at several different temperatures. These experiments allowed us to calculate rate constants and estimate the activation energies of the unthreading process. First we constructed thermal denaturation curves by measuring the extent of unthreading after 1 hour (Figure 2a) or 2 hours (Figure 2b) at a range of temperatures. This analysis was carried out on the four variants that are both stable at room temperature and able to unthread at higher temperatures: astexin-3 Y15F, astexin-2 C3 wild-type, astexin-2 C3 F15Y, and astexin-2 C3 R16W. The astexin-3 Y15F peptide exhibited a sigmoidal curve and the temperature at which half of the peptide was unthreaded at 1 hour ($T_{1/2}$) was 63±4 °C. The astexin-2 C3 variants were generally more stable than astexin-3 Y15F with wild-type astexin-2 C3 exhibiting a $T_{1/2}$ of 86±2 °C and astexin-2 C3 F15Y having $T_{1/2} = 87±2$ °C. The R16W astexin-2 C3 variant was more thermostable and achieved an extent of unthreading under 50 % after 1 hr giving an extrapolated $T_{1/2}$ of 110±20 °C. It is important to note that while these curves resemble the sigmoidal curves seen in protein thermal denaturation studies, the unthreading observed here is completely irreversible.

Next we examined the rate of unthreading of the same set of four peptides. This analysis was carried out at three different temperatures, 75 °C, 85 °C, and 95 °C. (Figure 2c). An example of a time course of lasso peptide unthreading is presented in Figure S3a for the astexin-3 Y15F variant. The unthreading rates correlate with the $T_{1/2}$ data presented above in that astexin-3 Y15F has the largest rate of unthreading at each of the temperatures tested and astexin-2 C3 R16W has the smallest rate of unthreading. The measurement of the unthreading rate at different temperatures allowed for estimation of an activation energy of

the unthreading process (Figure 2c, Figure S4). Consistent with the thermostability metrics, astexin-3 Y15F showed the smallest value of the unthreading activation energy.

Unthreading mechanism of astexin-3 Y15F

One of the more interesting observations from our mutagenesis studies is that the conservative substitution of tyrosine for phenylalanine in position 15 of astexin-3 converts the thermostable wild-type peptide into one that is nearly completely unthreaded after 15 minutes of treatment at 95 °C (Figure S3). This finding indicates that lasso peptide thermostability can be controlled by some differences as small as the presence or absence of a hydroxyl functional group. Furthermore, since the Tyr-15 position is on the "loop side" of the ring (above the ring in Figure 1), the thermolability of the Y15F astexin-3 peptide suggests that it may be unthreading by a mechanism in which the tail of the lasso peptide is being pulled, causing the loop to unthread through the ring. This unthreading by "tail pulling" requires some local twisting of the lasso peptide ring in contrast to "loop pulling" which does not require as much twisting (Figure 3). The size of the ring (28 atoms) and the presence of two glycine residues in the ring argue for the feasibility of ring twisting required in the tail pulling mechanism.

We probed the directionality of unthreading using molecular dynamics simulations. Potential of mean force (PMF) curves were calculated by biased molecular dynamics simulations for loop and tail pulling unthreading of WT astexin-3 (Figure 4a Figure 4b). Our simulations predict a free energy barrier for passage of the Tyr15 residue through the ring (G_{Tyr15}) of $\approx$ 130 k_BT and $G_{Trp16} \approx 160 \, k_BT$ for passage of the Trp16 residue (Movies S1–S2). The large free energy barriers for passage of either steric lock are consistent with the extreme thermostability of WT astexin-3 under experimental conditions. The ~30 k_BT lower free energy barrier for passage of Tyr15 (tail pulling) relative to Trp16 (loop pulling) suggests that the thermal unthreading of astexin-3 at elevated temperatures may proceed by tail pulling. PMF curves were also generated for the tail pulling and loop pulling unthreading of the Y15F variant of astexin-3 (Movies S3–S4). The free energy barrier of passage of Phe15 through the ring $\Delta G_{\mathrm{Phe15}}^{\mathrm{Y15F}} \approx 95 \, k_B T$, a much smaller barrier value than the loop pulling scenario: $\Delta G_{\mathrm{Trp16}}^{\mathrm{Y15F}} \approx 180 \, k_B T$ (Figure 4c). The reduced barrier to passage of Phe relative to Tyr can be understood as a result of its slightly smaller steric volume[30] and increased hydrophobic character[31] that reduces the free energy penalty associated with breaking hydrophilic hydrogen bonding associations with the surrounding water prior to passage through the ring. We present in Figure S5 illustrations of the unthreading pathways for the astexin-3 WT and Y15F mutants observed over the course of our biased tail pulling simulations. In each case we observe the Tyr15 or Phe15 to adopt a planar configuration pointing away from the direction of passage through the ring and the upper surface of the aromatic ring oriented towards and passing directly under the Met5 side chain. After complete passage, side chain rotameric freedom is restored, and the hydroxyl group of the Tyr15 in the Y15F mutant is observed to enter a weak electrostatic interaction with the Met5 side chain, whereas the Phe15 in the WT engages in ring stacking with Trp16. Lastly, we used the tail pulling PMF curves to estimate the mean free passage time (MFPT) of the wild-type and Y15F astexin-3 variants (see Methods for details). At room temperature the MFPT

of the wild-type peptide was 5×10^{46} s, while astexin-3 Y15F exhibited an MFPT of 3×10^{32} s, a vast difference. These simulations and calculations provide even more compelling evidence that the Y15F variant of astexin-3 unthreads via tail pulling.

We also set out to develop an experiment that would explicitly report on the directionality of unthreading in astexin-3 Y15F. Two new variants of astexin-3 Y15F were produced in *E. coli*: astexin-3 Y15F S12C (astexin-3 Y15F loop cysteine) and astexin-3 Y15F ΔC1 A23C (astexin-3 Y15F tail cysteine) (Figure 5). Both of these substitutions are conservative, but allow the installation of an orthogonally reactive thiol group into either the loop or tail of the peptide. As a control peptide that cannot unthread, we also expressed astexin-3 S12C carrying a thiol in its loop. Next, we attached a large sulfo-Cy3 chromophore (referred to as Cy3 for brevity's sake) to each of these peptides via a thiol-maleimide linkage (Figure 5). The bulky, rigid Cy3 group, which is much larger than any of the 20 amino acids, serves as a steric blocking group that cannot pass through the astexin-3 ring. Upon heating of the astexin-3 S12C Cy3 conjugate (astexin-3 loop Cy3) at 95 °C for 1 h, we observed a small change in retention time in the HPLC and a corresponding hydrolysis event in mass spec (Figures S6–S8). Hydrolysis of thiol-maleimide conjugates at elevated temperatures has been noted before (Figure S9), and the resulting conjugate is reported to be even more stable than the parental thiol-maleimide linkage.[32]

Next, we carried out an experiment to determine whether astexin-3 Y15F could unthread by tail pulling (Figure 3). As a control, the astexin-3 Y15F tail cysteine construct was unthreaded by heating at 95 °C for 1 h after which it was reacted with Cy3 maleimide and heated further to hydrolyze the maleimide (Figure 6). This reaction gives a conjugate with a retention time of 16.3 minutes. In a separate experiment, the astexin-3 Y15F tail Cy3 conjugate was heated at 95 °C for 1 h, also resulting in a retention time of 16.3 minutes. Carboxypeptidase treatment of these two species resulted in similar traces on the HPLC, providing further evidence that these conjugates are equivalent (Figure S10). This experiment demonstrates that astexin-3 Y15F variant can still unthread even when a large Cy3 group is added to its tail. This provides strong evidence that this peptide is not passing the C-terminal tail through the ring as it unthreads, and is instead unthreading via a tail pulling mechanism.

As an additional test of this hypothesis, we examined the unthreading behavior of the astexin-3 Y15F loop cysteine peptide. If the tail pulling hypothesis were true, we would expect the Cy3 conjugate of this peptide to remain threaded even upon heating. When Cy3 is conjugated to threaded astexin-3 Y15F loop cysteine and subsequently heated, the main peak observed has a retention time of 15.8 minutes. In contrast, when the peptide is unthreaded first and then reacted with Cy3 (followed by heating to hydrolyze the maleimide), the resulting conjugate has a retention time of 16.1 minutes (Figure 7). This change in retention time, while small, was consistently observed in three replicates of the unthreading experiments. As a further test to determine whether the species at 15.8 minutes and 16.1 minutes were indeed different, we carried out carboxypeptidase treatment of these two peptide-Cy3 conjugates. The HPLC analysis of the carboxypeptidase reactions showed that these two species were indeed distinct, and MALDI-MS analysis of these reactions suggested that the astexin-3 Y15F loop Cy3 conjugate was unable to unthread upon heating

(Figure S11). The fact that Cy3 is unable to be pulled through the ring of astexin-3 Y15F provides further compelling evidence that this peptide is unthreading via a tail pulling mechanism.

Physicochemical determinants of free energy barriers for passage through astexin-3 ring

To gain insight into the physicochemical properties of each amino acid that govern the free energy barrier for passage of each amino acid through the astexin-3 ring, we constructed a simple two-peptide test system comprising the nine-residue Gly-Pro-Thr-Pro-Met-Val-Gly-Leu-Asp astexin-3 "ring" excised from the native peptide and a seven-residue Ala-Ala-Ala-**X**-Ala-Ala-Ala "thread" (Figure 8a). We replaced **X** with each of the 20 natural amino acids and conducted umbrella sampling molecular dynamics simulations to pull the thread through the ring and measure the free energy for passage of the **X** residue, $\Delta G_{\mathrm{x}}^{\mathrm{thread}}$. By embedding each residue in the middle of a thread of sterically small Ala residues, these calculations provide a measure of the free energy barriers for passage of each amino acid free of the confounding effects associated with its local environment and non-local interactions within the astexin-3 tail (Figure 8b). It is the intent of these calculations to estimate the intrinsic cost for passage of each of the 20 natural amino acid through the astexin-3 ring as an estimate of the degree to which each residue can enhance or reduce the thermostability of astexin-3 relative to the WT Tyr15 steric lock. Nevertheless, only full simulations of loop and tail unthreading for the astexin-3 mutants similar to those described above would be required to fully account for the multi-body and sequence-dependent effects on the free energy barrier in the native lasso.

Inspection of the calculated free energy barriers reveals, as anticipated, that bulkier residues tend to be associated with higher barriers for passage. To both quantify this trend and attempt to discover other attributes of the amino acids that govern the free energy barrier, we constructed least squares multivariable linear regression (MLR) models to predict the measured values of $\Delta G_{\mathrm{x}}^{\mathrm{thread}}$ from subsets of 20 physicochemical descriptors for each amino acid. The particular descriptors were selected as physical and chemical attributes of possible importance in dictating the free energy barrier including residue size, hydrophobicity, and polarity and are listed in Table S3. After eliminating three descriptors as redundant with one or more others in the ensemble (see Supporting Information for details of the cleaning procedure), we performed variable selection by exhaustively constructing all $(2^{17}-1) = 131,071$ MLR models containing all possible subsets of descriptors, and selecting the optimal model containing a particular number of descriptors as that with the smallest root mean squared error (RMSE) in the prediction of $\Delta G_{\mathrm{x}}^{\mathrm{thread}}$ (Figure S12). Cross validation suggests that the data supports a model containing up to 14 descriptors, but the RMSE of models containing 6 or more variables is lower than the estimated ± 8.3 $k_B T$ uncertainty in the values of $\Delta G_{\mathrm{x}}^{\mathrm{thread}}$ computed in our simulations (Figure S12). Higher order models with lower RMSE cannot be justified based on the estimated uncertainty in the response variable, and we judge only those models containing 5 or fewer descriptors to be supported by the data (Figure S13).

By analyzing the selected descriptors in the 5 lowest order models together with the sign and magnitude of the associated regression coefficients, our MLR model building procedure identifies the physicochemical attributes that govern the height of the free energy barrier (Table S5). The single most important descriptor is the normalized van der Waals volume,[30] which is the single descriptor selected by the univariate linear regression model, and appears in all 5 models with a large positive regression coefficient. As anticipated, the size of the residue is the primary discriminant of the free energy barrier and with the barrier increasing quickly with residue size due to more steric clashes of the residue with the ring. The bivariate model selected in addition to the normalized van der Waals volume, the absolute value of sidechain hydropathy,[31] which was also selected by the tri, tetra, and pentavariate models. In all cases this descriptor was associated with a positive regression coefficient indicating that the free energy barrier height increases for strongly hydrophilic and strongly hydrophobic residues. We observe, however, that there is a moderate correlation between the normalized van der Waals volume and the sidechain hydropathy ($\rho_{\text{Pearson}} = 0.50$), so it is conceivable that this descriptor was largely identified as a secondary proxy for residue size. The trivariate model additionally identifies the absolute value of a second measure of side-chain hydropathy,[31] which also appears in the tetra and pentavariate models. The associated regression coefficient is negative but with smaller absolute value than that for the first hydropathy measure, suggesting that this descriptor may be acting as a second order correction. In addition to these three descriptors, the tetravariate model also identifies polarity,[33] and the pentavariate model minimum sidechain width[30] and residue bulkiness.[33] These descriptors are closely related to the previously identified normalized van der Waals volume and sidechain hydropathy, so appear not to correspond to any qualitatively new attributes.

In sum, our MLR model construction approach has identified residue size and hydrophilicity/hydrophobicity as the principal determinants of the free energy barrier, and the quantitative values of $\Delta G_{\text{x}}^{\text{thread}}$ can be used to guide the engineering of astexin-3 variants with tailored thermostability.

This paper provides the first examination of the rates, energetics, and directionality of lasso peptide unthreading using a combination of experiments and molecular dynamics simulations. We calculated denaturation temperatures, rate constants, and activation energies for the unthreading processes of variants of the lasso peptides astexin-2   C3 and astexin-3. For these peptides, the identity of the steric lock residues is the main driver of lasso peptide thermostability. We also note that lasso peptide thermostability can exist on a razor's edge: the seemingly conservative change of Tyr15 to phenylalanine converts astexin-3 from a peptide that remains threaded after 8 hrs at 95 °C to one that completely unthreads after 15 min at 95 °C. The other new insight gained from studying the Y15F variant of astexin-3 is that it unthreads by a tail pulling mechanism (Figure 3). Both simulations (Figure 4) and experiments that introduce a bulky Cy3 moiety into different regions of the peptide (Figures 6 and 7) corroborate this observation.

A study of the thermal unthreading behavior is now frequently one of the characterization steps upon the discovery of a new lasso peptide.[18] In addition, lasso peptides that are found

to be stable, such as capistruin,[16] caulosegnin II,[17, 22] and xanthomonin II[20] have been subjected to mutagenesis to find residues that contribute to the thermostability. In the mutagenesis studies of these peptides, amino acid substitutions are made almost exclusively in the tail region of the peptide, thus presupposing a loop pulling unthreading mechanism. In the case of caulosegnin II, changes to a proline residue in the ring of the peptide were required to generate a thermolabile variant.[22] This finding is in contrast to our work here where thermostability of astexins-2 and -3 depend entirely on the steric lock residues. A comparison of astexin-3 to these other thermostable lasso peptides (Figure S14) illustrates that variants of astexin-3 may be unique in their ability to unthread by tail pulling. Capistruin, caulosegnin II, and xanthomonin II all have a relatively large amino acid in the second position of the loop, whereas astexin-3 has no large amino acids in its loop. For xanthomonin II, the methionine residue in position 9 is considered "large" because this peptide has a smaller ring of only 7 aa, and its tail side steric lock is isoleucine. We propose that the presence of these large amino acids in the loops of capistruin, caulosegnin II, and xanthomonin II are serving as an impediment to their unthreading via tail pulling. In contrast, once the loop-side steric lock Tyr-15 is removed from astexin-3, it is free to unthread via tail pulling.

The tail pulling mechanism may also explain the thermal unthreading behavior of variants of the lasso peptide caulonodin VI (Figure S14).[21] Wild-type caulonodin VI has a putative loop-side steric lock of Arg (R15) and a tail-side steric lock of Tyr (Y16) from NOESY studies. The wild-type peptide is thermolabile, unthreading at 95 °C, as is the Y16W variant. However, the R15W variant is thermostable, exhibiting no unthreading at 95 °C for up to 4 h. For this peptide, we propose that the tail side steric lock, Y16, is sufficiently large to prevent unthreading by loop pulling, and that the wild-type peptide instead unthreads by tail pulling. The R15W substitution prevents unthreading by tail pulling rendering the peptide thermostable.

Since lasso peptides are being discovered from genome sequences rapidly, the insights from this paper are expected to enable quicker determinations of whether a new lasso peptide is thermostable or thermolabile. Our experiments as well as the simulations carried out on the model ring-thread system point to amino acid sidechain size as being the primary factor contributing to the resistance of lasso peptide unthreading. More generally, our work here provides foundations for thinking about the reptation of peptide chains through peptide rings or other polypeptide entanglements.

METHODS

Detailed methods are provided in the Supporting Information. Lasso peptides were purified from the cell pellets of recombinant *E. coli* harboring plasmid-borne gene clusters for the astexin variants. All lasso peptides and lasso peptide-Cy3 conjugates were purified to homogeneity using HPLC and the identity of the peptides was confirmed by MALDI-MS. Lasso peptide heating experiments were carried out in a Peltier-controlled thermocycler. Molecular dynamics simulations were performed in GROMACS 4.6[34] using the CHARMM27 force field[35] and SPC water model.[36] Following equilibration, umbrella sampling was performed to apply a series of artificial biasing potentials to perform staged

sampling of the unthreading pathway.[37] For unthreading of the WT and Y15F astexin-3 variants, we applied harmonic restraining potentials between the center of mass of the ring and the Cα of residue 15 (loop pulling) or residue 16 (tail pulling). For the ring-thread system, we applied harmonic restraints between the centers of mass of the excised ring and the Ala1 residue of the thread. In all cases free energy profiles were calculated from the biased simulation data using the weighted histogram analysis method (WHAM)[38, 39] implemented in the g_wham tool within GROMACS 4.6.[40] Peptides were visualized using either PyMol or VMD.[41]

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

References

1. Jamroz M, Niemyska W, Rawdon EJ, Stasiak A, Millett KC, Sulkowski P, Sulkowska JI. KnotProt: a database of proteins with knots and slipknots. Nucleic Acids Res. 2015; 43:D306–D314. [PubMed: 25361973]

2. Lim NCH, Jackson SE. Molecular knots in biology and chemistry. J Phys: Conds Matter. 2015; 27

3. Lim K, Zhang H, Tempczyk A, Krajewski W, Bonander N, Toedt J, Howard A, Eisenstein E, Herzberg O. Structure of the YibK methyltransferase from Haemophilus influenzae (HI0766): A cofactor bound at a site formed by a knot. Proteins: Struct Funct, Genet. 2003; 51:56–67. [PubMed: 12596263]

4. Wagner JR, Brunzelle JS, Forest KT, Vierstra RD. A light-sensing knot revealed by the structure of the chromophore-binding domain of phytochrome. Nature. 2005; 438:325–331. [PubMed: 16292304]

5. Sayre TC, Lee TM, King NP, Yeates TO. Protein stabilization in a highly knotted protein polymer. Protein Eng Des Sel. 2011; 24:627–630. [PubMed: 21669955]

6. He C, Genchev GZ, Lu H, Li H. Mechanically Untying a Protein Slipknot: Multiple Pathways Revealed by Force Spectroscopy and Steered Molecular Dynamics Simulations. J Am Chem Soc. 2012; 134:10428–10435. [PubMed: 22626004]

7. Lim NCH, Jackson SE. Mechanistic Insights into the Folding of Knotted Proteins In Vitro and In Vivo. J Mol Biol. 2015; 427:248–258. [PubMed: 25234087]

8. Mallam AL, Jackson SE. Knot formation in newly translated proteins is spontaneous and accelerated by chaperonins. Nat Chem Biol. 2012; 8:147–153.

9. Zhang FM, Basinski MB, Beals JM, Briggs SL, Churgay LM, Clawson DK, DiMarchi RD, Furman TC, Hale JE, Hsiung HM, Schoner BE, Smith DP, Zhang XY, Wery JP, Schevitz RW. Crystal structure of the obese protein leptin-E100. Nature. 1997; 387:206–209. [PubMed: 9144295]

10. Haglund E, Sulkowska JI, Noel JK, Lammert H, Onuchic JN, Jennings PA. Pierced Lasso Bundles Are a New Class of Knot-like Motifs. PLoS Comput Biol. 2014; 10:11.

11. Dabrowski-Tumanski P, Niemyska W, Pasznik P, Sulkowska JI. LassoProt: server to analyze biopolymers with lassos. Nucleic Acids Res. 2016; 44:W383–W389. [PubMed: 27131383]

12. Craik DJ, Cemazar M, Daly NL. The chemistry and biology of cyclotides. Curr Opin Drug Discovery Dev. 2007; 10:176–184.

13. McDonald NQ, Hendrickson WA. A Structural Superfamily of Growth Factors Containing a Cystine Knot Motif. Cell. 1993; 73:421–424. [PubMed: 8490958]

14. Maksimov MO, Pan SJ, Link AJ. Lasso peptides: structure, function, biosynthesis, and engineering. Nat Prod Rep. 2012; 29:996–1006. [PubMed: 22833149]

15. Hegemann JD, Zimmermann M, Xie X, Marahiel MA. Lasso Peptides: An Intriguing Class of Bacterial Natural Products. Acc Chem Res. 2015; 48:1909–1919. [PubMed: 26079760]

16. Knappe TA, Linne U, Robbel L, Marahiel MA. Insights into the Biosynthesis and Stability of the Lasso Peptide Capistruin. Chem Biol. 2009; 16:1290–1298. [PubMed: 20064439]

17. Hegemann JD, Zimmermann M, Xie XL, Marahiel MA. Caulosegnins I–III: A Highly Diverse Group of Lasso Peptides Derived from a Single Biosynthetic Gene Cluster. J Am Chem Soc. 2013; 135:210–222. [PubMed: 23214991]

18. Hegemann JD, Zimmermann M, Zhu SZ, Klug D, Marahiel MA. Lasso Peptides From Proteobacteria: Genome Mining Employing Heterologous Expression and Mass Spectrometry. Biopolymers. 2013; 100:527–542. [PubMed: 23897438]

19. Zimmermann M, Hegemann Julian D, Xie X, Marahiel Mohamed A. The Astexin-1 Lasso Peptides: Biosynthesis, Stability, and Structural Studies. Chem Biol. 2013; 20:558–569. [PubMed: 23601645]

20. Hegemann JD, Zimmermann M, Zhu SZ, Steuber H, Harms K, Xie XL, Marahiel MA. Xanthomonins I–III: A New Class of Lasso Peptides with a Seven-Residue Macrolactam Ring. Angew Chem-Int Edit. 2014; 53:2230–2234.

21. Zimmermann M, Hegemann JD, Xie X, Marahiel MA. Characterization of caulonodin lasso peptides revealed unprecedented N-terminal residues and a precursor motif essential for peptide maturation. Chem Sci. 2014; 5:4032–4043.

22. 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; 12:1106–1109. [PubMed: 26863937]

23. Maksimov MO, Link AJ. Discovery and Characterization of an Isopeptidase That Linearizes Lasso Peptides. J Am Chem Soc. 2013; 135:12038–12047. [PubMed: 23862624]

24. Wilson KA, Kalkum M, Ottesen J, Yuzenkova J, Chait BT, Landick R, Muir T, Severinov K, Darst SA. Structure of microcin J25, a peptide inhibitor of bacterial RNA polymerase, is a lassoed tail. J Am Chem Soc. 2003; 125:12475–12483. [PubMed: 14531691]

25. Zirah S, Afonso C, Linne U, Knappe TA, Marahiel MA, Rebuffat S, Tabet JC. Topoisomer Differentiation of Molecular Knots by FTICR MS: Lessons from Class II Lasso Peptides. J Am Soc Mass Spectrom. 2011; 22:467–479. [PubMed: 21472565]

26. Perot-Taillandier M, Afonso C, Enjalbert Q, Antoine R, Dugourd P, Cole RB, Tabet JC, Rebuffat S, Zirah S. Electron detachment/photodetachment dissociation of lasso peptides. Int J Mass Spectrom. 2015; 390:91–100.

27. Fouque KJD, Afonso C, Zirah S, Hegemann JD, Zimmermann M, Marahiel MA, Rebuffat S, Lavanant H. Ion Mobility-Mass Spectrometry of Lasso Peptides: Signature of a Rotaxane Topology. Anal Chem. 2015; 87:1166–1172. [PubMed: 25495527]

28. Maksimov MO, Pelczer I, Link AJ. Precursor-centric genome-mining approach for lasso peptide discovery. Proc Natl Acad Sci U S A. 2012; 109:15223–15228. [PubMed: 22949633]

29. Maksimov MO, Koos JD, Zong C, Lisko B, Link AJ. Elucidating the Specificity Determinants of the AtxE2 Lasso Peptide Isopeptidase. J Biol Chem. 2015; 290:30806–30812. [PubMed: 26534965]

30. Fauchere JL, Charton M, Kier LB, Verloop A, Pliska V. Amino-acid Side-chain Parameters for Correlation Studies in Biology and Pharmacology. Int J Pept Protein Res. 1988; 32:269–278. [PubMed: 3209351]

31. Roseman MA. Hydrophilicity of Polar Amino-acid Side-chains Is Markedly Reduced by Flanking Peptide-bonds. J Mol Biol. 1988; 200:513–522. [PubMed: 3398047]

32. Fontaine SD, Reid R, Robinson L, Ashley GW, Santi DV. Long-Term Stabilization of Maleimide-Thiol Conjugates. Bioconjug Chem. 2015; 26:145–152. [PubMed: 25494821]

33. Zimmerman JM, Eliezer N, Simha R. Characterization of Amino Acid Sequences in Proteins by Statistical Methods. J Theor Biol. 1968; 21:170–201. [PubMed: 5700434]

34. Hess B, Bekker H, Berendsen HJC, Fraaije J. LINCS: A linear constraint solver for molecular simulations. J Comput Chem. 1997; 18:1463–1472.

35. MacKerell AD, Bashford D, Bellott M, Dunbrack RL, Evanseck JD, Field MJ, Fischer S, Gao J, Guo H, Ha S, Joseph-McCarthy D, Kuchnir L, Kuczera K, Lau FTK, Mattos C, Michnick S, Ngo

T, Nguyen DT, Prodhom B, Reiher WE, Roux B, Schlenkrich M, Smith JC, Stote R, Straub J, Watanabe M, Wiorkiewicz-Kuczera J, Yin D, Karplus M. All-atom empirical potential for molecular modeling and dynamics studies of proteins. J Phys Chem B. 1998; 102:3586–3616. [PubMed: 24889800]

36. Berendsen, HJC.; Postma, JPM.; van Gunsteren, WF.; Hermans, J. Intermolecular Forces. Reidel; Dordrecht: 1981.

37. Torrie GM, Valleau JP. Nonphysical sampling distributions in Monte Carlo free-energy estimation: Umbrella sampling. J Comput Phys. 1977; 23:187–199.

38. Ferrenberg AM, Swendsen RH. Optimized Monte-Carlo data-analysis. Phys Rev Lett. 1989; 63:1195–1198. [PubMed: 10040500]

39. Kumar S, Bouzida D, Swendsen RH, Kollman PA, Rosenberg JM. The weighted histogram analysis method for free-energy calculations on biomolecules. I The method. J Comput Chem. 1992; 13:1011–1021.

40. Hub JS, De Groot BL, Van Der Spoel D. g_wham -- A Free Weighted Histogram Analysis Implementation Including Robust Error and Autocorrelation Estimates. J Chem Theory Comput. 2010; 6:3713–3720.

41. Humphrey W, Dalke A, Schulten K. VMD: Visual Molecular Dynamics. J Mol Graph. 1996; 14:33–38. [PubMed: 8744570]
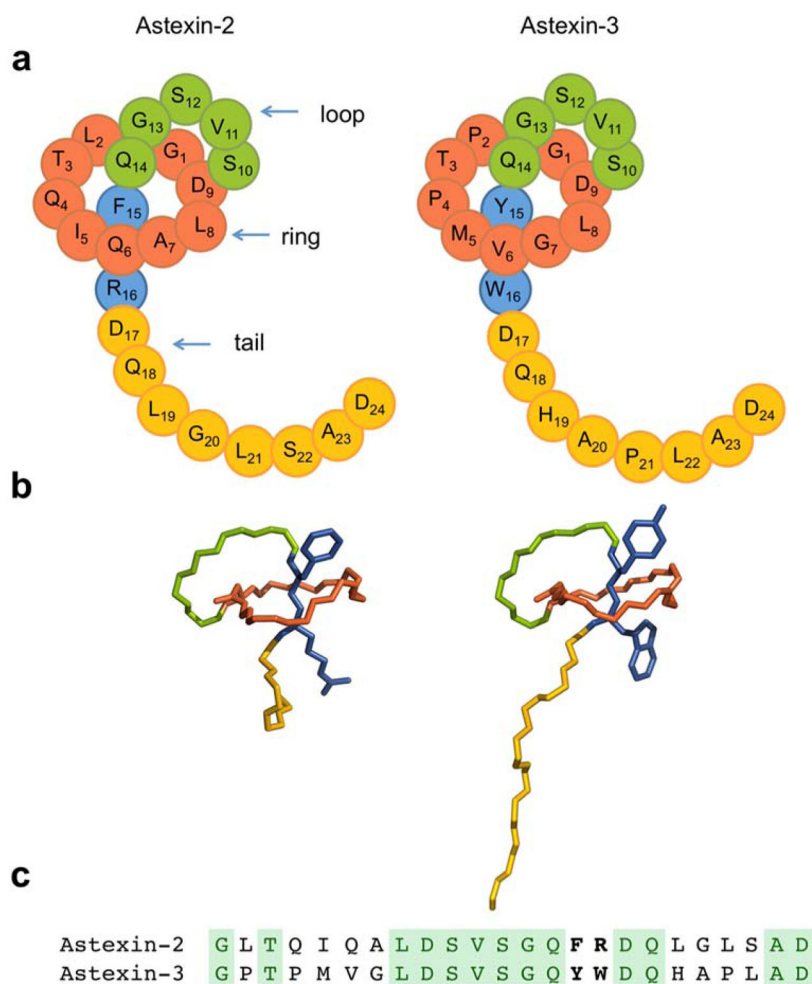
Figure 1. Sequence and structure of astexin-2 and astexin-3

(a) Cartoon representation of the peptides. Red represents the ring of the peptide, green is the loop, blue are steric lock residues, and orange is the C-terminal tail. (b) NMR structures of astexin-2 C4 (left), from PDB file 2N6U and full-length astexin-3, from PDB file 2N6V (right). Sidechains of steric lock residues are shown and the color scheme is as in part A. (c) Sequence comparison of astexins-2 and -3. Identities are highlighted in green, steric lock residues are in bold type.
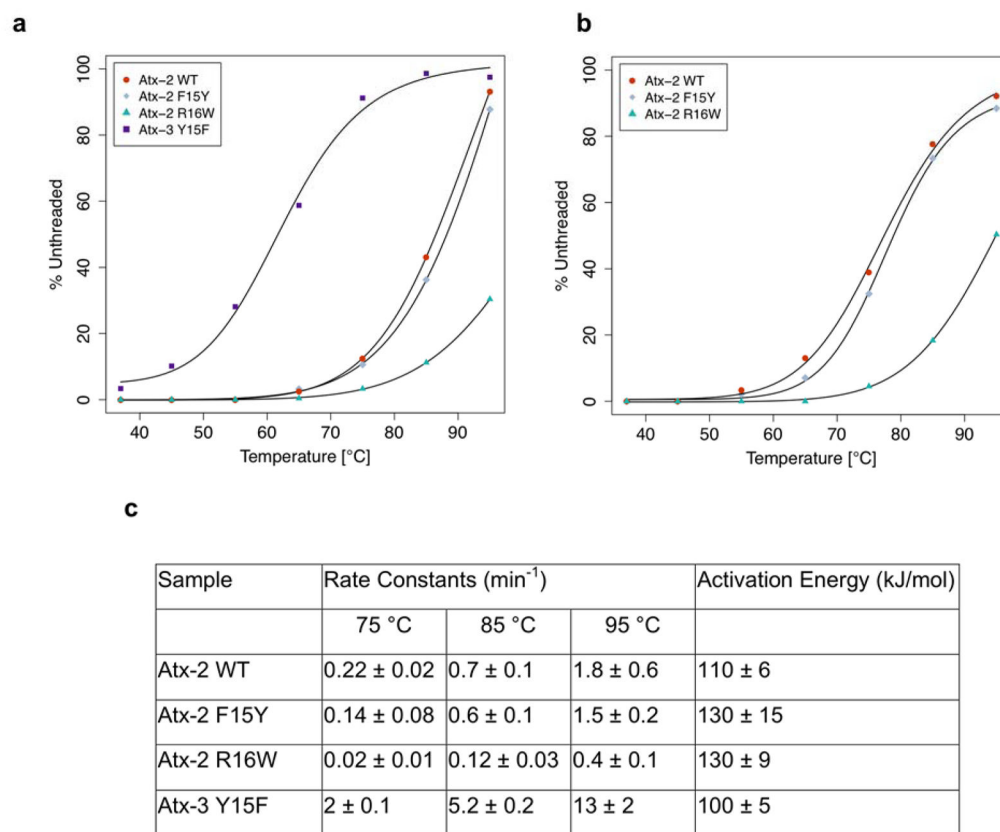
a

b

c

Sample	Rate Constants (min⁻¹)			Activation Energy (kJ/mol)
	75 °C	85 °C	95 °C	
Atx-2 WT	0.22 ± 0.02	0.7 ± 0.1	1.8 ± 0.6	110 ± 6
Atx-2 F15Y	0.14 ± 0.08	0.6 ± 0.1	1.5 ± 0.2	130 ± 15
Atx-2 R16W	0.02 ± 0.01	0.12 ± 0.03	0.4 ± 0.1	130 ± 9
Atx-3 Y15F	2 ± 0.1	5.2 ± 0.2	13 ± 2	100 ± 5

Figure 2. Thermal denaturation curves, rate constants, and activation energies for thermolabile astexin-2 C3 and -3 variants

Thermal denaturation curves for peptides heated for (a) one hour or (b) two hours, at the indicated temperatures and then analyzed by HPLC. The % unthreaded was determined by integrating the chromatograms. (c) Rate constants and activation energies
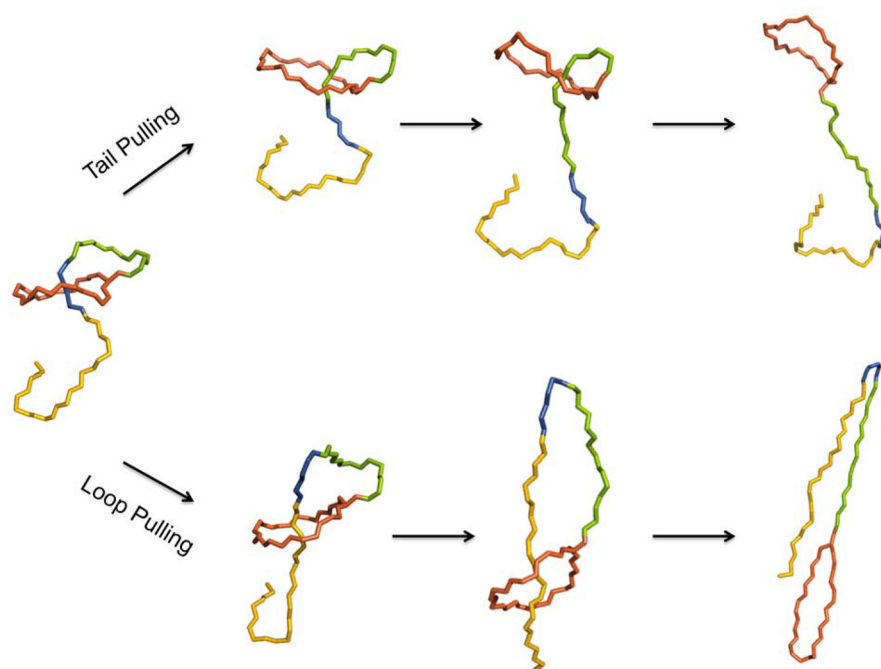
Figure 3. Schematic of two directions for lasso peptide unthreading

Top: unthreading via pulling on the C-terminal tail. Bottom: unthreading via loop pulling.

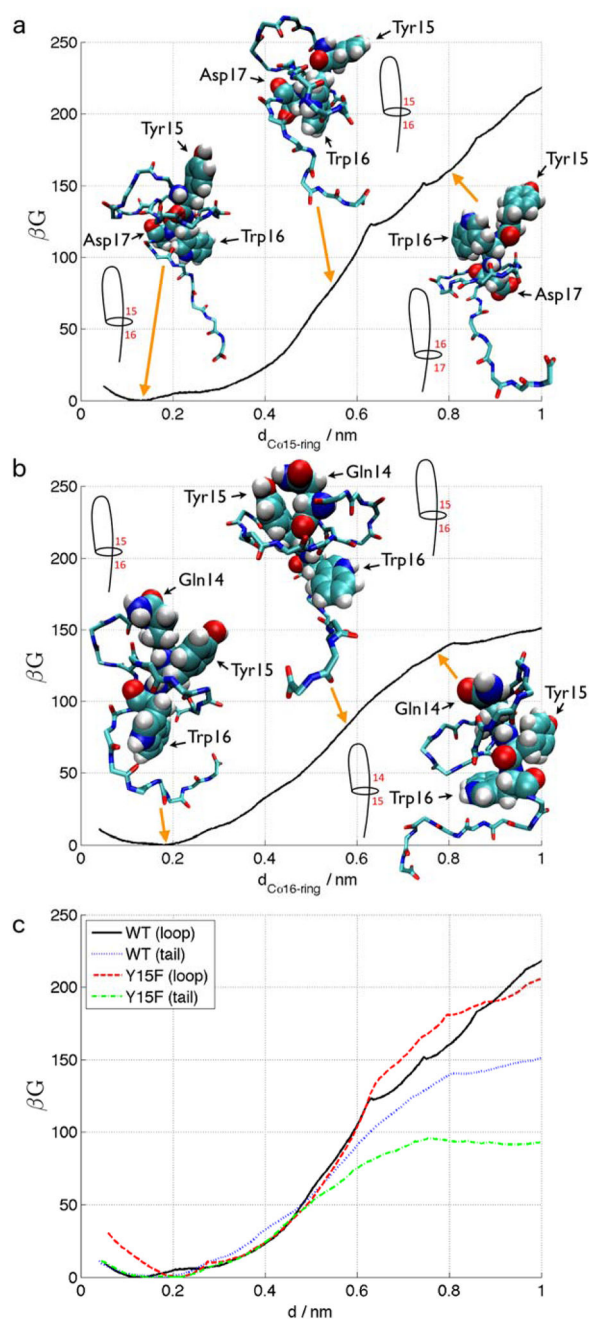The color scheme is as in Figure 1; position of steric lock residues are marked in blue.

Figure 4. Potential of mean force (PMF) computed by umbrella sampling molecular dynamics simulations for the unthreading of astexin-3 WT and Y15F mutants

(a) Unthreading of astexin-3 WT by loop pulling in which the Trp16 steric lock residue passes through the ring. Complete passage of Trp16 occurs at $d_{C\alpha15\text{-ring}} \approx 0.8$ nm. We report the PMF (Gibbs free energy) G rendered dimensionless by the reciprocal temperature $\beta = 1/k_BT$ where k_B is Boltzmann's constant and $T = 298$ K. The arbitrary zero of free energy of each PMF curve was specified to coincide with its global minimum, and uncertainties estimated by 100 rounds of bootstrap resampling are smaller than 1 k_BT. The PMF curves are superposed with representative snapshots of the peptide along the unthreading trajectory

rendered using VMD and schematic line diagrams to illustrate which tail residues are spanning the ring in each image. (b) Unthreading of astexin-3 WT by tail pulling in which the Tyr15 steric lock passes through the ring. Complete passage of Tyr15 occurs at $d_{C\alpha16\text{-ring}}$ ≈ 0.75 nm. (c) Composite plot of the PMF curves for loop and tail pulling of the astexin-3 wild-type (WT) and Y15F variant. d corresponds to $d_{C\alpha15\text{-ring}}$ in the case of loop pulling, and to $d_{C\alpha16\text{-ring}}$ in the case of tail pulling. For both variants of the peptide, the free energy barrier for tail pulling (slipping residue 15) is significantly lower than that for loop pulling

(slipping residue 16): $\Delta G_{\text{Tyr15}}^{\text{WT}} \approx 130\, k_B T$ vs. $\Delta G_{\text{Trp16}}^{\text{WT}} \approx 160\, k_B T$ for the WT, and

$\Delta G_{\text{Phe15}}^{\text{Y15F}} \approx 95\, k_B T$ vs. $\Delta G_{\text{Trp16}}^{\text{Y15F}} \approx 180\, k_B T$ for Y15F.
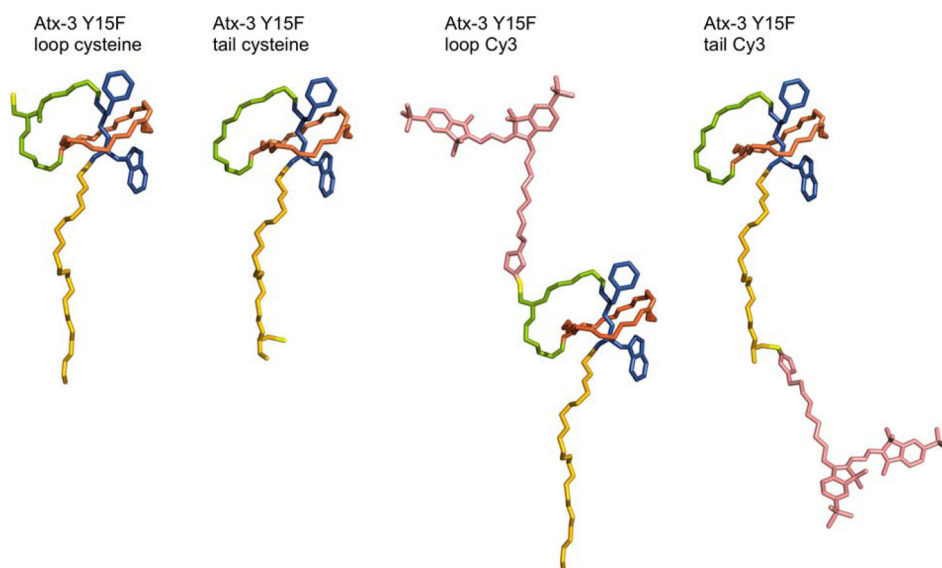
Figure 5. Astexin-3 Y15F constructs for studying unthreading directionality
The cysteine sidechains are highlighted in bright yellow and the steric lock sidechains are in blue as in Figure 1. The bulky sulfo-Cy3 moiety attached as a secondary steric lock is in light pink.
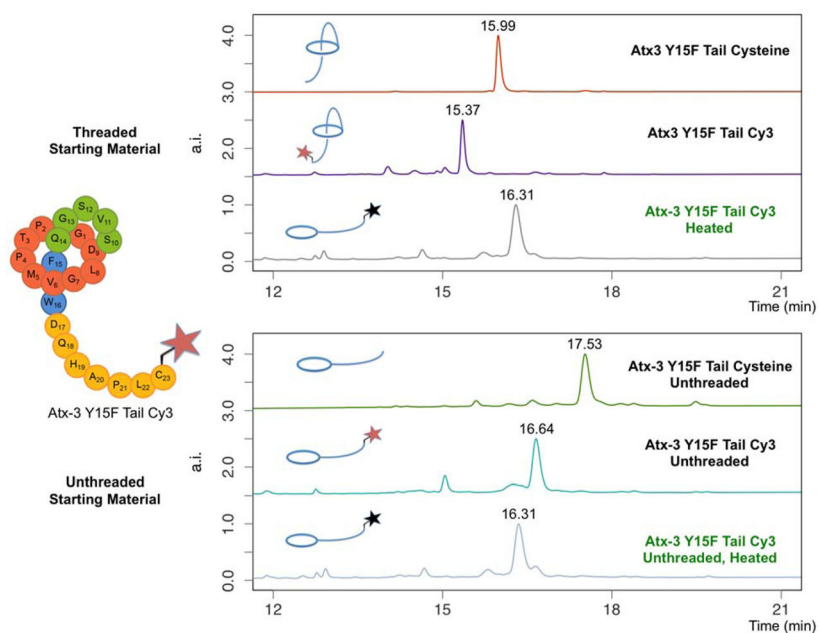
Figure 6. Probing unthreading of astexin-3 Y15F using tail-labeled peptide
Top panel: from top to bottom, astexin-3 Y15F tail cysteine is labeled first with Cy3 and then heated, resulting in an unthreaded species. Bottom panel: from top to bottom, astexin-3 Y15F tail cysteine is first unthreaded by heating then labeled with Cy3 followed by heating again to hydrolyze the maleimide. The final product is identical to what is observed in the top panel, indicating that the Cy3 group on the tail does not prevent unthreading. Red stars represent non-hydrolyzed Cy3 groups, black stars are maleimide hydrolyzed Cy3.
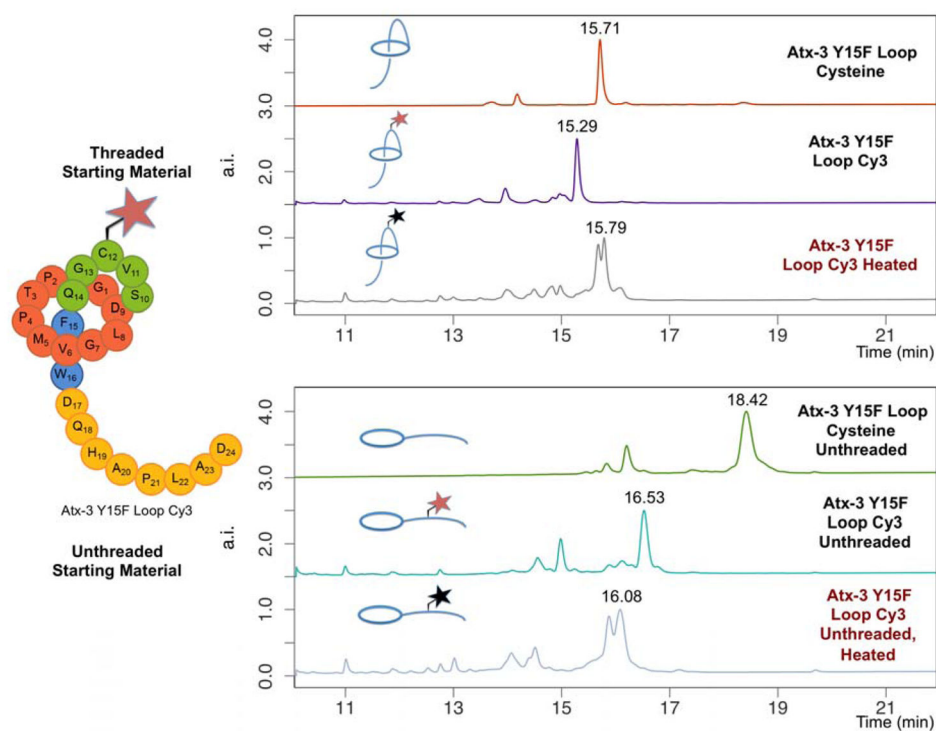
Figure 7. Probing unthreading of astexin-3 Y15F using loop-labeled peptide
Top panel: from top to bottom, astexin-3 Y15F loop cysteine was labeled with Cy3 and subsequently heated at 95 °C for 1 hr. Bottom panel: from top to bottom, astexin-3 Y15F loop cysteine was fully unthreaded by heating, followed by conjugation with Cy3 and further heating to hydrolyze the maleimide. In contrast to the results in Figure 6, these two treatments result in different species, indicating that conjugation of Cy3 in the loop region prevents full unthreading of the peptide. Red stars represent non-hydrolyzed Cy3 groups, black stars are maleimide-hydrolyzed Cy3.
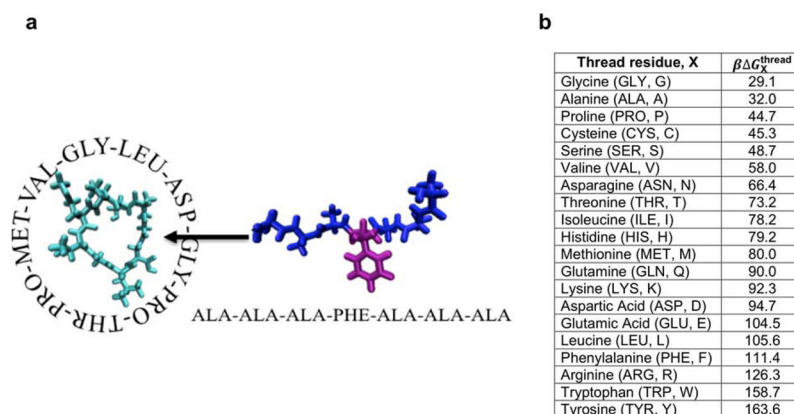
a

b

Thread residue, X	$\beta\Delta G_X^{thread}$
Glycine (GLY, G)	29.1
Alanine (ALA, A)	32.0
Proline (PRO, P)	44.7
Cysteine (CYS, C)	45.3
Serine (SER, S)	48.7
Valine (VAL, V)	58.0
Asparagine (ASN, N)	66.4
Threonine (THR, T)	73.2
Isoleucine (ILE, I)	78.2
Histidine (HIS, H)	79.2
Methionine (MET, M)	80.0
Glutamine (GLN, Q)	90.0
Lysine (LYS, K)	92.3
Aspartic Acid (ASP, D)	94.7
Glutamic Acid (GLU, E)	104.5
Leucine (LEU, L)	105.6
Phenylalanine (PHE, F)	111.4
Arginine (ARG, R)	126.3
Tryptophan (TRP, W)	158.7
Tyrosine (TYR, Y)	163.6

Figure 8. Ring-thread system comprising the nine-residue Gly-Pro-Thr-Pro-Met-Val-Gly-Leu-Asp astexin-3 "ring" excised from the native peptide and a seven-residue Ala-Ala-Ala-X-Ala-Ala-Ala "thread" in which X was exchanged for each of the 20 natural amino acids

(a) This image illustrates the **X** = Phe case. The thread was computationally docked and threaded through the ring using artificial umbrella biasing potentials, and the umbrella sampling data analyzed using the WHAM approach to estimate the free energy barrier for passage of each amino acid residue through the ring free of the confounding effects of its local environment within the astexin-3 peptide chain. (b) Calculated free energy barriers for passage of the **X** residue within the AAA**X**AAA thread through the excised Gly-Pro-Thr-Pro-Met-Val-Gly-Leu-Asp astexin-3 ring by umbrella sampling molecular dynamics simulations. Values are reported in units of $k_BT = \beta^{-1}$ where k_B is Boltzmann's constant and $T = 298$ K. Uncertainties in the free energy barriers were quantified by performing five independent runs of the AAA**F**AAA system from which we estimated an uncertainty in the measured barrier of ± 8.3 k_BT.