Biochem. J. (1986) 240, 289–292 (Printed in Great Britain)

Four classes of β-hairpins in proteins

E. James MILNER-WHITE* and Ron POET†
Departments of *Biochemistry and †Computing Science, University of Glasgow, Glasgow G12 8QQ, Scotland, U.K.

We show that β-hairpins can be divided into four classes, each with a number of members. Hairpins from a single class are readily interconverted by loss or gain of hydrogen bonds, but interconversion between classes requires complete unzipping and reformation of the entire β-hairpin. Sibanda & Thornton [(1985) Nature (London) **316**, 170–174] have classified β-hairpins as either two-residue, three-residue, four-residue etc., loops. We point out that their nomenclature, by itself, gives rise to ambiguities, but that, if the class (one of the four mentioned above) is also specified, the description of β-hairpins becomes straightforward. A range of proteins of known three-dimensional structure has been examined; it provides examples of hairpins of each of the four classes and give some indication of their frequency of occurrence. The distribution observed is substantially different from that described by Sibanda & Thornton [(1985) Nature (London) **316**, 170–174].

INTRODUCTION

In order to understand how the amino acid sequence of a newly synthesized polypeptide folds to the three-dimensional structure of the protein, it is necessary to comprehend the principles of the structures themselves. Well-known features include α-helices, β-sheets, β-bulges and β-turns (Richardson, 1981). Another feature that is also common, yet has received surprisingly little attention until recently (Sibanda & Thornton, 1985), is the β-hairpin. A β-hairpin is a polypeptide chain that folds back on itself so that the two halves constitute two adjacent strands of an antiparallel β-sheet. Sibanda & Thornton (1985) describe β-hairpins in terms of the number of 'loop residues'; thus they can be considered as either two-residue, three-residue, four-residue etc., loops. We show that this terminology, by itself, is defective in that, in many cases, it gives the same name to structurally different hairpins. However, our observation tht β-hairpins fall naturally into four classes can be used to provide a means of referring to hairpins that is unambiguous. A preliminary report along these lines has already been compiled (Milner-White, 1986). We have also developed computer-graphics displays (Milner-White & Poet, 1985; Poet & Milner-White, 1986) that are suitable for picturing these features.

METHODS

For a β-hairpin the criteria used in the present paper are that there must be at least two hydrogen bonds between main chain NH and CO groups arranged in the antiparallel β-sheet conformation as follows: either (i) both are between residues i and $(i+n)$ where $1 < n < 10$, or (ii) one is between residues i and $(i+n)$ and the other is between residues $(i-2)$ and $(i+n+2)$ where $1 < n < 10$.

Of course additional inter-main-chain hydrogen bonds frequently occur, but the presence of the two bonds specified above constitutes a minimum. Note also that, if within a β-hairpin one hydrogen bond is missing, the gap is ignored for the purposes of describing the hairpin, provided the appropriate flanking hydrogen bonds are present.

Hydrogen bonds are defined as described by Baker & Hubbard (1984). We used the slightly more rigorous definition that they recommended, rather than the one that they used. Briefly, the angle between the N, H and O atoms has to be more than 120°, the angle between the H, O and C atoms has to be more than 90°, and the distance between the H and O atoms has to be less than 0.25 nm. Where the position of the hydrogen atom is not defined, the sole criteria is that the distance between N and O atoms has to be less than 0.35 nm.

The proteins examined are those in the Protein Data Bank (Bernstein *et al.*, 1977) whose structures have been determined to a resolution of 0.2 nm or better by X-ray crystallography. They are listed in Table 1. A FORTRAN computer program was written that lists all the inter-main-chain hydrogen bonds in a protein using the above criteria. It also lists the main-chain dihedral angles.

RESULTS AND DISCUSSION

The diagrams in Fig. 1(*a*) are an attempt to portray all the possible types of hydrogen-bond arrangements that may be found at the loop end of β-hairpins. They are drawn so that adjacent hairpins in one row are related by the presence or absence of a single hydrogen bond. Each row forms a series that in principle extends infinitely to the right. All β-hairpins can be placed unambiguously in one, and only one, of the four rows in Fig. 1(*a*), which can be seen to correspond to four structurally distinct classes. They are distinct because any interconversion between classes requires complete rupture of all hydrogen bonds in the β-hairpin and reformation of new ones, between different pairs of amino acids. On the other hand, any interconversion between members of one class is expected to involve the retention of at least one of the existing hydrogen bonds. It may be predicted that, during evolution, one hairpin is more likely to evolve into another of the same class than into one of a different class. This is particularly likely because amino acid substitutions are more commonly found than insertions or deletions. Any of the latter, at the loop, would be expected to give rise to a change in class.

Fig. 1(*a*) also give the name (one-residue, etc.),

Table 1. Proteins examined

Below is a list of the 39 proteins that have been examined with regard to their β-hairpin content. Their Protein Data Bank (PDB) codes are also given.

Protein	PDB code
Cytochrome c_{551}	551C
Actinidin	2ACT
Penicillopepsin	2APP
Cytochrome b_5	2B5C
Phospholipase A_2	1BPL
Carbonic anhydrase	2CAB
Chymotrypsin (tosyl)	2CHA
Crambin	1CRN
Carboxypeptidase A	5CPA
Citrate synthase	2CTS
Cytochrome *c*	3CYT
Cytochrome *c* peroxidase	1CYP
Dihydrofolate reductase	5DFR
Erythrocruorin	1ECD
Immunoglobulin Fab′ fragment	3FAB
Ferredoxin	1FDX
Glutathione reductase	2GRS
Flavodoxin	3FXN
High-potential iron protein	1HIP
Deoxyhaemoglobin	4HHB
Haemerythrin	1HMQ
Insulin	1INS
Leghaemoglobin	1LH1
Lysozyme	1LYZ
Myoglobin	1MBD
Prealbumin	2PAB
Plastocyanin	1PCY
Kallikrein A	2PKA
Avian pancreatic polypeptide	1PPT
Pancreatic trypsin inhibitor	4PTI
Ribonuclease A	1RN3
Rubredoxin	2RXN
Streptomyces griseus proteinase A	2SGA
Ovomucoid domain 3	3SGB
Scorpion neurotoxin	1SN3
Staphylococcus nuclease	2SNS
Superoxide dismutase	2SOD
Trypsinogen	1TGN
Thermolysin	3TLN

according to the rules devised by Sibanda & Thornton (1985), of each of the hairpins illustrated. The four classes have been named 1,2,3 and 4 because the tightest members of each class are one-residue, two-residue, three-residue and four-residue hairpins, respectively. At first sight the existence of a class 5, corresponding to a five-residue hairpin, appears to be possible, but such a hairpin has to belong to class 1, because it is related, by partial unzipping, to a one-residue hairpin. This is why there are only four classes. Fig. 1(*a*) also reveals that many hairpins with the same name occur in two different classes. This anomaly is discussed below.

Sibanda & Thornton (1985) named hairpins according to the number of loop residues not involved in the β-ladder. The latter refers to the residues taking part in the antiparallel β-sheet and is defined precisely by Kabsch & Sander (1983). The nomenclature is straightforward for hairpins where the hydrogen-bonded amino acids nearest the loop end are linked by a pair of hydrogen bonds between both sets of CO and NH groups. The loop residues are simply those not involved in such hydrogen-bonding. However, where there is only a single hydrogen bond (that between the NH group of one amino acid and the CO group of an amino acid ahead of it in sequence) between the pair of amino acids at the loop end, it is not immediately obvious how to name the loop. The explanation is that Kabsch & Sander's (1983) β-ladder is defined so that such a pair of amino acids are regarded as loop residue rather than as part of the β-ladder.

According to the above nomenclature, a two-residue loop, which has a pair of hydrogen bonds between two amino acids at the loop end, becomes a four-residue loop when the bond nearest the loop end is lost. This change involves no alteration in class. However, the term 'four-residue loop' also applies to a hairpin with a pair of bonds at the loop end and belonging to a different class. Examination of Fig. 1 confirms that the two four-residue loops are indeed structurally different. The problem of having one name for two different loops also applies to all possible names with the exception of one-residue and two-residue hairpins. Clearly it will give rise to confusion. In the present paper we circumvent the problem by referring to an *n*-residue loop as being of a particular class.

Further evidence for the need to distinguish between all of the hairpins in Fig. 1(*a*) comes from a consideration of some of the data in Sibanda & Thornton's (1985) paper. They found that about half of all four-residue and five-residue hairpins form distinct structural families. The two families are based on the presence of a type I turn spanning loop residues L1 and L4 in both cases, plus a set of main-chain dihedral angles that are characteristic for each family. However, all of the eight examples of such five-residues hairpins they quote are class 3 structures (see Fig. 1). Sibanda & Thornton (1985) present six examples of their structurally distinct four-residue hairpins. Of these, five are of the class 4 type but one (papain) is a class 2 four-residue turn. However, examination of its dihedral angles show that the ϕ values are all negative and that the ϕ,ψ values for loop residues L1 and L4 are atypical, compared with others in the family; thus it cannot be said to belong to the family. These angles are plotted in Fig. 2(*c*) of Sibanda & Thornton's (1985) paper, but the above information is needed in order to know to which protein the angles refer. We conclude that the two structural families are specific to one only of the two types of four-residue hairpins and five-residue hairpins respectively. This confirms the need for a hairpin classification providing a unique name for each of the structures in Fig. 1(*a*).

Fig. 2 shows the frequency distribution of the different types of hairpins in Fig. 1 among the sample of 39 proteins whose three-dimensional structures have been determined at high resolution by X-ray crystallography. The criteria used have been described in the Methods section. Table 2 gives the names of hairpin-containing proteins and the relevant amino acid sequence numbers. Note that some proteins have as many as four of the same hairpin. Fig. 2 can be compared with Fig. 1 in Sibanda & Thornton (1985). They found that the order, in terms of numbers observed, for hairpins was: two-residue > five-residue > four-residue > others. We agree that two-residue hairpins are commonest, but are

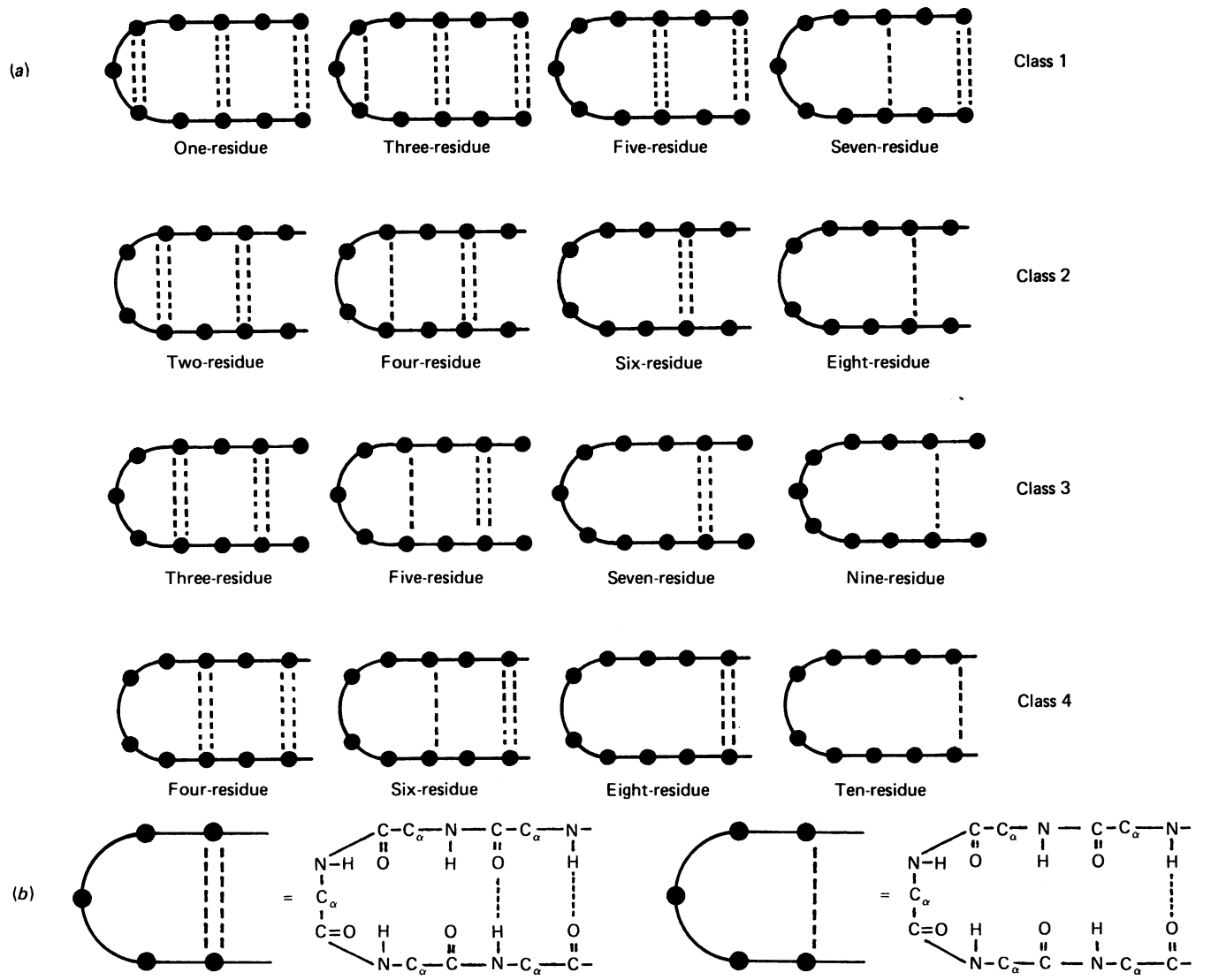

Fig. 1. Four classes of β-hairpins

In (*a*) the main chain is drawn as a series of dots, representing amino acids, linked together. The inter-main-chain hydrogen bonds are portrayed as broken lines between the amino acids they join. The four classes are illustrated by the four rows. The simplified drawing conventions used may be understood by reference to (*b*), which illustrates how the diagrams drawn in this Figure are related to more conventional atomic formulae. Hydrogen bonds are again drawn as broken lines.

able to state that the next commonest hairpins are of the class 3 five-residue type, rather than class 1 five-residue hairpins (see beow). The two types of four-residue hairpins are both fairly common, but no more so, as far as the data reveal, than class 3 three-residue, class 3 seven-residue, class 4 six-residue and class 4 eight-residue hairpins. Hence, by using the improved definitions of hairpins, a substantially different pattern of hairpin distribution has emerged.

The distribution of hairpins within classes suggests that certain members of each class are favoured compared with others. For example, in class 2, two-residue hairpins are especially frequent, whereas four-residue hairpins are less so. Reasons for the stability of two-residue hairpins have been discussed by Sibanda & Thornton (1985). Within class 3, on the other hand, five-residue, rather than three-residue, hairpins appear to be favoured. This is likely to be a consequence of the observation that the loop end of a class 3 five-residue hairpin often has the structure of a G1-type β-bulge (Richardson *et al.*, 1978; Sibanda & Thornton, 1985). Within class 4, four-residue and six-residue hairpins are about equally frequent, and it is worth noting that the loop ends of both can easily adopt the structure of a G1 type β-bulge [with one extra loop residue, as noted by Richardson *et al.* (1978), compared with class 3]. Within class 1 it is striking that five-residue hairpins are absent and there is only one seven-residue hairpin. However, there are three one-residue hairpins. The latter can be thought of as five-residue hairpins with a γ-turn, defined (Richardson, 1981) as a hydrogen bond between the CO group of loop residue 2 and the NH group of loop residue 4. (It should be added that, for two of these proteins, there is no hydrogen bond between the NH group of loop residue 2 and the CO group of loop residue 4.) It seems that one-residue hairpins are preferred to five-residue ones.

In describing secondary structure, the criteria used for

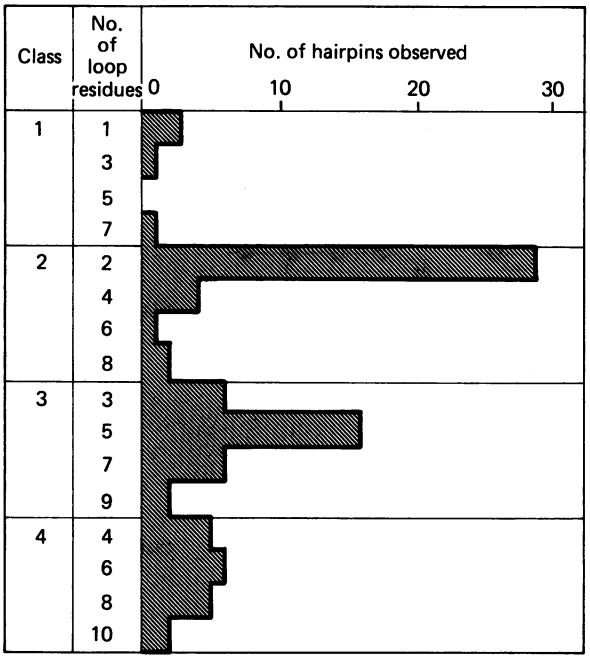

Fig. 2. Frequency of β-hairpins

The Figure indicates the number of β-hairpins, of the different types, that are found among the 39 proteins examined. They are listed in Table 2.

hydrogen bonds are critical. If these are stringent, fewer bonds appear to be present than if they are broad. This uncertainty means that the same hairpin can be described by two, or even more, of the structures in Fig. 1(*a*), depending on the criteria used. However, for a given hairpin, the structures must all belong to the same row or class. Thus the class of a hairpin is, to some extent, a more reliable feature than its categorization on the basis of the number of loop residues. Fig. 2 shows that examples of each of the four hairpin classes are commonly found. In the sample of 39 proteins studied, their frequencies, expressed as a precentage of the total, are: 6, 40, 34 and 20, for classes 1, 2, 3 and 4 respectively.

To conclude, an average protein in a random sample from the Protein Data Bank has at least two of the β-hairpins described. Therefore the idea that the different types of hairpins occur in four different classes, and the ability this confers to appreciate them and to give each type an unambiguous name, should prove useful.

REFERENCES

Baker, E. N. & Hubbard, R. E. (1984) Prog. Biophys. Mol. Biol. **44**, 97–179

Bernstein, F. C., Koetzle, T. K., Williams, G. J. B., Meyer, E. F., Brice, M. D., Rogers, J. R., Kennard, O., Simanouchi, T. & Mitsuo, T. (1977) J. Mol. Biol. **112**, 535–542

Kabsch, W. & Sander, C. (1983) Biopolymers **22**, 2577–2637

Milner-White, E. J. (1986) Biochem. Soc. Trans. **14**, 877

Milner-White, E. J. & Poet, R. (1985) Biochem. Soc. Trans. **13**, 793–795

Poet, R. & Milner-White, E. J. (1986) Computer Graphics Forum **5**, 211–215

Richardson, J. S. (1981) Adv. Protein Chem. **34**, 167–339

Richardson, J. S., Getzoff, E. D. & Richardson, D. C. (1978) Proc. Natl. Acad. Sci. U.S.A. **75**, 2574–2578

Sibanda, B. L. & Thornton, J. M. (1985) Nature (London) **316**, 170–174

Table 2. β-Hairpins observed

All the β-hairpins found (89) in the 39 proteins examined are listed below, according to type. They are referenced by their Protein Data Bank codes (omitting the number at the left-hand side) and the amino acid sequence numbers of the loop residues. The numbers are those used in the Protein Data Bank.

Class	No. of loop residues	Name of protein and amino acid sequence number of loop residues
1	1	GRS (372), SGA (100), TLN (26)
	3	TGN (37–9)
	5	
	7	GRS (133–139)
2	2	ACT (64–5), ACT (173–4), ACT (191–2), APP (24–5), APP (93–4), APP (201–2), APP (261–2), B5C (26–7), BPL (79–80), CAB (110–1), CHA (204–5), CPA (160–1), DFR (16–7), FAB (L67–8), FAB (L92–3), GRS (143–4), PAB (113–4), PKA (36–8), PKA (202–3), SGA (34–9), SGA (48C–D), SGA (84–5), SGA (202–7), SNS (95–6), SN3 (43–4), TGN (116–7), TGN (202–3), TLN (36–7), TLN (251–2)
	4	CAB (62–65), FAB (L199–202), SOD (23–26), TLN (116–9)
	6	SNS (26–31)
	8	GRS (255–62), PCY (6–13)
3	3	BPL (26–28), CTS (60–2), FXN (57–9), SGA (140–56), SGA (26–28), CTS (60–2), FXN (57–9), SGA (140–56), SGA (220–2), SOD (89–91)
	5	APP (75–9), CHA (35–39), CYP (216–20), CYP (225–9), FAB (L169–73), FAB (H52–6), LYZ (46–50), PAB (36–40), PKA (47–51), SGA (120–120D), SGA (171–5), SGB (25–9), SNS (18–22), SNS (83–7), TGN (47–51), TLN (12–6)
	7	APP (157–63), CHA (46–52), FAB (L37–43), FAB (H39–45), FAB (H174–80), SOD (9–15)
	9	CAB (231–9), PCY (84–92)
4	4	APP (240–3), APP (312–5), FAB (H204–9), PAB (18–23), RN3 (112–5)
	6	APP (11–16), APP (277–82), FDX (24–9), GRS (321–6), LYZ (53–58), PTI (24–9)
	8	CAB (80–87), CYP (180–7), DFR (145–52), FAB (H71–8), RXN (5–12)
	10	RN3 (63–72), RN3 (87–96)

Received 6 May 1986/9 September 1986; accepted 15 September 1986