Crystal and NMR structures of a Trp-cage mini-protein benchmark for computational fold prediction

Michele Scian^{a,1}, Jasper C. Lin^a, Isolde Le Trong^{b,c}, George I. Makhatadze^d, Ronald E. Stenkamp^{b,c,e,1}, and Niels H. Andersen^{a,1}

^aDepartment of Chemistry, University of Washington, Seattle, WA 98195; ^bDepartment of Biological Structure, University of Washington, Seattle, WA 98195; ^cBiomolecular Structure Center, University of Washington, Seattle, WA 98195; ^dDepartment of Biology, Rensselaer Polytechnic Institute, Troy, NY 12180; and ^eDepartment of Biochemistry, University of Washington, Seattle, WA 98195

Edited by Alan R. Fersht, Medical Research Council Laboratory of Molecular Biology, Cambridge, United Kingdom, and approved June 25, 2012 (received for review December 26, 2011)

To provide high-resolution X-ray crystallographic structures of a peptide with the Trp-cage fold, we prepared a cyclized version of this motif. Cyclized Trp-cage is remarkably stable and afforded two crystal forms suitable for X-ray diffraction. The resulting higher resolution crystal structures validate the prior NMR models and provide explanations for experimental observations that could not be rationalized by NMR structural data, including the structural basis for the increase in fold stability associated with motif cyclization and the manner in which a polar serine side chain is accommodated in the hydrophobic interior. A hexameric oligomer of the cyclic peptide is found in both crystal forms and indicates that under appropriate conditions, this minimized system may also serve as a model for protein-protein interactions.

protein-protein interaction | folding simulation target | protein cyclization | peptide oligomerization | H-bonding motifs

Trp-cage miniproteins (at 18–20 residue length) are among the smallest peptides to fold into a stable protein-like structure (1, 2); as a result, the Trp-cage fold has emerged as a protein folding paradigm (3, 4). Trp-cage species have been extensively studied experimentally (4–11) and computationally (12–19), but the conclusions regarding the folding process and results drawn from these studies are not in accord at the level that should be attainable for such a small system.

The structural model used for testing computational folding methods has, in almost all cases, been that derived by NMR for the initially reported Trp-cage (TC5b), which displays a rather low fold stability, $T_m = 43^\circ\text{C}$ (1). Additional Trp-cage structures based on NMR measurements are now available (2, 6, 10, 20–22). All of these appear to possess a common fold topology that is supported by an extensive set of structuring-induced shifts (CSDs, chemical shift deviations from random coil norms) that are largely the result of ring current effects. The uniformity of these CSDs suggests a well-defined common structure. However, other spectroscopic measurements of Trp-cage molecules have been interpreted as being consistent with a folded structure undergoing a greater degree of dynamic fluctuation (7, 8). In addition, “folded” structures from molecular dynamics simulations, with a few notable exceptions (18, 19), fail to predict the ring current effects observed by NMR (2). The failure of numerous calculations to reproduce the NMR structures and contrary conclusions from other experimental methods is problematic and points to outstanding issues concerning the continuing* use of a Trp-cage as a benchmark for molecular dynamics folding simulations. Further structural studies of Trp-cage species are required to validate the target and to ascertain the key interatomic interactions leading to the folded state.

Herein we report X-ray structure determinations for two crystal forms of a cyclized Trp-cage (23), cyclo-TC1 (-GDAYAQLADGGPSSGRPPPSG-). Raising additional questions about the oligomeric state of the molecule, a Trp-cage hexamer is found in both crystal forms. However, the single chain structures observed in the solid-state (nine unique molecular structures in two

crystal forms) are remarkably similar to solution-state NMR structures, validating this computational target and providing a more detailed view of the atomic interactions stabilizing the folded state. In addition, the hexameric state observed in the crystals may serve as a tractable model for computational studies of protein-protein interactions and oligomer formation.

Results

Models of Trp-cages derived from NMR measurements all display fraying at the N and C termini and considerable variability in the conformations of exposed side chains. Reasoning that this type of fluxionality might retard crystallization, we replaced Lys8 with a helix-favorable alanine residue and added a glycine residue to each terminus to set the stage for cyclization. The resulting acyclic peptide (TC1) could, analogously to an early study of BPTI (23), be cyclized using a peptide coupling agent under aqueous conditions. Cyclo-TC1 proved to be the most stable Trp-cage fold prepared to date, significantly more stable than our prior hyperstable construct (6) in which Gly10 and Gly15 had been mutated to D-Ala (TC16b, $T_m = 83^\circ\text{C}$, $\Delta G_U^{280} = 18.6$ kJ/mol). Cyclo-TC1 displayed significant NH exchange protection at pD 7–8, with the largest protection factor (Gly11-H_N) corresponding to $\Delta G_U = 20.5$ kJ/mol. NMR melts (see *SI Appendix*) support a two-state unfolding scenario with the folded state stabilized by *circa* 8 kJ/mol by the cyclizing loop.

Circular dichroism (CD) measurements provided an accurate determination of the melting point ($T_m = 95 \pm 1^\circ\text{C}$). This corresponds to a ΔT_m in excess of 35°C versus the acyclic control (TC1, $T_m = 58^\circ\text{C}$) of identical sequence. Existing calibrations allow increases in T_m , as measured by CD melts, to be converted into $\Delta\Delta G_U^{280}$ values (2, 5, 6): $\Delta\Delta G_U^{\text{cyc}} = 8.2$ kJ/mol. The additional fold stability of cyclo-TC1 was even greater in the presence of denaturants; the CD melts for cyclo-TC1 throughout a GdmCl denaturation experiment appear in *Fig. S1B* in *SI Appendix*. The fold stabilization observed upon cyclization far exceeds that expected from entropy effects for a cyclic peptide of this size (24) and thus requires some structural rationale.

Author contributions: N.H.A. designed research; M.S., J.C.L., I.L.T., and G.I.M. performed research; M.S., J.C.L., I.L.T., G.I.M., R.E.S., and N.H.A. analyzed data; and M.S., R.E.S. and N.H.A. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Data deposition: The atomic coordinates and structure factors have been deposited in the Protein Data Bank, www.pdb.org (PDB ID codes 3UC7, 3UC8, and 2LL5); the NMR chemical shifts have been deposited in the BioMagResBank, www.bmrb.wisc.edu (accession no. 18023).

*As detailed in the *SI Appendix*, at least 40 publications reporting either new experiments on Trp-cage species or the use of the Trp-cage as the target structure for computational folding or fold prediction methods appeared in 2011 and the first few months of 2012.

To whom correspondence may be addressed. E-mail: scian@u.washington.edu or stenkamp@u.washington.edu or andersen@chem.washington.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1121421109/-DCSupplemental.

Fig. 4. Side-chain conformation comparisons between the solution and solid-state structure ensembles of cyclo-TC1. The side chains of Asp1, Gln5, Asp9, and Arg16 are shown in cyan for the NMR ensemble and green for the X-ray structures, with a representative backbone also displayed with the same color-coding. The conformations of Asp1 and Asp9 differ between the NMR and X-ray ensembles, likely due to the protonation state difference noted in the text.

when the Arg side chain is oriented near the Asp9 carboxyl. We view this as evidence for a transient hydrogen bonding interaction that would be strengthened by a favorable Coulombic interaction at pH 7.

The trimers in the crystal structures consist of subunits related by a C_3 axis. A subunit–subunit interaction found in all trimers is an intermolecular hydrogen bond between Tyr-OH and O=C-Gly-1 (Fig. 5). Although the AUC studies indicate that cyclo-TC1 is monomeric in the NMR buffer over a wide range of concentrations (60 μ M to greater than 25 mM), solution-state trimers or hexamers may form under high ionic strength and low pH conditions, and they may be required for crystallization. Prior to solving the structure by molecular replacement, we carried out extensive crystallization trials on the Tyr $\rightarrow$ *para*-bromoPhe derivative of cyclo-TC1 to introduce a strong anomalous scatterer for solving the phase problem. No crystals were observed, consistent with the loss of the proton donor for the Tyr-Gly trimer-stabilizing hydrogen bond.

The presence of specific inter-residue interactions in the monomer–monomer interfaces suggests the trimer could be a stable structure at higher protein concentrations. An intermolecular hydrophobic cluster made up of Ala4 and Leu7 side chains (Fig. 5) accounts for significant apolar surface burial at the trimer–trimer interface. The strongest evidence for the existence of the hexamer is that the same arrangement of six subunits is found in each crystal form. The packing environments for the hexamers, however, are not equivalent in the two forms, and this suggests that hexamers are the molecular species that crystallize. The small size of this hexamer could be a benefit for computational studies providing a tractable model for simulations of protein–protein interactions and oligomer formation.

It is apparent that the overall structure of a Trp-cage is not changed by cyclization or by the extensive intermolecular interactions in the crystal lattice. The higher resolution crystal structures provide answers to a number of questions that remained even after detailed examination of Trp-cage NMR structures, including that of cyclo-TC1. The solid-state ensemble has ϕ/ψ values that conform to the established norms throughout the α -, 3_{10} -, and polyProII-helices, and these structures predict the ring current shifts observed in NMR studies of the monomer. Cyclization and residue mutations result in very minor changes in

Fig. 5. A more detailed view of one trimer unit. The hydrogen bonds from the Tyr hydroxyls appear at the surface closest to the reader in this view; the Tyr3O...O=C-Gly-1 distance is remarkably constant (2.68 ± 0.03 Å) in the crystal structures. Below these is the hydrophobic clustering of Ala4 and Leu7 side chains from different monomers.

the Trp-cage structure, but the key features that stabilize this protein-like fold produce a remarkably consistent structural motif that can serve to benchmark fold predictions. A previously reported Trp-cage structure (e.g., TC10b, ref. 2) or that observed for residues 3–19 in the crystal structures presented here can serve in that capacity. It remains for computational folding methods to achieve this degree of consistency in predicting the fold features and the protein–protein association interface interactions shown in Fig. 5.

Materials and Methods

Peptide Synthesis and Purification. The peptide was synthesized on an Applied Biosystem 433A synthesizer using standard Fmoc (9-fluorenylmethoxycarbonyl) solid-phase peptide synthesis methods. The linear peptide was purified by RP-HPLC, using C_{18} stationary phases and a water (0.1% trifluoroacetic acid, TFA)/acetonitrile (0.085% TFA) gradient as previously described (2). A Wang resin preloaded with the C-terminal amino acid was used for the synthesis. The peptide was cleaved from the resin using a 95/2.5/2.5 TFA/triisopropylsilane/water mixture.

Cyclo-Trp-cage1 (cyclo-TC1) can be formed by native chemical ligation with a preformed thioester prepared using the 4-sulfamylbutyryl AM “safety-catch” resin (31). However, folding-mediated cyclization of fully deprotected GDA-YAQLADGGPSSGRPPPSG provided a more reproducible synthesis; optimum results were obtained with 0.5 mM 1-ethyl-3-(3-dimethylaminopropyl)carbodiimide hydrochloride (EDC $\cdot$ HCl), 75 μ M N-hydroxysuccinimide (Sulfo-NHS) sodium salt at room temperature over a four-day period at moderately high dilution (50 μ M peptide in 25 mM 3-(N-morpholino)propanesulfonic acid (MOPS), pH 6.5). In the preparative scale reactions, a quench with 2-mercaptoethanol (5 mM) and hydroxylamine $\cdot$ HCl (150 μ M) was essential in order to prevent Trp-cage oligomerization during the reaction work up. The cyclic product was obtained in acceptable yield (10%) and unreacted acyclic material could be recovered (HPLC) and recycled to increase the net conversion.

Circular Dichroism Spectroscopy (CD). Peptide stock solutions of approximately 200 μ M were prepared using 50 mM phosphate buffer, pH 7.0. Accurate concentrations were determined by UV spectroscopy assuming standard molar absorptivities for the Trp and Tyr residues. The peptide was typically diluted to about 30 μ M, and CD spectra were recorded on a Jasco J720 spectropolarimeter using 0.1 cm pathlength cells over a UV range of 190–270 nm as previously described (22). For determining the melting point for the peptide, the CD melts were performed without added denaturant and with 1–7 M guanidinium chloride (GdmCl) present.

NMR Spectroscopy. Samples for 2D NMR spectral studies consisted of 1–1.5-mM peptide in 50-mM phosphate buffer at pH 7.0, 10% D_2O , with

4,4-dimethyl-4-silapentane-1-sulfonic acid (DSS) as an internal chemical shift reference. NMR experiments were collected at 500 or 750 MHz on Bruker DRX and AV spectrometers. Full ^1H spectral assignments were made by using a combination of 2D NOESY and TOCSY experiments. The resulting chemical shifts were converted to folding-associated shifts, given as chemical shift deviations (CSDs) from random coil norms, using the CSDb algorithm (32–34). The large diagnostic CSDs for cyclo-TC1 and its acyclic precursor (TC1) appear in Fig. 1 wherein they are compared to the values recorded for the prior hyperstable Trp-cage species (6). The CSDs were monitored over the 280–320 K temperature range to provide an alternative measure of melting and $\Delta\Delta G_U$ values (*SI Appendix*).

NH Exchange Protection. The NH exchange experiments were performed at 280 K by adding pre-cooled D_2O phosphate buffer (pD = 7.0 and 8.0) to the lyophilized peptide sample in a pre-cooled NMR tube and then recording 1D proton spectra at various time intervals. Exchange rates (k_{obs}) were obtained from 1D spectra as the slopes of plots of $\ln(\text{NH signal intensity})$ versus time. The reference random coil rate constant for exchange (k_{rc}) was calculated using the appropriate Molday factors (35), as previously described for TC5b and TC10b (1, 2); protection factors were calculated as $P_F = k_{\text{rc}}/k_{\text{obs}}$. The relationship, $\chi_F = 1 - (1/P_F)$, was employed to convert protection factors to an extent of folding measure (ΔG_U). Cyclo-TC1 displayed NH exchange protection factors ranging from $10^{3.52}$ – $10^{3.81}$ for the hydrogen bonded sites at the C terminus of the N-terminal helix (Trp6, Leu7, Ala8, $\Delta G_U = 19.5 \pm 0.6$ kJ/mol). As observed in all other Trp-cages, Gly11- H_N displayed the largest protection factor providing the measure of overall fold stability, $\Delta G_U = 20.5$ kJ/mol.

NMR Structure Ensemble Calculations. The CNS-based refinement procedure, as well as the derivation of NOE distance constraints from the NOESY intensities, followed the examples of prior Trp-cage structure determinations (2, 22, 31). Further details appear in *SI Appendix*. The NMR structure ensemble has been deposited with the Protein Data Bank (ID code 2LL5) and the Biological Magnetic Resonance Bank (entry 18023).

Crystallization, Crystal Structure Determination, and Refinement. Two crystal forms of cyclo-TC1 grew from sitting drop vapor-diffusion experiments. A monoclinic crystal form was obtained with 2- μL drops made by combining

1 μL of 40 mg/mL protein (in MilliQ water) and 1 μL of reservoir solution (0.16 M Tris $\cdot$ HCl, 0.16 M MgCl_2 , 1.75 M NaCl, pH 7.0). A tetragonal form resulted from changing the reservoir solution to 0.2 M Tris $\cdot$ HCl, 0.2 M MgCl_2 , 2.0 M NaCl, pH 7.0. Diffraction data for the monoclinic crystals were collected at beamline 11-1 at SSRL. Those for the tetragonal crystals were measured using a Rigaku Raxis-IV++ image plate system and processed using HKL2000 (36). Table 2 contains data set and final refinement statistics for both crystal forms.

The monoclinic form was solved using one of the NMR models of cyclo-TC1 for molecular replacement (model 1 from PDB ID code 2LL5). Phaser (25) placed five molecules in the asymmetric unit, and a sixth was added manually by fitting it in a difference electron density map. One of the trimers served as a test structure for molecular replacement in the tetragonal crystal form. Molrep (37) found a structure solution in space group P4_32_2 . Again, the tetragonal crystals contain an asymmetric hexamer; however, unlike the monoclinic form, the twofold axis is now a crystallographic symmetry axis.

Both structures have been refined using Refmac-5 (26) in the CCP4 program package (27). The models have been validated using Molprobity (38), and coordinates and structure factors have been deposited in the Protein Data Bank (39) (PDB ID codes 3UC7 and 3UC8). The quality of the model fit to electron density is illustrated in Fig. S3 in *SI Appendix*. Subunit–subunit interactions are listed in *SI Appendix, Table S4* and illustrated in Fig. S4 in *SI Appendix*.

All structural figures were made with the PyMol Molecular Graphic System, Version 1.3, Schrödinger LLC.

ACKNOWLEDGMENTS. We thank Maria M. Lopez for helping us collect the AUC data and Eric Larson for helpful discussion and suggestions. This work has been supported in part by National Institutes of Health (NIH) grants GM059658 and GM099889 (N.H.A., P.I.). Portions of this research were carried out at the Stanford Synchrotron Radiation Lightsource, a Directorate of SLAC National Accelerator Laboratory and an Office of Science User Facility operated for the U.S. Department of Energy Office of Science by Stanford University. The SSRL Structural Molecular Biology Program is supported by the DOE Office of Biological and Environmental Research, the NIH National Institute of General Medical Sciences (including P41GM103393), and the National Center for Research Resources (P41RR001209). The contents of this publication are solely the responsibility of the authors and do not necessarily represent the official view of NIGMS, NCR, or NIH.

- Neidigh JW, Fesinmeyer RM, Andersen NH (2002) Designing a 20-residue protein. *Nat Struct Biol* 9:425–430.
- Barua B, et al. (2008) The Trp-cage: Optimizing the stability of a globular miniprotein. *Protein Eng Des Sel* 21:171–185.
- Searle MS, Ciani B (2004) Design of beta-sheet systems for understanding the thermodynamics and kinetics of protein folding. *Curr Opin Struct Biol* 14:458–464.
- Mok KH, et al. (2007) A pre-existing hydrophobic collapse in the unfolded state of an ultrafast folding protein. *Nature* 447:106–109.
- Lin JC, Barua B, Andersen NH (2004) The helical alanine controversy: An (Ala)6 insertion dramatically increases helicity. *J Am Chem Soc* 126:13679–13684.
- Williams DV, Barua B, Andersen NH (2008) Hyperstable miniproteins: Additive effects of D- and L-Ala mutations. *Org Biomol Chem* 6:4287–4289.
- Ahmed Z, Beta IA, Mikhonin AV, Asher SA (2005) UV-resonance raman thermal unfolding study of Trp-cage shows that it is not a simple two-state miniprotein. *J Am Chem Soc* 127:10943–10950.
- Neuweiler H, Dose S, Sauer M (2005) A microscopic view of miniprotein folding: Enhanced folding efficiency through formation of an intermediate. *Proc Natl Acad Sci USA* 102:16650–16655.
- Bunagan MR, Yang X, Saven JG, Gai F (2006) Ultrafast folding of a computationally designed Trp-cage mutant: Trp2-cage. *J Phys Chem B* 110:3759–3763.
- Neuman RC, Jr, Gerig JT (2008) Solvent interactions with the Trp-cage peptide in 35% ethanol-water. *Biopolymers* 89:862–872.
- Wafer LN, Streicher WW, Makhatadze GI (2010) Thermodynamics of the Trp-cage miniprotein unfolding in urea. *Proteins* 78:1376–1381.
- Simmerling C, Strockbine B, Roitberg AE (2002) All-atom structure prediction and folding simulations of a stable protein. *J Am Chem Soc* 124:11258–11259.
- Snow CD, Zagrovic B, Pande VS (2002) The Trp cage: Folding kinetics and unfolded state topology via molecular dynamics simulations. *J Am Chem Soc* 124:14548–14549.
- Chowdhury S, Lee MC, Xiong G, Duan Y (2003) Ab initio folding simulation of the Trp-cage mini-protein approaches NMR resolution. *J Mol Biol* 327:711–717.
- Zhou R (2003) Trp-cage: Folding free energy landscape in explicit water. *Proc Natl Acad Sci USA* 100:13280–13285.
- Linhanta A, Boer J, MacKay I (2005) The equilibrium properties and folding kinetics of an all-atom Go model of the Trp-cage. *J Chem Phys* 122:114901.
- Juraszek J, Bolhuis PG (2006) Sampling the multiple folding mechanisms of Trp-cage in explicit solvent. *Proc Natl Acad Sci USA* 103:15859–15864.
- Xu W, Mu Y (2008) Ab initio folding simulation of Trpcage by replica exchange with hybrid Hamiltonian. *Biophys Chem* 137:116–125.
- Day R, Paschek D, Garcia AE (2010) Microsecond simulations of the folding/unfolding thermodynamics of the Trp-cage miniprotein. *Proteins* 78:1889–1899.
- Liu Y, Liu Z, Androphy E, Chen J, Baleja JD (2004) Design and characterization of helical peptides that inhibit the E6 protein of papillomavirus. *Biochemistry* 43:7421–7431.
- Hudaky P, et al. (2008) Cooperation between a salt bridge and the hydrophobic core triggers fold stabilization in a Trp-cage miniprotein. *Biochemistry* 47:1007–1016.
- Williams DV, Byrne A, Stewart J, Andersen NH (2011) Optimal salt bridge for Trp-cage stabilization. *Biochemistry* 50:1143–1152.
- Goldenberg DP, Creighton TE (1983) Circular and circularly permuted forms of bovine pancreatic trypsin inhibitor. *J Mol Biol* 165:407–413.
- Zhou HX (2003) Effect of backbone cyclization on protein folding stability: Chain entropies of both the unfolded and the folded states are restricted. *J Mol Biol* 332:257–264.
- McCoy AJ, et al. (2007) Phaser crystallographic software. *J Appl Crystallogr* 40:658–674.
- Murshudov GN, Vagin AA, Dodson EJ (1997) Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr D Biol Crystallogr* 53:240–255.
- Anonymous (1994) The CCP4 suite: Programs for protein crystallography. *Acta Crystallogr D Biol Crystallogr* 50:760–763.
- Lee B, Richards FM (1971) The interpretation of protein structures: Estimation of static accessibility. *J Mol Biol* 55:379–400.
- Streicher WW, Makhatadze GI (2007) Unfolding thermodynamics of Trp-cage, a 20 residue miniprotein, studied by differential scanning calorimetry and circular dichroism spectroscopy. *Biochemistry* 46:2876–2880.
- Makhatadze GI, Medvedkin VN, Privalov PL (1990) Partial molar volumes of polypeptides and their constituent groups in aqueous solution over a broad temperature range. *Biopolymers* 30:1001–1010.
- Lin JC (2007) Application of the Trp-cage motif to polypeptide folding questions. PhD Thesis (University of Washington, Seattle).
- Fesinmeyer RM, Hudson FM, Andersen NH (2004) Enhanced hairpin stability through loop design: The case of the protein G B1 domain hairpin. *J Am Chem Soc* 126:7238–7243.
- Fesinmeyer RM, et al. (2005) Chemical shifts provide fold populations and register of beta hairpins and beta sheets. *J Biomol NMR* 33:213–231.
- Eidenschink L, Crabbe E, Andersen NH (2009) Terminal sidechain packing of a designed beta-hairpin influences conformation and stability. *Biopolymers* 91:557–564.
- Bai Y, Milne JS, Mayne L, Englander SW (1993) Primary structure effects on peptide group hydrogen exchange. *Proteins* 17:75–86.
- Otwinski Z, Minor W (1997) Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol* 276:307–326.
- Vagin A, Teplov A (1997) MOLREP: An automated program for molecular replacement. *J Appl Crystallogr* 30:1022–1025.
- Davis IW, et al. (2007) MolProbity: All-atom contacts and structure validation for proteins and nucleic acids. *Nucleic Acids Res* 35:W375–383.
- Berman HM, et al. (2000) The Protein Data Bank. *Nucleic Acids Res* 28:235–242.