Retro operation on the Trp-cage miniprotein sequence produces an unstructured molecule capable of folding similar to the original only upon 2,2,2-trifluoroethanol addition

Jiří Vymětal1†, Sreenivas Reddy Bathula2,3, Jiří Černý4, Radka Chaloupková5, Lukáš Židek1-3, Vladimír Sklenár2,3 and Jiří Vondrášek1,6

1Academy of Sciences of the Czech Republic, Center for Biomolecules and Complex Molecular Systems, Institute of Organic Chemistry and Biochemistry, Flemingovo nám. 2, 16610 Prague 6, Czech Republic, 2National Center for Biomolecular Research, Faculty of Science, Masaryk University, Kamenice 5, 625 00 Brno, Czech Republic, 3CEITEC, Masaryk University, Kamenice 5, 625 00 Brno, Czech Republic, 4Institute of Biotechnology, Academy of Sciences of the Czech Republic, Videnska 1043, 142 20 Prague 4, Czech Republic and 5Loschmidt Laboratories, Department of Experimental Biology and Research Centre for Toxic Compounds in the Environment RECETOX, Faculty of Science, Masaryk University, Kamenice 5/A13, 625 00 Brno, Czech Republic

†To whom correspondence should be addressed.
E-mail: jiri.vondrasek@uochb.cas.cz

Received August 13, 2013; revised August 31, 2014; accepted September 12, 2014

Edited by Niels Andersen

Amino acid sequence and environment are the most important factors determining the structure, stability and dynamics of proteins. To evaluate their roles in the process of folding, we studied a retroversion of the well-described Trp-cage miniprotein in water and 2,2,2-trifluoroethanol (TFE) solution. We show, by circular dichroism spectroscopy and nuclear magnetic resonance (NMR) measurement, that the molecule has no stable structure under conditions in which the Trp-cage is folded. A detectable stable structure of the retro Trp-cage, with the architecture similar to that of the original Trp-cage, is established only upon addition of TFE to 30% of the total solvent volume. The retro Trp-cage structure shows a completely different pattern of stabilizing contacts between amino acid residues, involving the guanidinium group of arginine and the aromatic group of tryptophan. The commonly used online prediction methods for protein and peptide structures Robetta and PEP-FOLD failed to predict that the retro Trp-cage is unstructured under default prediction conditions. On the other hand, both methods provided structures with a fold similar to those of the experimentally determined NMR structure in water/TFE but with different contacts between amino acids.

Keywords: protein folding/protein-structure prediction/molecular dynamics/NMR methods/CD spectroscopy

Introduction

An enormous number of protein-folding scenarios have been evaluated, and this work has greatly increased our ability to predict protein structures and dynamics (Chan and Dill, 1993, 1998; Daggett and Fersht, 2003; Dill et al., 2008; Fersht, 2008; Daggett and Fersht, 2009; Bowman and Pande, 2010). Sosnick and Barrick recently reviewed many of the most intriguing problems regarding the folding processes of single-domain proteins (Sosnick and Barrick, 2010). Our knowledge about the properties of the denatured-state ensemble, the early events of folding, the detailed folding pathway, intermediates and desolvation remains incomplete, because an experimental setup capable of capturing the details of folding processes at an atomistic level is too difficult to implement (Ferguson and Fersht, 2003; Vendruscolo and Paci, 2003; Vendruscolo et al., 2003; Fersht, 2008; Vendruscolo, 2009).

Several single-domain miniproteins provide valuable information about the self-organization process and its thermodynamics and structural background. One of the best-known examples is the Trp-cage miniprotein, a 20-amino-acid construct that can fold spontaneously in <4 μs (Neidigh et al., 2002). Many groups have conducted experimental and theoretical studies on Trp-cage dynamics and folding pathway (Hu et al., 2008; Day et al., 2010; Halabis et al., 2012), its thermal stability (Barua et al., 2008; Hudak et al., 2008; Williams et al., 2011) and denaturation behavior. This includes mutation studies and thermodynamic studies (Garcia et al., 2009; Gattin et al., 2009; Heyda et al., 2011; Wu et al., 2011). Halabis et al. studied Trp cage behavior close to its melting temperature and uncovered an interesting feature of the miniprotein’s stability and dynamics (Halabis et al., 2012). They reported that most short-range contacts disappeared, while a ‘long-range interaction such as Trp6-Arg16 remains until 313K.’

As has been observed both experimentally and theoretically (Hu et al., 2008; Paschek et al., 2011; Wu et al., 2012), Trp-cage folding is cooperative and is driven by the hydrophobic effect, burying Trp6 into the protein interior. The protein is subsequently stabilized by interactions with a stretch of three proline residues (Pro17-Pro18-Pro19) (Biedermannova et al., 2008). The presence of a single well-defined secondary structure element (an α-helix) gives rise to the question of whether the reverse (retro) amino acid sequence of Trp-cage is able to form a stable 3D structure with native contacts similar to those of the original.

Retro transformation of a protein sequence, i.e. a change in sequence direction from N→C to C→N, is not a new idea (Ruvo and Fassina, 1996; Mittl et al., 2000). Retro transformation can lead either to a complete loss of folding ability and 3D structure (Olszewski et al., 1996; Lacroix et al., 1998) or to a change resulting in a stable, folded molecule with the same or a similar topology (Haack et al., 1997; Witte et al., 1998;
Pan et al., 1999). Most studies revealing stable folding of retro proteins have been performed with helical proteins.

Here, we use a retro Trp-cage model to test the mechanism of protein-folding pathways and evaluate the role of the key amino acid side chains responsible for Trp-cage structure formation and its compactness (e.g., Trp6, Tyr3, Pro17–19 and the salt bridge between Asp9 and Arg16). We also analyzed whether the overall α-helical fold, as well as the topology of native contacts, is preserved in the retro Trp-cage.

According to the literature, the structural stability of the original Trp-Cage sequence (and its variants) is strongly coupled with the stability of the helical N-terminal part. As Barua et al. showed (Barua et al., 2008), mutations in the N-terminal part, which stabilize or destabilize the N-terminal α-helix, determine the stability of the overall constructs. Hence, questions remain regarding how the retro Trp-cage stability is coupled with the stability of the expected C-terminal helix and how the stability of the C-terminal helix in water differs from its stability upon addition of 2,2,2-trifluoroethanol (TFE), which is known to induce helix formation.

We have made an interesting observation regarding the helix formation propensity for the corresponding amino acids of the Trp cage and its retro variant. We used AGADIR algorithm (Muñoz and Serano, 1994; Lacroix et al., 1998) on the isolated NLYIQWLKD (Trp cage) versus the DKLWQIYLN (retro Trp cage) sequences as well as on the whole length proteins. Obtained helical propensities for the corresponding amino acids in the Trp cage were slightly higher than those in the retro Trp cage (for the isolated NLYIQWLKD motif as well as for the whole protein sequence). On the contrary, helical propensities for the same amino acids were observed higher for the retro Trp cage sequence if the N-terminal was acetylated and the C-terminal methylated (isolated motif as well as for the whole protein sequence). The setup better mimicking our experimental conditions—the setup better mimicking our experimental conditions (see Supplementary information—AGADIR prediction of helical propensities of Trp cage and retro Trp cage amino acids).

Interestingly, prediction of the secondary structure content run by several algorithms on both proteins did not confirm the presence of the expected short helix neither in the Trp cage nor in its retro variant (see Supplementary Fig. S1).

We evaluated the structure of the retro Trp-cage and analyzed its properties both theoretically and experimentally. Far-ultraviolet (UV) CD spectrometry measurements were made to determine the presence of secondary structure elements in the construct. Detailed nuclear magnetic resonance (NMR) measurements of the protein in water and 30% TFE served as a tool to map the conformational properties of the miniprotein. We utilized two commonly used, freely available prediction servers—Robetta (Kim et al., 2004) and PEP-FOLD (Maupeut et al., 2009; Maupeut et al., 2010; Thévenet et al., 2012)—to predict the structure of the retro Trp-cage sequence, which has no known homolog. Independent equilibrium molecular dynamics simulations were performed to compare the dynamics/stability of the miniprotein in both environments to gain insight into retro Trp-cage stability.

**Methods**

**Synthesis of retro Trp-cage**

Retro Trp-cage (amino acid sequence SPPRGSSPGGDKLWQIYLN) was synthesized by Fmoc solid-phase chemistry in an ABI 433A Peptide Synthesizer (Applied Biosystems) in the form of peptidyl amides with acetylated N-termini. Peptides were purified using an RP-HPLC (Waters 600; column: Waters Symmetry 300, 19 × 150 mm, 5 μm) using a water/acetonitrile elution system containing 0.1% trifluoroacetic acid and characterized by ESI mass spectrometry on an LCQ Classic Finnigan Mat device (Thermo Finnigan).

**CD spectroscopy of retro Trp-cage**

Circular dichroism (CD) spectra of 0.07 mM (0.15 mg/ml) retro Trp-cage were recorded at 2°C using Jasco J-810 (Jasco, Japan) and Chirascan CD (Applied Photophysics, UK) spectrometers equipped with Peltier cell holders. The data were collected from 185 to 260 nm, at 100 nm/min, 1 s response time and 2 nm bandwidth using a 0.1-cm quartz cuvette. Each spectrum shown in Fig. 1 is the average of 10 individual scans corrected for the buffer absorbance. The CD data collected were expressed in terms of the mean residue ellipticity (Θ_{MRE}) using the equation:

\[
Θ_{MRE} = \frac{Θ_{obs}M_w100}{ncl},
\]

where Θ_{obs} is the observed ellipticity in degrees, M_w is the protein molecular weight, n is the number of residues, l is the cell-path length, c is the protein concentration and the factor 100 originates from the conversion of the molecular weight to mg/dmol. The secondary structure content was calculated using the CDSSSTR (Compton and Johnson, 1986; Manavalan and Johnson, 1987; Sreerama et al., 2000) and CONTIN (Provencher and Glockner, 1981; Vanstokkum et al., 1990) methods implemented on the DichroWeb server (Lobley et al., 2002; Whitmore and Wallace, 2004; Whitmore and Wallace, 2008). Singular value decomposition in the Octave program (J. Eaton, University of Wisconsin, Madison, USA) and the component spectra reported by Matsuo et al. (Matsuo et al., 2007) were used to estimate the α-helical content in the presence of polyproline II.

Thermal unfolding of retro Trp-cage in the presence and absence of 30% TFE was followed by monitoring the CD spectra from 185 to 250 nm and the ellipticity at 207 nm.

![](http://www.dx.doi.org/10.1093/peds/27.12.463)
where the largest changes of the CD signal were observed over a temperature range of 2–80 °C, with a resolution of 0.1 °C, at a heating rate of 1 °C/min. The thermal denaturation curves recorded (see Supplementary Fig. S2) were normalized to represent signal changes between ~1 and 0 and fitted to the sigmoidal curves described by the equation $A + BT + (A - C + (B - D)T)/(1 + exp((T - T_0)/T))$ using Octave software.

**NMR experiments**

For NMR measurements, 600 μl of 1 mM retro-Trp-cage in 9 mM phosphate buffer, pH 7.0 (uncorrected reading), containing 30% TFE, 10% deuterium oxide and 0.03% NaN$_3$, was prepared from 1.3 mg of lyophilized peptide. The samples were temperature-refolded before each measurement.

NOESY, TOCSY, COSY and $^{13}$C,$^1$H HSQC spectra were measured at 2 °C on a 500 MHz Bruker NMR spectrometer equipped with a $^1$H/$^{13}$C/$^{15}$N TXI probe head. All spectra were processed using NMRPipe software (Delaglio et al., 1995). A TOCSY spectrum recorded with a mixing time of 50 ms and a COSY spectrum were used to determine the sequential connectivity in combination with a NOESY spectrum acquired with a 200-ms mixing time. The latter spectrum also provided distance restraints for structure calculation.

The Sparky program (T.D. Goddard and J. Kneller, University of California, San Francisco) was used for peak picking and assignment, ARIA version 2.1 (Habeck et al., 2004) for Nuclear overhauser effect (NOE) assignment and CNS version 1.2 for structure calculations (Brunger et al., 1998). The CING program was used to validate the calculated NMR structures, which were deposited in the protein data bank (PDB) under accession code 2LUF.

**Robetta server modeling**

The full retro-Trp-cage sequence was subjected to BLAST against the sequences of the PDB (www.rcsb.org) and UNIPROT (www.uniprot.org) databases. We did not find any homologs of the retro-Trp-cage in the sets analyzed. We utilized the Robetta (Kim et al., 2004) server (http://robetta.bakerlab.org/), which provides both ab initio and comparative models of protein domains and uses the ROSETTA fragment-insertion method (Simons et al., 1997).

**PEP-FOLD modeling**

PEP-FOLD is a *de novo* approach aimed at predicting peptide structures from amino acid sequences (Mauget et al., 2009). We submitted the full length of the retro-Trp-cage to the online PEP-FOLD prediction server (http://bioserv.rpbs.univ-paris-diderot.fr/services/PEPFOLD/) with default options and obtained six lowest conformations identified as the most representative ones. We selected one of them satisfying criteria of expected fold and length of the helix.

**Molecular dynamics simulations**

**Simulation protocol.** All simulations were performed in the Gromacs Molecular Dynamics Package, version 4.5.3 (Berendsen et al., 1995; Lindahl et al., 2001). The NMR model was used as a starting structure for all simulations. Parameters were assigned by the pdb2gmx tool using the corresponding Gromacs parameter library. The experimental structures were placed in a 39 × 39 × 33 Å rectangular box and solvated with water (modeled by the TIP3P explicit water potential) or mixed solvent (30% aqueous TFE). To remove close contacts, short geometry optimization was conducted by the steepest descent algorithm. The solvent molecule positions were further optimized in equilibration simulations of 10-ns and 20-ns solvent dynamics in neat water and TFE solution, respectively. The structure of retro-Trp-cage was constrained during optimization and equilibration by means of a position restraint with a force constant of 1000 kJ mol$^{-1}$ nm$^{-2}$ placed on all non-hydrogen atoms. The data were collected from production simulations, each conducted for 200 ns.

**Simulation details.** The equations of motions were integrated with a time step of 2 fs. All bond lengths were constrained by the LINCS algorithm. Water molecule geometry was constrained by the SETTLE algorithm. The temperature (275K) was maintained by a v-rescale thermostat (Bussi et al., 2007) with a coupling constant of 1.0 ps. Berendsen’s weak coupling barostat (Berendsen et al., 1984) was used for pressure control with a coupling constant of 1.0 ps. Compressibility used in the barostat is compressibility of pure water, i.e. $4.5\times10^{-10}$ m$^2$ N$^{-1}$. Electrostatic interactions were treated by the Particle Mesh-Ewald algorithm (Darden et al., 1993) with a real-space cutoff of 10 Å and an fast fourier transform grid spacing of 10 Å. Non-bonded Lennard-Jones interactions were evaluated with a simple cutoff of 14 Å accompanied by a long-range correction to the energy and pressure tensor. Long-range corrections to energy and pressure tensor compensate for missing contribution of dispersion due to the finite cut-off. The dispersion interaction between particles beyond cutoff distance can be in Gromacs approximated assuming homogeneity of the system. No special treatment of water and TFE was attempted (See Tables SI–SV in the Supplementary material). In all cases, the value of correction to the total energy was almost constant in course of simulation with a magnitude of ~100 kJ mol$^{-1}$ for the whole system (Allen and Tildesley, 1997).

**List of simulations.** The NMR structure of retro-Trp-cage was simulated in two different force fields—Amber FF99SB-ILDN (Hornak et al., 2006; Lindorff-Larsen et al., 2010) and Amber FF03 (Duan et al., 2003) integrated in Gromacs. Four independent simulations were performed for each force field, using neat water or TFE solution as the solvent.

**Analysis of the trajectories.** All analyses were performed with tools from the Gromacs MD package. The secondary structure elements were identified by the DSSP algorithm.

**Results**

**CD spectroscopy of retro-Trp-cage**

CD spectra of the retro-Trp-cage in aqueous buffer (Fig. 1) revealed an unfolded structure, characterized by a single negative peak at 203 nm. The deconvoluted spectra showed 6% $\alpha$-helical content, 29% $\beta$-sheet, 17% turn and 48% disordered structures. Spectral features typical of a helical motif, characterized by two negative peaks at ~207 and 221 nm and one positive peak at ~192 nm, appeared only in the 30% TFE solution (Fig. 1), confirming the previously reported ability of TFE to induce $\alpha$-helix formation (Jaravine et al., 2001).

While qualitative interpretation of the observed changes in the CD spectra is clear, a quantitative analysis is less straightforward. The retro-Trp-cage contains a stretch of prolines. The DichroWeb server (Lobley et al., 2002; Whitmore and Wallace,
2004; Whitmore and Wallace, 2008) offers prediction of secondary structure motifs including the polyproline II conformation, but the CD spectra must cover the vacuum ultraviolet region (Matsuo et al., 2007). As we did not perform such measurements, our estimated α-helical content is likely biased by neglecting the polyproline II contribution. No further changes in the secondary structure of retro Trp-cage were observed when the TFE concentration was increased to 50% (Fig. 1).

Thermally induced denaturation of retro Trp-cage in the presence and absence of 30% TFE was tested to explore the miniprotein’s structural stability and the reversibility of its refolding. No significant changes in the CD spectra of retro Trp-cage in the absence of TFE were observed with increasing temperature, confirming a disordered conformation. In contrast, the CD spectra of the retro miniprotein in the presence of TFE changed with increasing temperature. The observed changes correspond to the melting curves, with an average transition temperature of 25.5°C (Supplementary Fig. S2). Following thermal denaturation, the refolded protein in 30% TFE showed a CD spectrum with shape and intensity similar to that of the protein before heating.

**NMR structure determination of retro Trp-cage**

After we identified conditions for folding of retro Trp-cage, we used NMR spectroscopy for further, detailed structural characterization of the miniprotein. First, folding and stability of the retro Trp-cage samples were examined. Quick precipitation occurred in 30% aqueous TFE at high peptide concentration, but a 1 mM sample survived for 3 days at low temperatures with no detectable signs of aggregation/association. However, no signal was observed in NMR spectra collected after 7 weeks, indicating major changes in the molecular ensemble (data not shown). Therefore, we used fresh 1 mM samples of the miniprotein in 30% TFE in our NMR experiments. Proton-resonance frequencies were assigned based on 2D NOESY and TOCSY spectra (Wüthrich, 1995). Only one set of peaks was observed in the spectra, indicating that the majority of the miniprotein is present in one folded form. This agrees with our CD spectroscopy results showing that the helical content did not change when the TFE concentration was increased from 30 to 50%. The spin systems of individual amino acids were identified in 2D TOCSY and COSY spectra. Sequential assignment was based on the observed short-range £H i-£H i+1 NOESY cross-

**Figure 2.** Region of the NOESY spectrum showing the cross-peaks of alpha and amide protons. The intraresidual cross-peaks are labeled with residue numbers. Sequential connectivities are indicated with the solid line, while medium-range cross-peaks, typical of alpha-helices, are connected by dashed lines.
peaks and confirmed by sequential NOEs between side chain protons. The assigned chemical shifts were deposited in the BioMagResBank (http://www.bmrb.wisc.edu) under accession code 18519. Medium-range (Fig. 2) and \(\text{H}^\text{N}-\text{H}_{\text{N}+1}\) (Supplementary Fig. S3) NOEs provided evidence that residues D12–L19 form an \(\alpha\)-helical structure. In agreement with the CD measurements, the \(\text{H}^\text{N}-\text{H}_{\text{N}+1}\) cross-peaks were very weak or lacking in the 2D spectrum of the miniprotein in the absence of TFE, whereas the intensities of the cross-peaks of aromatic and side chain amide protons were comparable at 0 and 30% TFE (Supplementary Figs S3 and S4). This indicates that the \(\alpha\)-helix was formed only in the presence of TFE.

Due to the limited stability of folded retro Trp-cage, it was important to rule out the possibility that our structure is biased by the presence of different protein conformations. The single set of signals observed in NMR spectra indicates that the sample does not contain a significant amount of other species in slow exchange with the major form of the miniprotein. To test whether the observed NMR data represent an average of several rapidly interconverting conformations, we analyzed the chemical shifts. Prediction of secondary-structure propensities by the SSP program (Marsh et al., 2006) based on \(\text{^{13}C}_a\) and \(\text{^{13}C}_\beta\) chemical shifts provided values ranging between 0.69 and 1.02 for residues 11–19 (Supplementary Fig. S5A). Such values are typical of well-ordered domains, indicating that the helical form of retro Trp-cage is the major conformation in 30% TFE. The \(\text{d}_2\)D program (Camilloni et al., 2012) provided similar values (Supplementary Fig. S5C). The somewhat lower propensity calculated by the SSP program from \(\text{^{13}C}_a\), \(\text{^{13}C}_\beta\) and \(\text{^4H}_\alpha\) chemical shifts (Supplementary Fig. S5B) may reflect the higher sensitivity of the \(\text{^4H}_\alpha\) frequencies to the experimental conditions, which were not precisely accounted for by the referencing. Comparing side chain chemical shifts with their average values in the BMRB database (Supplementary Table SVII) revealed a significant up-field shift of Arg5 side chain protons, suggesting that this side chain is likely present next to the face of the aromatic rings of Trp15 or Tyr18.

When the resonance assignment was completed, all NOESY cross-peaks were picked and assigned when possible. The intensities of unambiguously assigned NOESY cross-peaks corresponding to protons of known distances (the aromatic protons of Trp15 and Tyr18 and the side chain protons of Pro2, Pro3 and

---

**Figure 3.** 3D structure of retro Trp-cage determined by NMR. The packing of the miniprotein core is schematically shown in (A); the set of NMR structures is shown in (B).

**Figure 4.** PEP-FOLD (A) and Robetta (B) models of the retro Trp-cage and the experimental structure of the Trp-cage (C; PDB 1L2Y). The residues forming the interior of the miniprotein are shown in stick representation.
Pro4 and Gln16) were used to calibrate the peak intensities. As a result, a set of 299 distance restraints was obtained. NOEs were assigned with ARIA 2.1, using a set of 11 unambiguously assigned long-range cross-peaks. The obtained distance list was used as a restraint set for structure calculation by running restrained molecular dynamics in CNS 1.2, utilizing the standard protocol and input scripts written by the authors of the RECOORD database. A set of 300 structures was calculated in vacuo, and the 100 structures with the lowest energy were further refined in an explicit water solvent. All calculated structures adopted the same fold (Supplementary Fig. S6). The subset of the 10 structures with the lowest energy was validated using the CING program and deposited in the PDB under accession code 2LUF. The main structural features are shown in Fig. 3, and the structure determination statistics are summarized in the Supplementary material (Supplementary Table SVI).

The C-terminus forms a helix (Gly11–Gln16) that approximately corresponds to the N-terminal Trp-cage helix. The overall RMSD for the superposition with the Trp-cage structure is ∼3.3 Å. Most of the overall structural similarity stems from the corresponding helices, while the C- or N-termini in Trp-cage and retro Trp-cage, respectively, differ significantly. More importantly, the amino acid side chains forming protein cores in both molecules are quite different. There is no proline–tryptophan motif in the retro Trp-cage structure (Fig. 3) that would stabilize the Trp residue. In the Trp-cage, this stabilization is enhanced by interaction with Tyr3. A novel stacking interaction between Trp15 and Arg5 emerges in the retro Trp-cage, and stabilization can be assigned to the charged Arg guanidinium group and the aromatic heterocycle of Trp. These interactions are known to contribute significantly to the overall protein stability (Flocco and Mowbray, 1994; Tsou et al., 2002).

### Full-atom model determined by Robetta and PEP-FOLD prediction servers

The performance of Robetta and PEP-FOLD prediction servers was evaluated by prediction-time comparison and by superposition of the models and experimental structures: Trp-cage (1L2Y) or retro Trp-cage (2LUF) (Fig. 4). Whereas PEP-FOLD provided results within a few hours, the first run on the Robetta server took more than 2 weeks. Both PEP-FOLD and Robetta successfully predicted the C-terminal helix and an overall fold similar to that of the original Trp-cage (best C_r RMSD: 2.3 and 3.9 Å, respectively) and to the experimental retro Trp-cage structure (best C_r RMSD: 2.1 and 4.0 Å, respectively). None of the models possessed correct packing of the protein core. The panel in Fig. 4 clearly demonstrates preservation of the terminal helix in three structures—the Robetta model, the PEP-FOLD model and the experimental structure of Trp-cage. The experimental Trp-cage structure (C) is compact, and the fold-stabilizing interactions occur almost exclusively between the Trp residue and three Pro residues and one Tyr. None of the predicted models of retro Trp-cage shows a similar arrangement of these amino acids. The PEP-FOLD model includes only one interaction of Trp with the guanidinium group of Arg5 (A), similar to the NMR determined structure, whereas Robetta predicted a non-canonical interaction of Trp with three terminal prolines (B).

Most of the differences between the two models can be assigned to the N-terminus and its interactions, which likely reflects the different preferences for non-covalent (Trp...Pro-Pro-Pro) interactions in both methods.
MD simulations of the retro Trp-cage in different force fields

Characteristics of the retro Trp-cage structure during MD simulations are shown in Tables I and II. The chosen features utilized measures corresponding to the molecule as a whole (RMSD of the Cα atoms, the radius of gyration—Rg—and solvent accessible surface—SAS) or to a particular structural motif (helix and Trp-Arg contact), providing a more comprehensive picture of the molecule and its behavior during the simulation.

All simulations employing both force fields showed a fast deviation from the starting NMR structure (2LUF) in terms of Cα RMSD (Table I). Deviations below 2 Å can be considered the result of fluctuation around the experimental structure or non-critical inaccuracies of the force field. Higher RMSD values were connected with alteration in the structure of the N-terminal tail or core repacking. RMSD values exceeding 5 Å were observed for structures with a greatly loosened or unfolded C-terminal helix.

To completely analyse the MD simulations and characterize the NMR structure of retro Trp-cage (Fig. 3) and its behavior in all aspects possible, we decided to couple two features—the overall Cα RMSD and the distance between Arg and Trp—the core-forming residue located in the experimental structure (Fig. 5).

As shown in Fig. 5, there is no significant correlation between the Cα RMSD and Trp-Arg contact. It is not clear whether the differences resulted from insufficient sampling or actual differences between the force fields.

Discussion

The 20-amino-acid Trp-cage miniprotein, which is among the fastest-folding molecules, has been studied in detail experimentally and theoretically. Here, we studied its retro variant to shed light on the contributions of particular side chains, their interactions and corresponding backbone conformational states to the stability and fold of the molecule.

Both CD and NMR experiments show unambiguously that under normal conditions there is no stable conformational state of the retro Trp-cage. A well-defined helical structure was observed only in the presence of 30% TFE at low temperature. Because we used a chemically synthesized sample with natural isotope abundance, only 2D-NOESY spectra with limited resolution could be employed to derive structural data. Most of the
NOESY cross-peaks were assigned by an iterative structure calculation/peak assignment procedure using the ARIA program. However, the most important structural features were observed directly from the spectrum, namely the well-characterized pattern of cross-peaks defining the alpha helix (Fig. 2) and the unambiguously assigned cross-peaks corresponding to the short distances between Arg5 and Trp15 (Fig. 6). On the other hand, no cross-peaks were observed in the region of the spectrum where correlations between Trp15 and Pro2, Pro3 and Pro4 would appear. These observations confirm that the general fold of our retro Trp-cage is not an artifact of the restrained molecular dynamics used for structure calculation.

While interactions between tryptophan and prolines are crucial for Trp-cage folding and stability (Bendova-Biedermannova et al., 2008; Biedermannova et al., 2008), our data show that the cation–π interaction (Wu et al., 2012) between tryptophan and arginine dominantly stabilizes the retro Trp-cage structure in 30% TFE. These guanidinium-aromatic interactions are mostly stacking (Heyda et al., 2011; Vazdar et al., 2011). While there is a general preference for inter-planar contacts with the five-atom ring of tryptophan, arginine prefers a six-atom ring (Ma and Dougherty, 1997), as observed in our structures. Tsou et al. (2002) showed that the interaction between tryptophan and arginine has a stronger stabilizing effect on α-helices than salt bridges. Planar stacking interactions between arginine and aromatic side chains are important stabilizing contributions to protein stability (Flocco and Mowbray, 1994). Interactions with prolines likely prevent tryptophan from making cation–π interactions with arginine in the original Trp-cage structure.

The theoretical online predictors (Robetta and PEP-FOLD) provided semi-folded retro Trp-cage structures in aqueous solution, which does not correspond to our experimental findings. Recent work (Das, 2011) has shown that the most commonly used prediction methods require better parameterization for non-covalent interactions. Such parameterization would make it possible to reach atomistic accuracy and should be based on values obtained on model systems by high-level ab initio methods (Klusak et al., 2003; Řezáč et al., 2008; Berka et al., 2009; Vymetal and Vondrasek, 2011; Hamada, 2013).

We show that the equilibrium MD simulations differ depending on the type of force field utilized. This is not very surprising if one takes into account a recent study showing that ‘the folding mechanism and the properties of the unfolded state depend substantially on the choice of force field’ (Piana et al., 2011). Our analysis shows that selection of the force field in an MD simulation strongly influences interpretation of molecular behavior, especially in cases where a stable minimum of the molecule is not reported.

Our study revealed non-covalent interactions between amino acid side chains that were not present in the original Trp-cage molecule. The structural motif between Arg and Trp residues in the retro Trp-cage (Fig. 3) is an energy-rich interaction that could compete with the Trp...Pro-Pro-Pro or Trp...Tyr interactions present in the Trp-cage. We hypothesize that either the interaction energy sum of the latter non-covalent interactions is not higher than the Arg-Trp interaction or that the retro Trp-cage backbone does not allow to adopt a conformational state favoring these interactions which further stabilize helix and are necessary for proper folding. Our observations together with the predicted helical propensities and secondary structure predictions indicate that folding of the Trp-cage as well as its retro variant is a cooperative event and the Trp-burial is required to produce or significantly stabilize the helical segment in the folded molecules.

**Conclusions**

The amino acid sequence of proteins is a major factor determining their folding, dynamics and structural stability. Here, we show that retro operation on the Trp-cage sequence renders the existence of a stable 3D structure in aqueous solution impossible, even though the sequence composition and order of the original Trp-cage molecule are retained. On the other hand, in 30% TFE, the retro Trp-cage forms a relatively stable and defined structure with a new motif of core stabilization and overall fold similar to that of Trp-cage. Our results confirm that there is a direct
coupling between the stability of the C-terminal helix and the stability of the retro Trp-cage molecule. The Robetta and PEP-FOLD prediction servers provided results that did not reflect the environmental influences. In addition, the servers differed in prediction of native contacts and stabilization elements in the molecule. Our simulations in two commonly used force fields in water and 30% TFE unambiguously indicated that force field selection is crucial for a realistic description of the molecule, although additional development is needed.

The transition between an unfolded and folded state of retro Trp-cage caused by environmental changes indicates that a sequence might latently contain ‘a structure’ that forms upon a change in the environment.

Supplementary data
Supplementary data are available at PEDS online.

Acknowledgements
This work was supported by the Ministry of Education, Youth and Sports of the Czech Republic (Grant LH1 1020 and LO1214) and by the Czech Science Foundation (Grant 203/08/0114). Access to the computing and storage facilities owned by the parties and projects contributing to the National Grid Infrastructure MetaCentrum, provided under the program ‘Projects of Large Infrastructure for Research, Development, and Innovations’ (LM2010005), is gratefully acknowledged. The work at the Academy of Sciences was supported via Project Z40550506 and the work at Masaryk University was supported via Project Z40550506 and the work at Masaryk University was supported via Grant number CZ.1.05/1.1.00/02/0068. Niels Andersen, Marius Clore and Lars Baltzer have been selected as the PEDS Board Members whose interests lie closest to the subject of the paper.

References