Sequence Requirements of the GPNG \(\beta\)-Turn of the Ecballium elaterium Trypsin Inhibitor II Explored by Combinatorial Library Screening*

(Received for publication, March 16, 1999, and in revised form, May 12, 1999)

Alexander Wentzel‡, Andreas Christmann‡, Ralph Krätzner§, and Harald Kolmar¶¶

From the ‡Department of Molecular Genetics and Preparative Molecular Biology, Institute for Microbiology and Genetics and the §Institute for Anorganic Chemistry, Georg-August-University, 37077 Göttingen, Federal Republic of Germany

The Ecballium elaterium trypsin inhibitor II (EETI-II) contains 28 amino acids and three disulfides forming a cystine knot. Reduced EETI-II refolds spontaneously and quantitatively in vitro and regains its native structure. Due to its high propensity to form a reverse turn, the GPNG sequence of segment 22–25 comprising a \(\beta\)-turn in native EETI-II is a possible candidate for a folding initiation site. We generated a molecular repertoire of EETI-II variants with variegated 22–25 tetrapeptide sequences and presented these proteins on the outer membrane of Escherichia coli cells via fusion to the Iga\(_\text{a}\) autotransporter. Functional trypsin-binding variants were selected by combination of magnetic and fluorescence-activated cell sorting. At least 1–5% of all possible tetrapeptide sequences were compatible with formation of the correct three disulfides. Occurrence of amino acid residues in functional variants is positively correlated with their propensity to be generally found in \(\beta\)-turns. The folding pathway of two selected variants, EETI-\(\beta\)^{EDED}e and EETI-\(\beta\)^{NNNN}, was found to be indistinguishable from EETI-II and occurs through formation of a stable 2-disulfide intermediate. Substantial amounts of misfolded byproducts, however, were obtained upon refolding of these variants corroborating the importance of the wild type EETI-II GPNG sequence to direct quantitative formation of the cystine knot architecture.

Numerous small proteins, typically not longer than 40 residues in length, share a common structural motif consisting of a cystine knot and a small triple-stranded \(\beta\)-sheet. Members of the “knottin” family (1) of small proteins have a common architecture, but diverse biological activities and negligible amino acid sequence identity. Examples are (i) \(\omega\)-conotoxin MVIIa, a 26-residue polypeptide found in the venom of the cone snail Conus magus, which acts as a neurotoxin by its high affinity binding to voltage-gated Ca\(^{2+}\) channels (2); (ii) potato carboxypeptidase inhibitor (PCI), a 39-amino acid peptide (3); and (iii) EETI-II from the squirting cucumber Ecballium elaterium, a member of the squash family of protease inhibitors (4). These proteins are mainly stabilized by three intramolecular disulfide bonds, where the first cysteine residue in the polypeptide chain is connected with the fourth, the second with the fifth, and the third with the sixth (Fig. 1, top panel). The cystine knot formed by these three disulfide bonds is defined by a ring formed by the first and the second disulfide bond in the peptide sequence and the intervening polypeptide backbone, through which the third disulfide bond passes (Fig. 1, bottom panel). Structural alignment of several cystine knot proteins revealed that the second and third disulfide bond together with the three \(\beta\)-strands that are interconnected by these disulfide bridges superimpose very well (5–7). In contrast, the first disulfide bond shows greater structural variation. Therefore, it was proposed that the region stabilized by the other two bridges forms a 2-disulfide motif serving as a basic scaffold (5, 8). This view is supported by the finding that the cellulose binding domain of the fungal enzyme cellobiohydrolase from Trichoderma reesei displays the typical cystine knot fold, but contains the central two pairs of cysteine residues only, and lacks the cysteine residues forming the first disulfide bond (9).

Cucurbitaceae seeds have proven to be a rich source of trypsin inhibitors with cystine knot folding motif, thereby defining a family of serine protease inhibitors known as the squash family (10). All members possess an extended amino-terminal inhibitor loop, which is tethered at its amino and carboxyl termini to the cystine knot framework. A structurally and functionally well characterized member of the squash inhibitor family is the E. elaterium trypsin inhibitor II (EETI-II) (1, 4). The major features of the EETI-II secondary structure are a short \(3_{10}\)-helix for sequence 11–15, a \(\beta\)-turn 16–19, and a triple antiparallel \(\beta\)-sheet 20–28, with a \(\beta\)-turn formed by residues 22–25 (4) (Fig. 1). During the folding of EETI-II, formation of a rigid core, which contains two native disulfide bonds (C9-C21, C15-C27) precedes inhibitor loop anchoring. This core, which comprises residues 9–28, appears to be the direct precursor of the natural, fully oxidized product and is structurally closely related to it (8).

Two-dimensional NMR study of an EETI-II variant, where all six cysteine residues have been replaced by serine, revealed presence of native-like secondary structures for segments 10–15 (\(3_{10}\)-helix) and \(\beta\)-turns 16–19 and 22–25, but no native tertiary interactions were observed (12). Hence, it was hypothesized that these native-like local conformations could play a major role early in the folding of EETI-II (12). Folding of EETI-II was found to be a clean and quantitative process (8). Unlike any other protein of the cystine knot family, EETI-II contains the sequence GPNG in the 22–25 \(\beta\)-turn segment connecting \(\beta\)-strands 2 and 3 (Fig. 1). The quantitative folding of EETI-II has been mainly attributed to the high \(\beta\)-turn propensity of this tetrapeptide sequence (13, 14), which may facili-
FIG. 1. Primary and three-dimensional structure of EETI-II. A, amino acid sequence of the EETI-II peptide. The disulfide bonds found in the native peptide are indicated by solid lines linking the Cys residues. B, the ribbon diagram of mature EETI-II was drawn with the program MOLSCRIPT (11) from the atomic coordinates of one of the two-dimensional NMR solution structures (4). The β-turn segment of residues 22–25 is indicated as a bold ribbon.

ite association of β-strands in early folding stages of EETI-II followed by covalent fixation of the tertiary fold by disulfide bond formation (8). In order to investigate the influence of the 22–25 turn sequence on the folding of EETI-II, we have generated a molecular repertoire of EETI-II variants with variegated 22–25 turn sequences and displayed this repertoire on the surface of Escherichia coli cells. This library was then selected for binding to trypsin to assess whether we could generate EETI-II mutants that allow formation of the correctly disulfide-bonded cystine knot framework. We have obtained numerous inhibitory active EETI-II derivatives and analyzed the in vitro refolding kinetics of selected variants compared with the wild type protein.

EXPERIMENTAL PROCEDURES

Strains and Reagents—E. coli strain BMH71–18 dbA (F’ lacZam15, proA’ B’; lac-proAB, supE, thi, dbA) was constructed by P1 transduction (15) of dbAach1 from JC8609 (16) into BMH71–18. Biotinylated bovine trypsin was obtained from Sigma. Streptavidin, R-phycocerythrin conjugate was purchased from Molecular Probes. All other chemicals were of analytical grade and obtained from Sigma. Synthetic oligonucleotides were AW-blabio (biotin-5’-CTCACCCGAAAGGCTGTGTT, AW-End (5’-GGCGGGATCCAGCGGAG), betaturn (5’-GGCGGGATCCAGCGGAG), IgAselo (5’-GGCCCTCTAGATGAGACGGAATCGTGTAT), PLZMET (5’-GGCCGATCCTGGCGCAGGCAGGAAAGGAGCAAGCGGAGCAGCAGCC, N = A, C, G, T; S = G, C), IgAsup (5’-GGCGGATTCGCGGAAAGGCAGGAACGGAATCGTGTAT), G-terminus (5’-GCGCTCTAGATTAGAACGGAATCGTGTAT), RSPX (5’-GCGGATCCCTGTGCCCTGTGCAAGGGAGCAAGCGGAGCGGGAGCAGCAGCAGCCGAGCAGGAAAGGAGCAAGCGGAGCAGCAGCC), JuFoup (5’-GGCGGATTCGCGGAAAGGCAGGAACGGAATCGTGTAT), and RSFX (5’-GCGGATCCCTGTGCCCTGTGCAAGGGAGCAAGCGGAGCGGGAGCAGCAGCAGCCGAGCAGGAAAGGAGCAAGCGGAGCAGCAGCC). Folding Experiments—Reframing of the crude mixture, electrocompetent cells of E. coli as soluble periplasmic proteins via fusion to maltose-binding protein by cloning the respective EETI-II gene as an AvaI/BamHI fragment into similarly digested pMEETI-II plasmid. To obtain sufficient amounts of the respective variant for refolding experiments, the respective EETI-II gene was amplified by PCR using the primers PLZMET and RSFX, cleaved with EcoRI and XbaI, and ligated into similarly digested pLZPB1 (18). E. coli W3110 (18) was used as expression host. Samples were obtained from French press lysates of bacterial liquid cultures overnight grown at 37 °C in rich medium with 1% isopropyl-1-thio-β-galactopyranoside and ampicillin (100 µg/ml) added. These were dissolved in 15 ml of 70% formic acid/d/gelatin and inclusion bodies. After 24 h of incubation with cyanoen bromide (150 mg/ml of inclusion bodies), proteins were precipitated with two volumes of diethylether. The precipitate was dissolved by sonication in 8 ml urea, 100 mM NaCl, 100 mM sodium phosphate buffer, pH 8.0, and subjected to immobilized metal ion adsorption chromatography. To the EETI-II containing fractions from the imidazole step gradient elution, dithiothreitol was added at 50 mM final concentration and samples were subjected to reversed-phase HPLC using the following conditions. The HPLC column was Waters µBondapak C18 (3.9 × 150 mm); solvent A was water, containing 0.1% trifluoroacetic acid; solvent B was acetonitrile containing 0.1% trifluoroacetic acid. The gradient was 10% B to 37% B linear in 16 min at a flow rate of 1 ml/min. The detector wavelength was set to 217 nm. Samples were freeze-dried and dissolved in 10 mM HCl. Protein concentration was determined by derivatization of fully reduced EETI-II with 5,5’-dithio-bis-(2-nitrobenzoic acid) (19). The probability of obtaining a particular codon was then calculated as shown in Equation 1, where \( F_N \) is the scored number of nucleotide i codons and \( P_{i\text{MTK}} \) is the expected number.

\[
P_{i\text{MTK}} = \frac{N_i}{\sum_i N_i}
\]

(1)

(2)

P(z|N_1 S_1) = \frac{1}{\sqrt{2\pi} \sigma_1} e^{-\frac{(z-\mu_1)^2}{2 \sigma_1^2}} \frac{1}{\sqrt{2\pi} \sigma_2} e^{-\frac{(z-\mu_2)^2}{2 \sigma_2^2}} \frac{1}{\sqrt{2\pi} \sigma_3} e^{-\frac{(z-\mu_3)^2}{2 \sigma_3^2}}

(2)

The probability for a particular amino acid residue aa defined by i codons is given as shown in Equation 3.
R-phycoerythrin conjugate. 200,000 cells were run through a MoFlo cell sorter (Cytomation). The percentage of the cells that fell into that window is indicated above.

Three non-native cysteine residues were encoded in the obtained data set was found to be zero, it was increased to 0.5 so that \( F(aa) \) was never zero.

**RESULTS**

Display of EETI-II Variants on the Surface of E. coli Cells—In squash inhibitors, the inhibitor loop is held in place by amino- and carboxyl-terminal cysteine residues tethering it through disulfide bond formation to the structural framework. Removal of any of the three disulfide bonds in EETI-II by cysteine to serine replacement (C2S, C2S1S, C2S7S) was found to abolish cysteine knot formation and trypsin binding (data not shown). Likewise, misfolded variants of Cucurbita maxima trypsin inhibitor 1 (CMTI-I) containing three non-native intramolecular disulfide bonds had no inhibitory activity (21). These findings indicated that trypsin binding could be used as an indicative for manifestation of the correct cystine knot framework.

Initial experiments to present EETI-II and derived variants on the surface of phage in order to enrich functional variants via binding to immobilized trypsin were unsuccessful (data not shown). As a practicable alternative, we opted for the presentation of a repertoire of EETI-II variants on the outer membrane of E. coli cells. Cells displaying an EETI-II variant, which retains the ability to bind trypsin can then be isolated by labeling with biotinylated trypsin followed by specific fluorescence labeling with streptavidin, R-phycoerythrin conjugate and enrichment of binders by magnetic and fluorescence-activated cell sorting.

**Construction of an EETI-II Library and Selection of Trypsin-binding Proteins—** A library of EETI-II genes was generated by randomizing PCR, where the EETI-II codons 22–25 were randomized using a NN(G/C) coding scheme. These variant EETI-II genes were ligated to plasmid pHK-EETI-Iga under lac promoter/operator control (23). Efficient translocation of a passenger domain fused to Iga, which contains an intramolecular disulfide bond, was achieved only in an E. coli mutant carrying a defect in the dsbA gene encoding periplasmic disulfide oxidoreductase. This finding suggests that translocation through the outer membrane requires an unfolded conformation of the passenger domain and can be blocked by disulfide loop formation (24). Fluorescence microscopy of BHMH71–18 dsbA cells containing pHK-EETI-Iga treated with biotinylated trypsin, followed by incubation with streptavidin, R-phycoerythrin revealed that only a small fraction of the cell population (approximately 5–10%) were phycoerythrin-labeled (No fluorescent cells were detected (data not shown)). A majority of the cells that carried a phycoerythrin label could be simultaneously stained with propidium iodide (data not shown), which preferentially stains non-viable cells (25). Whether the high mortality of EETI-II presenting cells is the cause or the result of the cell surface presentation of the β-lactamase-EETI-II fusion protein is currently not clear. Nevertheless, the finding that EETI-II producing cells labeled with biotinylated trypsin/streptavidin, R-phycoerythrin conjugate were distinguishable by FACS from cells producing an EETI-II variant that is unable to bind trypsin (Fig. 2) prompted us to use indirect fluorescent trypsin labeling followed by magnetic and fluorescent-activated cell sorting to identify and isolate cells carrying a functional EETI-II protein.

**FIG. 2.** FACS histograms of recombinant BMH71–18 dsbA to determine E. coli cell surface presentation of EETI-II. Flow cytometric analysis of BMH71–18 dsbA harboring pHK-EETI-Iga (panel A) or pHK-EETI-Cy5 and-Iga (panel B). Cells displaying EETI-II wild type (A) or EETI-Cy5 (B), a variant that is unable to bind trypsin, were incubated with biotinylated trypsin and, after washing, labeled with streptavidin, R-phycoerythrin conjugate. 200,000 cells were run through a MoFlo cell sorter (Cytomation). The bar in each graph indicates a window ranging from 350 to 650 relative fluorescence units. The percentage of the cells that fall into that window is indicated above the bar.

\[
P(aa) = \sum_i P(N_i, S_i) \quad \text{(Eq. 3)}
\]

The score for a particular residue aa is given by Equation 4 (20),

\[
\text{Score}(aa) = \ln \left( \frac{F(aa)}{P(aa)} \right) \quad \text{(Eq. 4)}
\]

where \( F(aa) \) is the observed frequency of a residue found in region 22–25 of the set of trypsin-binding EETI-II variants. If the number of residue aa in the obtained data set was found to be zero, it was increased to 0.5 so that \( F(aa) \) was never zero.
positive event. The percentage of cells counted as positive event is indicated as a positive value overrepresentation. The 20 amino acids were ordered according to their overall amino acid, a positive value indicates underrepresentation of the respective residue in the initial unselected library. A negative value over indicates underrepresentation of the respective EETI-II variants (data not shown). A score was calculated for each residue, which reflects the deviation of the occurrence of a particular amino acid in the data set of residues 22–25 of the 22–25 region before and after enrichment of trypsin-binding activity. To obtain sufficient amounts of protein for the study of refolding kinetics, proteins were produced in soluble form via secretion into the E. coli periplasmic space by fusion to maltose-binding protein. These proteins were purified from osmotic shock fluid of the bacterial cells by immobilized metal ion adsorption chromatography. Very similar dissociation constants toward trypsin were measured for MalE-EETI-II wild type and the selected library variants (Table I).

We have further characterized the disulfide-coupled folding of EETI-II wild type, of the trypsin-binding variants EETI-II variants randomized in the segment 22–25. Cells were labeled with biotinylated trypsin and, after washing, successively incubated with streptavidin-coupled colloidal superparamagnetic microbeads and streptavidin, R-phycocyanin conjugate. Prior to FACS, labeled cells were resorted by magnetic separation by passage of the cell population through a high gradient magnetic separation column. Since the colloidal magnetic particles are too small to be detected by the flow cytometer, cells could be immediately subjected to FACS analysis. C, mixture of cells obtained from magnetic cell sorting after first round of fluorescence-activated cell sorting. D, mixture of cells from C, after second round of FACS.

To quantitatively assess the trypsin binding activity of the selected peptides, six EETI-II variants together with the EETI-II wild type protein were produced in soluble form via secretion into the E. coli periplasmic space by fusion to maltose-binding protein. These proteins were purified from osmotic shock fluid of the bacterial cells by immobilized metal ion adsorption chromatography. Very similar dissociation constants toward trypsin were measured for MalE-EETI-II wild type and the selected library variants (Table I).

We have further characterized the disulfide-coupled folding of EETI-II wild type, of the trypsin-binding variants EETI-II variants randomized in the segment 22–25. Cells were labeled with biotinylated trypsin and, after washing, successively incubated with streptavidin-coupled colloidal superparamagnetic microbeads and streptavidin, R-phycocyanin conjugate. Prior to FACS, labeled cells were resorted by magnetic separation by passage of the cell population through a high gradient magnetic separation column. Since the colloidal magnetic particles are too small to be detected by the flow cytometer, cells could be immediately subjected to FACS analysis. C, mixture of cells obtained from magnetic cell sorting after first round of fluorescence-activated cell sorting. D, mixture of cells from C, after second round of FACS.

To quantitatively assess the trypsin binding activity of the selected peptides, six EETI-II variants together with the EETI-II wild type protein were produced in soluble form via secretion into the E. coli periplasmic space by fusion to maltose-binding protein. These proteins were purified from osmotic shock fluid of the bacterial cells by immobilized metal ion adsorption chromatography. Very similar dissociation constants toward trypsin were measured for MalE-EETI-II wild type and the selected library variants (Table I).

We have further characterized the disulfide-coupled folding of EETI-II wild type, of the trypsin-binding variants EETI-II variants randomized in the segment 22–25. Cells were labeled with biotinylated trypsin and, after washing, successively incubated with streptavidin-coupled colloidal superparamagnetic microbeads and streptavidin, R-phycocyanin conjugate. Prior to FACS, labeled cells were resorted by magnetic separation by passage of the cell population through a high gradient magnetic separation column. Since the colloidal magnetic particles are too small to be detected by the flow cytometer, cells could be immediately subjected to FACS analysis. C, mixture of cells obtained from magnetic cell sorting after first round of fluorescence-activated cell sorting. D, mixture of cells from C, after second round of FACS.

![Figure 3](image-url) **Figure 3. Histogram data from FACS enrichment of trypsin-binding cells.** The bar in each graph represents the sorting gate defined as positive event. The percentage of cells counted as positive event is indicated above the bar. A, flow cytometric analysis of 200,000 cells from library of E. coli cells presenting EETI-II variants randomized in the segment 22–25. Cells were labeled with biotinylated trypsin and, after washing, successively incubated with streptavidin-coupled colloidal superparamagnetic microbeads and streptavidin, R-phycocyanin conjugate. B, cells after enrichment through magnetic cell sorting. Prior to FACS, labeled cells were resorted by magnetic separation by passage of the cell population through a high gradient magnetic separation column. C, mixture of cells obtained from magnetic cell sorting after first round of fluorescence-activated cell sorting. D, mixture of cells from C, after second round of FACS.

- Asparagine, which has the second highest overall β-turn potential (26/180). Exceptions to this correlation are cysteine and proline.

We have further characterized the disulfide-coupled folding of EETI-II wild type, of the trypsin-binding variants EETI-II variants randomized in the segment 22–25. Cells were labeled with biotinylated trypsin and, after washing, successively incubated with streptavidin-coupled colloidal superparamagnetic microbeads and streptavidin, R-phycocyanin conjugate. Prior to FACS, labeled cells were resorted by magnetic separation by passage of the cell population through a high gradient magnetic separation column. Since the colloidal magnetic particles are too small to be detected by the flow cytometer, cells could be immediately subjected to FACS analysis. C, mixture of cells obtained from magnetic cell sorting after first round of fluorescence-activated cell sorting. D, mixture of cells from C, after second round of FACS.
The combinatorial library of EETI-II variants differing in the amino acid sequence of the 22–25 turn region was presented on the surface of E. coli cells. Cells presenting an EETI-II variant capable of trypsin binding were trypsin-labeled and isolated by combination of magnetic and fluorescence-activated cell sorting. The amino acid sequences of 30 EETI-II variants were deduced from the nucleotide sequences of the respective EETI-II genes residing in cell surface display vector pHKBlα-EETI-Igαβ. The EETI-II variants displayed in the first column were produced as maltose-binding protein fusions and purified to homogeneity by metal chelate affinity chromatography. Varying amounts of MalE-EETI fusion protein were incubated with 5 nM trypsin for 10 min at 37 °C in 100 μl of 50 mM Tris-HCl, pH 7.5, 150 mM NaCl, 5 mM CaCl2. Residual trypsin activity was monitored by following the linear release at 412 nm of 4-nitroaniline from added Boc-Leu-Gly-Arg-pNA (0.4 μM). The dissociation constants \( K_d \) were calculated as described (27).

### TABLE I

| 22–25 | \( K_d \) |
|-------|----------|
| GPNG  | \( 6.09 \pm 1.97 \times 10^{-9} \) |
| NEDE  | \( 3.00 \pm 0.55 \times 10^{-9} \) |
| TNNK  | \( 7.53 \pm 3.37 \times 10^{-9} \) |
| LDET  | \( 9.97 \pm 0.91 \times 10^{-9} \) |
| NNTD  | \( 13.79 \pm 0.16 \times 10^{-9} \) |
| ERNE  | \( 21.14 \pm 8.05 \times 10^{-9} \) |
| HKRT  | \( 9.83 \pm 3.16 \times 10^{-9} \) |

### DISCUSSION

Work in other laboratories has previously demonstrated that several of the squash inhibitors can correctly form their disulfides in vitro with yields varying among different species (1, 21, 28). EETI-II folds quantitatively simply by air oxidation, while another inhibitor of the squash family under similar conditions yields only about 60–80% of native protein (21). We have recently shown that EETI-II folds in high yield in vivo when expressed in E. coli via secretion into the periplasmic space, where oxidizing conditions prevail.2 Comparison of the amino acid sequences of \( \beta \)-strands connecting turns of squash inhibitors and other cystine knot proteins reveals an unique feature of EETI-II, which is the GPNG sequence of segment 22–25. This segment forms a typical type I \( \beta \)-turn connecting \( \beta \)-strands 2 and 3. Type I is a frequently found \( \beta \)-turn type with near-helical \( \phi, \psi \) values that can be adopted by any amino acid (14). The structure of the corresponding region in C. maxima trypsin inhibitor I, \( \omega \)-conotoxin, and kalata B1, a 29-residue peptide from the tropical plant Oldenlandia affinis DC with cysteine knot fold shows similar linkage of \( \beta \)-strands 2 and 3 by a type I turn, where the central residues lie in the \( \alpha \)-region of a

**Fig. 4.** Occurrence of amino acid residues in the 22–25 segment of trypsin-binding EETI-II variants. For each residue, the score was calculated as described under “Experimental Procedures,” which reflects the preference to be found in the set of trypsin-binding EETI-II variants compared with the unselected library. Residues are averaged over a large data set of all classified turn types (14). These indicate the preference for each residue to occur in a \( \beta \)-turn generally and are averaged over a large data set of \( \beta \)-turns from various proteins and over all classified turn types (14).

To study the folding kinetics of EETI-II, EETI-\( \beta^{\text{NEDE}} \), EETI-\( \beta^{\text{TNNK}} \), and EETI-\( \beta^{\text{ATVF}} \), respectively, completely reduced variants were allowed to refold by air oxidation in 100 mM \( \text{NH}_4\text{HCO}_3 \), \( \text{pH} \) 9.1. Samples were withdrawn at various time points from a folding reaction, quenched by mixing with 1/20 volume of concentrated phosphoric acid, and subsequently analyzed by HPLC. Fig. 5 shows HPLC profiles of the disulfide-bonded forms trapped after various times of refolding. Under the chromatographic conditions used here, molecules with the largest fraction of exposed nonpolar surface area are expected to elute late from the column. In accordance with this expectation, the fully reduced peptide of each variant eluted later than any other species containing disulfide bonds, while the native form eluted first from the column. With EETI-II wild type, over 90% of the reduced form yielded the native peptide after overnight oxidation. EETI-\( \beta^{\text{NEDE}} \) and EETI-\( \beta^{\text{TNNK}} \) however, were only obtained in yields of approximately 65%. The non-trypsin-binding EETI-\( \beta^{\text{ATVF}} \) behaved in a completely different manner. This peptide eluted in its reduced form from the reversed-phase column in several overlapping peaks, which indicates aggregation and oligomer formation. Both the reduced and the oxidized forms of the protein absorbed to a large extent to walls of the reaction vial. Furthermore, refolding of EETI-\( \beta^{\text{ATVF}} \) resulted in a large number of various species, and no predominant HPLC peak corresponding to the native form was found (data not shown).

For EETI-II wild type refolding, the distribution of species was dominated within 5 min by a major component. This folding intermediate was identified by Castro and co-workers as dihydro-2,19 EETI-II, a species that lacks the (C2-C19) disulfide bond and is structurally closely related to the native molecule (8). The peak of the intermediate disappeared within the next 120 min to the benefit of the peak corresponding to the native product. A predominant folding intermediate with HPLC retention times very similar to the dihydro-2,19 intermediate of wild type EETI-II emerged in the same manner during the folding reaction of EETI-\( \beta^{\text{NEDE}} \) and EETI-\( \beta^{\text{TNNK}} \) together with several other species. This indicates that folding of these variants most likely proceeds through the same folding pathway as the wild type EETI-II molecule with the dihydro-2,19 species being the major intermediate. Of all three folding reactions, the rate-limiting step is the formation of the (C2-C19) disulfide bond. Compared with wild type EETI-II, EETI-\( \beta^{\text{NEDE}} \) folds slightly and EETI-\( \beta^{\text{TNNK}} \) exceedingly more sluggish, requiring overnight incubation for formation of the native product.
Ramachandran plot (5). Unlike these proteins and all other 20 known squash inhibitors, EETI-II contains at position $i + 1$ of the $\beta$-turn a proline residue. Proline is by far the most favored residue at the second position because of the restriction of its $\phi$ angle to about $-60^\circ$. The other preferred residues at the second position, glutamic acid and serine, can stabilize the $\phi$ angles by forming a hydrogen bond between their side chain oxygen atoms and the main chain amide (14). Indeed, many of the squash inhibitors contain a glutamic acid or serine residue at the respective position (11/21).

The high turn forming propensity of the GPNG sequence was corroborated in the EETI-II sequence context by the finding that in our experimental system only about 5–10% of positive clones with their propensity to be generally found in $\beta$-turns. Hutchinson and Thornton have identified and classified 3899 $\beta$-turns using a nonhomologous data set of 205 protein chains (14). These data were used to derive $\beta$-turn overall and positional potentials for the different turn types. As shown in Fig. 4, a correlation between the overall $\beta$-turn potential of a particular residue in the whole data base of high resolution structures and its occurrence in the $\beta$-turn of folded EETI-II variants exists. Exceptions to this rule are cysteine and proline. No cysteine residue was found in the sequenced data set of functional EETI-II variants. Sequence analysis of unselected clones revealed the presence of a variant containing the sequence N\textsuperscript{3}EGRRH\textsuperscript{3}, which proved to be nonfunctional (data not shown). This is not unexpected, since the additional cysteine residue in the $\beta$-turn raises the number of possible combinations of three disulfide pairs in a molecule from 15 to 48, thereby expanding the possibilities for misfolded species with non-native disulfide bonds considerably. Proline was only found twice in the sequences of trypsin-binding variants, less frequently than expected from its overall $\beta$-turn potential. In type I turns, proline residue is frequently found at position $i + 1$ of the loop, but occurs rarely at positions $i + 3$ and $i + 4$ (14). In the 22–25 segment of EETI-II variants, proline residue in the $i + 1$ position is obviously not necessarily required and may be even disfavored in positions $i + 3$ and $i + 4$, which might in total account for its relatively low occurrence.

We have investigated the folding kinetics of EETI-II wild type and variants EETI-$\beta$\textsuperscript{NEDE} and EETI-$\beta$\textsuperscript{TNNK} by in vitro refolding the purified fully reduced variants and subsequent acid trapping the products at various time points. Differences in the yields of native protein notwithstanding, the folding pathway of the EETI-$\beta$\textsuperscript{NEDE} and EETI-$\beta$\textsuperscript{TNNK} variants is fundamentally indistinguishable from that of EETI-II wild type protein. With all three proteins, a predominant folding intermediate emerges during early stages of folding. The EETI-II...
Folding of the E. elaterium Trypsin Inhibitor

wild type intermediate has been shown to be the stable (C9-C21, C15-C27) intermediate lacking the (C2-C19) disulfide bridge (8). Two-dimensional NMR studies showed that the intermediate is very similar in structure to the native EETI-II (8). Our finding that dihydro-2,19-EETI-II and the EETI-βNEDE intermediate isolated by acid trapping were both converted to fully folded EETI-II without accumulation of any other intermediate supports the notion that this form is the direct precursor of the natural product (data not shown). Dihydro-2,19-EETI-II was the predominant folding intermediate already after 0.5 min of refolding (data not shown). However, progression from the fully reduced state to the main folding intermediate and to the native product occurs more sluggishly with the two EETI-II β-turn derivatives. The process of (C2-C19) disulfide bond formation represents the major rate-limiting step in the folding of EETI-II and the variants with different β-turn sequences. This may be attributed to the fact that loop crossing has to occur at this last step in the folding pathway. In native EETI-II, the macrocycle made up of two disulfides linking the sequences of C9KQDSDC15 and C21GPNGFC27 is penetrated by the disulfide bridge (C15-C27) thus forming a tight pseudo-knot structure (5). Fixation of the amino-terminal 6-residue inhibitor loop (3–8) by (C2-C19) disulfide bond formation occurs after formation of the structural framework. As a consequence, formation of the cystine knot fold is largely independent of the length and amino acid sequence of the inhibitor loop. This renders EETI-II an ideal scaffold for the presentation of a repertoire of conformationally constrained peptides aimed at isolating variants with novel binding characteristics.2

The folding pathway of EETI-II contrasts remarkably with that observed with other proteins of the cystine knot family, such as the conotoxin MVIIA and PCI (29–32). The folding mechanism of PCI can be dissected into two steps (32). The sequential flow of fully reduced PCI through equilibrated one- and two-disulfide species results in formation of equilibrated scrambled species. In the final and major rate-limiting step, PCI reorganizes to attain the native disulfide structure. From these findings, it was concluded that there is not a predominant folding pathway for PCI (32). Experiments performed with toxins derived from the Conus species reached similar results (29–31). These conotoxins are able to refold correctly with efficiencies ranging from 15% to 50%. The distribution of equilibrated disulfide-bonded species compared with the native form observed under optimal refolding conditions indicate that the stability of the native conformation relative to other forms is only marginal. Furthermore, in contrast to EETI-II, the forms with two native disulfides appear to be largely devoid of folded structure. Both in PCI and in conotoxins, there appears to be little specificity on the formation of the initial disulfides during the folding reaction. Again, the non-native two-disulfide forms are able to form scrambled non-native three-disulfide species which can only interconvert through partial reduction. Hence, without any thiol reshuffling reagents like GSH present, folding progresses to fully oxidized scrambled species which become trapped due to the concomitant loss of free thiols acting as reshuffling catalysts. Folding of EETI-II, however, occurs by air oxidation without accumulation of scrambled species and does not require addition of thiol compounds as catalysts.

Our data show that several thousand of the 160,000 possible tetrapeptide combinations in the 22–25 β-turn segment of EETI-II are compatible with the formation of the correct cystine knot fold. In contrast to EETI-II wild type, however, which refolds nearly quantitatively, the 24-h samples of EETI-βNEDE and EETI-βTNKK refolded by air oxidation contain substantial amounts of misfolded byproducts and yields of correctly folded proteins are only around 65%. This is strikingly similar to the refolding by air oxidation of CMTI-I, which contains a LEHG β-turn sequence (21). In conclusion, the GPNG sequence of EETI-II directs the clean and quantitative formation of the dihydro-2,19 EETI-II intermediate, which is finally oxidized to the native three-disulfide form. Hence, it may be interesting to see whether transplantation of that sequence into the sequence context of other cystine knot protein like PCI or conotoxins might influence their folding pathway and the yield of correctly folded protein.

REFERENCES

1. Le-Nguyen, D., Heitz, A., Chiche, L., Castro, B., Boigegrain, R. A., Favel, A., and Coletti-Previero, M. A. (1996) Biochimie 72, 431–435
2. Kohno, T., Kim, J. I., Kobayashi, K., Kodera, Y., Maeda, T., and Sato, K. (1995) Biochemistry 34, 10256–10265
3. Reeves, D. C., and Lipscomb, W. N. (1982) Proc. Natl. Acad. Sci. U. S. A. 77, 4633–4637
4. Heitz, A., Chiche, L., Le-Nguyen, D., and Castro, B. (1989) Biochemistry 28, 2392–2398
5. Pallaghy, P. K., Nielsen, K. E., Craik, D. J., and Norton, R. (1994) Protein Sci. 3, 1833–1839
6. Isaacs, N. W. (1995) Curr. Biol. 5, 391–395
7. Tamaoki, H., Miura, R., Kasunoki, M., Koyogoku, Y., Kobayashi, Y., and Moroder, L. (1998) Protein Eng. 11, 649–659
8. Le-Nguyen, D., Heitz, A., Chiche, L., El Hajji, M., and Castro, B. (1993) Protein Sci. 2, 165–174
9. Kraulis, P. J., Clore, G. M., Nilges, M., Jones, T. A., Pettersson, G., Knowles, J., and Gronenborn, A. M. (1989) Biochemistry 28, 7241–7257
10. Oltewski, J. (1990) Biochim. Biophys. Acta 1071, 23–28
11. Kraulis, P. J. (1991) J. Appl. Crystallogr. 24, 946–956
12. Heitz, A., Chiche, L., Le-Nguyen, D., and Castro, B. (1995) Eur. J. Biochem. 233, 837–846
13. Chou, P. Y., and Fasman, G. D. (1977) J. Mol. Biol. 115, 135–175
14. Hutchinson, E. G., and Thornton, J. M. (1994) Protein Sci. 3, 2207–2216
15. Miller, J. H. (1972) Experiments in Molecular Genetics, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY
16. Bardwell, J. C., McGovern, K., and Beckwith, J. (1991) Cell 67, 581–589
17. Kramer, W., and Fritz, H.-J. (1987) Methods Enzymol. 155, 350–367
18. Preuss, K. D., Krausha, E., and Scheit, K. H. (1990) Biochim. Biophys. Acta 1015, 215–222
19. Eillman, G. L. (1955) Arch. Biochem. Biophys. 62, 70–77
20. Bowie, J. U., Luthy, R., and Eisenberg, D. (1991) Science 253, 164–170
21. Kojima, S., Katsunori, M., and Miura, K. (1996) Protein Eng. 9, 1241–1246
22. Klauser, T., Pohlner, J., and Meyer, T. F. (1993) BioEssays 15, 649–659
23. Komano, H., Ferrando, E., Hennecke, F., Wippler, J., and Fritz, H.-J. (1992) J. Mol. Biol. 228, 359–365
24. Jose, J., Krauser, J., Krausser, K., and Zuber, T. F. (1996) Gene (Amst.) 178, 107–110
25. Lopez-Amoros, R., Comas, J., and Vives-Rego, J. (1996) Appl. Environ. Microbiol. 61, 2521–2526
26. Millenyi, S., and Pfuger, E. (1992) in Flow Cytometry and Cell Sorting (Radbruch, A., ed) pp. 141–152, Springer Laboratory, Berlin
27. Bietti, J. G. (1984) Biochem. Mol. Biol. 32, 387–397
28. Chen, X.-M., Qian, Y.-W., and Ch., C.-W., Gan, K.-D., Zhang, M.-F., and Chen, C.-Q. (1992) J. Biochem. (Tokyo) 112, 45–51
29. Price-Carter, M., Gray, W. R., and Goldenberg, D. P. (1996) Biochemistry 35, 15537–15546
30. Price-Carter, M., Gray, W. R., and Goldenberg, D. P. (1996) Biochemistry 35, 15547–15557
31. Price-Carter, M., Hull, M. S., and Goldenberg, D. P. (1998) Biochemistry 37, 9851–9861
32. Chang, J.-Y., Canals, F., Schindler, P., Querol, E., and Aviles, F. X. (1994) J. Biol. Chem. 269, 22067–22094