# CONSERVATION OF CHAIN REVERSAL REGIONS IN PROTEINS

PETER Y. CHOU, *Department of Chemistry, Worcester Polytechnic Institute, Worcester, Massachusetts 01609*
GERALD D. FASMAN, *Graduate Department of Biochemistry, Brandeis University, Waltham, Massachusetts 02154 U.S.A.*

ABSTRACT Using the bend frequencies based on 29 proteins in the previous paper (Chou and Fasman, 1979), $\beta$-turn probability profiles were calculated for the C-peptides of 10 mammalian proinsulins, for 7 proteinase inhibitors, and for 12 species of pancreatic ribonucleases. $\beta$-turn correlation coefficient matrix tables were also computed to obtain the statistical mean between 45 pairs of proinsulin C-peptides, $<C_t> = 0.57 \pm 0.31$; 21 pairs of proteinase inhibitors, $<C_t> = 0.73 \pm 0.13$; and 66 pairs of ribonucleases, $<C_t> = 0.83 \pm 0.08$. Despite relatively low sequence conservation in these three sets of proteins, $\beta$-turns were predicted to be highly conserved: 33% sequence vs. 78% bend for the proinsulins, 20% sequence vs. 85% bend for the proteinase inhibitors, and 65% sequence vs. 92% bend for the ribonucleases. These results suggest that chain reversal regions play an essential role in keeping the active structural domains in hormones and enzymes intact for their specific biological function.

## INTRODUCTION

Studies of amino acid sequence homologies in proteins have yielded much information in comparative biochemistry and molecular evolution (Dayhoff, 1972). With the rapid increase of known three dimensional structures of proteins determined from x-ray crystallography, recent attention has been focused on the comparison of tertiary structures. Although only 39% of the amino acid sequences are identical in $\alpha$-chymotrypsin and elastase (Shotten and Hartley, 1970), a comparison of diagonal distance maps of the two enzymes showed 85% homology in tertiary structure (Nishikawa and Ooi, 1974). Despite great variability in primary structure of the globins (e.g., only 24% of the sequences of glycera hemoglobin and myoglobin were found to be homologous), a close resemblance in their conformation was observed (Padlan and Love, 1974; Ptitsyn, 1974). Aside from the structural resemblances found within the protein families such as serine proteases and globins, similarity in structural domains were discovered between apparently diverse proteins. A common heme binding pocket in globins and cytochrome $b_5$ was located (Rossman and Argos, 1975), and similarity was observed in the folding pattern of the $\beta$-strands between the immunoglobulin molecule and superoxide dismustase (Richardson et al., 1976). More recently, a domain with four roughly parallel $\alpha$-helices was observed in hemerythrin tobacco mosaic virus protein, and tyrosyl-t-RNA synthetase (Argos et al., 1977). A comparison of the chain reversal regions in elastase and $\alpha$-chymotrypsin showed that 21 of the 27 $\beta$-turns in the former enzyme were

---

Dr. Fasman is the Rosenfield Professor of Biochemistry.

conserved in the latter (Chou and Fasman, 1977). A more detailed x-ray crystallographic analysis between these two serine proteases have shown their remarkable structural similarities (Sawyer et al., 1978).

In addition to the comparative studies of known x-ray structures mentioned above, secondary structural predictions have proved to be helpful in the recognition of structural domains. Using the Chou and Fasman (1974) predictive rules with the aid of homologous sequence comparisons, coenzyme-binding domains of glutamate dehydrogenases were located (Wooton, 1974). Conformational homologies were also elucidated in growth hormones (Pena et al., 1975), immunoglobulins (Low et al., 1976), and neurotoxins (Smythies, 1975; Menez, 1978). In the present paper, the $\beta$-turn frequencies based on 29 proteins (Chou and Fasman, 1979) have been used to predict the chain reversal regions in 10 mammalian proinsulin C-peptides, 7 proteinase inhibitors, and 12 pancreatic ribonucleases. Correlation coefficient matrix tables were computed to examine quantitatively the extent of $\beta$-turn conservation between species within each family of proteins.

## METHODS

### *Prediction of β-Turns*

The bend frequencies based on 29 proteins (Table 1 of Chou and Fasman, 1979) were used in the automated bend prediction computer program as described in the previous paper. $\beta$-turn probability profiles were obtained for 10 species of proinsulin C-peptides, 7 species of proteinase inhibitors, and 12 species of pancreatic ribonucleases. Some of the redundant b-turns were omitted after predictions for helical and $\beta$-sheet regions using the $P_\alpha$ and $P_\beta$ parameters derived from 29 proteins (Chou and Fasman, 1978). The final $\beta$-turns predicted after this more refined analysis are numbered in the graphical bend plots (Figs. 1–4).

### *Calculation of β-Turn Correlation Coefficient Matrix*

The same correlation coefficient, $C_t$, used in comparing predicted bends with observed bends from x-ray (see Eq. 7 of previous paper; Chou and Fasman, 1979), may also be used in comparing bend homology between proteins. Thus, to find the bend correlation coefficient between proteins $a$ and $b$ we may use:

$$C_t = \frac{(p/N - \bar{P}_a \bar{P}_b)}{[\bar{P}_a \bar{P}_b(1 - \bar{P}_a)(1 - \bar{P}_b)]^{1/2}}, \quad (1)$$

where p is the number of identical turn residues predicted in both proteins, $N$ is the total number of residues in the protein. $\bar{P}_a$ and $\bar{P}_b$ are, respectively, the fraction of protein $a$ and $b$ predicted to be in $\beta$-turns. All pairs of species within the three sets of proteins were compared using Eq. 1. Hence, for the C-peptides of proinsulin, $1/2(10 \times 9) = 45$ correlations were obtained; for the ribonucleases, $1/2(12 \times 11) = 66$ correlations, and for the proteinase inhibitors $1/2(7 \times 6) = 21$ correlations, or $1/2(8 \times 7) = 28$ correlations including bromelain inhibitor. These correlations were averaged to obtain the mean $\langle C_t \rangle$ and standard deviation for each series of comparisons and for the entire set as a whole. To obtain a symmetrical cross correlation matrix table (i.e., $C_t$ should be the same when $a$ is compared to $b$ or when $b$ is compared to $a$), $N$ was held constant within each set. Hence $N = 31$ for

TABLE I
FRACTION OF β-TURN RESIDUES PREDICTED FOR THREE HOMOLOGOUS SETS OF PROTEINS*

| Pancreatic ribonucleases | $n_{t/N}$ | $\bar{P}$ | Proinsulin C-peptides | $n_{t/N}$ | $\bar{P}$ | Proteinase inhibitors | $n_{t/N}$ | $\bar{P}$ |
|---|---|---|---|---|---|---|---|---|
| Bovine | 30/124 | 0.24 | Human | 9/31 | 0.29 | BPTI | 28/58 | 0.48 |
| Pig | 37/124 | 0.30 | Monkey | 9/31 | 0.29 | Bov. col.‡ | 24/67 | 0.36 |
| Rat | 46/127 | 0.35 | Guinea pig | 4/29 | 0.14 | Snail K | 20/58 | 0.35 |
| Sheep | 32/124 | 0.26 | Chinchilla | 4/31 | 0.13 | Turtle egg | 28/64 | 0.44 |
| Horse | 32/125 | 0.26 | Dog | 6/23 | 0.26 | Viper toxin | 24/60 | 0.40 |
| Giraffe | 29/124 | 0.23 | Pig | 4/29 | 0.14 | Toxin I | 16/60 | 0.27 |
| Roe deer | 33/124 | 0.27 | Ox | 9/26 | 0.35 | Toxin K | 16/57 | 0.28 |
| Red deer | 33/124 | 0.27 | Rat 1 | 7/31 | 0.23 | Bromelain inhibitor | 16/52 | 0.31 |
| Guinea pig A | 34/124 | 0.27 | Rat 2 | 9/31 | 0.29 | | | |
| Guinea pig B | 34/128 | 0.27 | Horse | 7/31 | 0.23 | | | |
| Chinchilla | 36/124 | 0.29 | | | | | | |
| Coypu | 39/128 | 0.30 | | | | | | |

*$(f_t) = \bar{P} = n_{t/N}$, where $n_t$ is the number of residues predicted in the β-turn conformation, and $N$ is the total number of residues in the protein.
‡Bovine colostrum inhibitor.

proinsulin C-peptides, $N = 58$ for proteinase inhibitors, and $N = 124$ for ribonucleases. However, $N$ was allowed to vary for each of the species when their $\bar{P}$ values were calculated. It should be noted that these adjustments were necessitated by the differences in sequence length between species of the same protein. Such a problem did not arise in correlations between predicted and observed x-ray structures inasmuch as $N$ was identical for such comparisons. The fraction of bend residues predicted $\bar{P} = n_t/N$ for the three sets of proteins are given in Table I.

## RESULTS AND DISCUSSION

### *Reverse Chain Folding in the C-Peptide of Proinsulin*

Insulin is synthesized as a single chain precursor, proinsulin, in which residue B30 of the B-chain is joined to A1 of the A-chain by connecting peptide: $NH_2$-(B-chain)-Arg Arg-(C-peptide)-Lys Arg-(A-chain)-COOH. The C-peptide is the connecting peptide minus the four basic amino acids that are removed by proteolysis in the conversion of proinsulin to insulin (Steiner and Oyer, 1967; Chance et al., 1968). The chain length in the C-peptides varies from 26 residues in the bovine to 29 in porcine and 31 in human and rat C-peptides (Markussen and Sundby, 1972). The self-association of proinsulin was found to be essentially the same for insulin, suggesting that the C-peptide does not interfere with dimer and hexamer formation of insulin, and that the insulin moiety of proinsulin exists in the same conformation as insulin itself (Frank and Veros, 1968; Pekar and Frank, 1972). Using this assumption and the empirical rules of Chou and Fasman (1974), Snell and Smyth (1975) predicted the C-peptide conformation and were able to join this fragment to the known x-ray structure of insulin (Blundell et al., 1972). They examined the C-peptide sequences of 10 mammalian species in which only 10 residues are invariant. Although only 1/3 of the residues in the

C-peptide are conserved (as compared to 24/51, $\simeq 1/2$, of insulin residues [sequence alignment from 18 species in Dayhoff, 1972]), there is nevertheless a remarkable conservation in its conformation as deduced from the Chou-Fasman predictive method. Snell and Smyth (1975) predicted a $\beta$-turn at residues C15–18 flanked by two helices C6–11 and C21–27, and proposed a three dimensional structure of proinsulin based on model building. Vogt et al. (1976) also applied the conformational parameters of Chou and Fasman (1974) and predicted a $\beta$-turn at C13–16 flanked by helical segments C6–12 and C21–28 in the porcine proinsulin C-peptide. Furthermore, their circular dichroism (CD) studies of synthetic C-peptides appear to correlate with the predicted structure. In particular, it was shown that the helical potential in C21–28 may be induced in the fragment C9-31 and C14–31 by 1% sodium dodecylsulfate, and that the CD spectra of fragments C9–22 and C14–22 (where $\beta$-turns may occur) where characterized by a random type spectra with positive ellipticity at 213 nm.

Using the $\beta$-turn frequencies from 29 proteins (Chou and Fasman, 1979), the probabilities of bend occurrence for the entire C-peptide of all 10 species were calculated and compared graphically in Fig. 1. With the exception of the chinchilla (which has a high bend potential at C11–14), all the species have a high $\beta$-turn potential in the C12–20 region, and, more importantly, none outside of it. Whereas a single chain reversal appears likely for porcine at C12–15 and for guinea pig at C17–20, multibends at C12–15, C15–18, and C17–20 are possible for human, monkey, bovine, and rat C-peptides (Fig. 1).

To assess the bend homology between species in a more quantitative manner, the $\beta$-turn correlation coefficients (Eq. 1) were computed for all nine pairs of proinsulins. The $10 \times 10$ cross correlation matrix for the C-peptides is shown in Table II. A value of $C_t = 1.00$ indicates perfect conservation of bends between species, whereas $C_t = -1.00$ shows total dissimilarity in bend positionings. Hence, the identical bends predicted for human, monkey, and rat-2 C-peptides are reflected by the $C_t = 1.00$ values between these species. Although the bends for ox were also predicted at identical positions (i.e., 12, 15, 17) its $C_t$ value was calculated as 0.88 instead of 1.00 when compared to human, monkey, or rat 2. This arose due to a 5-residue deletion at C22–26 in the ox C-peptide, and the use of $N = 26$ resulted in different $\bar{P}$ values for Eq. 1. Similarly, the bend correlation coefficients for dog and pig in Table II appear poorer than the graphical bend profiles (Fig. 1) because of deletions of eight and two residues, respectively, in the sequence of these two species. Nevertheless, the overall mean of matrix, $\langle C_t \rangle = 0.57 \pm 0.31$ indicates that $\simeq 78\%$ of the $\beta$-turns are conserved in these 10 species of proinsulin C-peptides whose sequence homology showed only 33% conservation.

Using the $P_\alpha$ and $P_\beta$ parameters based on 29 proteins (Chou and Fasman, 1978), residues C6–12 and C21–27 are predicted as helical in porcine C-peptide, agreeing with the earlier helical predictions using data based on 15 proteins (Snell and Smyth, 1975; Vogt et al., 1976). Only one helix is predicted for bovine at C6–12, whereas helices at C1–12 and C21–27 are predicted for human C-peptide. The x-ray crystallographic studies of proinsulin is still at a preliminary stage (Fullerton et al., 1970). Although earlier CD studies showed that the C-peptide region of proinsulin to be essentially devoid of secondary structure (Frank and Veros, 1968), recent laser Raman spectroscopy studies (Yu et al., 1972) revealed a considerable fraction of $\alpha$-helical structure. A recent conformational analysis of CD spectra by nonlinear regression (Markussen and Volund, 1975) showed that proinsulin has four more

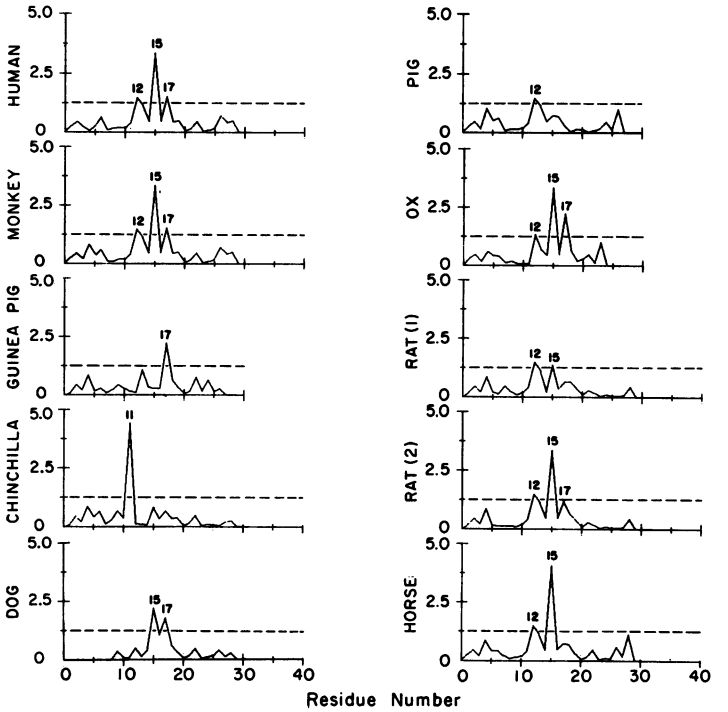


FIGURE 1 Probability of $\beta$-turn occurrence in C-peptides of proinsulin in 10 mammalian species. The dashed horizontal lines correspond to an arbitrary cutoff value of $p_t = 1.25 \times 10^{-4}$. The numbered probability peaks were predicted as $\beta$-turns.

TABLE II
$\beta$-TURN CORRELATION COEFFICIENT MATRIX OF PROINSULIN C-PEPTIDES

| | Human | Monkey | Rat 2 | Rat 1 | Horse | Ox | Pig | Dog | Guinea pig | Chinchilla |
|---|---|---|---|---|---|---|---|---|---|---|
| Human | — | 1.00 | 1.00 | 0.84 | 0.84 | 0.88 | 0.57 | 0.59 | 0.57 | 0.39 |
| Monkey | 1.00 | — | 1.00 | 0.84 | 0.84 | 0.88 | 0.57 | 0.59 | 0.57 | 0.39 |
| Rat 2 | 1.00 | 1.00 | — | 0.84 | 0.84 | 0.88 | 0.57 | 0.59 | 0.57 | 0.39 |
| Rat 1 | 0.84 | 0.84 | 0.84 | — | 1.00 | 0.74 | 0.68 | 0.38 | 0.23 | 0.48 |
| Horse | 0.84 | 0.84 | 0.84 | 1.00 | — | 0.74 | 0.68 | 0.38 | 0.23 | 0.48 |
| Ox | 0.88 | 0.88 | 0.88 | 0.74 | 0.74 | — | 0.50 | 0.49 | 0.50 | 0.33 |
| Pig | 0.57 | 0.57 | 0.57 | 0.68 | 0.68 | 0.50 | — | −0.02 | −0.16 | 0.68 |
| Dog | 0.59 | 0.59 | 0.59 | 0.38 | 0.38 | 0.49 | −0.02 | — | 0.61 | −0.23 |
| Guinea pig | 0.57 | 0.57 | 0.57 | 0.23 | 0.23 | 0.50 | −0.16 | 0.61 | — | −0.15 |
| Chinchilla | 0.39 | 0.39 | 0.39 | 0.48 | 0.48 | 0.33 | 0.68 | −0.23 | −0.15 | — |
| Mean $<C_t>$ | 0.74 | 0.74 | 0.74 | 0.67 | 0.67 | 0.66 | 0.45 | 0.38 | 0.33 | 0.31 |
| SD | 0.22 | 0.22 | 0.22 | 0.25 | 0.25 | 0.21 | 0.32 | 0.30 | 0.31 | 0.30 |

Mean of matrix, 0.57 (78% turns/33% sequence conserved). SD of matrix, 0.31.

residues in the β-structure than does insulin, and that 5 M urea did not change the conformation of the C-peptide of proinsulin. Although no biological role has been assigned to the C-peptide, the predictive method shows that the β-turn conservation in proinsulin is probably required for directing the proper folding of the C-peptide helices. It is possible that this conformation serves to mask the receptor binding region of insulin (Pullen et al., 1976), thus making the precursor proinsulin inactive.

*Conformational Homology in Proteinase Inhibitors*

Laskowski et al. (1974) and Creighton (1975*a*) have pointed out the sequence homology of seven proteinase inhibitors. Despite the large phylogenetic distance between the various proteins (bovine pancreatic trypsin inhibitor, bovine colostrum inhibitor, turtle egg white inhibitor, snail inhibitor K, Russell's viper toxin, black mamba toxins K and I), all 6 Cys residues may be alligned without any internal deletions or insertions. There are only five other invariant residues (Gly 12, Tyr 23, Gly 37, Asn 43, and Phe 45), so that only 20% of the residues in these inhibitors are conserved. However, here again, as in the case of the proinsulin C-peptides, conformational homology was observed from the bend probability profiles in Fig. 2 which shows the conservation of β-turns. The x-ray structure of bovine pancreatic trypsin inhibitor (BPTI) shows β-turns initiating chain reversals at residues 12, 25, 37, and 41 (Deisenhofer and Steigemann, 1975). These same regions appear with high bend probability in the other six species. It is noteworthy that of the 11 invariant residues, 4 are found (Gly 12, Cys 14, Gly 37, Cys 38) to occupy the β-turn regions 12–15 and 37–40 of trypsin inhibitor. Three other invariant residues (Tyr 23, Cys 30, and Phe 45) are located adjacent to β-turns 25–28 and 41–44.

Reddy et al. (1975) have proposed that the bromelain inhibitor from pineapple stem may be homologous to the proteinase inhibitors by connecting the 41 residue A-chain to the 11 residue B-chain with a bridge hexapeptide. However, there are 10 Cys residues in the bromelain inhibitor as compared to 6 in the proteinase inhibitors, and only 4 of these could be aligned precisely. The β-turn probability profile of bromelain inhibitor, shown at the bottom of Fig. 2, bears little resemblance to the other inhibitors, suggesting a lack of conformational homology. This is more apparent in the predicted secondary structure of the bromelain inhibitor when compared with the other 7 proteinase inhibitors shown in the schemetic diagrams of Fig. 3. The anti-parallel β-sheets 16–24 and 29–35, the last β-turn 41–44, and the COOH-terminal helix 48–54, as determined in the x-ray structure of trypsin inhibitor (Deisenhofer and Steigemann, 1975), are found to be preserved in the predicted conformation of the proteinase inhibitors but not in bromelain inhibitor. These differences in secondary structures determined from the Chou and Fasman predictive model (1974, 1978) reinforce the earlier conclusion of Creighton (1975a), based on sequence comparisons, that the bromelain inhibitor is not structurally homologous to the proteinase inhibitors.

Table III shows the β-turn correlation coefficients among seven proteinase inhibitors. Although only 20% of the sequence is preserved, there is a remarkable 85% conservation in the chain reversal regions of these inhibitors. The cross correlation mean of matrix, $\langle C_t \rangle = 0.73 \pm 0.13$ showed a better agreement and lower standard deviation than that for the proinsulin C-peptide. This may be due to the deletions of residues in the primary sequences of the proinsulins that are totally absent in the inhibitors. A comparison of the bend homology

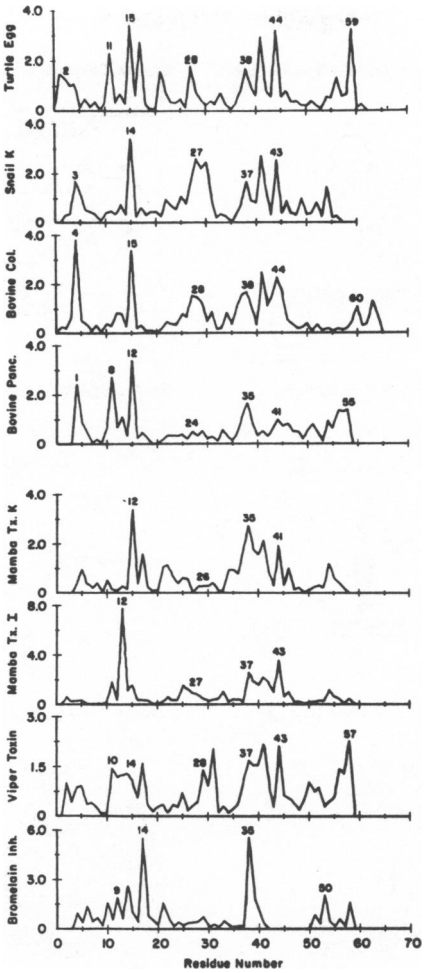


FIGURE 2 Probability of $\beta$-turn occurrence in proteinase inhibitors. The numbered probability peaks were predicted as $\beta$-turns. The amino acid sequences of the 8 species were aligned homologously according to Laskowski et al. (1974) and Reddy et al. (1975), and this sequence alignment is preserved in the bend probability plots. However, the residue number denotes the amino acid position along the polypeptide chain for each species, respectively, and not to the residue number of pancreatic trypsin inhibitor.

between bromelain inhibitors and other inhibitors gave the following values of $C_t$: 0.17 (turtle egg), 0.05 (bovine colostrum), 0.07 (snail K), 0.10 (BPTI), 0.22 (Russell's viper toxin), 0.19 (black mamba toxin I), and 0.17 (toxin K). The low mean bend correlation of $\overline{C}_t = 0.14 \pm 0.06$ for bromelain inhibitor indicates that it is a poor structural homologue of the proteinase inhibitors, agreeing with the qualitative conclusions drawn from Figs. 2 and 3.

A comparison of the $\beta$-turn assignments in the various proteinase inhibitors, shown in Fig. 3, suggests that they may play an essential role in disulfide formation. Creighton (1975*b*) proposed the folding pathway of pancreatic trypsin inhibitor to involve the linkages of Cys 30–51 followed by Cys 5-55 and finally Cys 14–38. It is interesting to note that the first step in the folding of trypsin inhibitor involves a $\beta$-$\alpha$ interation. This is, $\beta$-region 16–24 must be

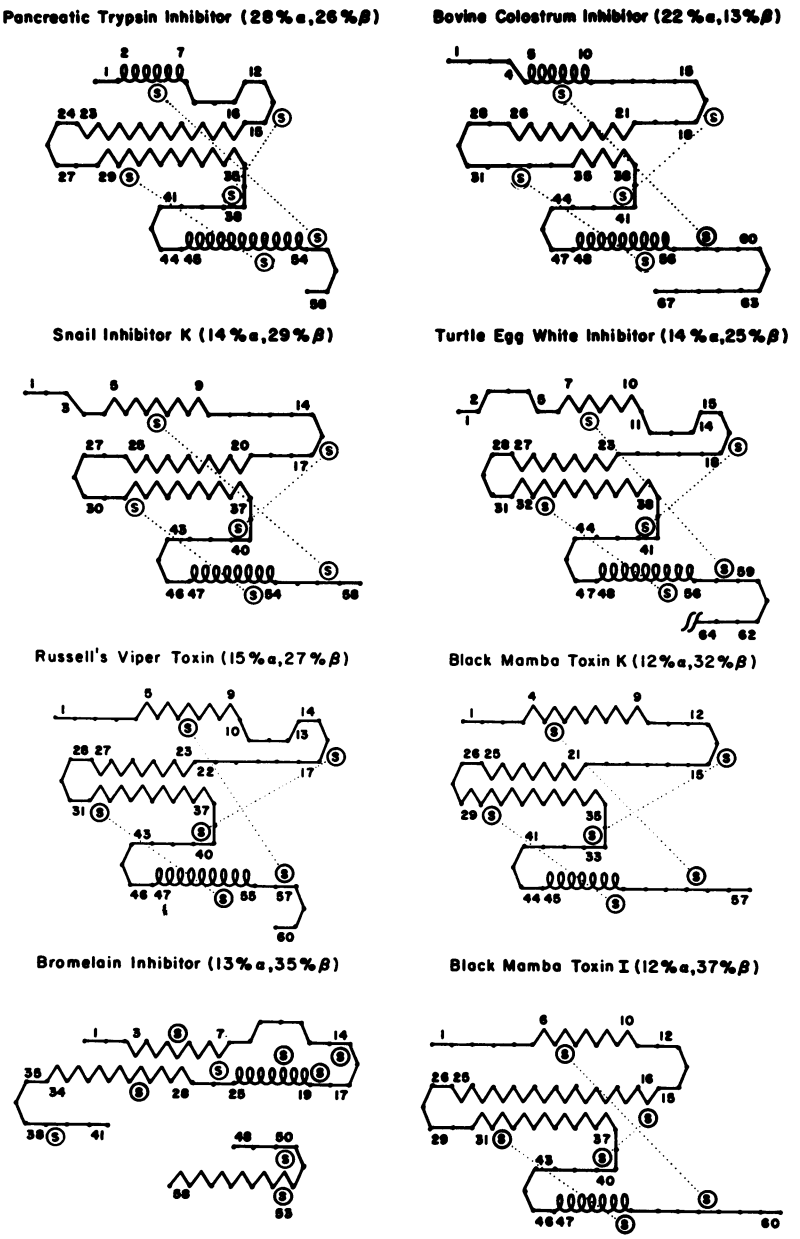


FIGURE 3 Schematic diagrams of predicted secondary structures in proteinase inhibitors. The predicted helical and $\beta$-sheet content are shown in parentheses. Residues are represented in helical (ℓ), $\beta$-sheet ( $\wedge$ ) and coil (•——•) conformations. $\beta$-turn tetrapeptides are denoted by chain reversals. Conformational boundary residues are numbered, as are the Cys residues, by Ⓢ. The disulfide bridges (denoted by dotted lines) are assumed to be identical to those found in trypsin inhibitor. The predicted conformation of bromelain inhibitor is also shown for comparison.

TABLE III
β-TURN CORRELATION COEFFICIENT MATRIX OF PROTEINASE INHIBITORS

| | Turtle egg | Bov. col. | Snail K | BPTI | Viper toxin | Toxin I | Toxin K |
|---|---|---|---|---|---|---|---|
| Turtle egg | — | 0.94 | 0.82 | 0.96 | 0.84 | 0.73 | 0.61 |
| Bov. col. | 0.94 | — | 0.97 | 0.79 | 0.64 | 0.69 | 0.73 |
| Snail K | 0.82 | 0.97 | — | 0.68 | 0.52 | 0.71 | 0.76 |
| BPTI | 0.96 | 0.79 | 0.68 | — | 0.76 | 0.59 | 0.47 |
| Viper toxin | 0.84 | 0.64 | 0.52 | 0.76 | — | 0.70 | 0.74 |
| Toxin I | 0.73 | 0.69 | 0.71 | 0.59 | 0.70 | — | 0.75 |
| Toxin K | 0.61 | 0.73 | 0.76 | 0.47 | 0.74 | 0.75 | — |
| Mean $<C_t>$ | 0.82 | 0.79 | 0.74 | 0.71 | 0.70 | 0.69 | 0.68 |
| SD | 0.13 | 0.13 | 0.15 | 0.17 | 0.11 | 0.06 | 0.12 |

Mean of matrix, 0.73 (85% turn/20% sequence conserved). SD of matrix, 0.13.

brought into close proximity to helix 48–54 to allow the S-S bridge of residues 30 and 51 to form. The second fold (Cys 5–55) involves a helix-coil contact (helix 3–6 with coil region 55–58) followed by a turn-turn interaction (bend 12–15 and bend 37–40) forming the last disulfide bridge (Cys 14–38). The other inhibitors, as well as the snake toxins shown in Fig. 3, also have a β-α interaction involving the Cys 30–51 disulfide. This suggests that the kinetics of protein folding is governed by strong secondary structural contacts. It is also significant that the Cys 14–38 disulfide, which occurs last in the folding process (Creighton, 1975*c*), is found at the β-turn in all the proteinase inhibitors. This may also explain why these latter S-S bonds are not important for stabilizing the native conformation. The secondary structural arrangement for trypsin inhibitor and colostrum inhibitor are identical ($\alpha\beta\beta\alpha$). However, for snail inhibitor K, turtle egg white inhibitor, and the three snake toxins, the order of topological packing is predicted as ($\beta\beta\beta\alpha$). Hence the short helix at the $NH_2$-terminal of trypsin inhibitor (3–7 α, x-ray; 2–7 α, predicted) may not be conserved in these five homologues, but is replaced by a short β-region. The structural arrangement in bromelain inhibitor is ($\beta\alpha\beta$-$\beta$) and does not show conformational homology to the other 7 inhibitors as already evidenced by the β-turn probability profiles of Fig. 2.

The x-ray analysis, at 2.8 Å resolution, of the trypsin-trypsin inhibitor complex shows the Lys 15 side chain of the inhibitor in the specificity pocket of the enzyme (Rühlmann et al., 1973). It is interesting that β-turn is predicted at residues 12–15 of the trypsin inhibitor (Fig. 3) in agreement with the x-ray data, and that this tetrapeptide sequence (Gly-Pro-Cys-Lys) is conserved in the other three inhibitors and also in toxin K. Despite the differences in residues found in this region for viper toxin, 14–17 (Gly-Arg-Cys-Arg), and toxin I, 12–15 (Asn-Pro-Gly-Arg), β-turns are still predicted for these tetrapeptides (Figs. 2 and 3), thus placing Arg instead of Lys in the 4th position of the β-turn. The only other tetrapeptide sequence that is conserved in the inhibitors and viper toxin is 35–38 (Tyr-Gly-Gly-Cys). Again, this region is predicted as a β-turn, in close agreement with the chain reversal found at residues 37–40 by x-ray analysis. Although the homologous sequences of toxin K, 35–38 (Tyr-Ser-Gly-Cys), and toxin I, 37-40 (Trp-Ser-Gly-Cys), show different residues at this site, the bend probability peaks (Fig. 2) indicate that chain reversals have been conserved in these regions.

*Bend homologies in ribonucleases*

Using the x-ray atomic coordinates of bovine pancreatic ribonuclease S (Richards and Wyckoff, 1973), the $C_1^\alpha$–$C_4^\alpha$ and $O_1$–$N_4$ distances were calculated, and in conjunction with the $\varphi$, $\psi$ dihedral angles, 11 $\beta$-turns were elucidated and classified (Chou and Fasman, 1977): 16–19 (bend type II), 23–26 (VII), 34–37 (VII), 36–39 (I), 65–68 (III), 66–69 (II), 75–78 (IV), 87–90 (II), 91–94 (VI), 92–95 (I), and 112–115 (VI). It is interesting to note that six different bend types, without their mirror-image counterparts, are represented in ribonuclease. To determine whether these chain reversal regions are conserved, the amino acid sequence of bovine ribonuclease A (Smyth et al. 1963) was aligned with the sequences of ribonucleases of horse, sheep, roe deer, red deer, pig, rat (Scheffer and Beintema, 1973), giraffe (Gaastra et al., 1974), coypu, chinchilla, and guinea pigs A and B (Van den Berg and Beintema, 1975). Of these 12 species, 8 have chain lengths of 124 residues. The exceptions are rat, with three extra residues at the $NH_2$-terminus; horse, with a deletion at position 39 and two extra residues at the C-terminus (Scheffer and Beintema, 1973); coypu and guinea pig B, with four additional residues at the COOH-terminus. Ribonuclease sequences from different mammalian orders vary $\simeq$ 25% with the exception of rat ribonuclease which differs $\simeq$ 35% from all the other species (Van den Berg and Beintema, 1975). The invariability of the distribution of hydrophobic residues in stabilizing the secondary structures of 24 homologous ribonucleases (Lenstra et al., 1977) and the molecular evolution of pancreatic ribonucleases (Beintema et al., 1977) have been recently discussed.

Using the bend frequencies (Table I of Chou and Fasman, 1979) the probability of $\beta$-turn occurrences in the 12 species of pancreatic ribonuclease were calculated. The bend probability profiles are shown in Fig. 4, with the 11 $\beta$-turns found from the x-ray structure of bovine ribonuclease S represented by arrows in the bottom profile. The computer prediction missed 2 of the 36 bend residues while overpredicting 28 bend residues in bovine ribonuclease with $\%_t = 94\%$, $\%_{nt} = 68\%$, $Q_t = 81\%$, $\%_{t+nt} = 76\%$, and $C_t = 0.57$ (see Tables IV and V of Chou and Fasman, 1979). However, some of the redundant $\beta$-turns (i.e., 12–15, 32–35, 59–62, 29–72, 70–73, and 82–85) may be eliminated when the helical and $\beta$-sheet predictions of ribonuclease (Chou and Fasman, 1974) are considered. The remaining $\beta$-turns predicted for the 12 species, after this more refined analysis, are numbered at the ith bend position of the probability peaks shown in Fig. 4. It should be noticed that the numbering for the predicted bends in rat ribonuclease appears to be shifted by +3 compared with the other sequences. This is due to the three extra residues (Gly-Glu-Ser) at the $NH_2$-terminus for rat ribonuclease that were not used in calculating the bend probability profile. Instead, tetrapeptide 4–7 (Arg-Glu-Ser-Ser-) was treated as if it were tetrapeptide 1–4, residues 5–8 as tetrapeptide 2–5, etc., to achieve alignment with the bovine sequence. Because of a deletion at residue position 39 for horse ribonuclease, the bend probability profile, after this position, appears to be off by −1 residue when compared to the other sequences in Fig. 4.

A 12 × 12 cross correlation matrix for the $\beta$-turns predicted in the pancreatic ribonucleases is presented in Table IV. The matrix mean of $<C_t> = 0.83 \pm 0.08$ indicates that despite 65% sequence conservation, 92% of the bends in the 12 species of ribonuclease are conserved.

Of the 11 $\beta$-turns observed in the x-ray structure of bovine ribonuclease, only bend 65–68 has identical residues for all the 12 species. This tetrapeptide, Cys-Lys-Asn-Gly, with $<P_t>$

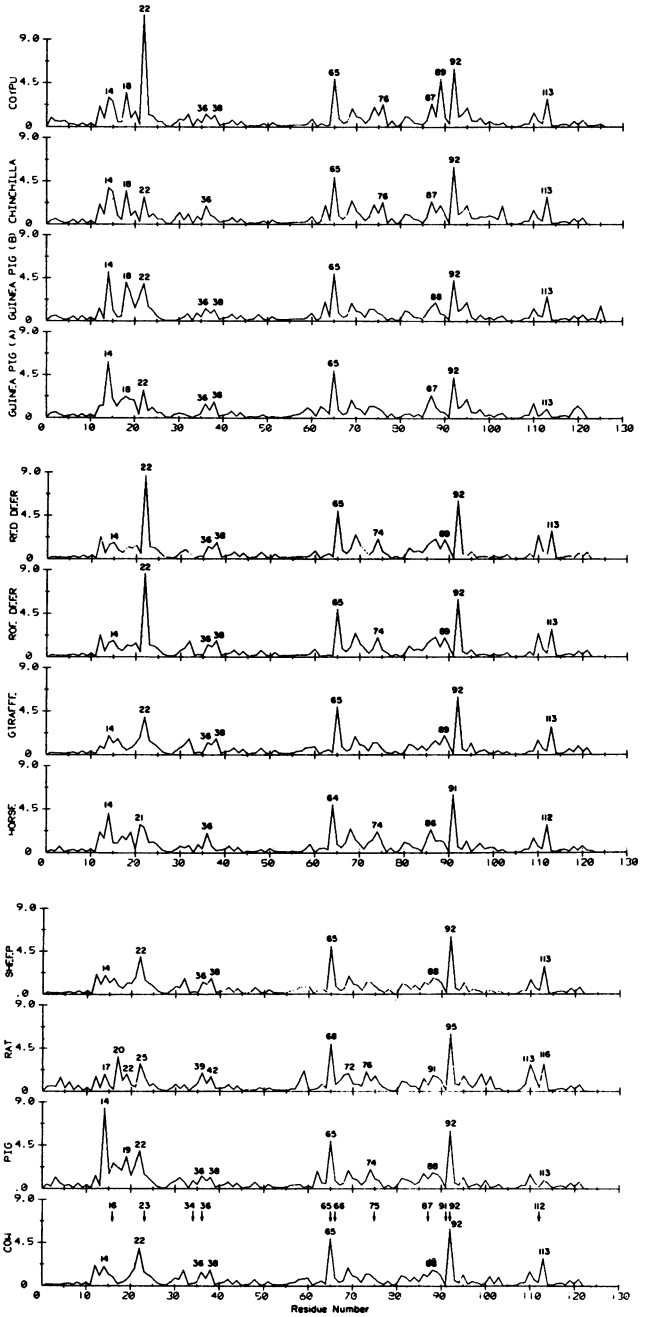


FIGURE 4 Probability of $\beta$-turn occurrences in 12 species of pancreatic ribonuclease. There are three extra residues at the $NH_2$-terminus of rat ribonuclease after homologous alignment with the bovine sequence so that the bend peaks appear to be shifted by +3 residues for rat when compared to the other species. Due to a deletion at residue position 39 for horse ribonuclease, the bend peaks after this position appear to be off by −1 residue when compared to the other profiles. The numbered peaks indicate the predicted $\beta$-turns in the 12 species after consideration of helical and $\beta$-sheet predictions in these ribonucleases.

TABLE IV

β-TURN CORRELATION COEFFICIENT MATRIX OF PANCREATIC RIBONUCLEASES

| | Bovine | Pig | Guinea pig B | Guinea pig A | Coypu | Giraffe | Roe deer | Red deer | Chin-chilla | Horse | Sheep | Rat |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Bovine | — | 0.87 | 0.94 | 0.88 | 0.85 | 0.98 | 0.90 | 0.90 | 0.76 | 0.79 | 0.91 | 0.75 |
| Pig | 0.87 | — | 0.92 | 0.86 | 0.91 | 0.85 | 0.92 | 0.92 | 0.83 | 0.83 | 0.78 | 0.83 |
| Guinea pig B | 0.94 | 0.92 | — | 0.98 | 0.95 | 0.92 | 0.84 | 0.84 | 0.86 | 0.78 | 0.86 | 0.84 |
| Guinea pig A | 0.88 | 0.86 | 0.98 | — | 0.93 | 0.86 | 0.78 | 0.78 | 0.88 | 0.80 | 0.79 | 0.78 |
| Coypu | 0.85 | 0.91 | 0.95 | 0.93 | — | 0.83 | 0.83 | 0.83 | 0.97 | 0.86 | 0.77 | 0.78 |
| Giraffe | 0.98 | 0.85 | 0.92 | 0.86 | 0.83 | — | 0.92 | 0.92 | 0.74 | 0.77 | 0.89 | 0.73 |
| Roe deer | 0.90 | 0.92 | 0.84 | 0.78 | 0.83 | 0.92 | — | 1.00 | 0.74 | 0.82 | 0.81 | 0.76 |
| Red deer | 0.90 | 0.92 | 0.84 | 0.78 | 0.83 | 0.92 | 1.00 | — | 0.74 | 0.82 | 0.81 | 0.76 |
| Chinchilla | 0.76 | 0.83 | 0.86 | 0.88 | 0.97 | 0.74 | 0.74 | 0.74 | — | 0.89 | 0.68 | 0.70 |
| Horse | 0.79 | 0.83 | 0.78 | 0.80 | 0.86 | 0.77 | 0.82 | 0.82 | 0.89 | — | 0.71 | 0.67 |
| Sheep | 0.91 | 0.78 | 0.86 | 0.79 | 0.77 | 0.89 | 0.81 | 0.81 | 0.68 | 0.71 | — | 0.67 |
| Rat | 0.75 | 0.83 | 0.84 | 0.78 | 0.78 | 0.73 | 0.76 | 0.76 | 0.70 | 0.67 | 0.67 | — |
| Mean $<C_t>$ | 0.87 | 0.87 | 0.88 | 0.85 | 0.86 | 0.85 | 0.85 | 0.85 | 0.80 | 0.79 | 0.79 | 0.75 |
| SD | 0.07 | 0.05 | 0.06 | 0.07 | 0.07 | 0.08 | 0.08 | 0.08 | 0.09 | 0.06 | 0.08 | 0.06 |

Mean of matrix, 0.83 (92% turns/65% sequence conserved). SD of matrix, 0.08.

$= 1.33 > <P_\alpha> = 0.77$ and $<P_\beta> = 0.89$ has a bend probability ($p_t = 5.0 \times 10^{-4}$) 10 times greater than the average probability of β-turn occurrence. An even higher bend probability occurs at residues 92–95 with $p_t = 6.0 \times 10^{-4}$ and $<P_\alpha> = 0.66 < <P_t> = 1.35 > <P_\beta> = 1.03$. This tetrapeptide sequence (Tyr-Pro-Asn-Cys) is conserved for all 12 ribonucleases with the exception of guinea pigs A and B (Phe-Pro-Asn-Cys). The conservative substitution Tyr → Phe, resulting in $p_t = 4.3 \times 10^{-4}$, indicates that β-turn 92–95 is preserved in all the species. Another high bend probability is found at position 22–25 (Ser-Ser-Asn-Tyr) for bovine, giraffe, sheep, pig, and guinea pig B ribonuclease with $p_t = 4.0 \times 10^{-4}$. Even higher bend probabilities are found (Fig. 4) for roe and red deer at residues 22–25 (Ser-Pro-Asn-Tyr) with $p_t = 8.6 \times 10^{-4}$, and for coypu at residues 22–25 (Asn-Pro-Asn-Tyr) with $p_t = 11.6 \times 10^{-4}$ because of the Ser → Pro substitution at the 2nd bend position. Although only 16 of the 36 β-turn residues (44%) observed from x-ray are conserved from sequence homology, the bend probability profiles of Fig. 4 show that the chain reversal regions are remarkably conserved in the 12 ribonucleases. This is mostly due to substitutions of residues that have similar bend positional potentials. For example, β-turn 36–39 (Thr-Lys-Asp-Arg), a type I bend found at the surface of bovine ribonuclease, has only Thr common in all the 12 species. The bend frequencies of Lys 37 in bovine, rat, and guinea pig A ($[f_{i+1}]_{Lys} = 0.115$) and Gln 37 in the other 9 species ($[f_{i+1}]_{Gln} = 0.098$) and both above the average $<f_j> = 0.086$ and are similar in magnitude. The same can be said of Asp 38 in bovine, sheep, guinea pig A, and roe and red deer ($[f_{i+2}]_{Asp} = 0.191$) and Gly 38 in the other 7 species ($[f_{i+2}]_{Gly} = 0.190$). Arg 39 appears in 9 of the species and has an average bend frequency $(f_{i+3})_{Arg} = 0.085$. Although there is a deletion at position 39 in horse ribonuclease, for sequence alignment purposes, Cys 40 becomes effectually the 4th bend residue with $(f_{i+3})_{Cys} = 0.128$. The homologous position for rat is Ser 42 ($[f_{i+3}]_{Ser} = 0.106$) and for chinchilla is Tyr 39 ($[f_{i+3}]_{Tyr} = 0.125$). Hence, the β-turn 36–39 appears to have enhanced

stability in the horse, rat, and chinchilla due to higher bend frequency residues at the 4th position.

It is striking that there are no tetrapeptides in the 12 species that have a high bend probability between the bend peaks 38 and 65 shown in Fig. 4. An examination of the bovine ribonuclease x-ray structure shows that region 42–65 contains helix 50–59 which is preceded and followed by $\beta$-sheets 42–48 and 60–65. The low bend probability in the $\beta\alpha\beta$ domain 42–65 is parallelled by the almost 0% accessibility of the backbone in this region of bovine ribonuclease to solvent molecules (Richards and Wyckoff, 1973). On the other hand, residues having high percent accessibility are mostly found in the $\beta$-turns, indicating that chain reversal regions are predominantly on the surface of proteins. An examination of the 8 Cys residues shows Cys 26, 65, and 95 involved in $\beta$-turns 23–26, 65–68, and 92–95, respectively. Furthermore, Cys 40, 72, 84, and 110 are found in the proximity of $\beta$-turns 36–39, 75–78, 87–90, and 112–115, respectively. It is interesting to note that the S-S bridge between the $\beta$-sheet residues Cys 65 and 72 is made possible by two adjacent bends 65–68 and 66–69. Hence, as in the case for the proteinase inhibitors, the $\beta$-turns in the ribonucleases appear to induce the correct tertiary folding before formation of disulfide bridges. From the above studies on proinsulins, proteinase inhibitors, and ribonucleases, it is apparent that $\beta$-turns are important in keeping certain structural domains within hormones and enzymes intact for their specific biological functions.

We appreciate the help of Mr. Tim Hickey and Mr. John M. Kennedy for assistance in computer programming.

This research was generously supported in part by grants from the U.S. Public Health Service (GM 17533), National Science Foundation (PCM 76-21856), and the American Cancer Society (P-577). This is publication No. 1250 from the Graduate Department of Biochemistry, Brandeis University, Waltham, Mass. 02154.

*Received for publication 10 January 1979.*

## REFERENCES

ARGOS, P., M. G. ROSSMAN, and J. E. JOHNSON. 1977. Four-helical super-secondary structure. *Biochem. Biophys. Res. Commun.* **75**:83.

BEINTEMA, J. J., W. GAASTRA, J. A. LENSTRA, G. W. WELLING, and W. M. FITCH. 1977. Molecular evolution of pancreatic ribonuclease. *J. Mol. Evol.* **10**:49.

BLUNDELL, T., G. DODSON, D. HODGKIN, and D. MERCOLA. 1972. Insulin: the structure in the crystal and its reflection in chemistry and biology. *Adv. Protein Chem.* **26**:279.

CHANCE, R. E., R. M. ELLIS, and W. W. BROMER. 1968. Procine proinsulin: characterization and amino acid sequence. *Science (Wash. D.C.).* **161**:165.

CHOU, P. Y., A. J. ADLER, and G. D. FASMAN. 1975. Conformational prediction and circular dichroism studies on the *lac* repressor. *J. Mol. Biol.* **96**:29.

CHOU, P. Y., and G. D. FASMAN. 1974. Prediction of protein conformation. *Biochemistry.* **13**:222.

CHOU, P. Y., and G. D. FASMAN. 1977. $\beta$-Turns in Proteins. *J. Mol. Biol.* **115**:135.

CHOU, P. Y., and G. D. FASMAN. 1978. Prediction of the secondary structure of proteins from their amino acid sequence. *Advances in Enzymol.* **47**:45.

CHOU, P. Y., and G. D. FASMAN. 1979. Prediction of $\beta$-Turns. *Biophys. J.* **26**:367.

CREIGHTON, T. E. 1975*a*. Homology of protein structures: proteinase inhibitors. *Nature (Lond.).* **255**:743.

CREIGHTON, T. E. 1975*b* The two-disulphide intermediates and the folding pathway of reduced pancreatic trypsin inhibitor. *J. Mol. Biol.* **95**:167.

CREIGHTON, T. E. 1975*c*. Interactions between cysteine residues as probes of protein conformation: the disulphide bond between Cys-14 and Cys-38 of the pancreatic trypsin inhibitor. *J. Mol. Biol.* **96**:767.

DAYHOFF, M. O. 1972. Atlas of Protein Sequence and Structure, National Biomedical Research Foundation, Georgetown University Medical Center, Washington, D.C. **5**:D-187.

DEISENHOFER, J., and W. STEIGEMANN. 1975. Crystallographic refinement of the structure of bovine pancreatic trypsin inhibitor at 1.5 Å resolution. *Acta Cryst allogr. Sect. B Struct. Crystallogr. Cryst. Chem.* **B31**:238.

FRANK, B. H., and A. J. VEROS. 1968. Physical studies on proinsulin—association behavior and conformation in solution. *Biochem. Biophys. Res. Commun.* **32**:155.

FULLERTON, W. W., R. POTTER, and B. W. LOW. 1970. Proinsulin: crystallization and preliminary x-ray diffraction studies. *Proc. Natl. Acad. Sci. U.S.A.* **66**:1213.

GAASTRA, W., G. GROEN, G. W. WELLING and J. J. BEINTEMA. 1974. The Primary Structure of Giraffe Pancreatic Ribonuclease. *FEBS Fed. Eur. Bio Lett.* **41**:227.

LASKOWSKI, M., Jr., I. KATO, T. R. LEARY, J. SCHRODE, and R. W. SEALOCK. 1974. Evolution of specificity of protein proteinase inhibitors. *In* Proteinase Inhibitors. H. FRITZ, H. TSCHESCHE, L. J. GREENE, and E. TRUSCHEIT, editors. Springer-Verlag, Berlin. pp. 597.

LENSTRA, J. A., J. HOFSTEENGE, and J. J. BEINTEMA. 1977. Invariant features of the structure of pancreatic ribonuclease. A test of different predictive models. *J. Mol. Biol.* **109**:185.

LOW, T. L. K., Y. S. V. LIU, and F. W. PUTNAM. 1976. Structure, function, and evolutionary relationships of $F_c$ domains of human immunoglobulins A, G, M, E. *Science (Wash. D.C.)* **191**:390.

MARKUSSEN, J., and F. SUNDBY. 1972. Rat-proinsulin C-peptides. amino-acid sequences. *Eur. J. Biochem.* **25**:153.

MARKUSSEN, J., and A. VØLUND. 1975. Conformational analysis of circular dichroism spectra of insulin, proinsulin and C-peptides by non-linear regression. *Int. J. Pept. Protein Res.* **7**:47.

MATTHEWS, B. W. 1975. Comparison of the predicted and observed secondary structure of T4 phage lysozyme. *Biochim. Biophys. Acta.* **405**:442.

MENEZ, A., G. LANGLET, N. TAMIYA, and P. FROMAGEOT. 1978. Conformation of snake toxic polypeptides studied by a method of prediction and circular dichroism. *Biochimie. (Paris).* **60**: 505.

NISHIKAWA, K., and T. OOI. 1974. Comparison homologous tertiary structure of proteins. *J. Theor. Biol.* **43**:351.

PADLAN, E. A., and W. E. LOVE. 1974. Three-dimensional structure of haemoglobin from the polychaete annelid, *Glycera dibranchiata*, at 2.5 Å resolution. *J. Biol. Chem.* **249**:4067.

PEKAR, A. H., and B. H. FRANK. 1972. Conformation of proinsulin. A comparison of insulin and proinsulin self-association at neutral pH. *Biochemistry.* **11**:4013.

PENA, C., J. M. SEWART, A. C. PALADINI, J. M. DELLACHA, and J. A. SANTOME. 1975. Conformational analysis of growth hormones and synthesis of a fragment of human growth hormone. *In* Peptides: Chemistry, Structure, and Biology. R. Walter and J. Meienhofer, editors. Ann Arbor Science, Publishers, Inc., Ann Arbor, Mich. 523–528.

PTITSYN, O. B. 1974. Invariant features of globin primary structure and coding of their secondary structure. *J. Mol. Biol.* **88**:287.

PULLEN, R. A., D. G. LINDSAY, S. P. WOOD, I. J. TICKLE, T. L. BLUNDELL, A. WOLLMER, G. KRAIL, D. BRANDENBURG, H. ZAHN, J. GLIEMANN, and S. GAMMELTOFT. 1976. Receptor-binding region of insulin. *Nature (Lond.)* **259**:369.

REDDY, M. N., P. S. KEIM, R. L. HEINRIKSON, and F. J. KEZDY. 1975. Primary structural analysis of sulfhydryl protease inhibitors from pineapple stem. *J. Biol. Chem.* **250**:1741.

RICHARDS, F. M., and H. W. WYCKOFF. 1973. Atlas of Molecular Structures in Biology-Ribonuclease S. Clarendon Press, Oxford, England.

RICHARDSON, J. S., D. C. RICHARDSON, K. A. THOMAS, E. W. SILVERTON, and D. R. DAVIES. 1976. Similarity of three-dimensional structure between the immunoglobulin domain and the copper, zinc superoxide dismutase subunit. *J. Mol. Biol.* **102**:221.

ROSSMAN, M. G., and P. ARGOS. 1975. A comparison of the heme binding pocket in globulins and cytochrome $b_5$. *J. Biol. Chem.* **250**:7525.

RÜHLMANN, A., D. KUKLA, P. SCHWAGER, K. BARTELS, and R. HUBER. 1973. Structure of the complex formed by bovine trypsin and bovine pancreatic trypsin inhibitor. Crystal structure determination and stereochemistry of the contact region. *J. Mol. Biol.* **77**:417.

SAWYER, L., D. M. SHOTTON, J. W. CAMPBELL, P. L. WENDELL, H. MUIRHEAD, H. C. WATSON, R. DIAMOND, and R. C. LADNER. 1978. The atomic structure of crystalline porcine pancreatic elastase at 2.5 Å resolution: comparisons with the structure of $\alpha$-chymotrypsin. *J. Mol. Biol.* **118**:137.

SCHEFFER, A. J., and J. J. BEINTEMA. 1974. Horse pancreatic ribonuclease. *Eur. J. Biochem.* **46**:221.

SHOTTON, D. M., and B. S. HARTLEY. 1970. Three-dimensional structure of tosyl-elastase. *Nature (Lond.).* **225**:802.

SMYTH, D. G., W. H. STEIN, and S. MOORE. 1963. The sequence of amino acid residues in bovine pancreatic ribonuclease: revisions and confirmations. *J. Biol. Chem.* **238**:227.

SMYTHIES, J. R., F. BENINGTON, R. J. BRADLEY, W. F. BRIDGES, R. D. MORIN, and W. O. ROMINE, Jr. 1975. The molecular structure of the receptor-ionophore complex at the neuromuscular junction. *J. Theor. Biol.* **51**:111.

SNELL, C. R., and D. G. SMYTH. 1975. Proinsulin: a proposed three-dimensional structure. *J. Biol. Chem.* **250**:6291.

STEINER, D. F., and P. OYER. 1967. The biosynthesis of insulin and a probable precursor of insulin by a human islet cell adenoma. *Proc. Natl. Acad. Sci. U.S.A.* **57**:473.

VAN DEN BERG, A., and J. J. BEINTEMA. 1975. Non-constant evolution rates in amino acid sequences of guinea pig, chinchilla and coypu pancreatic ribonucleases. *Nature (Lond.)*. **253**:207.

VOGT, H-P., A. WOLLMER, V. K. NAITHANI, and H. ZAHN. 1976. The conformational potential of porcine proinsulin C-peptide. *Hoppe-Seyler's Z. Physiol. Chem.* **357**:107.

WOOTON, J. C. 1974. The coenzyme-binding domains of glutamate dehydrogenases. *Nature (Lond.)*. **252**:542.

YU, N-T., C. S. LIU, and D. C. O'SHEA. 1972. Laser Raman spectroscopy and the conformation of insulin and proinsulin. *J. Mol. Biol.* **70**:117.