DOI: 10.1002/cbic.200600565

In Vivo Screening Identifies a Highly Folded β-Hairpin Peptide with a Structured Extension

Zihao Cheng, Mark Miskolzie, and Robert E. Campbell*[a]

The conventional approach for developing folded peptides involves chemical synthesis of systematically modified peptide variants and individual characterization by circular dichroism and NMR spectroscopy.[1–5] The use of synthetic peptides and reliance on low-throughput characterization techniques necessarily restricts the sequence diversity that can be explored, though efforts have been made to overcome this limitation.[6–8] We have recently described an approach that enables us to asses the ability of peptides to fold into β-hairpins in the cytoplasm of live cells.[9] Our strategy entails the recombinant expression of a peptide gene fused in frame between flanking genes encoding a cyan fluorescent protein (CFP) and a yellow fluorescent protein (YFP). If a particular peptide sequence adopts a folded structure, CFP and YFP are brought into closer proximity and exhibit a higher efficiency of fluorescence resonance energy transfer (FRET). Higher FRET efficiency enhances the YFP (acceptor) fluorescence at the expense of the CFP (donor) fluorescence. Imaging of plates harboring colonies of transformed bacteria provides the YFP and CFP fluorescence emission intensities for each colony, and thus peptides that are highly folded in vivo can be distinguished from those that are not.

As will be described in this manuscript, this FRET-based approach also provides a versatile method for screening large libraries of peptide sequences for highly structured variants. We have applied this strategy to the development of a version of a "tryptophan zipper" (trpzip)-type β-hairpin[1] with a structured extension. Trpzips are a class of highly folded β-hairpins that are defined by the presence of cross-strand diagonally oriented Trp/Trp pairs on one face of the hairpin.[1,5,10] The minimal and highly stable trpzip structure has emerged as a preferred model system for computational and experimental studies of protein folding.[11–13] We have proposed that trpzip-type peptides (or tandem fusions of such peptides) could serve as a minimal protein scaffold for molecular recognition in the cytoplasm of live cells.[9,14] With this goal in mind, we sought to employ our screening strategy to identify a candidate trpzip-type peptide with high fold stability in vivo. A similar approach has previously been used to increase the thermal stability of engineered immunoglobulin V_L domains.[15]

The template for our initial "extended trpzip" library was a 20-mer version of the highly folded 16-mer trpzip **HP5W4.**[10] The 20-mer contained two additional pairs of residues, one random and one threonine, genetically inserted after what would otherwise have been the second and 14th residues of **HP5W4**. The sequence of the resulting "peptide" portion of the 400-member protein library was $K^1\,K^2\underline{X}^3\underline{T}^4W^5T^6W^7N^8P^9\,A^{10}T^{11}G^{12}\,K^{13}W^{14}T^{15}W^{16}\underline{T}^{17}\underline{X}^{18}Q^{19}E^{20}$; here X represents all 20 amino acids, and the residues inserted relative to **HP5W4** are underlined. Assuming β-strand conformation, the randomized positions would be directed towards the face of **HP5W4** that harbors the interdigitated Trp side chains.

Escherichia coli was transformed with the gene library, and ~6×10^3 colonies on 10 Petri dishes were subject to fluorescence imaging in order to identify those exhibiting the highest ratio of YFP-to-CFP fluorescence emission. We have previously demonstrated the ability of this imaging system to reliably distinguish colonies that express FRET constructs of various FRET efficiencies.[9] The fluorescent brightness of each colony, considered commensurate with peptide solubility,[16] was determined by direct excitation and imaging of YFP. Five individual colonies that exhibited both a high ratio of YFP-to-CFP fluorescence emission and high brightness were cultured overnight, and the plasmid DNA was purified. DNA sequencing revealed a striking consensus at the randomized positions: in all five sequences, position 3 was Trp and position 18 was either Lys or Arg (Table 1). This consensus sequence was used as the new template for a second library in which positions 2 and 19 were similarly randomized. Library screening and DNA sequencing revealed a strong preference for Asn at position 19 and a weaker preference for small side chains, such as Ala and Ser, at position 2. An iterative third library with randomization of posi-

Table 1. Peptide sequences identified by library screening

	Library[a]	Sequences identified[b]			
		X^3	**X^{18}**		**#**
1	$K^1K^2\mathbf{X}^3T^4\ldots T^{17}\mathbf{X}^{18}Q^{19}E^{20}$	W	R		3
		W	K		2
		X^2	**$(K/R)^{18}$**	**X^{19}**	**#**
2	$K^1\mathbf{X}^2W^3T^4\ldots T^{17}\mathbf{(K/R)}^{18}\mathbf{X}^{19}E^{20}$	A	K	N	2
		S	K	N	2
		P	K	N	1
		Q	R	N	1
		$(A/S)^2$	**X^4**	**X^{17}**	**#**
3	$K^1\mathbf{(A/S)}^2W^3\mathbf{X}^4\ldots\mathbf{X}^{17}K^{18}N^{19}E^{20}$	A	T	R	3[c]
		A	S	K	2
		A	S	R	1
		A	S	V	1

[a] "…" denotes residues 5 to 16 of the peptide. Positions marked "X" were subject to saturation mutagenesis by using the codon "NNK" where N = adenine (A), guanine (G), cytosine (C), or thymine (T), and K = G or T. In round 2, position 18 was mutated to either Lys or Arg by using the codon "ARG", where R = A and G. In round 3, position 2 was mutated to either Ala or Ser by using the codon "KCC". [b] Peptide sequence determined by DNA sequencing of plasmid DNA. [c] This sequence was designated as **xtz1**.

[a] *Z. Cheng, M. Miskolzie, Prof. R. E. Campbell*
Department of Chemistry, University of Alberta
Edmonton, AB T6G 2G2 (Canada)
Fax: (+1) 780-492-8231
E-mail: robert.e.campbell@ualberta.ca

Supporting information for this article is available on the WWW under http://www.chembiochem.org or from the author.

tions 4 and 17 was constructed and screened as above. The CFP–YFP fusion protein containing the sequence identified most often in this third library was designated extended trpzip 1 (**xtz1**) and subject to further investigation.

Three additional CFP–YFP fusion proteins were investigated as comparisons and controls (Table 2). The first of these has been previously described as **tz1**[9] and contains a 16-mer peptide identical to **HP5W4**.[5] The second contains a 20-mer peptide with a potential third Trp/Trp cross-strand pair (**xtz2**, equivalent to **xtz1** with $K^2W^3T^4$ and $T^{17}W^{18}Q^{19}$). The third is a control protein that contains an unstructured 20-mer peptide (**xtz3**, equivalent to **xtz2** with all six Trp replaced with Gly). The in vitro FRET efficiencies for **tz1**, **xtz1**, **xtz2**, and **xtz3** were determined to be 59, 68, 62, and 44%, respectively. We have previously established[9] that the peptide portion of **tz1** is highly structured and the higher FRET efficiencies for **xtz1** and **xtz2** suggest that their peptides are also highly structured. The FRET efficiency of **xtz3** is consistent with previous work in which it was shown that a CFP–YFP fusion protein with an unstructured 16-mer peptide of similar composition had a FRET efficiency of 47%.[9] Additional characterization of **xtz2** and **xtz3** by CD and resistance to proteolysis[14] has provided further support for the conclusion that the peptide portion of **xtz2** is highly structured, and that the peptide portion of **xtz3** is unstructured, in the context of the CFP–YFP fusion protein.

Table 2. Summary of proteins and peptides used in this work

Protein name	Peptide name	Peptide sequence (flanked by CFP and YFP in protein)
tz1	**HP5W4**	KK––WTWNPATGKWTW––QE
xtz1	**xtz1**-peptide	KAWTWTWNPATGKWTWRKNE
xtz2	**xtz2**-peptide	KKWTWTWNPATGKWTWTWQE
xtz3	not investigated	KKGTGTGNPATGKGTGTGQE

The isolated peptide portions of **tz1** (**HP5W4**), **xtz1** (**xtz1**-peptide), and **xtz2** (**xtz2**-peptide) were produced either through peptide synthesis or by recombinant expression and purification. As previously reported for highly folded trpzip peptides,[1] both **HP5W4** and **xtz1**-peptide are soluble in water at millimolar concentrations and have characteristic exciton-coupled bands in their CD spectra (Figure 1A). In contrast, **xtz2**-peptide is soluble in water at only micromolar concentrations and in the absence of trifluoroethanol (TFE), its CD spectrum showed no significant exciton peaks. We attribute the lack of solubility in water to hydrophobic character conferred by the additional Trp side chains present in **xtz2**-peptide. It is apparent that under conditions more favorable to its dissolution, such as 50% TFE or when CFP and YFP are fused to the termini, the peptide can indeed fold into a trpzip-type structure.

To obtain further insight into the structure of **xtz1**-peptide, we acquired 2D NMR spectra and employed NOE distance constraints to generate a structure ensemble (Figure 2). As expected, **xtz1**-peptide adopts a typical trpzip-type β-hairpin conformation with the indole rings of the four **HP5W4**-derived Trp residues tightly packed on one face of the structure. The relative orientation of the diagonally paired Trp residues closest to the turn (Trp7/Trp14) is very similar to the edge-to-face (EtF) interaction observed in trpzip4, **HP5W4**, and HP7 (Figure S1).[1,10] The middle pair of indole rings (Trp5/Trp14) displays a parallel-displaced orientation similar to that observed in **HP5W4**.[10] It has been reported that the diagonal Trp pair closest to the termini (Trp5/Trp16) displays an EtF interaction in trpzip4 but a parallel-displaced interaction in **HP5W4**.[10] In **xtz1**-peptide, Trp5/Trp16 are oriented in what is best described as an EtF interaction. We attribute this adjustment in packing interaction (relative to **HP5W4**) to the additional conformational restrictions enforced by the additional cross-strand interactions associated with the inserted residues.

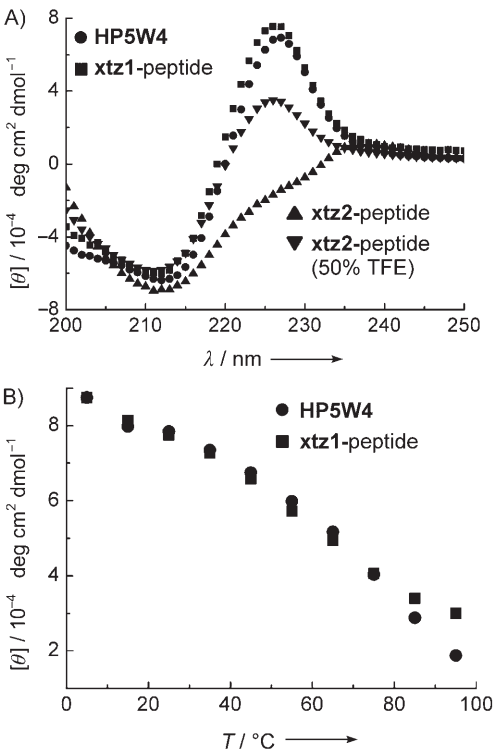

Figure 1. A) CD spectra of identical concentrations of **HP5W4**, **xtz1**-peptide, and **xtz2**-peptide. The strong peak at 227 nm is a characteristic feature of trpzip-type peptides and is attributed to exciton coupling of the closely packed tryptophan residues.[1] Only in the presence of trifluoroethanol (TFE) does **xtz2**-peptide exhibit a significant exciton peak. B) CD melting curves at 227 nm for **HP5W4** and **xtz1**-peptide.

During the three rounds of peptide screening, the ratio of YFP-to-CFP fluorescence emission intensity for the most abundant variants increased from 7.0 to 11.0 to 11.5. The structure of **xtz1**-peptide provides insight into the specific molecular interactions that are associated with these increases in FRET efficiency. For example, in round 1, we found that the peptide sequences that gave rise to the highest FRET efficiency had Trp at position 3 and Lys or Arg at the cross-strand diagonal position 18. This preference implied the presence of a Lys3/Trp18 cation–π interaction in **xtz1**-peptide.[17–21] Indeed, analysis of ^{1}H chemical-shift deviations in **xtz1**-peptide revealed that the side chain of Lys18 is significantly more shielded than any other

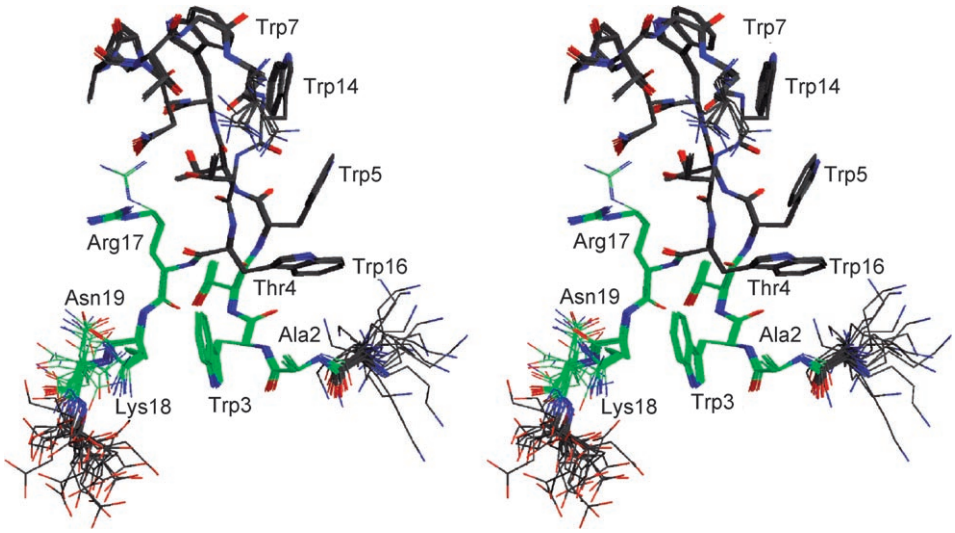

Figure 2. Stereoview of the NMR structure ensemble of **xtz1**-peptide with labels indicating the four tryptophan residues derived from **HP5W4** and the six residues identified at randomized positions. Coordinates have been deposited in the Protein Data Bank (ID: 2ORU)

non-Trp residue in the peptide (Figure S2). This result is consistent with the observed packing of the Lys18 side chain against the indole of Trp3 in a canonical cation–π interaction (Figure 3A).[17,18] In round 2, there was a strong preference for

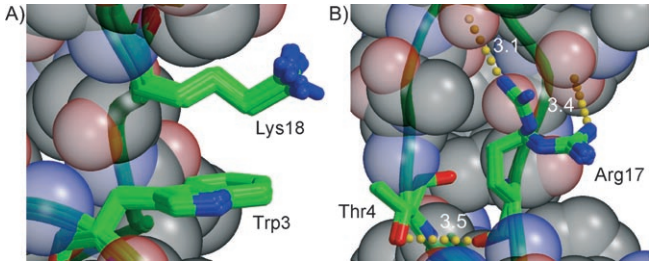

Figure 3. A) Detail of the Trp3/Lys18 cation–π interaction in all 20 structures of the ensemble. B) Detail of the Thr4/Arg17 pair with important hydrogen-bond interactions indicated with a dashed line. Typical O–O or O–N distances [Å] are shown in white. The remainder of the peptide is shown as a single representative space-filling structure.

small residues (such as Ala and Ser) at position 2 and Asn at position 19. Rationalization of the position 19 preference is difficult, since the side chain is disordered in the NMR ensemble (Figure 2). For position 2, the structure suggests that the preference could be attributed to steric hindrance between the larger side chains and the indole rings of Trp3 and Trp16 (Figure 2). In round 3, a strong preference for Thr or Ser at position 4 and Arg or Lys at position 17 was observed. We speculate that only Thr or Ser at position 4 would be able to participate in an apparent cross-strand hydrogen bond between the side-chain hydroxyl and the main-chain carbonyl of Arg17 (3.5 Å, Figure 3B). The side chain of Arg17 extends along the groove between the two strands of the β-hairpin, making hydrophobic contacts with Thr4 and Thr6 and hydrogen bonds to the main-chain carbonyl of Trp7 (3.1 Å, Figure 3B) and the side-chain hydroxyl of Thr15 (3.4 Å, Figure 3B). A lysine at this position could presumably make similar contacts.

We surmise that the combination of residues selected at the targeted positions of **xtz1**-peptide is particularly effective at holding the ends of the peptide in close proximity and minimizing the average distance between the fused CFP and YFP so that they exhibit more efficient FRET. In other words, these residues are participating in specific molecular interactions that stabilize a well-folded conformation of the "extension" that we have appended onto the parent hairpin, **HP5W4**. Why then does this additional stabilization not manifest itself as a higher melting temperature for **xtz1**-peptide relative to **HP5W4** (Figure 1B)? We speculate that the answer lies with the fact that the fitness criterion used during library screening was high FRET efficiency, not overall thermal stability of the hairpin fold. Consequently, if the fraction of peptide folded at ambient temperature is not significantly diminished, a partially destabilized hairpin could still be "improved" by the FRET fitness criterion. We expect that this situation is occurring in **xtz1**-peptide; overall fold stability has been partially sacrificed in order to maximize FRET efficiency. Specifically, the Lys3/Trp18 cation–π interaction has introduced conformational restrictions that force the indole side chains of Trp5 and Trp14 to pack in a conformation that is of higher energy than the conformation they adopt in **HP5W4**. It has previously been suggested that there are intrinsic limits on β-strand length for some sequences.[22] Since the extension present in **xtz1**-peptide is not in a β-sheet conformation, our results neither support nor controvert this suggestion.

Having established that **xtz1**-peptide was highly folded in *E. coli* and in vitro, we transfected HeLa cells with **tz1**, **xtz1**, **xtz2**, and **xtz3** to see whether the conformational stability would be retained in mammalian cells. Fluorescence imaging revealed **tz1** and **xtz1** to be evenly distributed throughout the cell (Figure 4A and B) and exhibiting FRET efficiencies that closely paralleled our in vitro results (Figure 4D). In contrast, much of the **xtz2** protein formed a low FRET efficiency (comparable to the unstructured peptide **xtz3**) aggregate in the apparent vicinity of the endoplasmic reticulum (Figure 4C). The low FRET efficiency of the aggregated protein might be due to either a lack of peptide structure in the fusion protein or a loss of FRET due to proteolysis. The **xtz2** excluded from the aggregate had a FRET efficiency similar to the in vitro value.

This conclusive demonstration of the efficacy of in vivo screening provides researchers with a powerful new tool in the search for folded proteins.[6–8] In particular, we expect this ap-

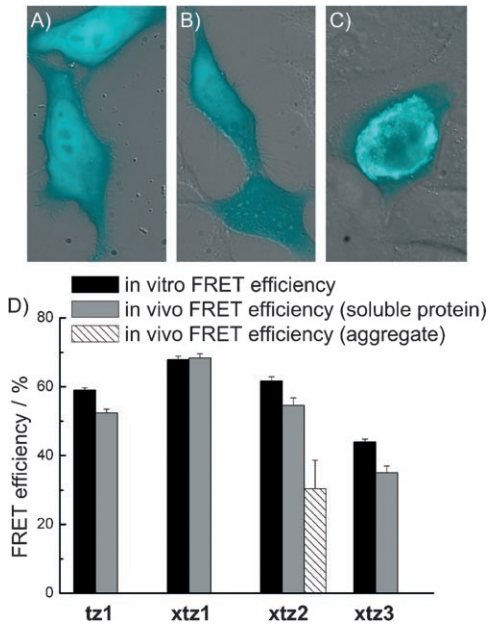

Figure 4. Fluorescence imaging of the FRET constructs A) **tz1**, B) **xtz1**, and C) **xtz2** expressed in HeLa cells. Each panel is a composite of the CFP fluorescence and a differential interference contrast image. While the fluorescence of **tz1** and **xtz1** is distributed evenly throughout the cell, that of **xtz2** appears to aggregate in the vicinity of the nuclear membrane. D) FRET efficiencies determined for proteins in vitro and in HeLa cells.

proach will facilitate the discovery of new scaffolds for intracellular molecular recognition and provide further insight into the structural determinants of peptides' in vivo conformational stability.

Acknowledgements

This research was supported by grants from the University of Alberta, CFI, NSERC, and Alberta Ingenuity. We thank Wayne Moffat, Stephen Neal, and the University of Alberta MBSU for technical assistance. The authors are grateful to N. H. Anderson and R. M. Fesinmeyer for providing us with the stereochemical assignments for HP5W4. R.E.C. holds a Canada Research Chair in Bioanalytical Chemistry.

Keywords: biophysical methods · fluorescence · FRET · peptides · protein engineering

[1] A. G. Cochran, N. J. Skelton, M. A. Starovasnik, *Proc. Natl. Acad. Sci. USA* **2001**, *98*, 5578 – 5583.
[2] N. J. Skelton, T. Blandl, S. J. Russell, M. A. Starovasnik, A. G. Cochran, *Spectroscopy (Amsterdam, Neth.)* **2003**, *17*, 213 – 230.
[3] S. J. Russell, T. Blandl, N. J. Skelton, A. G. Cochran, *J. Am. Chem. Soc.* **2003**, *125*, 388 – 395.
[4] B. Ciani, M. Jourdan, M. S. Searle, *J. Am. Chem. Soc.* **2003**, *125*, 9038 – 9047.
[5] R. M. Fesinmeyer, F. M. Hudson, N. H. Andersen, *J. Am. Chem. Soc.* **2004**, *126*, 7238 – 7243.
[6] M. T. Pastor, M. Lopez de La Paz, E. Lacroix, L. Serrano, E. Perez-Paya, *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 614 – 619.
[7] T. J. Magliery, L. Regan, *Eur. J. Biochem.* **2004**, *271*, 1595 – 1608.
[8] A. L. Watters, D. Baker, *Eur. J. Biochem.* **2004**, *271*, 1615 – 1622.
[9] Z. Cheng, R. E. Campbell, *ChemBioChem* **2006**, *7*, 1147 – 1150.
[10] 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.
[11] J. P. Ulmschneider, W. L. Jorgensen, *J. Am. Chem. Soc.* **2004**, *126*, 1849 – 1857.
[12] W. W. Streicher, G. I. Makhatadze, *J. Am. Chem. Soc.* **2006**, *128*, 30 – 31.
[13] C. D. Snow, L. Qiu, D. Du, F. Gai, S. J. Hagen, V. S. Pande, *Proc. Natl. Acad. Sci. USA* **2004**, *101*, 4077 – 4082.
[14] Z. Cheng, R. E. Campbell, *SPIE* (Eds.: S. Achilefu, D. J. Bornhop, R. Raghavachari, A. P. Savitsky, R. M. Wachter), *Vol. 6449*, **2007**, pp. 64490S.
[15] B. Philipps, J. Hennecke, R. Glockshuber, *J. Mol. Biol.* **2003**, *327*, 239 – 249.
[16] G. S. Waldo, B. M. Standish, J. Berendzen, T. C. Terwilliger, *Nat. Biotechnol.* **1999**, *17*, 691 – 695.
[17] J. P. Gallivan, D. A. Dougherty, *Proc. Natl. Acad. Sci. USA* **1999**, *96*, 9459 – 9464.
[18] C. D. Tatko, M. L. Waters, *Protein Sci.* **2003**, *12*, 2443 – 2452.
[19] C. D. Tatko, M. L. Waters, *J. Am. Chem. Soc.* **2004**, *126*, 2028 – 2034.
[20] R. M. Hughes, M. L. Waters, *J. Am. Chem. Soc.* **2005**, *127*, 6518 – 6519.
[21] R. M. Hughes, M. L. Waters, *J. Am. Chem. Soc.* **2006**, *128*, 13586 – 13591.
[22] H. E. Stanger, F. A. Syud, J. F. Espinosa, I. Giriatt, T. Muir, S. H. Gellman, *Proc. Natl. Acad. Sci. USA* **2001**, *98*, 12015 – 12020.

Received: December 23, 2006
Published online on April 24, 2007

