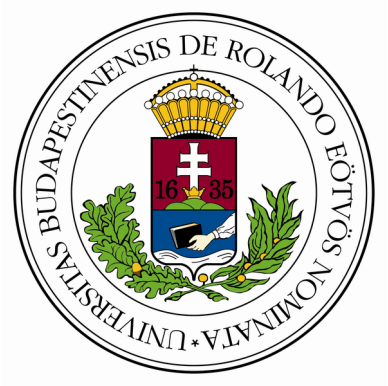
Trp-cage miniproteins
Structure, Dynamics and Folding

Theses of the PhD dissertation



Petra Rovó

Supervisor: Dr. András Perczel, D. Sc

Chemistry Doctoral School
Head: Dr. György Inzelt, D. Sc
Synthetic Chemistry, Materials Science and Biomolecular Chemistry
Head: Dr. András Perczel, D. Sc

Eötvös Loránd University
Faculty of Natural Sciences, Insitute of Chemistry
Laboratory of Structual Chemistry and Biology

2014

Introduction

The 20-residue long, *de novo* designed Trp-cage miniprotein has been the subject of several *in vitro* and *in silico* studies. This doctoral thesis deals with the investigation of the structure, dynamics and folding process of this model protein family.

Trp-cage is an excellent protein model since its small size reduces the parameter space to a handful of independent variables which influences the proteins physico-chemical properties. Although Trp-cage is small, it resembles to larger globular proteins in many of its characteristics: it has a well-defined permanent three-dimensional structure under physiological conditions with a hydrophobic core and multiple secondary structural elements; it has a cooperative melting profile with apparent melting temperatures around $40 - 50$ °C; and its folding is governed by the same rules that is common for single-domain proteins.

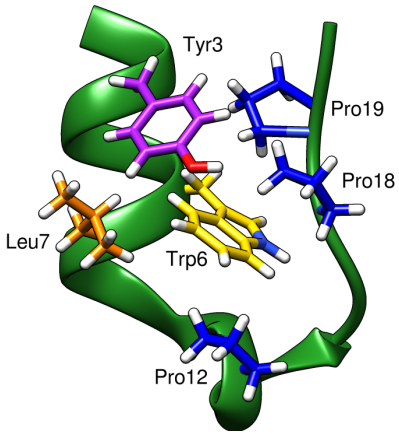


Figure 1. Ribbon representation of Trp-cage miniprotein (PDB: 1L2Y).

Beyond theoretical significance, the study of Trp-cage miniproteins has practical relevance since the 39 residue long polypeptide (Exendin-4) from which Trp-cage has been derived is a peptide drug in use to treat type two diabetes. By using the optimised Trp-cage structure it is possible to improve the physico-chemical properties of the antidiabetic drug as discussed in the present doctoral thesis.

Aims

The aims of my doctoral work was the study of the structure, dynamics and folding process of the Trp-cage miniproteins with solution state NMR spectroscopy.

In this study I investigate three Trp-cage groups.

1. The first group includes the $Asp^{9} - Arg^{16}$ salt-bridge variants. The analysis of these peptides provides insights into the nature of the electrostatic interaction, how it stabilises the miniproteins and in what other interactions are the salt-bridging side-chains involved. I apply two approaches to reduce the attractive force of the salt-bridge; on one hand I use the Asp $\rightarrow$ Asn, Ser and Arg $\rightarrow$ Ala mutations to eliminate the charge of one of the salt-bridging partners, on the other hand I measure the samples at acidic pH thereby protonating the Asp or Glu side-chains and thus reducing the electrostatic nature of the interaction.

2. The objects of my second study will be the phosphoserine variants of Trp-cage. There are three serines in the Trp-cage sequence—Ser^{13}, Ser^{14} és Ser^{20}—all of which have been replaced with glutamine or glutamate residue. Here, I compare the structural properties of these peptides with their phosphoserine analogues that is where the very same serine has been modified to a phosphoserine residue. In biology it is a common rule of thumb to use glutamate to mimic phosphoserine moieties. With this comparative work I would like to explore this assumption from a structural perspective.

3. The third Trp-cage group in my studies includes the sequences that were elongated at their N-terminus with the corresponding amino acids of Exendin-4. The major goal of this work is to create a polypeptide which has the same tertiary structure that Exendin-4 has, which is more helical than Ex4, less prone to oligomerisation and meanwhile which has a well-preserved biological activity. If these conditions are fulfilled then the designed molecule could be a perfect lead compound for further drug discovery.

To characterise these three Trp-cage miniprotein groups I use different NMR spectroscopic techniques.

1. With the help of homonuclear TOCSY, COSY and NOESY spectra I can gather information about the relative stability of the miniproteins. To do so I use the characteristic secondary chemical shifts of Trp-cage. With these chemical shift values I could detect subtle relative changes in the stability of the compared Trp-cage miniproteins. Thus, without demanding temperature-dependent or calorimetric measurements I could undoubtedly determine which variants are more stable than the others. With this method I analyse all Trp-cage sequences presented in this work.

2. When adequate number of long-range NOE restraints are available in the NOESY spectra I am going to perform an NOE-based structure ensemble calculation with a molecular dynamics program. The quality of the calculated structures will be assessed by the root mean square deviation of the backbone and side-chain atoms.

3. To enable heteronuclear backbone relaxation measurements (^{15}N R_1, R_2 and heteronculear NOE) and temperature-dependent HSQC (heteronuclear single-quantum coherence spectroscopy) folding studies I intended to produce selected sequences in a ^{15}N or ^{15}N and ^{13}C labeled form. With the temperature-dependent measurements I can characterise the folding route of the miniproteins while with the help of the relaxation measurements I can gain insights into the ps-ns time-scale dynamics of the Trp-cage backbone.

Methods

Most of the unlabeled samples were produced at the University of Szeged, Department of Medical Chemistry in Gábor Tóth's lab using solid phase peptid synthesis. The rest of the unlabeled samples and the ^{15}N and ^{13}C, ^{15}N labeled samples were prepared using recombinant techniques and expressed as a fusion protein in *E. coli* cells . The expressed proteins were purified by Ni-affinity chromatography and reveres-phase HPLC and their identity were confirmed using a mass spectrometer equipped with electron-spray ion source.

The presented ECD spectra were recorded on a Jasco J810 dichrograph in a $185-260$ nm wavelength range. The NMR spectra were recorded on the Bruker DRX 500 and 700 MHz spectrometers using a 5 mm triple resonance probe on peptide samples of $0.5-1.5$ mM in a solution of 10% D_2O:90% H_2O or 30% d_2-trifluoroethanol :70% H_2O. 4,4-dimethyl-4-silapentane-1-sulfonic acid (DSS) was used as the internal proton reference standard set to 0.0 ppm for all conditions. Secondary chemical shift was expressed as the difference of the measured and the random coil chemical shift.

The assignment of the homonuclear spectra was based on TOCSY, COSY and NOESY spectra. For structure calculations I used the NOE peak intensities and converted them to spacial restraints. The assignment of the heteronuclear spectra were based on 3D HSQC-TOCSY, HSQC-NOESY and HNCA, HNCOCA, HNCACB and CCCONH. Backbone ^{15}N longitudinal (R_1) and transverse (R_2) and $R_{1\rho}$ relaxation rates and heteronuclear cross-relaxation rate constants (^{1}H-^{15}N NOE) were acquired using common ^{1}H-^{15}N experiments . To analyse the relaxation data I used the Lipari-Szabo model-free formalism and the Reduces Spectral Density Mapping approach. To investigate the thermal unfolding I have developed a method with which both the chemical shifts of the pure states that emerge during the unfolding process both the thermodynamic parameters of the transition steps could be determined.

Thesis

The investigation of the Trp-cage salt-bridge variants helped me to understand what is the role of a weak electrostatic interaction in the structure formation and in stability.

1. I found that the Arg^{16} side-chain has more structure-stabilizing power than the Asp^{9} side-chain has since the Arg forms a salt-bridge and meanwhile protects the hydrophobic core.
2. The backbone relaxation studies showed the $Asp^{9} \rightarrow Asn^{9}$ mutaion does not affect significantly the dynamics of the miniprotien motion, on the other hand, the $Asp^{9} \rightarrow Ser^{9}$ side-chain modification gives large spacial freedom to the C-terminal part of the molecule.

The structural investigation of the Trp-cage variants where one of the serines has been replaced to glutamine, glutamate or phosphoserine helped to understand under what circumstances can the glutamate/phosphoserine mimicry be applied.

1. These studies revealed that the serine at position 14 is a structurally important site since both the oxygen and the hydrogen of its hydroxyl group form H-bonds and therefore, the serine $\rightarrow$ glutamine/glutamate mutation results in substantial destabilisation.
2. The mutation at position 13 is well-tolerated, both the S13Q and S13E variants have comparable stability as the original Trp-cage has. An unexpected result is that the $Ser^{13} \rightarrow Glu^{13}$ mutation creates an alternative conformer in which the $Trp^{6} - Pro^{12} - Pro^{18}$ side-chains are arranged in a twisted position while the other part of the molecule retains the Trp-cage structure.
3. I determined that the Ser^{20} of Trp-cage can be replaced with amino acids with longer side-chain without major structural destabilisation. In the presented instance, the Ser $\rightarrow$ Glu modification was even structure stabilising while on the other hand, the phosphoserine mutation resulted in a completely destabilised form. In conclusion I can state that in proteins with marginal structural sta-

bility the glutamate/phosphoserine mimicry is not applicable since both the length of the side-chain and their charge differ and these differences can result in opposite effects.

4. During the investigation of the phosphoserine variants I discovered a new, attractive, structure stabilising interaction which is is formed between a negatively charged carboxylic group (in this case a Glu^{20}) and a partially positively charged nodal plane of an aromatic ring (Tyr^{3}). In the future I will explore this anion-aromatic interaction in more details using quantumchemical calculations.

With the investigation of the N-terminal elongated variants I understood what is the connection between the structure, dynamics and the folding process of the Trp-cage miniproteins. These studies provided solid ground for the rational redesign of the antidiabetic drug, Exendin-4.

1. The N-terminal Asn $\rightarrow$ Arg mutation substantially destabilised the miniprotein, however with the further helix-elongation the stability of the protein is restored, the 25 amino acid long Trp-cage variant has comparable stability to the 20 residue long Tc5b.

2. I found that the C-terminal part of the α-helix is more stable and more rigid than its N-terminal part, since the former is protected by the PPII region of the molecule while the latter is a free α-helix.

3. The 39 amino acid long Exendin-4 analogue Trp-cage variant (E19) has remarkably similar structure and backbone dynamics as Exendin-4 has in 30% trifluoroethanol solution. By comparing E19 with Exendin-4 under the same solution I revealed that E19 is $20-30\%$ more helical than Exendin-4, and it is less prone to aggregation or oligomerisation but meanwhile they have comparable biological activity.

Using temperature-dependent NMR measurements I was able to characterise the complex folding pathway of Trp-cage.

1. I established that the folding pathway of Trp-cage can not be described with a simple two-state model which assumes that only the folded and the unfolded states populate the conformational space but the folding of Trp-cage – depending on the medium – involves one or even more intermediate states, both structured and unstructured ones.

2. For the analysis of the HSQC spectra I used a *Bezier* type curve fitting. With

this method I could precisely evaluate in what extent the temperature dependent change of the observed chemical shift differs from the „ideal" linear trend, i.e., how curved is the HSQC trend.

3. To quantify the the folding process I fitted the chemical shifts of the ^{13}C and ^{15}N HSQC spectra with appropriate thermodynamic equations. Using a non-linear iterative fitting approach I was able to derive both the thermodynamic parameters of the multi-step folding process ($\Delta H^{\text{F-I}}$, $T_m^{\text{F-I}}$, $\Delta C_p^{\text{F-I}}$ és $\Delta H^{\text{I-U}}$, $T_m^{\text{I-U}}$, $\Delta C_p^{\text{I-U}}$) both the NMR chemical shifts of the pure states.

4. From the chemical shifts of the pure states I was able to estimate the structure of the intermediate state and I determined that this state is a result of a $3_{10} \rightleftharpoons \alpha$-helix structural rearrangement in the $\text{Gly}^{10} - \text{Gly}^{15}$ segment.

5. With temperature-dependent measurements under acidic conditions I detected an alternative intermediate state (U') which is in slow-exchange with the major conformers (folded, intermediate and unfolded). This form is highly unstructured in which the $\text{Gly}^{11} - \text{Pro}^{12}$ peptid bond exists in both *cis* and *trans* conformation.

6. I repeated the temperature-dependent measurements with the salt-bridge variants and with the N-terminal elongated variants and I found that the multi-step folding process is common among the Trp-cage variants. With the curvature analyis I determined that the intermediate state population is higher in the samples in which an operative salt-bridge stabilise the native structure. I concluded that the intermediate state is an on-pathway intermediate, that is it situated between the folded and unfolded state in the folding funnel.

Conclusions

Although Trp-cage consists of only 20 residues and represents the smallest and simplest protein model ever investigated I found that nothing is simple neither for a model protein. The structural analysis revealed that a dense, well-balanced non-covalent, weak interaction network stabilises the miniprotein's structure; H-bonds, salt-bridges, dispersion forces act together to maintain the three-dimensional fold. A small perturbation to any of these interaction has dramatic effect on the miniprotein's stability. With the expansion of computer-aided chemistry it becomes feasible to calculate the interaction energies of even larger proteins which knowledge facilitates the interpretation of the *in vitro* results and provides a bases for future protein design.

Another important aspect what I encountered with is the fact that the folding/unfolding pathway of Trp-cage is more complex than expected for such a small model system. During the folding process structured and unstructured intermediates exchange with the major, folded state at different time-scales and all possess substantially different dynamic properties. Therefore, the question arises if the structure-stabilising interactions, the folding process and the time-scale of atomic motions of the simplest model system are such complicated then what shall we expect for a real protein consisting of hundreds of amino acid residues? To what extend can we apply the assumed simplifications (e.g., two-state folding model,or separable time-scale of motions) to describe the proteins' behavior? Much should be learned about the factors that influences the interactions which define the proteins' structure and internal dynamics and governs their folding processes. Hence, Trp-cage miniprotein will be excellent testing ground both for future experimental and computational studies.

The ability to detect differences in the atomic-scale structure, backbone dynamics and multi-step folding route of model peptides provides an excellent opportunity to gain new insights into the energetic aspects of the energy landscape of protein folding. Our new experimental data offer exceptional testing ground for further computational simulations.

Publications

Publications emerging from graduate work.

1. **Structural Insights into the Trp-Cage Folding Intermediate Formation**
   Petra Rovó, Pál Stráner, András Láng, István Bartha, Kristóf Huszár, László Nyitray and András Perczel
   Chem. Eur. J., **2013**, 19: 2628-2640. doi: 10.1002/chem.201203764, **IF**: 5.831

2. **Cooperativity network of Trp-cage miniproteins: Probing salt-bridges**
   Petra Rovó, Viktor Farkas, Orsolya Hegyi, Orsolya Szolomájer-Csikós, Gábor K. Tóth and András Perczel
   J. Peptid Sci., **2011**, 17: 610-619 doi: 10.1002/psc.1377, **IF**: 1.799

3. **Rational design of α-helix stabilized Exendin-4 analogues**
   Petra Rovó, Viktor Farkas, Pál Stráner, Ágnes Jermendy, Orsolya Hegyi, Gábor Tóth, András Perczel
   Biochemistry, **2014**, submitted manuscript

4. **Phosphorylation as conformational switch to amyloid state: Trp-cage as protein aggregation model**
   József Kardos, Bence Kiss, András Micsonai, János Kovács, Petra Rovó, Györgyi Váradi, Gábor Tóth, András Perczel
   under submission