



ACADEMIC
PRESS

Available online at www.sciencedirect.com

SCIENCE @ DIRECT®

Bioorganic Chemistry 31 (2003) 412–424

**BIOORGANIC
CHEMISTRY**

www.elsevier.com/locate/bioorg

Effect of linker sequence on the stability of circularly permuted variants of ribonuclease T1

James B. Garrett, Leisha S. Mullins, and Frank M. Raushel*

Department of Chemistry, Texas A&M University, P.O. Box 30012, College Station, TX 77842-3012, USA

Received 1 May 2003

Abstract

Circularly permuted variants of ribonuclease T1 were constructed with a library of residues covalently linking the original amino and carboxyl terminal ends of the wild-type protein. The library of linking peptides consisted of three amino acids containing any combination of proline, aspartate, asparagine, serine, threonine, tyrosine, alanine, and histidine. Forty two unique linker sequences were isolated and 10 of these mutants were further characterized with regard to catalytic activity and overall thermodynamic stability. The 10 mutants with the different linking sequences (HPD, TPH, DTD, TPD, PYH, PAT, PHP, DSS, SPP, and TPS), in addition to GGG and GPG, were 4.0–6.2 kcal/mol less stable than the wild-type ribonuclease T1. However, these circularly permuted variants were only 0.4–2.6 kcal/mol less stable than the direct parent protein that is missing the disulfide bond connecting residues 2 and 10. The most stable linking peptide was HPD.

© 2003 Elsevier Science (USA). All rights reserved.

1. Introduction

The creation of circular permutations within protein sequences facilitates the direct investigation into the precise role of the linear arrangement of the primary amino acid sequence on the kinetics and thermodynamics of the protein folding process.

* Corresponding author. Fax: 1-979-845-9452.

E-mail address: raushel@tamu.edu (F.M. Raushel).

This strategy, initially developed by Goldenberg and Creighton, involves the reordering of the primary amino acid structure by covalently linking the original amino and carboxy termini of the native protein while forming new termini at alternate sites within the primary structure [1]. Since the pioneering work of Goldenberg and Creighton on bovine pancreatic trypsin inhibitor (BPTI), many circularly permuted protein variants have been constructed and characterized [2–18]. While most of these examples have involved single domain monomeric proteins, circular variants have also been made of multidomain proteins such as T4 lysozyme [5] and β -lactamase [14] and multimeric proteins such as aspartate transcarbamoylase [6] and β -B2-crystallin [18]. Regardless of the original type of topographical fold, all of the circularly permuted protein variants reported to date have adopted the native tertiary structure of the parent protein.

Systematic efforts to explore and determine the structural contexts for the successful insertion of new termini sites have been performed on several proteins including ribonuclease T1¹ [12], aspartate transcarbamoylase [15], the SH3 domain of α -spectrin [11] and 1,3-1,4- β -glucanase H [17]. These studies have found that a remarkably large number of sites exist within these proteins that can structurally accommodate new termini. These investigations have also highlighted the regions within the folded protein that appear to be important for the maintenance and/or initiation of the folding process. For example, every loop and turn within the native fold of ribonuclease T1 can be exploited for the construction of a unique circularly permuted variant with the single exception of the tight β -turn that connects residue Asn-83 with residue Asn-84 [12]. There is indirect evidence to suggest that this site may be involved in the nucleation of the folding cascade for ribonuclease T1 [19].

It is also apparent that there is a great deal of variability in the overall global thermodynamic stability within the family of circularly permuted proteins that can be constructed from a parent protein. It can be clearly demonstrated that the difference in the overall structural stability within the nested set of circularly permuted variants from a single protein is due primarily to the location of the new termini. However, the difference in thermodynamic stability between an individual member within the family of circularly permuted variants and the parent protein cannot entirely be attributed to the change in location of the new termini. In the creation of a circular variant, a new loop or turn must be introduced into the protein when the original amino and carboxy terminal ends are linked together. The new loop or turn can potentially introduce a significant thermodynamic contribution of its own, either positive or negative, to the overall global stability of the circular variant. The modulation of the kinetic and thermodynamic properties for the folding of the new circular variant is thus coupled to the perturbation induced by the connection made to the original termini [20]. It can be presupposed that the best linker connecting the original termini would be one that has no net affect on the overall thermodynamic stability of

¹ *Abbreviations used:* RNase T1, ribonuclease T1; (C2,10A), a mutant of the wild-type RNase T1 where cysteine residues 2 and 10 have been mutated to alanine; cp35S1, a circular permuted variant of RNase where Ser-35 now serves as the amino terminal residue.

the circular variant. However, since a neutral linker cannot be designed a priori, we have now used a combinatorial approach to create a library of linkers that have been introduced into the most stable circular variant of RNase T1.

RNase T1 is a small, globular protein containing 104 residues which cleaves RNA 3' to guanosine sites [21]. The native structure, as determined by X-ray crystallography, consists of two β -sheets and a 4.5 turn α -helix and is stabilized by two disulfide bonds [22]. Initially, two different linker sets, Gly–Pro–Gly and Gly–Gly–Gly were used to create two circularly permuted proteins, cp35S1-gpg and cp35S1-ggg, respectively [7]. The stability of cp35S1-ggg was found to be greater than that of the homolog with the Gly–Pro–Gly linker. For this reason, the Gly–Gly–Gly linker was subsequently used in the creation of all other circular variants. Since all of the circularly permuted proteins were found to be less stable than the parent protein, (C2,10A), a new study was begun to determine the effect of the linker sequence on the global stability of the circularly permuted proteins. The results of this investigation show that the amino acid content of the linker sequence produces a demonstrable effect on the thermodynamic stability of the circular variant into which they are incorporated. Thus, the linker sequence will significantly influence the selection of permissible locations for the insertion of new termini.

2. Materials and methods

2.1. Materials

Urea, GpC, all buffers, Type II-C ribonucleic acid core (RNA), and Type IV Torula yeast RNA were purchased from Sigma. All other reagents and restriction enzymes were obtained from either Promega, Perkin–Elmer, Stratagene, or United States Biochemical Corp. The oligonucleotides that were used for mutagenesis and sequencing were synthesized by the Gene Technology Laboratory of Texas A&M University. The plasmid pMc5TPRTQ and *Escherichia coli* strain WK6 [23] were gifts from Professor C.N. Pace of Texas A&M University.

2.2. Construction of circularly permuted linker variants of RNase T1

The gene for RNase T1 is encoded in the plasmid, pKW01 [7], immediately downstream from the signal peptide portion of the alkaline phosphatase gene, *phoA*, and under the transcriptional control of the *tac* promoter. In this plasmid, the RNase T1 gene is bounded by a *Hind*III site at the 3' end of the gene and a *Sty*I restriction site at the 5' end of the gene near the end of the sequence for the *phoA* leader peptide. To facilitate circular permutations, another plasmid, pKW02, was created in which the disulfide bond connecting residues 2 and 10 in the wild-type protein was removed by replacement of the codons for cysteine with those for alanine [7]. The protein resulting from this construction has been designated (C2,10A) and was used for direct comparison with the newly created circularly permuted proteins since it served as the immediate progenitor for these mutants.

The procedure to construct the genes encoding the circularly permuted variants of RNase T1 used four oligonucleotide primers in three PCR overlap extension steps [12]. This procedure is detailed in Fig. 1. The cp35A and cp35D primers are the circularly permuting oligonucleotides encoding the new amino and carboxyl termini, respectively, while the Linker-B3 and Linker-C3 primers encode the residues which will connect the original amino and carboxyl termini with a variable linker sequence. The cp35A primer consists of a portion of the *phoA* leader sequence and a *Sty*I restriction site followed by a codon for an extra alanine and then the codons for residues 35–39 of the wild-type protein. The extra alanine was added to ensure proper processing of the mature circularly permuted protein from the *phoA* fusion product [12]. The Linker-B3 primer contains the codons for Gly-97 through Thr-104 of the wild-type protein, the codons for the 3 residue variable linker, and the codons for the first two residues of the (C2,10A) protein. The complementary primer, Linker-C3, contained the codons for Cys-103 through Thr-104, the codons for the 3 residue variable linker, and the codons for the first 8 residues of the (C2,10A) protein. The cp35D primer contained the codons for residues 30–34 of wild-type RNase T1, a stop codon, and the *Hind*III restriction site.

As illustrated in Fig. 1, the first PCR step involved the amplification of the pKW02 plasmid with the primers cp35A and Linker-B3 to create a DNA fragment encoding a region from the new amino terminus through the linker. In the second PCR step, pKW02 was amplified with the cp35D and Linker-C3 primers yielding a DNA fragment encoding the linker region to the new carboxyl terminus. The amplified fragments from the first two PCR steps were then isolated and combined in a third PCR step with the cp35A and cp35D primers to create an amplified fragment containing the entire circularly permuted sequence. The circularly permuted

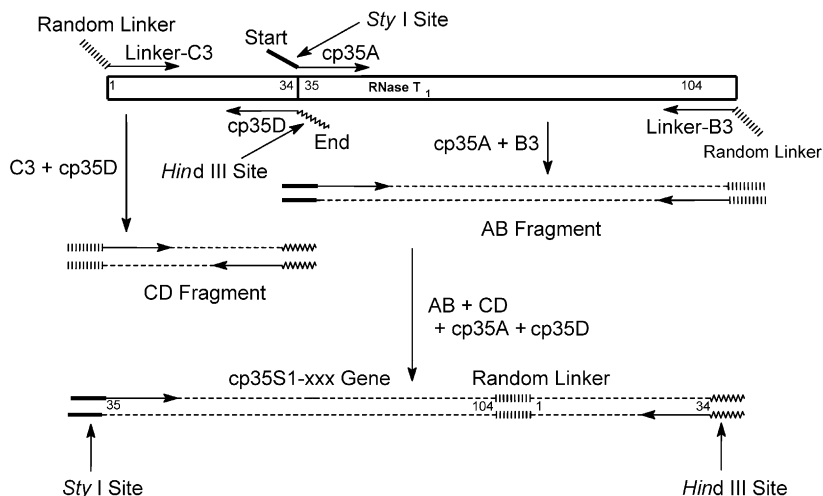


Fig. 1. Protocol for the creation of circularly permuted genes with a variable linker. Additional details are given in the text.

fragment was then restricted with *StyI* and *HindIII*, purified, and subcloned into the pKW01 vector that had also been restricted with these same enzymes and purified.

2.3. Selection methods

A colorimetric plate assay was used for screening protein expression, catalytic activity, and temperature sensitivity [24]. The 1.5% agar LB plates contained 20 µg/mL toluidine blue-O dye, 4 mg/mL Type IV *Torula* yeast RNA, 20 µg/mL chloramphenicol and 30 µg/mL IPTG. The recombinant plasmids containing the genes for the library of circularly permuted variants of RNase T1 were transformed into the *E. coli* strain XL1-Blue. Bacterial colonies were selected at random from a plain 1.5% agar LB plate and the associated plasmids purified and partially sequenced. The amino acid sequences for 42 of these colonies are shown in Table 1. All of the plasmids that had been partially sequenced were then used to transform the cell line WK6. The transformed cells were then spread onto two toluidine blue-O plates and incubated at 25 or 37 °C. The plates were removed from the incubators after 14 h of incubation and allowed to sit at room temperature. The colonies were then scored at 2-h intervals according to how long they took to develop a pink halo around several of the colonies on the plate. Ten colonies from the plates that scored the highest at both temperatures were then used to transform XL1-Blue cells. Single colonies were selected from these plates and grown in culture tubes containing 6 mL of LB media and 20 µg/mL chloramphenicol. The plasmids were identified (pJG35-SPP, pJG35-PAT, pJG35-PAH, pJG35-TPS, pJG35-TPD, pJG35-DSS, pJG35-HPD, pJG35-TPH, pJG35-PHP, and pJG35-DTP) and completely sequenced to verify that no extraneous mutations were present in the circularly permuted genes.

2.4. Cell growth and enzyme purification

Wild-type RNase T1 and the cp35S1 linker variants were expressed and purified to apparent homogeneity (SDS-PAGE) [23,25]. The transformed cells were grown in TB media supplemented with 20 µg/mL chloramphenicol at 37 °C until induction with IPTG. The flasks were allowed to equilibrate at 25 °C for overnight growth. All purification steps were performed at 4 °C unless otherwise noted. After an initial

Table 1
Sequences of linkers found for circular permuted variants of RNase T1

ADH	APA ^a	DAP ^a	DDS	DSS
DST	DTD	DTP	DYP	HAT ^a
HHP ^a	HPD	HAS	NPA	NTY
PAH	PAP	PAT	PHP	PNP
PPD	PPH	PPN	PPP	PPS
PPT	PSS	PTN	PTS	PTT
PYH	PYN	SPH	SPP	SSS
TAA	TDP	TPD	TPH	TPP
TPS	TTY			

^a Indicates the colony had no RNase T1 activity in the colorimetric plate assay.

centrifugation, the cell paste was resuspended in 800 mL of 50 mM Tris–HCl, pH 7.5, 20% sucrose, 10 mM EDTA, and then stirred for 45 min. The suspension was centrifuged (30 min, 8000g) and the supernatant fluid was saved. The liquid fraction was diluted to 4 L with 400 mL of 0.5 M sodium phosphate buffer and the pH was adjusted to 7.1. The solution was applied to a Zetaprep QAE cartridge pre-equilibrated with 50 mM sodium phosphate buffer, pH 7.1. Upon completion of sample loading, the cartridge was washed with 5 volumes of 50 mM sodium phosphate buffer, pH 7.1, and then the protein was eluted with 100 mM sodium phosphate buffer, pH 2.7. Fractions containing RNase T1 activity were pooled and then concentrated using diafiltration. The samples were desalted by continual diafiltration against deionized H₂O. The protein was assayed, concentrated, and stored at 4 °C.

2.5. Assay of enzyme activity

Enzyme activity was determined by the catalytic hydrolysis of Type II ribonucleic acid core using a continuous spectrophotometric assay at 25 °C [26]. The concentration of wild-type RNase T1 and of the circularly permuted variants were determined using an extinction coefficient of 1.67 at 278 nm for a 1 mg/mL solution in a 1.0 cm path length cell [27].

2.6. Thermodynamic measurements

Urea denaturation curves were determined for wild-type RNase T1, (C2,10A), and all of the cp35S1 linker variants by measuring the intrinsic fluorescence intensity of tryptophan with a thermostated SLM fluorimeter. The fluorescence intensity (290 nm excitation and 320 nm emission) of solutions containing approximately 3 μM protein were measured after incubation in denaturant (urea) for 24 h. The free energy of unfolding was calculated by the linear extrapolation method [28] and fit to Eq. (1) using iterative calculations. In this equation, Y_O is the observed fluorescence intensity, Y_N is the fluorescence intensity of the native conformation, Y_U is the fluorescence intensity of the unfolded molecule, m_N is the slope of the increase in fluorescence intensity of the native molecule, m_U is the slope of the increase in fluorescence intensity of the unfolded molecule, $[D]$ is the denaturant concentration, ΔG_{N-U} is the change in free energy in going from the folded to unfolded state, m_G is the slope of the ΔG_{N-U} extrapolation, R is the universal gas constant, and T is the absolute temperature. The estimated error for the protein stability measurements is ± 0.6 kcal/mol

$$Y_O = \frac{(Y_N + m_N[D]) + (Y_U + m_U[D]) \exp(-(\Delta G_{N-U} + m_G[D])/RT)}{1 + \exp(-(\Delta G_{N-U} + m_G[D])/RT)}. \quad (1)$$

2.7. Circular dichroism

The proteins were analyzed on an Aviv 62ADF circular dichroism spectrometer using the software Igor Pro by Wavemetrics. All proteins were analyzed in deionized H₂O using a 1 mm cell. The collected data were averaged over 3 acquisitions with a 0.5 nm wavelength increment and a 2 nm bandwidth.

3. Results

3.1. Design and construction of random linker library

The three dimensional structure of RNase T1 is shown in Fig. 2A. In the wild-type protein, the amino and carboxy termini are approximately 11 Å apart. This distance necessitates a five amino acid linker to bridge the distance between these termini without introducing undue torsional strain [7]. Based on simple molecular modeling, it was estimated that removal of the disulfide bond between Cys-2 and Cys-10 would reduce the conformational constraints placed on the two termini, thus allowing the use of a three amino acid linker [7]. Previous characterizations show that RNase T1 will fold and adopt a native-like conformation when the disulfide bond between Cys-2 and Cys-10 is absent [29]. Therefore, the previously characterized circular variants of RNase T1 used, as a parent protein, a mutant (C2,10A) in which Cys-2 and Cys-10 were replaced with alanine thereby eliminating the disulfide bond. As seen in Fig. 2A, a small anti-parallel β -sheet is directly adjacent to the amino and carboxy termini of RNase T1. This lead to the supposition that the original termini would best be linked with residues that could form a reverse turn. Reverse turns are characterized by the geometry of the residues within the turn and a sharp reversal in the direction of the peptide chain at the β -turn. Because there are special requirements on the residues at the second and third position of the turn, proline and glycine residues predominate in these two positions [30]. In addition to proline and glycine, aspartate, asparagine, glutamate, serine, threonine, and tyrosine are also frequently found in β -turns [31].

A combinatorial approach was used to generate a random linker consisting of six of the amino acids with high frequencies for occurring in β -turns. In contrast to the original linkers (Gly–Pro–Gly and Gly–Gly–Gly) the new linkers contained residues

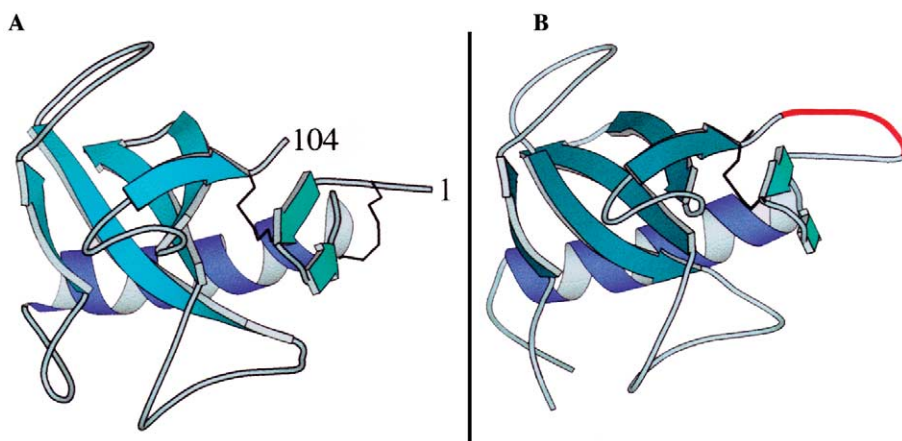


Fig. 2. (A) A ribbon representation of the structure of the wild-type RNase T1. (B) A model for the structure of the circularly permuted form of RNase T1, cp35S1-ggg.

with side chains that could participate in stabilizing interactions in the folded conformation. A partially variable codon was designed that would allow the random incorporation of the desired residues into a three amino-acid linker. The first codon position was allowed to vary so that all four nucleotide bases could be introduced. The second codon position was restricted to only adenine or cytosine. The last position of the codon was restricted to cytosine, which limited the possible amino acid residues to six of the eight residues with a high normalized frequency for occurring in β -turns. The degenerate codon designed in this manner allowed for the random incorporation of proline, aspartate, asparagine, serine, threonine, and tyrosine. Histidine and alanine were also selected by default since their codons were included in the possible combinations specified by the degenerate linker codons. This procedure was used to create a degenerate set of primers which encoded a library of variable linker combinations. The degenerate primers were used in the protocol described by Mullins et al. [7] for the formation of a set of circular variants which had new amino and carboxy termini between residues Gly-34 and Ser-35. This procedure, produced a library of plasmids encoding a circular permuted RNase T1, cp35S1-xxx, having 512 (8^3) possible linker sequence combinations. Partial sequencing of a subset of the isolated plasmids yielded a total of 42 unique linker sequences, which are shown in Table 1.

3.2. Selection and characterization of circularly permuted proteins with random linkers

The bacterial system used to produce RNase T1 is not dependent on the *in vivo* expression of the protein. Therefore, the initial screen of the plasmids containing the 42 unique linker sequences was done randomly and was not based on *in vivo* protein expression or catalytic activity. In order to assess the *in vivo* expression of the circularly permuted proteins, the colorimetric plate assay of Quaas et al. [32] was used. This plate assay allows the determination of bacterial colonies that express catalytically active RNase T1. Since the cp35S1 circular variant of RNase T1 was previously shown to be catalytically active, it could be assumed that many of the linker variants of cp35S1 would also be catalytically active. The colorimetric plate assay was used as an initial measure of the relative thermodynamic stability and/or expression of the variable linker proteins. Nearly 90% of the new linker variants exhibited measurable *in vivo* catalytic activity (Table 1). Only the linker sequences His–His–Pro, Ala–Pro–Ala, Asp–Ala–Pro, and His–Ala–Thr showed no measurable activity indicating that there is a wide range of permissible combinations of amino acids that can be utilized as linker residues. The plasmids were isolated for 17 of the 38 colonies showing catalytic activity and the RNase T1 gene was completely sequenced. From this set of 17 plasmids, 10 linker sequences were chosen for further characterization: His–Pro–Asp, Thr–Pro–His, Asp–Thr–Pro, Thr–Pro–Asp, Pro–Tyr–His, Pro–Ala–Thr, Pro–His–Pro, Asp–Ser–Ser, Ser–Pro–Pro, and Thr–Pro–Ser.

The 10 variable linker circularly permuted proteins were purified and compared to cp35S1-ggg, cp35S1-gpg (C2,10A), and wild-type RNase T1 with regard to CD spectrum, catalytic activity and thermodynamic stability. The thermodynamic stability was the most distinguishing characteristic of the variable linker circularly permuted

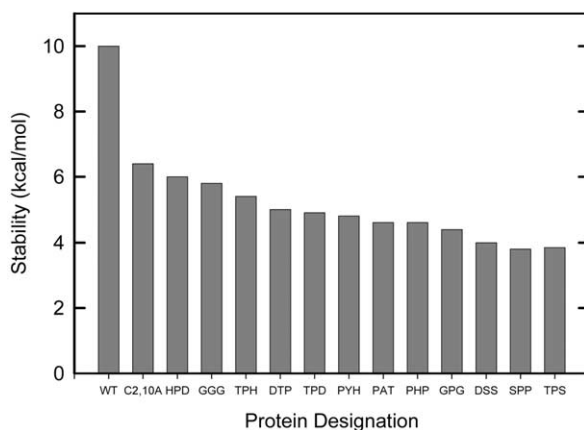


Fig. 3. The relative thermodynamic stabilities for the folding of the circular variants of RNase T1, cp35S1-xxx, containing alternative linking residues. Additional details are given in the text.

cp35S1 proteins. The free energies of unfolding (ΔG) ranged from a high of 6.0 kcal/mol (His–Pro–Asp) to a low of 3.8 kcal/mol (Thr–Pro–Ser). A graphical comparison of the experimentally determined ΔG values is shown in Fig. 3. One of the new linker cp35S1 proteins, His–Pro–Asp, was found to be more stable than the previously constructed cp35S1 protein with the Gly–Gly–Gly linker. The protein with the His–Pro–Asp linker was only 0.4 kcal/mol less stable than the (C2,10A) parent protein. In contrast, the original cp35S1 with the Gly–Pro–Gly linker was only 0.5 kcal/mol more stable than the least stable linker variant.

The catalytic activities of all the variable linker circularly permuted proteins were very uniform even though the thermodynamic stabilities varied by 2.2 kcal/mol. This was not unexpected because each of the variable linker proteins was preselected for catalytic activity. A comparison of the catalytic activities with a model substrate is shown in Fig. 4. The circular variants at the extreme ranges of catalytic activity contained the linkers Thr–Pro–His and Gly–Pro–Gly.

All 14 proteins showed similar CD spectra (data not shown). This, in conjunction with the similar catalytic activity, indicates that the variable linker proteins share a similar tertiary structure with both (C2,10A) and the native protein. Since all the proteins appear to have similar tertiary structures, the range in thermodynamic stability indicates that the various linkers adopt different local conformations that affect the global stability of the proteins.

4. Discussion

A series of circularly permuted variants of RNase T1 has been constructed and characterized in an attempt to more fully understand the role of turns and surface loops during the process of protein folding [12]. These initial studies have

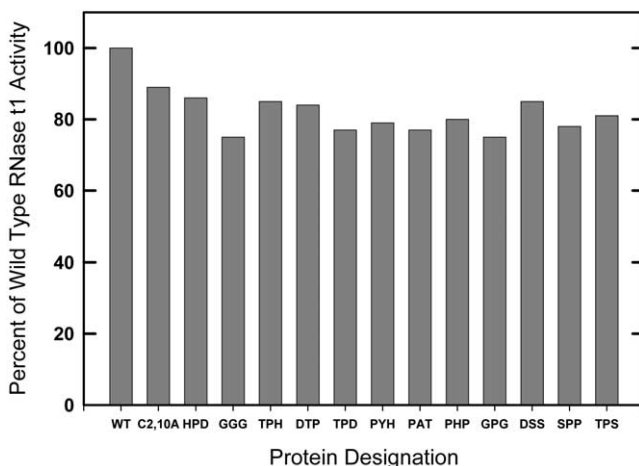


Fig. 4. The relative catalytic activities of the circular variants of RNase T1, cp35S1-xxx containing alternative linking residues. Additional details are given in the text.

demonstrated that at least four of the five loops and turns within the native sequence of RNase T1 are ultimately not required for proper folding of this protein. However, these investigations have clearly demonstrated that significant differences can occur in the overall thermodynamic stability of the folded forms of RNase T1 when alternative sequences of amino acid are used to join the amino and carboxy termini of the native enzyme. In order to help identify the fundamental factors that influence the stability differences in these circular permuted variants of RNase T1, we have constructed a small library of linking residues within these constructs by combinatorial methods. Since the amino and carboxyl terminal ends of the native RNase T1 were thought to best be joined by a reverse β -turn of three amino acid residues, we biased the library in favor of those residues known to be predominantly found in such structures.

Several stable geometries are possible for residues which can form a β -turn within the protein backbone. These conformations have been characterized and grouped into several categories. The most common β -turns, Type I, II, and their mirror images I', II' are recognizable by a hydrogen bond between the backbone carbonyl oxygen of residue 1 of the β -turn and the backbone amide proton of residue 4 [33]. There are several residues which frequently appear in the second and third positions of β -turns because they are able to comply with the geometrical requirements of the turn. In Type I turns, any residue can be accommodated by the sequence, but proline cannot occupy the 3rd position. In Type I' turns, glycine is preferred for positions 2 and 3. Type II and II' turns prefer glycine in positions 3 and 2, respectively, but there are no other required residues. Proline is preferred at position 2 in both Type I and Type II turns.

The linker in the original circular permuted variant of RNase T1, Gly-Pro-Gly, was designed to facilitate the formation of a β -turn by placing a proline into the second position of the turn [7]. In a second experiment intended to improve the design

of the linker, a Gly–Gly–Gly linker was used to connect the original termini [12]. A model of the cp35S1-ggg circularly permuted variant of RNase T1 is shown in Fig. 2B. In a general survey of turns, it was noted that two adjacent glycine residues would provide a more generalized approach to the creation of a β -turn [30]. It was also reasoned that the linker which contained three consecutive glycines would allow more flexibility with regard to which pair of adjacent glycines initiated the turn. In a comparison of the thermodynamic stabilities resulting from the incorporation of each of the linkers into the cp35S1 framework, the Gly–Gly–Gly linker was 1.6 kcal/mol more stable than the Gly–Pro–Gly linker.

Because the two original linkers examined displayed a significant difference in thermodynamic stability and were less stable than the parent (C2,10A) protein, it was hypothesized that an even more stable linker design was feasible. In addition, circularly permuted proteins with a greater thermodynamic stability than the parent have been created [18]. To further examine the possibility of a better design, a combinatorial approach was used to generate a linker sequence with a moderate number of amino acids (Pro, Asp, Asn, Thr, Ser, and Tyr) that are known to have high frequencies for occurring in β -turns [31]. Two others residues, His and Ala, were also incorporated into this library as a result of having codons specified by the degenerate DNA primers. In contrast to the linkers with only proline and glycine, the new linkers contained residues with side chains that could participate in stabilizing interactions in the folded conformation.

The cp35S1 protein with the His–Pro–Asp linker proved to be the most thermodynamically stable of all the linker variants assayed. This was interesting since histidine has the lowest normalized frequency of occurrence (0.69) in a β -turn of all the residues that were incorporated into the artificial linker [31]. An examination of crystal structures with the same sequence as the HPD linker revealed that in addition to the backbone hydrogen-bond found between the first and fourth residues of the reverse turn another hydrogen bond was also present. In a structural alignment of the hi-potential iron–sulfur protein (1HIP),² galactose binding protein (3GPB), and subtilisin B1 (1ST3), the histidine side chain hydrogen-bonded to the amide nitrogen in the backbone of the aspartate residue. Comparison of the location of the backbone atoms by superposition showed a 0.56 Å rms deviation. An examination of the crystal structures of rat mast cell protease (3RP2) and NADH peroxidase (1NPX) revealed that the structures were stabilized by a hydrogen-bond from the aspartate side chain to the backbone amide proton of the residue flanking the turn and the amide proton of its own backbone, respectively. Because all of the crystal structures that were examined with this sequence initiated a reverse turn with the histidine residue, it is reasonable to assume that this is the more stable conformation for this particular linker.

The circularly permuted protein with the Gly–Gly–Gly linker sequence was the second most stable variant examined. Unlike the HPD linker, glycine does not have any side chains with which to hydrogen bond and stabilize the linker or predispose the chain

² These abbreviations refer to the file code from the Protein Data Bank (PDB) for the designated protein structure.

to a particular orientation. The Gly–Gly–Gly linker was originally created with the intent of making a generalized β -turn. It was noted by Chou and Fasman [33] that glycine was preferred in the second position of type I' and II' β -turns and also preferred in the third position of types I' and II. Therefore, a designed linker which had two adjacent glycines would fit a generic pattern for a β -turn. Including three consecutive glycines in the linker would potentially allow two conformations for a β -turn in which adjacent glycines could populate the second and third positions. The crystal structures of four proteins with β -turns were examined to determine if one conformational geometry was preferred over another. The crystal structures of achromobacter protease I (1ARB), phosphoglycerate kinase (1PHP), fatty acid binding protein (1MDC), and pea lectin (1RIN) were inspected. The turn in the achromobacter protease I structure started with the threonine preceding the glycines while the 3-phosphoglycerate kinase structure started the turn with the first of the consecutive glycines. Fatty acid binding protein and pea lectin both started the β -turn with the first glycine residue. Since the Gly–Gly–Gly linker has no side chains to stabilize a β -turn, it appears that the stability of the cp35S1 protein containing this sequence arises from the ability of the linker to adopt a variety of conformations and minimize steric constraints.

Six crystal structures were inspected which contained the Gly–Pro–Gly linker sequence. The structures of mouse N-cadherin (1NCG), cytochrome C2 (1CRY), basic fibroblast growth factor (2BFH), cyclophilin A (2CYH), 2-satellite tobacco necrosis virus (2STV), and uridylate kinase (1UKY) were very similar. In each of the crystal structures, the proline occurred in the second position of the turn. A superposition of the backbone of six crystal structures (1NCG, 1CRY, 2BFH, 2CYH, 2STV, and 1UKY) with a Gly–Pro–Gly linker aligned very closely with almost no deviation in the peptide backbone (data not shown). Because glycine and proline do not have functional side chains, no other hydrogen-bonds were present in the structures to enhance local stability or define a particular geometry. Frequencies of occurrence of these residues is the highest of all the linkers designed and the residues are favorably positioned, so it is not surprising that a β -turn should result from this particular sequence [31]. Because side chain interactions cannot occur, the constraints on the position of the β -turn are dictated by the geometry of the residues rather than the position or H-bonding status of the side chains.

The thermodynamic stability of the cp35S1 protein was influenced by the linker used to connect the original termini of wild-type RNase T1. In comparison to the known three dimensional structures of other proteins, linkers which appear to readily accommodate β -turns also stabilize the cp35S1 mutants to a significant extent. However, the conformational states of surface residues are often very mobile. Therefore, the geometries of the linkers identified in this investigation must await a determination by NMR or X-ray crystallography.

Acknowledgments

This work was supported in part by the NIH (GM 49706 and GM 33894) and the Robert A. Welch Foundation (A-840).

References

- [1] D.P. Goldenberg, T.E. Creighton, *J. Mol. Biol.* 165 (1983) 407–413.
- [2] K. Luger, U. Hommel, M. Herold, J. Hofsteenge, K. Kirschner, *Science* 243 (1989) 206–210.
- [3] A. Buchwalder, H. Szadkowski, K. Kirschner, *Biochemistry* 31 (1992) 1621–1630.
- [4] R.A. Horlick, R. Horuk, *Protein Eng.* 5 (1992) 427–431.
- [5] T. Zhang, E. Bertelsen, D. Benvegnu, T. Alber, *Biochemistry* 32 (1993) 12311–12318.
- [6] Y.R. Yang, H.K. Schachman, *Proc. Natl. Acad. Sci. USA* 90 (1993) 11980–11984.
- [7] L.S. Mullins, K. Wesseling, J.M. Kuo, J.B. Garrett, F.M. Raushel, *J. Am. Chem. Soc.* 116 (1994) 5529–5533.
- [8] R. Kreitman, R.K. Puri, I. Pastan, *Proc. Natl. Acad. Sci. USA* 91 (1994) 6889–6893.
- [9] N.Y. Protosova, A.T. Gudkov, *Protein Eng.* 7 (1994) 1373–1377.
- [10] M.L. Vignais, C. Corbier, G. Mulliert, C. Branlant, G. Branlant, *Protein Sci.* 4 (1995) 994–1000.
- [11] A.R. Viguera, F. Blanco, L. Serrano, *J. Mol. Biol.* 247 (1995) 670–681.
- [12] J.B. Garrett, L.S. Mullins, F.M. Raushel, *Protein Sci.* 5 (1996) 204–211.
- [13] V.N. Uversky, A.T. Gudkov, *Protein Sci.* 5 (1996) 1844–1851.
- [14] U. Pieper, K. Hayakawa, Z. Li, O. Herzberg, *Biochemistry* 36 (1997) 8767–8774.
- [15] R. Graf, H.K. Schachman, *Proc. Natl. Acad. Sci. USA* 93 (1997) 11591–11596.
- [16] M. Boissinot, S. Karnas, J.R. Lepock, D.E. Cabelli, J.A. Tainer, E.D. Getzoff, R.A. Hallewell, *EMBO J.* 16 (1997) 2171–2178.
- [17] J. Ay, M. Hahn, K. Decanniere, K. Piotukh, R. Borriss, U. Heinemann, *Proteins* 130 (1998) 155–167.
- [18] K. Wieligmann, B. Norledge, R. Jaenicke, E.M. Mayr, *J. Mol. Biol.* 280 (1998) 721–729.
- [19] L.S. Mullins, C.N. Pace, F.M. Raushel, *Biochemistry* 32 (1993) 6152–6156.
- [20] J.L. Johnson, F.M. Raushel, *Biochemistry* 35 (1996) 10223–10233.
- [21] F.G. Walz, H.L. Osterman, C. Libertin, *Arch. Biochem. Biophys.* 195 (1979) 95–102.
- [22] C.N. Pace, G.R. Grimsley, J.A. Thompson, B.J. Barnett, *J. Biol. Chem.* 263 (1988) 11820–11825.
- [23] B.A. Shirley, D.V. Laurents, *J. Biochem. Biophys. Methods* 20 (1990) 181–188.
- [24] R. Quaas, O. Landt, H. Grunert, M. Beineke, U. Hahn, *Nucleic Acids Res.* 17 (1989) 3318.
- [25] L.M. Mayr, F.X. Schmid, *Protein Express. Purif.* 4 (1993) 52–58.
- [26] T. Oshima, N. Uenishi, K. Imahori, *Anal. Biochem.* 71 (1976) 632–634.
- [27] C.Q. Hu, J.M. Sturtevant, R.E. Erickson, C.N. Pace, *Biochemistry* 31 (1992) 4876–4882.
- [28] M.M. Santoro, D.W. Bolen, *Biochemistry* 27 (1988) 8063–8068.
- [29] C.N. Pace, T.E. Creighton, *J. Mol. Biol.* 188 (1986) 11820–11825.
- [30] T.E. Creighton, in: *Proteins, Structures and Molecular Properties*, W.H. Freeman & Company, New York, 1993, pp. 199–264.
- [31] M. Levitt, *Biochemistry* 17 (1978) 4277–4285.
- [32] R. Quaas, Y. McKeown, P. Stanssens, R. Frank, R. Blocker, U. Hahn, *Eur. J. Biochem.* 173 (1988) 617–622.
- [33] P.Y. Chou, G.D. Fasman, *J. Mol. Biol.* 115 (1977) 135–175.