Journal of Computer-Aided Molecular Design, **13:** 469–483, 1999.
KLUWER/ESCOM

Design criteria for molecular mimics of fragments of the β-turn. 1. C_α atom analysis

S.L. Garland* & P.M. Dean
Drug Design Group, Department of Pharmacology, University of Cambridge, Tennis Court Road, Cambridge CB2 1QJ, U.K.

Received 16 June 1998; Accepted 15 December 1998

Key words: beta turn, data mining, drug design, molecular similarity, peptidomimetics

Summary

Peptides represent an extensive class of biologically active molecules. They may be used as leads in the development of novel therapeutic agents provided the pharmacophoric information present within them can be translated into non-peptide analogs that lack the peptide backbone and are stable to proteolysis. This is the rationale for peptidomimetic drug design. Frequently, the β-turn has been implicated as a conformation important for biological recognition of peptides. Empirical evidence from known peptidomimetics, coupled with a theoretical model of peptide binding and the observation that glycine and proline residues are common within the β-turn, has suggested the design of molecules to mimic placement of between two and four of the side-chains. The moderate number of different β-turn conformations, combined with the combinatoric nature of side-chain selection complicates the procedure. In this paper, cluster analysis has been used to classify the arrangement of C_α atoms about the various fragments of the β-turn. Recombination of the observed patterns provides a general model for the β-turn which may be used as an effective screen for potential peptidomimetic scaffolds in chemical databases.

Introduction

Many hundreds of biologically active peptides are known [1]. In principle, treatment with these molecules could provide a range of therapeutically useful effects. In practice, however, the peptides may not be used directly. The peptide backbone is proteolytically labile and reduces the bioavailability drastically [2, 3]. A number of researchers have sought to design molecules where the peptide backbone is replaced, thus circumventing the drug delivery problems experienced with the native peptides and providing structures with potential clinical value [4–6]. In so doing, it is implicitly assumed that it is the side-chains of the peptide and not the backbone that form the key interactions with the receptor. This is clearly a simplification and there will be cases where the backbone interactions are highly important or even dominate, but it is often possible to develop biologically active peptidomimetics on this basis [7, 8].

Variations in the structures developed to date have suggested that it is often possible to obtain active peptides and peptide mimetics that reproduce only a small number of the native amino acid side-chains. Typically structures are designed to reproduce the disposition of between two and four side-chains in three dimensions. This observation is in accordance with a theoretical model of peptide binding, where Farmer has postulated that three side-chains are sufficient for nanomolar affinity and discrimination for one molecule in approximately half a million at the receptor [9]. These values generally correspond to those observed with the native peptide and are typical for known drugs.

Information regarding which residues are key for the binding interaction at the receptor may be obtained experimentally through procedures such as the alanine scan [10–12]. Arguably it ought to be possible to establish whether any conformational change occurs through the use of physical techniques such as X-ray crystallography or nuclear magnetic resonance (NMR) [13]. The application of these procedures to peptides

*Present address: SmithKline Beecham Pharmaceuticals plc, Harlow, Essex CM19 5AD, U.K.

is far from simple and extension to the all-important receptor-bound conformation extremely difficult. This explains the use of the alanine scan. There would be no need to synthesize numerous homologues and infer details relating to binding if the bioactive conformation of the peptide could readily be established by direct experiment. Another approach has been to use presumed bioactive conformations. The β-turn has repeatedly been implicated as an important element for the biological recognition of proteins and peptides [14, 15]. A number of clinically interesting peptides have had the β-turn proposed for the biologically active conformation, including somatostatin [16], LHRH [17], GnRH [18], oxytocin [19], bradykinin [20] and enkephalin [21]. Attempts have been made to classify the β-turn for use in drug design [22]. The design of mimics for this structure provides molecules with which to validate the importance of the conformation for these and other peptides, as well as providing structures of potential therapeutic value in those instances where the hypothesis is proved correct [23].

A general approach to the design of β-turn peptidomimetics has been sought. Since between two and four side-chains are thought to be required for activity and the β-turn comprises four residues, design constraints have been derived for all 2-, 3- and 4-side-chain mimetics. In this, the first of two papers, the positions of the C_α atoms about the various turns are examined for all such fragments. A general model of the β-turn emerges that may be used as a rapid screen for peptidomimetic scaffolds in databases of known structures. A second analysis, of the C_α–C_β bond vectors, identifies patterns in the way in which side-chains are oriented in three dimensions [24]. These arrangements may be used as design constraints for the creation of novel peptidomimetics de novo, both for specific peptidal targets and for the creation of a small number of general combinatorial screening libraries based on peptide structure.

Structure of the β-turn

The β-turn is a tetrapeptidal segment that reverses the overall direction of a peptide chain (Figure 1). The exact conformation of the β-turn is subject to considerable variation, although a number of favourable, 'idealized' conformations have been identified [14, 25–27] (Table 1). Of these, 11 are well-defined (types IV and VII are excluded) and a series of design constraints have been sought for the selection of mimetics.

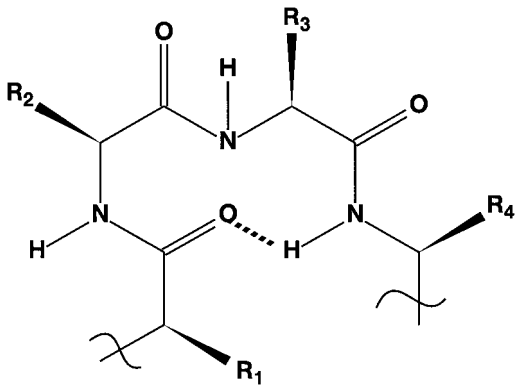

Figure 1. Schematic representation of the β-turn, including the optional hydrogen bond between the carbonyl oxygen of residue i (1) and the amide hydrogen of residue i + 3 (4).

Table 1. The backbone torsion angles of the various identified idealized β-turn types

	$i+1$		$i+2$	
	φ	ψ	φ	ψ
I	−60	−30	−90	0
I′	60	30	90	0
II	−60	120	80	0
II′	60	−120	−80	0
III	−60	−30	−60	−30
III′	60	30	60	30
IV	Turns with 2 or more angles that differ by at least 40° from the definitions of turn types I, I′, II, II′ , III and III′			
V	−80	80	80	−80
V′	80	−80	−80	80
VIa (*cis*)	−60	120	−90	0
VIb (*cis*)	−120	120	−60	0
VII	≈180	−60 to 60	−60 to 60	≈180
VIII	−60	−30	−120	120

The tight steric constraints of the β-turn mean that glycine and proline are commonly included in the sequence [14]. The type II turn, for instance, is almost invariably found with glycine in the third (i + 2) position and the type VI turns generally require proline to stabilize the intervening cis peptide bond. The use of the alanine scan to determine the recognition sequence of a peptide under these circumstances might well result in a significant decrease in biological activity, incorrectly identifying them as important for the binding interaction. Rather, it could well be the different steric characteristics of these residues that leads to a decrease in the binding affinity in the scan. Moreover, in almost all the peptidomimetics developed to

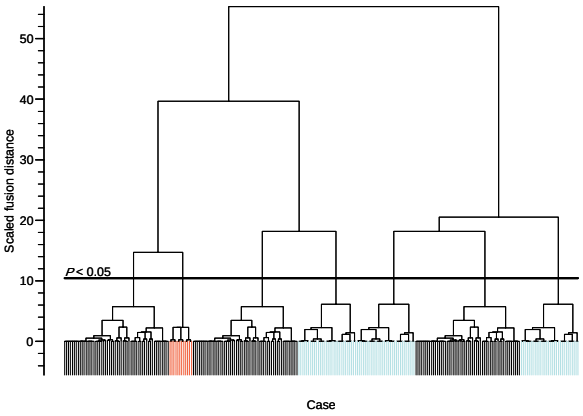

Figure 2. Cluster analysis of C_α atom positions for atom triplet fragments. Stopping rules determine the presence of seven significantly different classes. Close inspection of the membership of the classes reveals symmetry within the dendrogram. The first, third and sixth classes (numbered from left to right) are related and are coloured black in the dendrogram. Classes four, five and seven are also related and are coloured cyan. The second class, coloured red, is independent and has no associate classes but has a three-fold symmetry within the cluster.

date, it is assumed that it is the side-chains that form the majority of the binding interaction at the receptor. Since glycine has no side-chain whatsoever and proline very little with which to affect an interaction, there is little point constraining the design of mimetics for those positions.

A combination of this feature with empirical and theoretical considerations of peptide binding suggests the design of molecules that place between two and four of the amino acid side-chains about the β-turn. The moderately large number of different turn types, coupled with the combinatoric nature of side-chain selection, makes doing so prohibitive without simplification. Cluster analysis of the distances between C_α atoms in the various different β-turn types has been performed. This has revealed distinct patterns in the distribution of these atoms in 3-D space allowing us to define a small number of distance constraints that may be used to search for potential peptidomimetic scaffolds of this type.

Method

The search for molecules capable of correctly placing the C_α atoms of only two residues is simple. Two atoms are sought that are the correct distance apart and that may be potential positions for amino acid substitution. These distances are given in Table 2. It

Table 2. The distance (in Å) between C_α atom pairs for 11 idealized β-turn types

	1–2	1–3	1–4	2–3	2–4	3–4
I	3.8	5.2	4.6	3.8	5.3	3.8
I′	3.8	5.2	4.6	3.8	5.3	3.8
II	3.8	6.1	4.7	3.8	5.2	3.8
II′	3.8	6.1	4.7	3.8	5.2	3.8
III	3.8	5.5	5.4	3.8	5.2	3.8
III′	3.8	5.5	5.4	3.8	5.2	3.8
V	3.8	5.7	6.1	3.8	5.7	3.8
V′	3.8	5.7	6.1	3.8	5.7	3.8
VIa (*cis*)	3.8	5.5	4.1	2.8	4.7	3.8
VIb (*cis*)	3.8	5.6	2.7	2.8	3.6	3.8
VIII	3.8	5	6.7	3.8	6.6	3.8

may be seen that the distance between the C_α atom of any residue and the C_α atom of a neighbouring residue is 3.8 Å, with the exception of the distance between the C_α atoms of residues 2 and 3 for the type VIa and VIb turns, where the intervening peptide bond is cis rather than trans and the corresponding distance is 2.8 Å. The other C_α–C_α distances vary according to turn type, although it may be noted that the distances for a particular turn type and its inverse are exactly the same; the inversion of the turn alters its direction, but does not affect the distances between the C_α atoms.

A database of chemical structures may be rapidly screened for molecules capable of satisfying these criteria. Potential 2-residue peptidomimetics will have atoms capable of bearing an amino acid side-chain at distances of 2.7, 2.8, 3.6, 3.8, 4.1, 4.6, 4.7, 5.0, 5.2, 5.3, 5.4, 5.5, 5.6, 5.7, 6.1, 6.6 or 6.7 Å from one another. Since these pharmacophore-type searches are generally performed with suitable tolerances attached, these distances could be generalized to the seven search criteria 2.75 ± 0.05, 3.7 ± 0.1, 4.1, 4.65 ± 0.05, 5.35 ± 0.35, 6.1 and 6.65 ± 0.05 Å. The tolerances on all the searches will need to be enlarged by some user-defined factor, since the fit need not be perfect and a much larger number of molecules will be found by allowing some variance in the distances.

This style of analysis may be extended to cases in which three out of the four side-chains are thought to be involved in the binding interaction at the receptor. The selection of $r = 3$ C_α atoms from $n = 4$ may be achieved in a combination of four ways (nCr). These are the atom triples, 123, 124, 134 and 234, where the numbers indicate the position of the residue about

the turn from which the C_α atom is taken. The search for mimetics based on these distances would require a search for 4 different distance arrangements for each of the 11 different turn types, a total of 44 searches in all. Even allowing for simplification owing to the similarity between turn types and their inverses, this still constitutes a relatively large number of searches.

The number may be substantially reduced by examining the similarity that exists between a pair of atom triples. By only considering three of the four C_α atom positions, the different β-turn types fall into a small number of classes. The differences inherent between the different turn types are lost as portions of the β-turn are ignored and a general model emerges for the relationship between C_α atom positions about the β-turn.

C_α atom triple analysis

Tetra-alanyl peptide segments were constructed within Sybyl [28] using the various torsion angles of the different idealized β-turn types. The arrangement of C_α atoms to be used in the analysis was generated. For each of the four different atom triples, there are six ($Pn = 3!$) different ways of arranging the three atoms. Since the similarity between a given pair of atom triples may be greatest when the atoms are considered in an alternative order, all six possible orientations (permutations) need to be considered. Thus, the datafile contains twenty-four different arrangements for three C_α atoms.

Each triplet of atoms for each turn type is taken in order and optimally superposed [29] onto every atom triple, of every turn type, according to the relationship described by the arrangement given in the datafile. For instance, when comparing the atom triple abc of turn type X and the atom triple ijk of turn type Y, the C_α atom of residue a of the first turn type (X) is superposed onto the C_α atom of residue i of the second turn type (Y), whilst simultaneously superposing the C_α atom of residue b of turn type X onto the C_α atom of residue j of turn type Y and the C_α atom of residue c of turn type X onto the C_α atom of residue k of turn type Y. This leads to a total of 24×11 (264) atom triples being compared (the number of atom arrangements multiplied by the number of different conformations of the β-turn), a total of 264^2 (69 696) superpositions, 34 716 ($264 \times 263/2$) of which are unique, since the order in which a pair of C_α atom triples are selected does not alter the result of the fit.

Table 3. Distance matrix for atom triplet superposition. Atom triplets are indexed in the form A-*ijk*, where A is the turn type and *i*, *j* and *k* are the C_α atoms that are retained

	I-123	I-124	I-134	...	VIII-432
I-123	0.000	0.128	0.309	...	0.563
I-124	0.128	0.000	0.298	...	0.488
I-134	0.309	0.298	0.000	...	0.336
...	...	...	...	...	...
VIII-432	0.563	0.488	0.336	...	0.000

The results for the superpositions were stored in the appropriate location within a distance matrix of the form shown in Table 3. It should be noted that the matrix is symmetric; the diagonal elements are zero since the particular orientation of an atom triple for a given turn type will always fit perfectly onto itself and the order in which a pair of atom triples are selected does not affect the result of their fit, although the order in which the atoms are arranged within any particular triple is significant, altering the nature of the superposition. This feature explains the need for each orientation of each atom triple.

Upon completion of the superposition process, the integrity of the distance matrix is confirmed. All elements within the matrix must have been set, with all elements along the diagonal set to zero (or very near to zero) and the difference between elements reflecting one another in the diagonal less than 0.01 Å (a user-defined acceptable level of error in the value obtained for the minimum rms deviation). The elements of the distance matrix are then clustered [30].

Results

Cluster analysis of the superposition data obtained for all the orientations of the atom triples for the eleven specified turn types yields the dendrogram shown in Figure 2. Each of the different β-turn fragments is represented by a vertical bar along the x-axis. As the diagram is ascended, the elements fuse together forming a cluster, the level of similarity between them decreasing as the graph is ascended. Thus, identical elements fuse at $y = 0$ and more dissimilar elements fuse at increasing y values until eventually the level of similarity required for two elements to be regarded as falling into the same cluster is sufficiently low that all the elements fall into a single class. At some point between $y = 0$ and the scaled fusion distance where all

Table 4. Membership of classes 1, 3 and 6 obtained through cluster analysis of C_α atom positions about 11 idealized β-turn types where each atom triplet is specified using the scheme described in the legend of Table 3

Class 1		Class 3		Class 6	
I	123	I	132	I	213
I	321	I	312	I	231
I	234	I	243	I	324
I	432	I	423	I	342
I′	123	I′	132	I′	213
I′	321	I′	312	I′	231
I′	234	I′	243	I′	324
I′	432	I′	423	I′	342
II	123	II	132	II	213
II	321	II	312	II	231
II	143	II	134	II	324
II	341	II	314	II	342
II	234	II	243	II	413
II	432	II	423	II	431
II′	123	II′	132	II′	213
II′	321	II′	312	II′	231
II′	143	II′	134	II′	324
II′	341	II′	314	II′	342
II′	234	II′	243	II′	413
II′	432	II′	423	II′	431
III	123	III	132	III	213
III	321	III	312	III	231
III	234	III	243	III	324
III	432	III	423	III	342
III′	123	III′	132	III′	213
III′	321	III′	312	III′	231
III′	234	III′	243	III′	324
III′	432	III′	423	III′	342
V	123	V	132	V	213
V	321	V	312	V	231
V	234	V	243	V	324
V	432	V	423	V	342
V′	123	V′	132	V′	213
V′	321	V′	312	V′	231
V′	234	V′	243	V′	324
V′	432	V′	423	V′	342
VIa	123	VIa	132	VIa	124
VIa	321	VIa	312	VIa	142
VIa	143	VIa	134	VIa	213
VIa	341	VIa	314	VIa	231
VIa	214	VIa	241	VIa	324
VIa	412	VIa	421	VIa	342
VIa	234	VIa	243	VIa	413
VIa	432	VIa	423	VIa	431
VIb	123	VIb	132	VIb	213
VIb	321	VIb	312	VIb	231
VIb	143	VIb	134	VIb	413
VIb	341	VIb	314	VIb	431
VIII	123	VIII	132	VIII	213
VIII	321	VIII	312	VIII	231
VIII	134	VIII	143	VIII	314
VIII	431	VIII	413	VIII	324
VIII	234	VIII	243	VIII	341
VIII	432	VIII	423	VIII	342

elements fall into a single cluster, the data fall into a number of significantly different classes [31]. This has been determined for a significance level of $p < 0.05$ and is represented by the horizontal line labelled as such in the dendrogram. This line cuts the dendrogram at seven points, indicating that the data fall into seven different clusters at this level of significance.

Further analysis of the data shows that several of the classes exhibit a strong relationship to one another. The members of the first, third and sixth classes contain the same fragments following a re-orientation of the elements. An identical pattern is observed with the fourth, fifth and seventh classes. The exception to this pattern is the second class which has no associate classes, although very careful inspection of the dendrogram reveals a three-fold symmetry within the cluster. Close visual inspection also shows all associated classes to be identical in the way their elements

have clustered. This relationship may be understood by examination of the fragments present in each class. The class membership of classes 1, 3 and 6 (coloured black in the dendrogram) are given in Table 4, the membership of classes 4, 5 and 7 (coloured cyan in the dendrogram) is shown in Table 5 and the remaining elements of class 3 (coloured red in the dendrogram) are given in Table 6.

The elements in Tables 4 and 5 have been arranged to reveal the similarities that exist across the classes. Spacing is used to delimit elements with the same basic turn type backbone conformation. Within these blocks, the elements have been arranged according to their orientational group. That is, if an alternative orientation for a particular fragment belongs to the same class, the two have been grouped together. It can be seen that for each of the classes 1, 3 and 6, exactly two of the six possible orientations for a particular fragment are present in each class. The situation is repeated for classes 4, 5 and 7.

The membership of class 2 is given in Table 6. It is comprised of two sets of type VIb turn fragments with all six possible orientations present in the same class.

These classifications provide clear patterns in the distances between the C_α atoms. The average for each of the three distances is given for each of the classes in Table 7. Classes 1, 3 and 6 form an isosceles-type arrangement, with two sides of length 3.81 Å and one of 5.54 Å. Classes 4, 5 and 7 form a near-isosceles arrangement, with one side of length 3.82 Å and one each of 5.42 Å and 5.44 Å. Furthermore, class 2 is equilateral with sides of length 3.37 Å. The simplified triangular relations may be used to formulate a general description for the distance constraint between C_α atoms within a β-turn ('The Triangle Relation', below).

The quality of the superposition may be seen by examining Figure 3. One rotation of each of the class types is shown for classes 1, 2 and 4. The C_α positions are tightly grouped in the relationships determined by the cluster analysis. The C_β atoms are spread out much more widely, but this is to be expected since the fragments were clustered on the basis of C_α atom position alone. The implication is that whilst the C_α atom analysis has provided a much simpler method for screening for potential peptidomimetic scaffolds, the placement of the side-chains in 3-D has not been accounted for. A molecule may be selected from a database that places the side-chains the correct distance apart in space, but which does not position the C_β atom of the side-chain bond vector in the right

Table 5. Membership of classes 4, 5 and 7 obtained through cluster analysis of C_α atom positions about 11 idealized β-turn types where each atom triplet is specified using the scheme described in the legend of Table 3

Class 4		Class 5		Class 7	
I	412	I	124	I	142
I	421	I	214	I	241
I	134	I	341	I	314
I	143	I	431	I	413
I′	412	I′	124	I′	142
I′	421	I′	214	I′	241
I′	134	I′	341	I′	314
I′	143	I′	431	I′	413
II	412	II	124	II	142
II	421	II	214	II	241
II′	412	II′	124	II′	142
II′	421	II′	214	II′	241
III	412	III	124	III	142
III	421	III	214	III	241
III	134	III	341	III	314
III	143	III	431	III	413
III′	412	III′	124	III′	142
III′	421	III′	214	III′	241
III′	134	III′	341	III′	314
III′	143	III′	431	III′	413
V	412	V	124	V	142
V	421	V	214	V	241
V	134	V	341	V	314
V	143	V	431	V	413
V′	412	V′	124	V′	142
V′	421	V′	214	V′	241
V′	134	V′	341	V′	314
V′	143	V′	431	V′	413
VIII	412	VIII	124	VIII	142
VIII	421	VIII	214	VIII	241

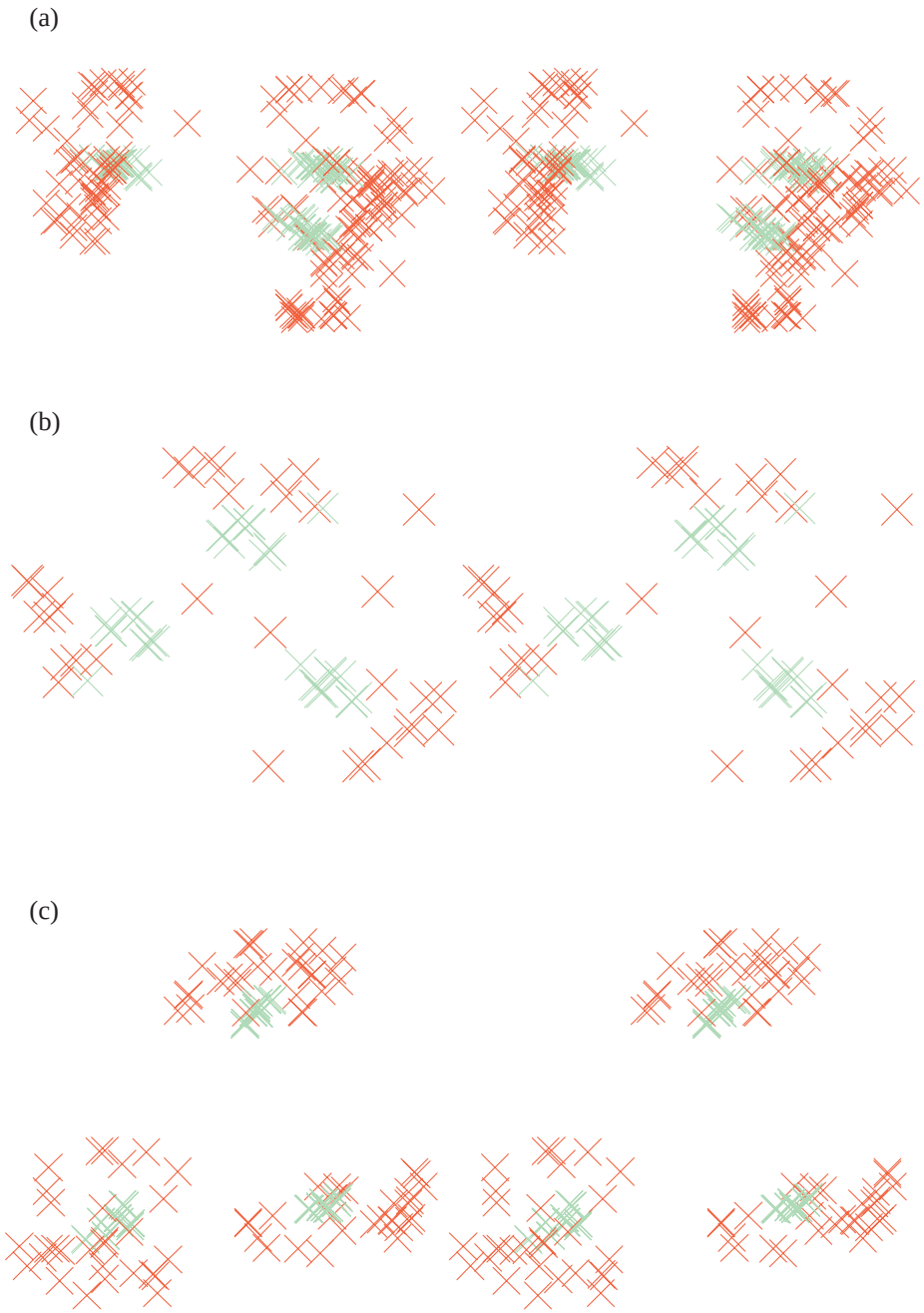

Figure 3. Stereo views of the superposition of C_α atoms (green) and C_β atoms (red) for (a) class 1, (b) class 2, and (c) class 4 (numbered left to right as described in the legend to Figure 2).

Table 6. Class membership of class 2 where each atom triplet is specified using the scheme described in the legend of Table 3

	Class 2
VIb	124
VIb	214
VIb	241
VIb	142
VIb	412
VIb	421
VIb	234
VIb	243
VIb	324
VIb	342
VIb	423
VIb	432

Table 7. The mean and standard deviations of the distances (in Å) between C_α atoms in the various classes derived through cluster analysis

Class	Distance 1	Distance 2	Distance 3
1	3.81 ± 0.37	3.81 ± 0.37	5.54 ± 0.49
2	3.37 ± 0.48	3.37 ± 0.48	3.37 ± 0.48
3	5.54 ± 0.49	3.81 ± 0.37	3.81 ± 0.37
4	5.42 ± 0.57	3.82 ± 0.01	5.44 ± 0.55
5	3.82 ± 0.01	5.42 ± 0.57	5.44 ± 0.55
6	3.81 ± 0.37	5.54 ± 0.49	3.81 ± 0.37
7	5.44 ± 0.55	5.42 ± 0.57	3.82 ± 0.01

direction for a good interaction at the receptor. This means that selected molecules must be analysed further for their ability to place the amino acid side-chains correctly in 3-D. This lack of 3-D information is incorporated by C_α–C_β bond vector analysis of the β-turn [24].

The triangle relation

The clustering of C_α atom positions in β-turn triplet fragments has revealed certain symmetries. The seven clusters form three groups, two with three alternate rotations and one independent class. These may be described in terms of the triangles shown in Figure 4, the

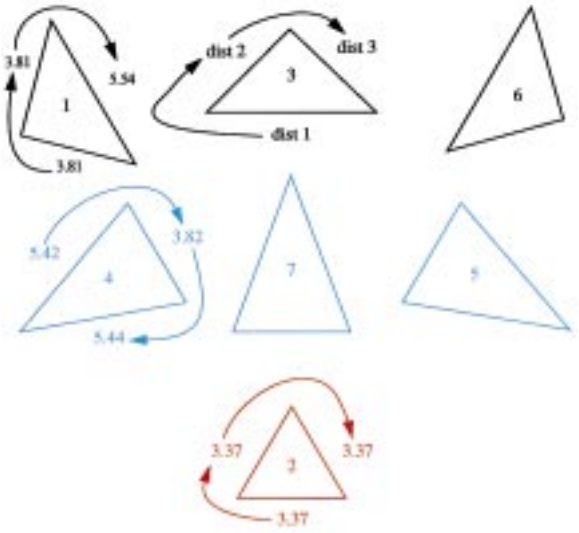

Figure 4. Triangular representation of the classes derived from clustering the C_α atoms. Three related groups are found (colour-coded according to the same scheme as Figure 2). Classes 1, 3 and 6 are isosceles triangles with two sides of length 3.81 Å and one of 5.54 Å and are related by a rotation that corresponds to the rotation of the β-turn fragments within the classes. A similar relationship is observed between the classes 4, 7 and 5 with the near-isosceles arrangement of sides of length 3.82 Å, 5.42 Å and 5.44 Å. The final class is independent and shows an equilateral arrangement of sides of length 3.37 Å. Within this class, two β-turn fragments are present in all six possible relative rotations.

lengths of the sides being the mean distances between C_α atoms determined earlier.

The distance between a pair of C_α atoms is conventionally taken to be 3.8 Å (when the intervening bond is trans), corresponding very well with the value observed in this analysis. Thus, the sides of length 3.8 Å correspond to the distance between C_α atoms arranged in sequence. The distance of approximately 5.4 Å in length (5.42 Å, 5.44 Å and 5.54 Å) corresponds to the case where a C_α atom is skipped in the sequence. Re-examination of the classification of the turn fragments in Table 4 and Table 5 shows that the vast majority obey this relationship. All those in Table 4 have three residues in sequence about the turn with the exception of the turn fragments II-134, II′-134, VIa-134, VIa-124, VIb-134 and VIII-134 (using the notation described in Table 3) in their various orientations. Within Table 5, all the fragments contain three residues that are non-sequential about the turn.

There are six possible combinations of C_α atom triplets that reconstruct the original turn: 123 and 124, 123 and 134, 123 and 234, 124 and 134, 124 and 234, and 134 and 234. Using the generalized triangular

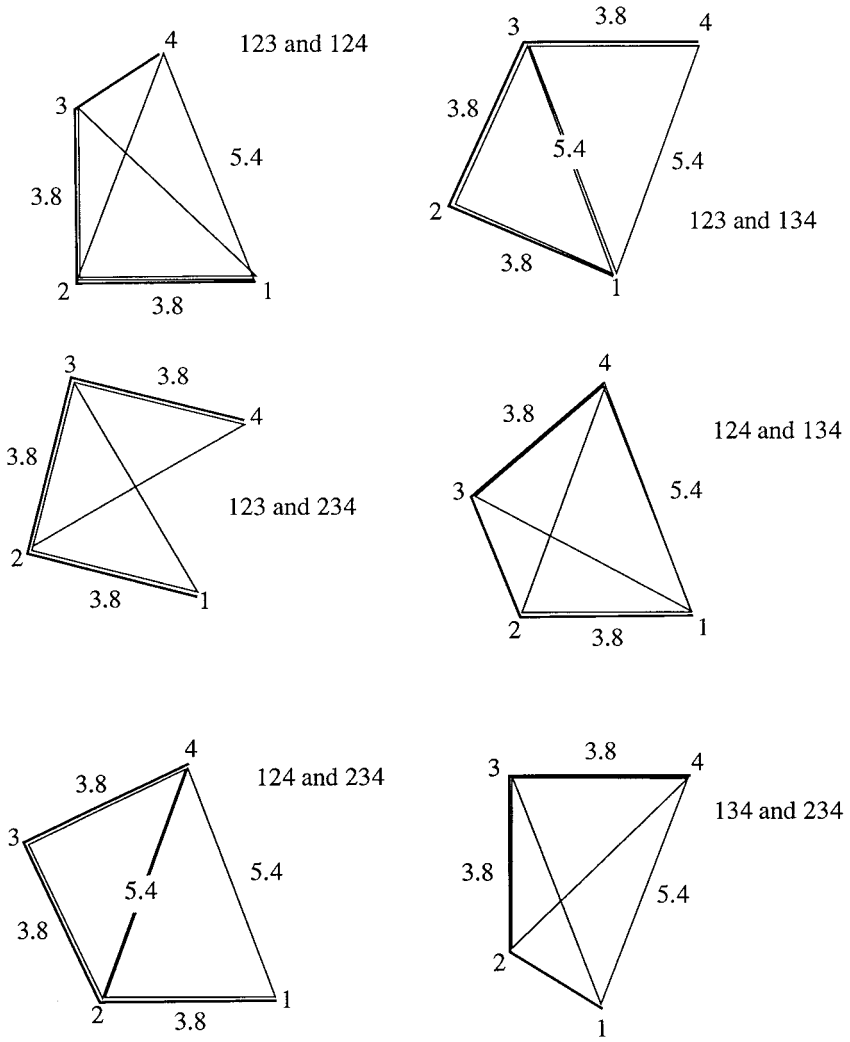

Figure 5. The relationship between the six different combinations of C_α atom triples (123 and 124, 123 and 134, 123 and 234, 124 and 134, 124 and 234, and 134 and 234) that may be used to recreate the β-turn positions of adjacent and non-adjacent residues.

relationship shown above provides the combinations shown in Figure 5. The combinations 123 and 124 and 134 and 234 are related, as are the pairs 123 and 134 and 124 and 234. All four are assembled by combining one of each of the two types of isosceles triangle. In the first pair, it is a shorter (3.8 Å) side that is in common, whereas the second pair share a longer (5.4 Å) side. The other two relationships are derived through the combination of two triangles of the same type.

The description of the turns by the triangles shown in Figure 5 is not complete. With the 123 and 124 fusion, for instance, the distance between C_α atoms 3 and 4 is not fixed. Treating the two triangles as coplanar would result in a distance of less than 3.8 Å. Making the bond distance 3.8 Å creates a 3-D arrangement of the atoms. This may be achieved in two ways, one such that atom 4 is above the plane of atoms 1, 2 and 3, the other where atom 4 is put below the plane. These two situations are depicted in Figure 6.

Close inspection of the distorted tetrahedra reveals that they are in fact the combination of all the triangular fusions of Figure 5. The undefined distances (2–3 for 124 and 134 and 1–2 for 134 and 234) and the undefined angles between the planes of the triangles (in all cases) are now fixed. The difference between the two tetrahedra of Figure 6 corresponds to the difference between a regular turn type and its inverse. Comparison with the superposition of regular and in-

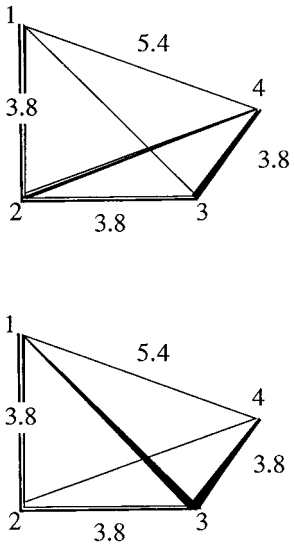

Figure 6. Three-dimensional arrangement of the triangle relations. Fixing the out-of-plane $C_\alpha(3)$–$C_\alpha(4)$ distance to the typical neighbouring C_α atom bonded distance of 3.8 Å defines all distances and angles between the C_α atoms. This may be achieved in two ways, one such that $C_\alpha(4)$ is above the plane of atoms $C_\alpha(1)$, $C_\alpha(2)$ and $C_\alpha(3)$ (left), the other such that $C_\alpha(4)$ is below the plane of the other C_α atoms (right).

Table 8. The distance constraints (in Å) for molecules to mimic C_α atom positions of the β-turn, where the rows and columns denote the residue position about the turn

	R_1	R_2	R_3	R_4
R_1	0.0	3.8	5.4	5.4± 1.6
R_2	3.8	0.0	3.8	5.4
R_3	5.4	3.8	0.0	3.8
R_4	5.4 ±1.6	5.4	3.8	0.0

Table 9. Mean and standard deviation (in Å) of the inter C_α–C_α distances observed in a sample set of 10 245 β-turns taken from the Brookhaven DataBank (PDB) where R_1–R_4 refer to the residue numbers about the turn

	R_1	R_2	R_3	R_4
R_1				
R_2	3.82 ± 0.60			
R_3	5.69 ± 1.13	3.82 ± 1.12		
R_4	5.57 ± 0.90	5.60 ± 0.87	3.84 ± 1.09	

verse turn types when superposed using their C_α atoms reveals this relationship (Figure 7).

Tolerance on the distance constraints

It is probably no coincidence that the $C_\alpha(1)$–$C_\alpha(4)$ distance is found to be, on average, 5.4 Å. This figure is exactly halfway between the standard, bonded, C_α–C_α distance (3.8 Å) and the defined upper limit for the β-turn (7.0 Å) [26, 27]. Using this information provides a convenient route to the definition of distance constraints for the selection of potential mimetic scaffolds. The simplified tetrahedral representation of C_α atom positions is used in conjunction with the upper and lower $C_\alpha(1)$–$C_\alpha(4)$ distances of 7.0 Å and 3.8 Å, respectively. This provides acceptable distance constraints between all C_α atom pairs and allows the facile selection of structures from a molecular database.

The acceptable ranges of distance constraints have been calculated and are given in Table 8. Neighbouring C_α atoms are constrained by the standard C_α–C_α trans bonded distance of 3.8 Å. The distances $C_\alpha(1)$–$C_\alpha(3)$ and $C_\alpha(2)$–$C_\alpha(4)$ are set at 5.4 Å to reflect that they still have the next-nearest neighbour relationship and the distance $C_\alpha(1)$–$C_\alpha(4)$ is constrained within the range specified (3.8–7.0 Å). The tetrahedra are related by the change in angle between the planes of the two triangles formed by points 1–2–3 and 2–3–4 (Figure 8). When the distance constraint is at its minimum, the two triangles are co-planar and the four atoms lie at the corners of a square.

Extension to irregular β-turns

The analysis performed thus far has been restricted to idealized β-turn types. The results may be extended to incorporate the irregular turns found in protein structures by examination of the variation in distances between C_α atoms. A search was performed on the 31

Table 10. The distance (in Å) between C_α atom surrogates in the benzodiazepine β-turn mimetic

	R_1	R_2	R_3	R_4
R_1	0.00	3.84	6.47	5.23
R_2	3.84	0.00	3.96	4.86
R_3	6.47	3.96	0.00	3.85
R_4	5.23	4.86	3.85	0.00

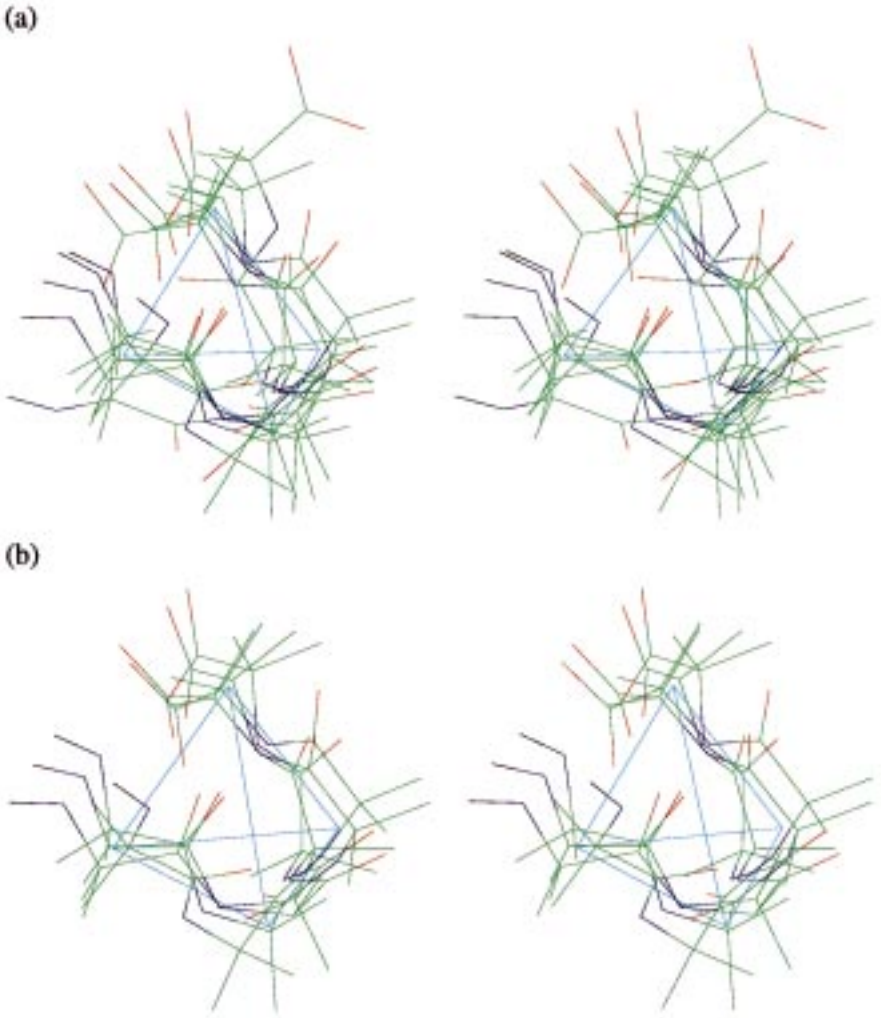

Figure 7. Tetrahedral arrangement of C_α atoms in (a) regular turn types and (b) inverse turn types.

July 1997 release of the Brookhaven Protein DataBank (PDB) [32] for proteins of 2.4 Å resolution or better which have portions of the backbone defined as being in a β-turn conformation. Proteins that were duplicated through the addition of more recent crystallographic determinations were removed by examination of the PDB codes. This yielded a total of 1132 proteins containing 10 245 β-turns. The β-turns were extracted from these structures and the distances between the different C_α atoms determined. The mean and standard deviation of these are given in Table 9. It can be seen that the mean values lie extremely close to those determined using the idealized β-turns of the previous analysis. The neighbouring C_α atoms lie 3.82, 3.83 and 3.84 Å from one another on average. The next-nearest neighbours have average distances of 5.69 and 5.60 Å, marginally greater than the predicted distance of 5.4 Å, as is the $C_\alpha(1)$–$C_\alpha(4)$ distance at 5.57 Å.

Searching for peptidomimetics

The distance constraints identified in the previous section may be used as a rapid screen for peptidomimetic molecules. A classic β-turn mimetic is the benzodiazepine scaffold. This has been shown to act as a good surrogate for a variety of different β-turn types [33] and somewhat unusually positions all four side-chains. The structure shown in Figure 9 was generated within Sybyl [28] (with R set to Me in all cases) and energy minimized using the standard Tripos molecular force field with Gasteiger–Hückel charges.

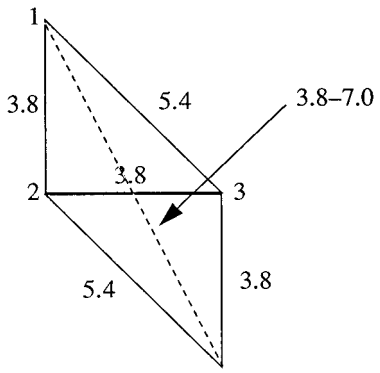

Figure 8. Derivation of the C_α–C_α distance constraints for mimetics of non-ideal β-turns from maximal and minimal $C_\alpha(1)$–$C_\alpha(4)$ distances.

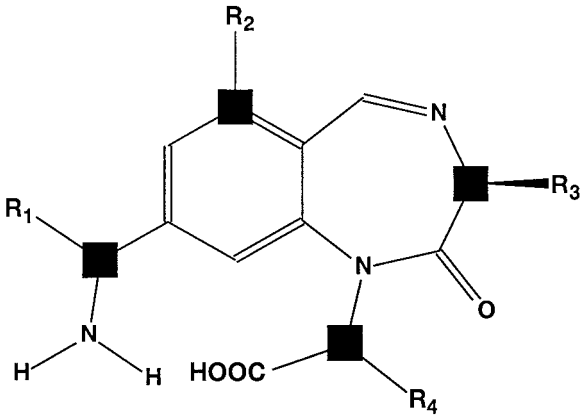

Figure 9. The benzodiazepine scaffold. Atoms marked with a square are the C_α atom surrogates when the molecule acts as a mimetic for the β-turn.

The distances between the C_α atom surrogates of the benzodiazepine scaffold were measured and are given in Table 10. By comparison with the distance matrix for the idealized general β-turn mimetic constraints (Table 8), it can be seen that the majority of distances lie very close to the ideal values. The neighbouring amide bond distance (ideally 3.8 Å) between adjacent side-chains is 3.84, 3.96 or 3.85 Å, a difference of no more than 0.16 Å. The $C_\alpha(1)$–$C_\alpha(4)$ distance, at 5.23 Å, is also very close to the ideal distance of 5.4 Å. The greatest variation is seen between next-nearest-neighbour C_α atom positions. The idealized distance from the general β-turn model derived above is 5.4 Å. The observed distances for the benzodiazepine scaffold are 6.47 and 4.86 Å for the $C_\alpha(1)$–$C_\alpha(3)$ and $C_\alpha(2)$–$C_\alpha(4)$ distances, a variation from the ideal of 1.07 and 0.56 Å, respectively.

The neighbouring C_α atom distances and the $C_\alpha(1)$–$C_\alpha(4)$ distance are also well within the range observed with non-idealized turns. For the next-nearest neighbour distances, it may be speculated that having one pair of C_α atom surrogates further apart than required and the other pair closer together, may go some way towards compensating for the variation from the idealized values and hence still provide a good fit for a large range of β-turn types.

Novel peptidomimetic templates from pharmacophore-type searches

Further potential peptidomimetic scaffolds may be obtained by performing pharmacophore-type searches on databases of known chemical structure. This has been demonstrated using the Available Chemicals Directory (ACD). A 1000 compound subset was taken and idealized conformations generated from the SMILES code [34, 35] using CONCORD [36]. These were searched for structures containing atoms with the correct distance arrangement to act as mimetics for three out of four of the side-chain positions about the β-turn.

Two searches were performed, the first for 3.8–3.8–5.4 Å atom triples, the second for 3.8–5.4–5.4 Å. Each was carried out ignoring any hydrogen atoms in the molecule and with a 0.1 Å tolerance on these values. Even at this low level of tolerance, a number of scaffolds were found with the correct atom arrangement, examples of which are shown in Figure 10. Interestingly, the steroid scaffold was found. This template has already been employed as a peptidomimetic [37], although only placing two side-chains in that case and using different points of substitution to those observed here. This analysis has shown that with a different substitution pattern, it is possible that the same scaffold may in fact place three side-chains the correct distance apart in space.

Even with a small number of compounds in the test set and using only one representative conformation with tight constraints on the acceptable distances between C_α atom surrogates, a fairly large number of novel peptidomimetic scaffolds are suggested. Many more would be found upon examination of the whole ACD and other databases, especially once the conformational profile of the molecules is taken into account. The analysis also does not take into account the synthetic viability of the molecules. Searching a database of known chemical structures ensures that the core template is synthetically viable in its own right, but

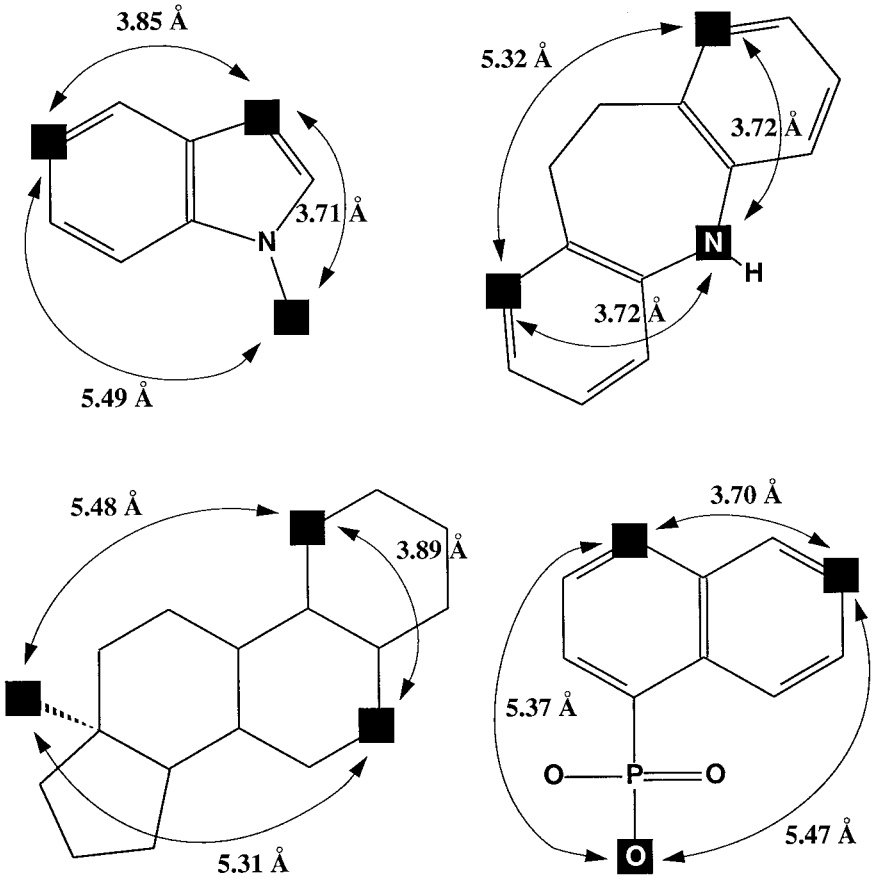

Figure 10. Example templates found within the Available Chemicals Directory that have the correct inter-atom distances for placement of three C_α atoms for mimics of β-turns. Atoms that duplicate the positions of the C_α atoms are marked with squares.

there is no guarantee that amino acid side-chains may be introduced easily at the desired positions.

Discussion and conclusions

Analysis of the similarity between three-residue fragments of idealized β-turns, measured in terms of the distance between C_α atoms, has identified clear patterns. A search of the Brookhaven Protein Databank also suggests that these results extend to non-idealized turns. Recombination of the observed triangle relations provides a complete generalized description for the β-turn. This defines all distances between C_α atoms and may be used as a search criterion for potential peptidomimetic scaffolds that will place between two and four side-chains the correct distance apart in space.

A similar classification of observed β-turn types has already been described [22]. It was empirically determined that the relative positions of bond 1, bond 2 and atom $C_\alpha(3)$ (Figure 11) remain similar across a broad range of β-turn types (both idealized and non-ideal). Likewise, the relative positions of atom $C_\alpha(3)$, bond 3 and bond 4 vary only slightly between the different turn types. The authors suggest that the only significant difference between the vari-

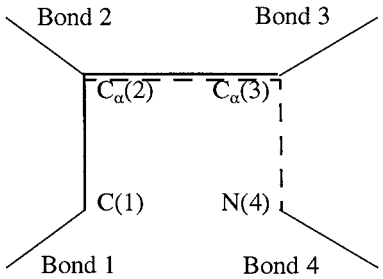

Figure 11. Topographical simplification of the β-turn. The structural units C(1)–C_α(2)–C_α(3) (solid line) and C_α(2)–C_α(3)–N(4) (dashed) are conformationally invariant. Derived from Ball et al. [22].

ous conformations of the β-turn is the dihedral angle between these two conformationally invariant units. This was directly attributed to the planar trans nature of intervening peptide bonds (the analysis does not encompass cis-proline containing type VI turns). Thus, the authors defined the variable b to describe the dihedral angle C(1)–C_α(2)-C_α(3)–N(4), which bears a close relationship to the results obtained through the previous analysis. The similarity of the two descriptions may be established through comparison of Figure 11 and Figure 8. It should be noted, however, that one was derived from the empirical observation of idealized and non-idealized β-turn types whilst the other came as a result of a statistical analysis of the C_α atom positions in only the idealized β-turn types. Also, the definition of the angle b uses the carbonyl carbon and the nitrogen atom rather than the C_α atom for residues 1 and 4, respectively.

The determination of the triangle relationship for C_α atom positions provides an extremely quick and efficient method for searching databases of chemical structure for potential peptidomimetic scaffolds. The use of three distances to define a pharmacophoric recognition element is standard practice and a number of software packages incorporate tools that have been developed on this basis (e.g., Chem-X [38]). The distance criteria may be used as a rapid screen for compounds that act as scaffolds in peptidomimetic drug design. This has been demonstrated with the example of the benzodiazepine scaffold. Provided sufficient variation in the distances is allowed, good mimetics may be extracted from chemical databases.

The generalized model for the turn also allows for the selection of molecules designed to mimic fragments of the β-turn rather than the structure as a whole. Since all distances are defined, searches may be performed for atom triplets or doublets that lie within some user-defined range of the idealized distances from one another. The molecules may act as good mimetics for three and two side-chain β-turn fragments, respectively. A test analysis on the ACD has shown that a number of possible peptidomimetic scaffolds may be retrieved, most of which would appear to be novel. Whilst the steroid nucleus has already been employed as a mimetic [37], only two positions were derivatised in that case and the positions of substitution were different to those observed here. Similarly, a phospho-naphthyl system has recently been described but with an intervening ethylene linker and only two points of attachment used [39].

Further analysis is probably required following identification of possible scaffolds since the placement of three atoms in space does not ensure that the side-chains are correctly orientated and the template may exceed the normal van der Waals envelope of the β-turn, potentially resulting in a steric clash at the receptor. The structures must also be checked for the synthetic accessibility of the amino acid substituted products. It seems likely, however, that the analysis will provide a number of novel peptidomimetic templates, primarily because it is far easier to find structures that are capable of positioning just two or three of the side-chains rather than all four about the turn, owing to the tight steric constraints that the conformation imposes.

The speed of the method is also of particular importance given the increasing size of chemical databases. To take full advantage of the broad range of chemical structures present, it is necessary to be able to search quickly for molecules that might possibly fulfil the design constraints, rapidly removing those that clearly cannot. The selected molecules may be examined more carefully later for the quality of fit to the target design constraints. The problems associated with the incorrect orientation of amino acid side-chains has been addressed through a similar analysis of the C_α–C_β bond vectors rather than C_α atoms alone [24]. This has provided design criteria that are readily applied to de novo structure generation procedures, yielding scaffolds that are both currently known (and would be found through database searches of the type described above) as well as those which have yet to be synthesized.

Acknowledgements

We wish to thank the Medical Research Council (S.L.G.) and the Wellcome Trust through the PRF scheme (P.M.D.) for personal financial support. Part of this work was carried out in the Cambridge Centre for Molecular Recognition, funded by the BBSRC.

References

1. König, W., Peptide and Protein Hormones: Structure, Regulation, Activity, VCH, Weinheim, 1993.
2. Taylor, M.D. and Amidon, G.L. (Eds) Peptide-Based Drug Design: Controlling Transport and Metabolism, ACS, Washington, DC, 1995.
3. Stephenson, S.L. and Kenny, A.J., Biochem. J., 241 (1987) 237.

4. Kahn, M., Synlett., 11 (1993) 821.
5. Giannis, A. and Kolter, T., Angew. Chem. Int. Ed. Engl., 32 (1993) 1244.
6. Gante, J., Angew. Chem. Int. Ed. Engl., 33 (1994) 1699.
7. Damewood, J.R., Rev. Comp. Chem., 9 (1996) 1.
8. Goodman, M. and Ro, S., In Wolf, M.E. (Ed.) Peptidomimetics for drug design, Vol. 1, Burger's Medicinal Chemistry and Drug Discovery, pp. 803–861.
9. Farmer, P.S. (1980) In Ariëns, E. J., Drug Design, Vol. 10, Academic Press, New York, NY, 1980, pp. 119–143.
10. Beck-Sickinger, A.G., Wieland, H.A., Wittneben, H., Willim, K.D., Rudolf, K. and Jung, G., Eur. J. Biochem., 225 (1994) 947.
11. Peeters, T.L., Macielag, M.J., Depoortere, I., Konteatis, Z., Vantrappen, G., Lessor, R. and Florance, J., Regul. Pept., 40 (1992) 226.
12. Tam, J.P., Liu, W., Zhang, J.-W., Galantino, M., Bertolero, F., Christiani, C., Vaghi, F. and de Castiglione, R., Peptides, 15 (1994) 703.
13. Wüthrich, K., NMR of Proteins and Nucleic Acids, John Wiley, New York, NY, 1991.
14. Rose, G.D., Gierasch, L.M. and Smith, J.A., Adv. Protein Chem., 37 (1985) 1.
15. Dyson, H.J., Cross, K.J., Houghton, R.A., Wilson, I.A., Wright, P.E. and Lerner, R.A., Nature, 318 (1985) 480.
16. Veber, D.F., Freidinger, R.M., Perlow, D.S., Paleveda, W.J., Holly, F.W., Strachen, R.G., Nutt, R.F., Arison, B.H., Homnick, C., Randall, W.C., Glitzer, M.S., Saperstein, R. and Hirschmann, R., Nature, 292 (1981) 55.
17. Kopple, K.D., In Rich, D.H. and Gross, E. (Eds) Peptides: Synthesis, Structure, Function, Pierce Chemical Co., Rockford, IL, 1981, pp. 295–298.
18. Momany, F.A., J. Am. Chem. Soc., 98 (1976) 2990.
19. Walter, R., Fed. Proc., 36 (1977) 1872.
20. Fox, J.W., Vavarek, R.J., Tu, A.T. and Stewart, J.M., Peptides, 1 (1982) 193.
21. Bradbury, A.F., Smythe, D.G. and Snell, C.R., Nature, 260 (1976) 165.
22. Ball, J.B., Hughes, R.A., Alewood, P.F. and Andrews, P.R., Tetrahedron, 49 (1993) 3467.
23. Ball, J.B. and Alewood, P.F., J. Mol. Recogn., 3 (1990) 55.
24. Garland, S.L. and Dean, P.M., J. Comput.-Aided Mol. Design, 13 (1999) 485.
25. Milner-White, E.J., Trends Pharmacol. Sci., 10 (1989) 70.
26. Wilmot, C.M. and Thornton, J.M., J. Mol. Biol., 203 (1988) 221.
27. Wilmot, C.M. and Thornton, J.M., Protein Eng., 3 (1990) 479.
28. SYBYL 6.4, 1997, Tripos Associates Inc., St. Louis, MO.
29. McLachlan, A.D., J. Mol. Biol., 128 (1979) 49.
30. Ward, J.H., J. Am. Stat. Assoc., 58 (1963) 236.
31. Mojena, R., Comput. J., 20 (1977) 359.
32. Abola, E.E., Bernstein, F.C., Bryant, S.H., Koetzle, T.F. and Weng, J., In Allen, F.H., Bergerhoff, G. and Sievers, R. (Eds), Crystallographic Databases: Information Content, Software Systems, Scientific Applications, Data Commission of the International Union of Crystallography, Cambridge, 1987, pp. 107–132.
33. Ripka, W.C., de Lucca, G.V., Bach, A.C., Pottorf, R.S. and Blaney, J.M., Tetrahedron, 49 (1993) 3609.
34. Weininger, D., J. Chem. Inf. Comput. Sci., 28 (1988) 31.
35. Weininger, D., Weininger, A. and Weininger, J. L., J. Chem. Inf. Comput. Sci., 29 (1989) 97.
36. Rusinko, A., Skell, J.M., Balducci, R., McGarity, C.M. and Pearlman, R.S., Concord: a program for the rapid generation of high-quality approximate 3-D molecular structures, Tripos Associates, St. Louis, MO, 1988.
37. Hirschmann, R., Sprengeler, P.A., Kawasaki, T., Leahy, J.W., Shakespeare, W.C. and Smith, A.B., Tetrahedron, 49 (1993) 3665.
38. Chemical Design Ltd., Chipping Norton, U.K.
39. Hoyt, M.J. and Bartlett, P.A., J. Am. Chem. Soc., 120 (1998) 4600.