Supporting Information

for

In vivo screening identifies a highly folded β-hairpin peptide with a structured extension

Zihao Cheng, Mark Miskolzie and Robert E. Campbell

Experimental procedures

General methods and materials. All synthetic DNA oligonucleotides were purchased from Sigma-Genosys Canada or Integrated DNA Technologies. Pfu polymerase (Fermentas) was used for all PCR amplifications in the buffer supplied by the manufacturer. PCR products and products of restriction digest were routinely purified using the QIAquick PCR purification kit (Qiagen). Plasmids were purified using GeneJet plasmid miniprep kit (Fermentas) according to the manufacturers' protocols. Restriction enzymes were purchased from either Invitrogen or New England Biolabs. The identity of all cDNA constructs was confirmed by dye terminator cycle sequencing using the DYEnamic ET kit (Amersham Biosciences). All sequencing reactions were analyzed at the University of Alberta Molecular Biology Service Unit. Protein samples for all *in vitro* spectral characterization experiments were in 50mM Tris, pH 7.5.

Library construction and screening. The CFP-peptide-YFP libraries were expressed using the previously described vector pZC1.[1] This vector can be used to encode proteins of the general structure Nterm-His_6-EK-CFP(1-230)-TSGAQ-peptide-GTSAE-YFP(5-238)-Cterm, where 'His_6' is six consecutive histidine residues that facilitate metal affinity purification,[2] and 'EK' is the recognition sequence for the protease enterokinase. The DNA encoding the TSGAQ sequence (actagtggagctcag) contains both a Spe1 and a Sac1 restriction site. The DNA encoding the GTSAE sequence (ggtacctctgcagag) contains both a Kpn1 and a Pst1 restriction site. In our previous work,[1] 'peptide' represented a 16mer from the **tz** series. In this work, 'peptide' represents either a 20mer library or an individual member of the **xtz** series.

To create libraries of 20mer peptides, double stranded (ds) DNA encoding the peptide library and with appropriate sticky ends were created by slow cooling pairs of complementary single stranded oligonucleotides from 95 °C to room temperature. This

synthetic annealed dsDNA was ligated into the CFP-peptide-YFP expression vectors that had been digested with Sac1 and Kpn1. Ligated product was used to transform *E. coli* DH10B (Invitrogen) by electroporation. Appropriate dilutions of transformed bacteria were plated on LB/agar containing ampicillin (0.1 mg/ml) and arabinose (0.2%). After overnight incubation at 37 °C, single colonies were picked and used to inoculate 2 mL of LB media containing ampicillin (0.1 mg/mL). Following overnight incubation at 37 °C with shaking at 225 rpm, plasmid DNA was isolated and used in subsequent procedures.

The system for imaging the fluorescence of bacterial colonies grown on 10 cm Petri dish was previously been described in detail.[1] Individual colonies were automatically identified within the digital images and the pixel intensities are summed in each of 3 images corresponding to the CFP channel, the FRET channel, and direct YFP channel. For each colony the ratio of YFP to CFP fluorescence emission intensity is calculated by dividing the average intensity in the FRET channel by the average intensity in the CFP channel. Colonies that had the highest ratio of YFP to CFP emission and the highest brightness were picked from the plate and cultured as described above. The sequence of the peptide portion of all selected proteins was determined by DNA sequencing with a forward primer (5'-CCCTCGTGACCACCCTGACCTGG-3') that anneals to the chromophore region of the gene for CFP.

Protein and peptides purification. The pBAD/His B vector used for all cloning and expression allows proteins to be expressed at high levels under control of the arabinose operon. A typical protein purification procedure started with the inoculation of 1 L LB media containing ampicillin (0.1 mg/mL) and arabinose (0.2%) with a single colony of *E. coli* DH10B expressing the protein of interest. Cultures were grown overnight at 37 °C with shaking at 225 rpm. Cultures were cooled to 4 °C on ice and harvested by centrifugation (10 min, 5000 rpm). The cell pellet was resuspended in PBS buffer (130 mM NaCl, 2.7 mM KCl, 10 mM phosphate, pH 7.6) and the cells lysed by a single passage through a French Press (Thermo Electron). Insoluble cell debris was pelleted by centrifugation at 4 °C (10 min, 10 000 rpm) and 1 ml of Ni-NTA resin (Qiagen) was added to the supernatant. Following one hour of gentle mixing at 4 °C, the supernatant was loaded onto a 6 ml polypropylene column, washed, and gravity eluted in 100 mM imidazole (pH 7.5) according to the manufacturers protocol. FRET constructs containing

both CFP and YFP were further purified on an AKTAdesign chromatography system (Amersham Biosciences) equipped with a Hiload 16/60 Superdex 75 prepgrade gel filtration column (Amersham Biosciences) that was equilibrated with 50 mM Tris pH 7.5. Isolated proteins were concentrated with a Centricon centrifugal filter YM-30 (Millipore) and stored at 4 °C.

The isolated peptide portions of **tz1** and **xtz1**, **HP5W4** and **xtz1-peptide** respectively, were purchased from Sigma-Genosys Canada. The **xtz2-peptide** was isolated using a previously described bacterial expression system.[1] All peptides were further purified by reversed-phase chromatography on a Prosphere HP C18 300A column (Alltech Associate, Inc.). A linear gradient of increasing acetonitrile in H_2O with 0.1% trifluoroacetic acid was used to elute the target peptides. The fraction containing the target peptide was lyophilized to provide a fluffy white powder. Calculated extinction coefficients of 22,300 $M^{-1}cm^{-1}$, 27,875 $M^{-1}cm^{-1}$ and 33,450 $M^{-1}cm^{-1}$ at 280 nm were used to determine the concentration of **HP5W4**, **xtz1-peptide** and **xtz2-peptide** respectively.

Spectroscopy. Steady-state fluorescence spectra for all proteins were recorded on a QuantaMaster spectrofluorometer (Photon Technology International) equipped with a Xenon arc lamp. The protein concentration for all fluorescence measurements was 40 nM. All CD spectra were obtained with 40 μM peptide in a 1 mm path length cuvette on an Olis DSM 17 CD spectrometer (Olis).

Protein expression and imaging in HeLa cells. To create the CFP-peptide-YFP mammalian expression plasmids, restriction sites encoding an Xho1 and an EcoR1 restrion site were introduced into pcDNA3 (Invitrogen) by a PCR-based method. The pBAD/His B plasmid containing the gene encoding CFP-peptide-YFP was digested with Xho1 and EcoR1 and ligated into the modified pcDNA3 plasmid digested with the same two enzymes. All DNA for mammalian cell transfection was purified by Plasmid Midi kit (Qiagen). HeLa cells were cultured at 37 °C in DMEM (Dulbecco's modified Eagle's medium; Invitrogen) supplemented with 10% (v/v) FBS (Sigma). Cells in 35 mm imaging dishes were transfected with 4 μg plasmid DNA mixed with 10 μg poly(ethylenimine) (linear, molecular weight ~25,000, Polysciences Inc.) in 0.5 ml OptiMEM (Invitrogen) and serum was added after 3 hours. Approximately 14 to 24 h later the medium was exchanged for Hanks' Balanced Salt solution (HBSS) containing no calcium chloride,

magnesium chloride, magnesium sulfate, or phenol red (Invitrogen). Cells were imaged with a Zeiss Axiovert 200M epi-fluorescence inverted microscope (Zeiss) equipped with a monochrome Retiga 2000R 12-bit cooled CCD camera (QImaging). The light from a xenon arc lamp is passed through an external filter wheel (Sutter) holding either a 426 nm to 446 nm bandpass filter (Chroma Technology Corp.) or a 490 nm to 510 nm bandpass filter for excitation of CFP or YFP respectively. The fluorescence emission is filter through a second filter wheel holding either a 460 nm to 500 nm bandpass filter or a 515 nm to 555 nm bandpass filter for CFP or YFP respectively.

FRET efficiencies were measured by measuring the intensity of the CFP signal (direct excitation) before and after bleaching of the YFP acceptor. FRET efficiencies reported in Figure 4d represent the average of at least 15 independent single cell measurements.

NMR Spectroscopy and Structure calculations. Synthetic **xtz1-peptide** (Sigma-Genosys) was dissolved in 90% H_2O / 10% D_2O to a final concentration of 2 mM. The pH was determined to be between 5 and 6. To this solution was added 0.010 ml of 0.1% the chemical shift reference 2,2-dimethyl-2-silapentane-5-sulfonic acid (DSS). All chemical shifts are quoted relative to the DSS reference set at 0.00 ppm.

Spectra were recorded on a Varian Inova 600MHz spectrometer. Data acquisition and initial Fourier transformations were performed using VNMR 6.1C. Two-dimensional (2D) total correlation spectroscopy (TOCSY),[3, 4] gradient enhanced double quantum filter correlation spectroscopy (gDQF-COSY),[5] and nuclear Overhauser enhancement spectroscopy (NOESY)[6] were carried out. Solvent suppression was achieved by presaturation of the water peak. Spectral widths were set to be equivalent in both dimensions for all 2D spectra. All spectra were acquired in the phase sensitive mode using the States-Haberkorn-Ruben method,[7] with the exception of the gDQF-COSY spectrum which was acquired in the absolute value mode selecting N-type signals. The gDQF-COSY and NOESY spectra employed 8192 data points in the directly detected dimension, whereas the TOCSY used 4096 data points in the directly detected dimension. The NOESY spectrum was recorded with 512 experiments in the indirectly detected dimension, and 32 scans were acquired per experiment, whereas the TOCSY and gDQF-COSY experiments were acquired with 256 experiments in the indirectly

detected dimension with 16 scans per experiment. All spectra employed squared sine bell and shifted squared sine bell functions as weighting functions in both dimensions. The TOCSY experiments employed a spin lock mixing time of 100ms and the mixing time for the NOESY experiment was 150ms.

The choice of temperature for all experiments was based on a series of one-dimensional spectra acquired at 5 °C, 15 °C, and 27 °C. The amide region of these spectra was examined to determine at which temperature contained the least spectral overlap for all amide resonances. 15 °C was found to be the most suitable temperature. $J_{HN\text{-}H\alpha}$ coupling constants were obtained from the 1D spectrum at 15 °C where possible.

Spectra were viewed in NMRview v.6.4[8] and chemical shift assignments (Table S1) were manually determined. Automated NOESY assignment and structure ensemble calculation was performed with CYANA v.2.1.[9, 10] A total of 431 unassigned NOESY peaks with intensity greater than ~1% of the most intense NOESY peak were input to CYANA. At less than 1% of the most intense NOESY peak, spectral noise was significant and peak assignments were unreliable. CYANA assigned 403 of 431 (93.5%) peaks, of which 309 were used as distance constraints in the structure calculation (Table S2).

Table S1. Chemical shifts for **xtz1-peptide**

Residue	HN[a]	Hα	Hβ,β'	Others
Lys 1	exchange	4.064	1.957 1.896	Hγ,γ' 1.531; Hδ,δ' 1.714; Hε,ε' 3.059
Ala 2	exchange	4.473	1.476	
Trp 3	8.329 (8.2)	5.010	2.832	Hδ1 7.197; Hε1 10.218; Hε3 7.041; Hζ2 7.484; Hζ3 7.617; Hη2 7.256
Thr 4	9.091 (8.2)	4.880	4.131	Hγ 1.302
Trp 5	9.169	5.060	3.273 2.591	Hδ1 6.937; Hε1 9.451; Hε3 6.361; Hζ2 7.159; Hζ3 6.794; Hη2 7.107
Thr 6	9.503 (9)	4.849	3.980	Hγ 1.228
Trp 7	8.782 (5.5)	3.792	2.477 1.300	Hδ1 6.497; Hε1 9.976; Hε3 5.646; Hζ2 7.369; Hζ3 6.670; Hη2 7.022
Asn 8	7.315	4.850	3.110 2.389	Hδ,δ' 7.303, 7.265
Pro 9		3.571	2.391 1.959	Hγ,γ' 2.043; Hδ,δ' 3.598, 3.767
Ala 10	7.752 (6.2)	4.166	1.433	
Thr 11	6.838 (9.7)	4.336	4.193	Hγ 1.045
Gly 12	8.019	3.832 3.258		
Lys 13	6.711 (9.4)	4.668	1.740 1.801	Hγ,γ' 1.329, 1.472; Hδ,δ' 1.639; Hε,ε' 3.055
Trp 14	8.700 (8.3)	4.889	2.228 2.792	Hδ1 7.458; Hε1 10.140; Hε3 6.897; Hζ2 7.209; Hζ3 7.150; Hη2 7.129
Thr 15	8.960 (5.8)	4.657	4.134	Hγ 1.152
Trp 16	8.732 (7.2)	5.160	3.107 2.701	Hδ1 7.176; Hε1 9.797; Hε3 6.979; Hζ2 7.255; Hζ3 7.055; Hη2 7.146
Arg 17	9.120 (8.5)	4.744	1.919 1.741	Hγ,γ' 1.620; Hδ,δ' 3.186; Hε 7.167
Lys 18	8.475	3.722	1.257 0.854	Hγ,γ' 0.741, 0.650; Hδ,δ' 1.374, 1.313; Hε,ε' 2.697
Asn 19	8.439 (7.4)	4.712	2.839 2.652	Hδ,δ' not observed
Glu 20	8.056 (7.4)	4.124	2.068 1.881	Hγ,γ' 2.186

[a] Numbers in parenthesis are $J_{HN-H\alpha}$ in hertz (Hz).

Table S2. Statistics for the 20 model NMR solution structures of **xtz1-peptide**

Number of residues	20
Average CYANA target function (Å^2)	0.22
Distance restraints (#)	
All	309
Short-range, $\|i - j\| \leq 1$	158
Medium-range, $1 < \|i - j\| < 5$	47
Long-range, $\|i - j\| \geq 5$	104
Distance Bounds (# restraints)	
- 2.99	12 (3.9 %)
3.00 - 3.99	94 (30.4 %)
4.00 - 4.99	140 (45.3 %)
5.00 - 5.99	63 (20.4 %)
6.00 -	0 (0 %)
RMSD	
All residues, backbone (Å)	0.27 +/- 0.09
All residues, heavy atoms (Å)	0.74 +/- 0.15
Ramachandran plot statistics	
Most favored regions	56.6 **%**
Additional allowed regions	43.4 **%**
Generously allowed regions	0.0 **%**
Disallowed regions	0.0 **%**

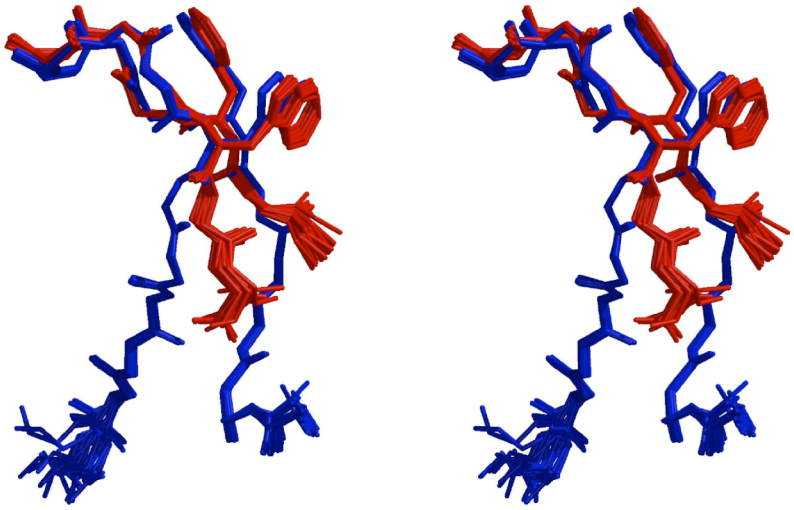

Figure S1. Stereoview of a superposition of the NMR structure ensembles of HP7 (red) and **xtz1-peptide** (blue). All atoms from Thr2 to Thr11 of HP7 (PDB ID 2EVQ)[11] were aligned with all atoms from Thr6 to Thr15 of **xtz1-peptide** using the least squares fit function of XtalView v.4.0.[12]

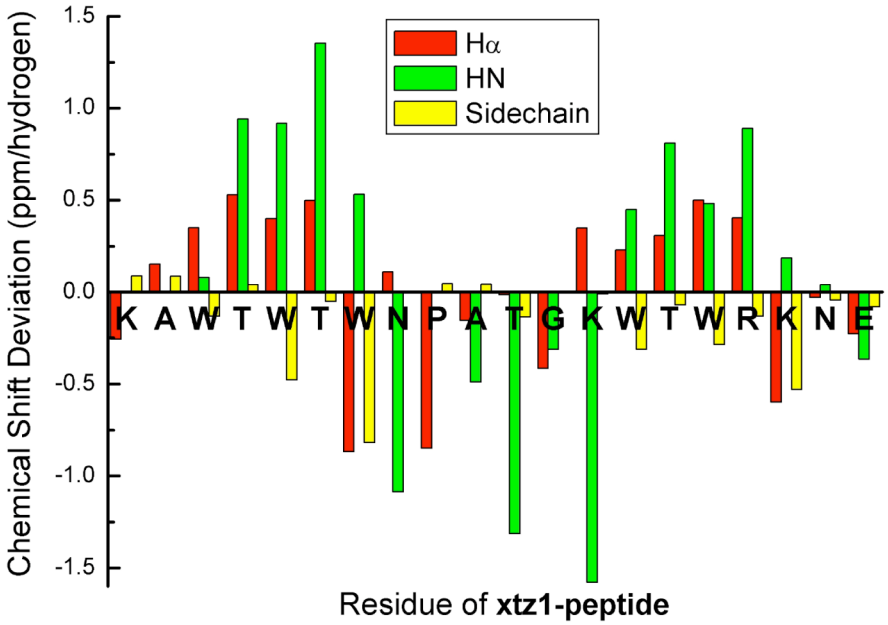

Figure S2. Chemical Shift Deviation (CSD) analysis of Hα, HN, and side chain hydrogen atoms of **xtz1-peptide**. CSDs were calculated by subtracting the residue-specific random coil chemical shift[13] from the observed chemical shift. For the side chain values, CSDs were summed over all hydrogen atoms in the residue and then divided by the total number of hydrogen atoms. The sign (positive = downfield shift; negative = upfield shift) and magnitude of the CSDs are in excellent agreement with the values previously reported for similar peptides.[11, 14] Note the strong upfield shift of 0.5 ppm per hydrogen over the complete side chain of K18.

References for Supporting Information

[1] Z. Cheng, R. E. Campbell, *ChemBioChem* **2006**, *7*, 1147-1150.
[2] E. Hochuli, W. Bannwarth, H. Dobeli, R. Gentz, D. Stuber, *Bio-Technology* **1988**, *6*, 1321-1325.
[3] A. Bax, D. G. Davis, *J. Magn. Reson.* **1985**, *65*, 355-360.
[4] M. H. Levitt, R. Freeman, T. Frenkiel, *J. Magn. Reson.* **1982**, *47*, 328-330.
[5] R. E. Hurd, *J. Magn. Reson.* **1990**, *87*, 422-428.
[6] S. Macura, R. R. Ernst, *Mol. Phys.* **1980**, *41*, 95-117.
[7] D. J. States, R. A. Haberkorn, D. J. Ruben, *J. Magn. Reson.* **1982**, *48*, 286-292.
[8] B. A. Johnson, *Methods Mol. Biol.* **2004**, *278*, 313-352.
[9] P. Guntert, C. Mumenthaler, K. Wuthrich, *J. Mol. Biol.* **1997**, *273*, 283-298.
[10] T. Herrmann, P. Guntert, K. Wuthrich, *J. Mol. Biol.* **2002**, *319*, 209-227.
[11] N. H. Andersen, K. A. Olsen, R. M. Fesinmeyer, X. Tan, F. M. Hudson, L. A. Eidenschink, S. R. Farazi, *J. Am. Chem. Soc.* **2006**, *128*, 6101-6110.
[12] D. E. McRee, *J. Struct. Biol.* **1999**, *125*, 156-165.
[13] D. S. Wishart, D. A. Case, *Methods Enzymol.* **2001**, *338*, 3-34.
[14] R. M. Fesinmeyer, F. M. Hudson, N. H. Andersen, *J. Am. Chem. Soc.* **2004**, *126*, 7238-7243.