Journal of Computer-Aided Molecular Design, **13:** 485–498, 1999.
KLUWER/ESCOM
© 1999 *Kluwer Academic Publishers. Printed in the Netherlands.*

Design criteria for molecular mimics of fragments of the β-turn. 2. C_α–C_β bond vector 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, drug design, molecular similarity, peptidomimetics

Summary

In a previous paper, we have shown the utility of cluster analysis for identifying patterns in the way the C_α atoms of fragments of the β-turn are distributed in three dimensions. This work has been extended to the C_α–C_β bond vectors of 2- and 3-side-chain fragments. Again, distinct patterns emerge and 10 and 12 classes of vector orientation have been identified for the 2- and 3-vector problem, respectively. These clusters of vector distribution provide an optimal reduced set of design criteria for the de novo generation of novel peptidomimetics for fragments of the β-turn.

Introduction

The exceptional number of known biologically active peptides makes the design of drugs based on them extremely desirable [1]. Proteolytic degradation generally renders the peptides themselves useless as therapeutic agents [2]; however, the native structure can provide a lead from which to develop a novel drug. The molecules are recognized by the endogenous receptors and must encode the necessary chemical information to elicit the desired biological response. If the pharmacophoric recognition elements of the peptide may be identified experimentally, alternative scaffolds may be designed that position the necessary chemical features correctly in three dimensions and also bind at the receptor. A range of different structures could be envisaged, from which choices could be made to select candidates with greatest probability of appropriate pharmacokinetic and pharmacodynamic characteristics. Methods such as the alanine scan have proved useful in identifying the chemical features present that give rise to activity [3–5]. The sequential substitution of the amino acids by alanine reveals which side-chains are predominant in forming the binding interaction. The technique implicitly assumes that it is the side-chains of the peptide and not the backbone that form the interaction. Whilst this is an approximation, the success of peptide mimetics designed on this basis [6–10] would lead us to believe that it is the side-chain interaction that dominates in a large number of cases. Greater difficulty remains, however, in the determination of the conformation of the peptide when bound at the receptor. Conventional techniques such as X-ray crystallography and nuclear magnetic resonance (NMR) spectroscopy are difficult to apply to the peptide, let alone the peptide-receptor complex [11]. The inherent flexibility of peptides generally presents problems. Moreover, in at least one instance when it has proved possible to examine both the peptide in solution and when bound to its putative receptor, it has been seen that the solution conformation may be quite distinct from that when complexed [12–14, 20].

Consequently, the bioactive conformation of peptides must often be inferred from biological data obtained with constrained analogs. In a number of cases, the β-turn has been proposed [15–20] as the active unit and potent analogs have been produced on that basis. The β-turn has also been considered more generally as a recognition element in light of its importance in protein structure and its predominance as a surface feature [21]. Given the difficulties inherent in determining the true bioactive conformation of a chosen

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

Table 1. Combinatorics of the β-turn side-chain vector problem

	2 side-chain	3 side-chain
Conformations	11	11
Combinations	6	4
Permutations	2	6
Stereochemistry	4	8
Total	528	2112

peptide, it would seem reasonable to start with a postulated β-turn when no other information is available. Synthesis and testing of analogs on this basis would provide molecules that could both help to identify the actual bound conformation as well as providing potential candidate drug molecules. β-Turn classification in proteins has received extensive attention and has led to the definition of a set of idealized conformations [21–23]. Although β-turns have been analysed for the purposes of drug design [24], little attention has focused on the classes of C_α–C_β bond vectors for subsets of residues about the turn. This approach has been adopted for the design of novel peptidomimetics as described in this paper. Examination of known structures, coupled with a theoretical model of peptide binding [25] and the observation that glycine and proline residues are common in β-turns [21], has led to the proposed development of mimetics for 2- and 3-side-chain fragments of the β-turn rather than the structure as a whole. This has already been used as a rationale for the development of distance constraints for the extraction of potential mimetic scaffolds from databases of known chemical structure [26]. One key problem identified with the technique is the lack of orientational information encompassed in the analysis. Molecules may be selected that place side-chains the correct distance apart in space but are not angled for good interaction at the receptor. An analysis has been performed on the C_α–C_β bond vectors present around the turn for all 2- and 3-side-chain fragments that provides an optimal reduced set of design constraints for 2- and 3-side-chain mimetics of all 11 well-defined, idealized β-turns (types I, I′, II, II′, III, III′, V, V′, VIa, VIb and VIII).

Combinatorics of the drug design problem

The design of molecules to mimic fragments of the β-turn presents difficulties. The problem is combinatoric in nature and the number of β-turn fragments is too large for the design of structures for each fragment to be practical (Table 1). In both the 2- and 3-side-chain problems, there are 11 different idealized conformations to consider for the peptide backbone. As with the C_α atom analysis [26], the type IV and VII turns are excluded because their backbone conformation is subject to much greater variation than the other turn types. The number of ways in which the two or three bond vectors may be selected must be considered. There are six combinations (nCr) when selecting two objects from four ($r = 2$, $n = 4$) and four combinations for three objects from four ($r = 3$, $n = 4$). Furthermore, the fit between any pair of turn fragments must be considered for every possible orientation (all permutations of each combination of C_α–C_β vectors). For the 2-vector problem, there are two permutations (2!) (e.g., 12 and 21, where the numbers indicate the position about the turn from which the C_α–C_β bond vector is taken, 12 being the bond vectors for side-chains i and $i + 1$) for each of the six combinations and for the 3-vector problem, there are six (3!) permutations (e.g., 123, 321, 132, 231, 213, 312) for each of the four combinations. In addition, it has proved useful to consider the alternative stereochemical possibilities at each position about the turn. Partly this has been done for completeness, but also because certain of the β-turn conformations, most notably the inverse turns (types I′, II′, III′ and V′), are only really observed when D amino acids are present in the peptide. With two different states (L and D) there are four (2^2) different arrangements (LL, LD, DL and DD) for the 2-vector problem and eight (2^3) (LLL, LLD, LDL, DLL, LDD, DLD, DDL and DDD) for the 3-vector case.

Multiplying these various factors together gives the total number of β-turn fragments that need to be considered in each case. The totals come to 528 and 2112 fragments for the 2-vector and 3-vector problems, respectively. It is not feasible to consider the design of over 2500 different mimetics for fragments of the β-turn. It will be shown that it is not necessary to design so many different structures. By calculating the similarity between β-turn fragments, it is possible to reduce the total number of design constraints in a way that optimally covers the original conformational space of the various β-turn types.

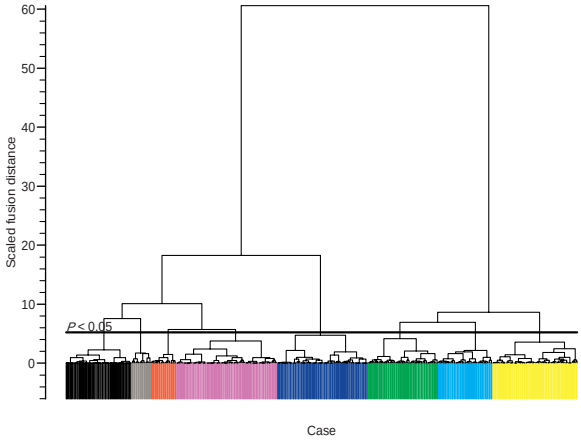

Figure 1. Dendrogram for cluster analysis of the distance matrix obtained in the 2-vector problem.

Method

Tetra-αMe-alanyl segments were generated and their backbone geometry was set to that of the selected idealized β-turn types within the program Sybyl [27]. Programs were written in FORTRAN to perform the superposition of all possible 2- and 3-vector β-turn fragments using a predefined list of the 11 turn molecule files and a datafile listing the different bond vectors to be retained in each relevant combination and permutation. The retained vectors are indexed by number 1–4 about the turn and D amino acids denoted by adding 4 to the value. Thus, 5 represents the use of the first C_α–C_β bond vector about the turn, but with D stereochemistry. These values are translated within the programs to refer to the relevant atoms in the molecule files. The bond vectors are selected in the appropriate order and superposed onto another set of bond vectors according to the defined orientation. Use of the MATFIT algorithm [28] for the superposition provided unreliable results, primarily when fitting vector doublets where the two were nearly co-planar. In order to overcome this problem, a second algorithm, QUATFIT, was found [29] that utilized the method of quaternions combined with the use of double precision representation for atom coordinates. Once derived, this data allows the construction of the distance matrix comparing all the different β-turn fragments with every other β-turn fragment. The matrix consists of 528 columns and 528 rows for the 2-vector problem (2112 square for the 3-vector case), each fragment being listed at the head of both a column and a row. Any pair of fragments may be selected and their optimal rms deviation read from the table. The matrix is symmetric since the fit between a fragment and itself must be zero and the fit between a given pair of fragments is the same, irrespective of the order in which they are selected, but the order in which the bond vectors are arranged within any given fragment is significant, explaining the need for each permutation of vectors in the analysis.

Results for the 2-vector problem

Cluster analysis of the 2-vector distance matrix using Ward's method [30] produces the dendrogram shown in Figure 1. The number of significantly different clusters was determined using Mojena's stopping rules [31] and a significance level of $p < 0.05$. The class membership of all the β-turn fragments is given in Figure 2. The elements in the array are coloured following the scheme shown in the dendrogram. These coloured tables provide a clear means by which to identify members of the various classes and to relate them to the original classification shown in the dendrogram.

For each of the charts, the 11 different idealized turn types are given up the Y-axis with each of the C_α–C_β bond vector pairs listed along the X-axis. The vector pairs are arranged such that the alternative orientation is placed next to the original. This provides greater clarity for the relationship between a pair of vector doublets where the stereochemistry of the retained vectors is the same. In Figure 2b, the vector doublet pairs are arranged such that the fragment is LD in the first column; flipping to provide the alternative permutation means that the fragment becomes DL in the second column. Thus, the chart is of LD/DL pairs of vector doublets and reciprocates the data shown in Figure 2c, which is for the DL/LD pairs. Plotting the data in this fashion also reveals patterns important in the equivalency removal process described below.

Equivalency removal

The classifications shown in Figure 2 have a certain redundancy. It was required earlier in the analysis that both permutations of each combination of vectors were retained, since it is quite possible that the alternative relative orientation of the vector pair may provide a better fit with other β-turn fragments. Now this data is unnecessary, however. A molecule that is capable of mimicking a I-12 fragment is clearly also

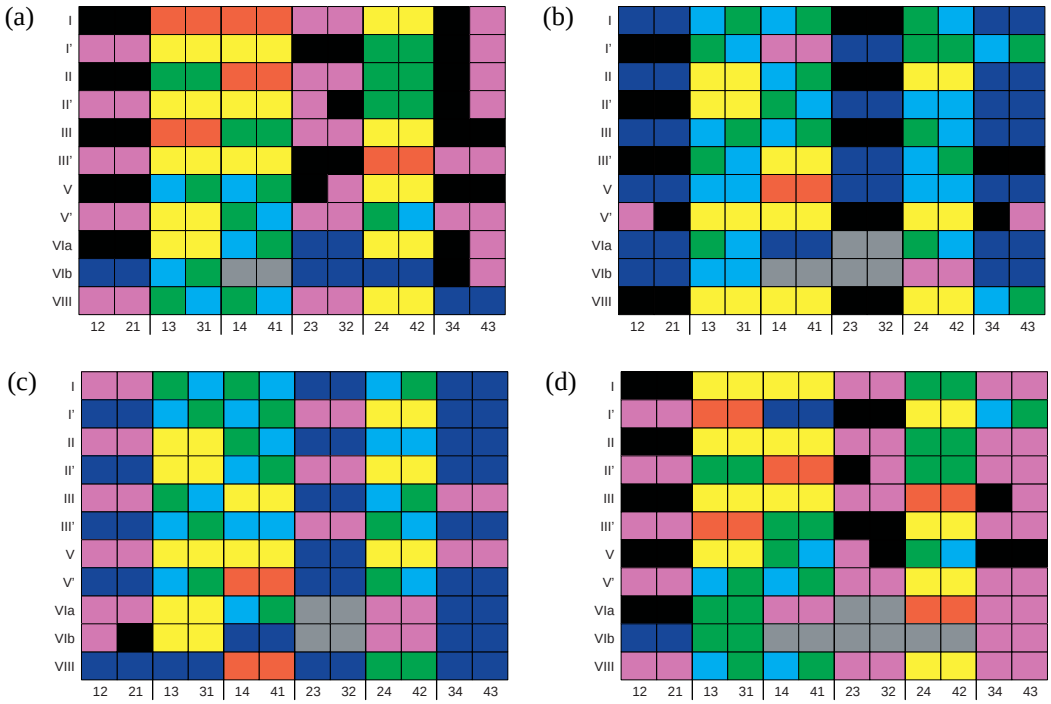

Figure 2. Class membership of (a) LL (b) LD/DL (c) DL/LD and (d) DD β-turn vector doublet fragments, colour-coded according to the class membership of the dendrogram (Figure 1).

capable of mimicking the I-21 fragment. This can be achieved in one of two ways: (1) it may merely be rotated, or (2) the amino acids may be substituted onto the scaffold the other way round. The first case is not entirely equivalent; the molecule will probably be asymmetric. The second case is, however, entirely equivalent. Thus, it is pointless to aim to mimic both permutations of each vector doublet and a method is required to remove the equivalencies in the data.

Close inspection of Figure 2 reveals that the classification of alternative permutations of vector doublets behaves in one of two ways. Commonly, both orientations of the fragment are found in the same class. This is because one of the driving forces in the clustering is the distance between the mid-points of the two vectors and this remains unchanged between the two alternative orientations. In cases where the extra component, related to the angular disposition of the two side-chains, is less significant, the two fragments are found in the same class. Thus, for those cases, the number of fragments present in each class may be halved by selectively retaining only one of the two orientations.

This is achieved by arbitrarily keeping the first vector doublet to occur in each case. For class 1, coloured black in the dendrogram, the first occurrence in Figure 2 is for the I-12 fragment (starting with the LL fragments). This element is retained and its alternative orientation, I-21, discarded. For each successive case where a vector doublet and its alternative orientation belong to the same class, the rms deviation between this first 'representative' fragment and the two alternative orientations is examined. The orientation with the lower deviation is retained and the other discarded. Thus, the next occurrence of a vector doublet and its alternative orientation in class 1 is with the II-12/II-21 pair. The rms deviation between I-12 and each of these fragments is examined and it is found that the superposition with II-12 has the lower value so it is this fragment that is retained.

For those cases where the angular disposition of the vectors has greater bearing on the way in which the vectors become clustered, the two alternative orientations of the fragment are found to lie in different classes. Examination of the classification of the fragments (Figure 2) shows that this happens in a small number of cases and that there are only two pairs of classes involved (coloured cyan-green and magenta-black).

Table 2. Distance (in Å) between the midpoints of the mean vector for each of the 10 identified significantly different vector doublets

Class	Colour	Distance
1	Black	4.34
2	Red	5.02
3	Dark green	6.05
4	Blue	4.47
5	Cyan	5.93
6	Magenta	4.40
7	Yellow	6.71
8	Grey	2.87
9	Light red	4.33
10	Light green	5.88

The classification of fragments which lie in different classes in their alternative orientations proceeds as follows: the first occurrence of such a case is detected and the orientation with the vector numbered in ascending order selected and placed into a new class and its alternative orientation discarded. All subsequent examples are found and the orientation is retained that was originally clustered with the representative fragment. For example, the I-34/I-43 pair is classified over two classes depending upon orientation. The fragment I-34 is retained since that gives the vector numbers in ascending order. The next occurrence of the pattern is observed with the I′-34/I′-43 pair. The element I′-34 is retained since this provides the fragment in the correct relative orientation; the I′-34 fragment belonged to the same initial class as the I-34 representative fragment.

Continuing the equivalency removal procedure described above provides the classifications shown in Figure 3. It can be seen that half the elements of the array are now empty; only one orientation of each fragment remains in each class. Where the alternative orientations lie in different classes in the initial analysis, new classes are derived. This happens in two cases and creates the classes coloured light green and light red in the diagram.

The presence of only two cases where interconversion is allowed indicates that the inter-midpoint distance between the two vectors is still an important criterion in the clustering process and this is borne out by the fact that the classes between which the elements are allowed to interconvert have similar mean inter-midpoint distances (Table 2).

Nature of the superpositions

Visualisation of the classes obtained through the clustering procedure shows the vector doublets to be generally well superposed. This is particularly true with the elements placed in class 1. Superposing the fragments according to the relation determined via cluster analysis provides the result shown in Figure 4. Bearing in mind that the bond vectors shown are each 1.54 Å long (the standard length of a carbon-carbon single bond), the C_α–C_β bond vectors are clearly very well superposed, with minimal variation; one can easily imagine that a molecule designed to span the mean positions of these target vectors will yield a structure that may efficiently mimic all members of the class.

One of the key differences between the various classes of vector doublets is the distance between the vector midpoints. This phenomenon is most clearly demonstrated in the case of the 3-vector problem. However, close inspection of the various 2-vector classes shows that the vectors tend to be centred around their midpoints with a range of slightly different orientations, the degree of variation depending on the compactness of the original cluster. This may be shown most clearly by displaying all the vector doublets in their optimal superposition relative to a reference fragment (Figure 5). The clusters of vectors belonging to the same class (coloured the same) are close to one another along a line that joins the midpoints of all the vectors. Around this line, the vectors spread with a range of different orientations, although the spread within a particular class is much lower than that of the group as a whole.

Hence, one of the features that characterizes each cluster is the mean distance between the vector midpoints. These have been measured for all the vector doublet classes and are given in Table 2. The two new classes generated from the equivalency-removal process were derived from the black-magenta and dark green-cyan pairs, with inter-midpoint distances of 4.43 and 4.40 Å and 5.93 and 6.05 Å, respectively. As was stated earlier, the elements that do not remain within the same class upon change in orientation, switch classes because the orientational component of the fit favours the transfer to a new class, but the new class must also share a similar inter-midpoint distance between the vector pairs. For these two cases, it may be seen that the differences are only 0.03 and 0.12 Å, respectively, and that the transfer from one class to another may be achieved very easily whilst retaining a good fit to the other members of the class.

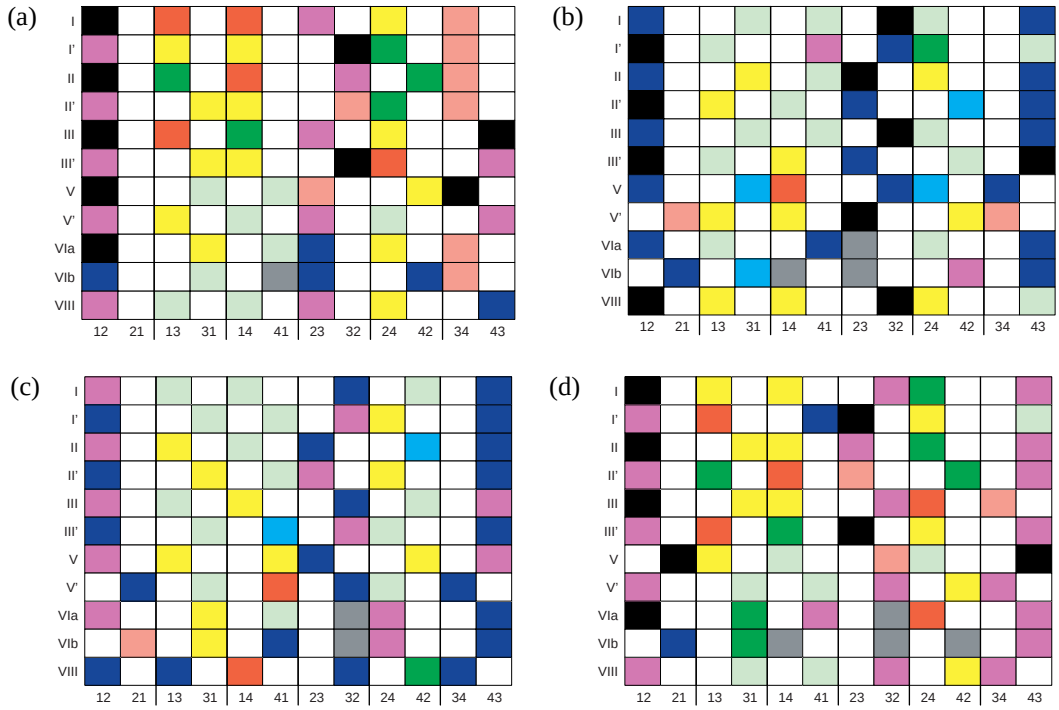

Figure 3. Classification of (a) LL, (b) LD/DL, (c) DL/LD, and (d) DD vector doublets after equivalency removal.

Figure 4. Stereo diagram of the optimal superposition of the vector doublets in class 1 after equivalency removal.

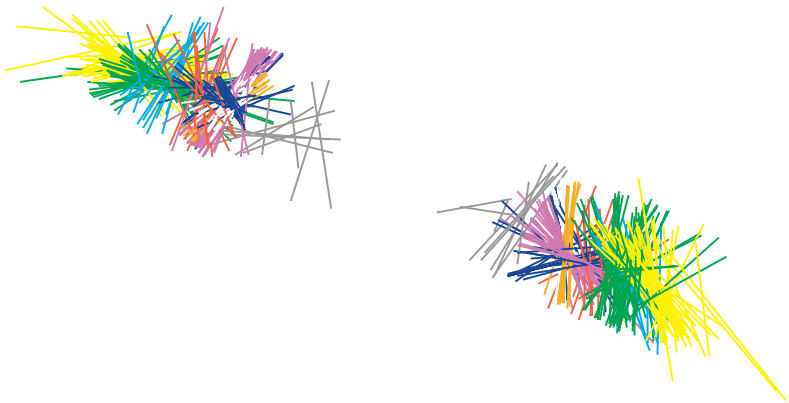

Figure 5. The superposition of all C_α–C_β bond vector doublets following equivalency removal onto a common reference structure (the I-12 fragment). The vectors are coloured according to the same scheme as used in the dendrogram and classification tables (Figure 1 and Figure 3).

Table 3. Mean and standard deviation of the rms fit for the superposition of all vector doublets within each class measured in Ångstroms

Class	Mean and standard deviation
1	0.16 ± 0.16
2	0.67 ± 0.38
3	0.92 ± 0.34
4	0.69 ± 0.39
5	1.16 ± 0.79
6	0.54 ± 0.25
7	1.33 ± 0.54
8	1.07 ± 0.60
9	0.35 ± 0.11
10	1.06 ± 0.44

The quality of the superposition obtained within each class varies according to the range of different orientations adopted by the member fragments. This may be measured in terms of the mean rms deviation for the optimal fit obtained between all member vector doublets for a particular class (Table 3). It may be seen that the exceptionally good fit obtained with the vector doublets of the first class (Figure 4) has a mean rms deviation substantially below that of the other classes.

Results for the 3-vector problem

Having determined the constraints for the design of molecules that mimic two positions about the β-turn, it is important to extend the analysis to the 3-vector problem. The Farmer hypothesis [25] and consideration of known mimetic structures suggest that this is possibly the optimal number of side-chains for the design of active species. Certainly, a stronger binding interaction is to be expected with a molecule that contains three functional elements for binding over one with only two.

Once again, the optimal superpositions of the different β-turn C_α–C_β bond vector triplets are performed and the rms deviations determined. These are then read in and a distance matrix is created. The matrix is clustered using Ward's method [30] and the number of different clusters determined using Mojena's stopping rules [31] with a significance level of $p < 0.05$. Sixteen clusters are found, the dendrogram for the cluster analysis being shown in Figure 6.

There is a complex symmetry arrangement that serves to explain the way in which the elements of the dendrogram fuse together. Close inspection of the diagram reveals that the first four classes form a subset of elements whose class membership is either symmetric or entirely non-symmetric upon change in orientation. The remaining 12 classes form three separate blocks within the dendrogram and in fact represent elements with a threefold symmetry axis; the elements interconvert class membership in an ordered fashion (this relationship is similar to that observed when clustering the C_α atom triples [26]). Thus, the dendrogram is the combination of that of a system with threefold symmetry with an additional, non-degenerate set of data that fuses at the bottom left-hand corner.

The class membership of each triplet fragment is given in Figure 7. It may be seen that the distribution of a β-turn triplet fragment and its five alternative orientations is more complex than for the 2-vector problem. The inter-midpoint distances are no longer such a simple criterion for clustering. There are now three such distances and their variation upon change in orientation gives rise to the pattern in the dendrogram described above.

Examples of the differences may be seen in Figure 7a. The I-123 fragment (colour-coded black), for instance, lies in the first cluster, in each of its six orientations. This pattern is also observed with the III-123, III′-234, VIb-234 and I-431 fragments. Elements with a three-fold symmetry appear with two orientations in each of three different classes. This is exemplified by the I′-431, II′-431, III′-431, V-431, V′-431, I′-124, II′-124, III-124, V-124, V′-124, VIa-124 and VIII-124 fragments which appear with the colour pattern violet-violet-black-black-yellow-yellow across their different orientations. There are other less obvious examples within the data, most notably the I′-123, II′-123, III′-123, III-234, VIa-234 group, colour-coded blue-blue-mauve-green-green-mauve.

The vast majority of the remaining elements lie in groups which show no such symmetry. In these cases, the six alternative orientations of the fragments lie in six different classes. There are exceptions, however; the VIb-124 fragment has four orientations in one class and the other two in two further classes. The VIb-431 fragment is also more complex in its class membership; there are two pairs of orientations present in two classes with the remaining two orientations placed in a further two classes.

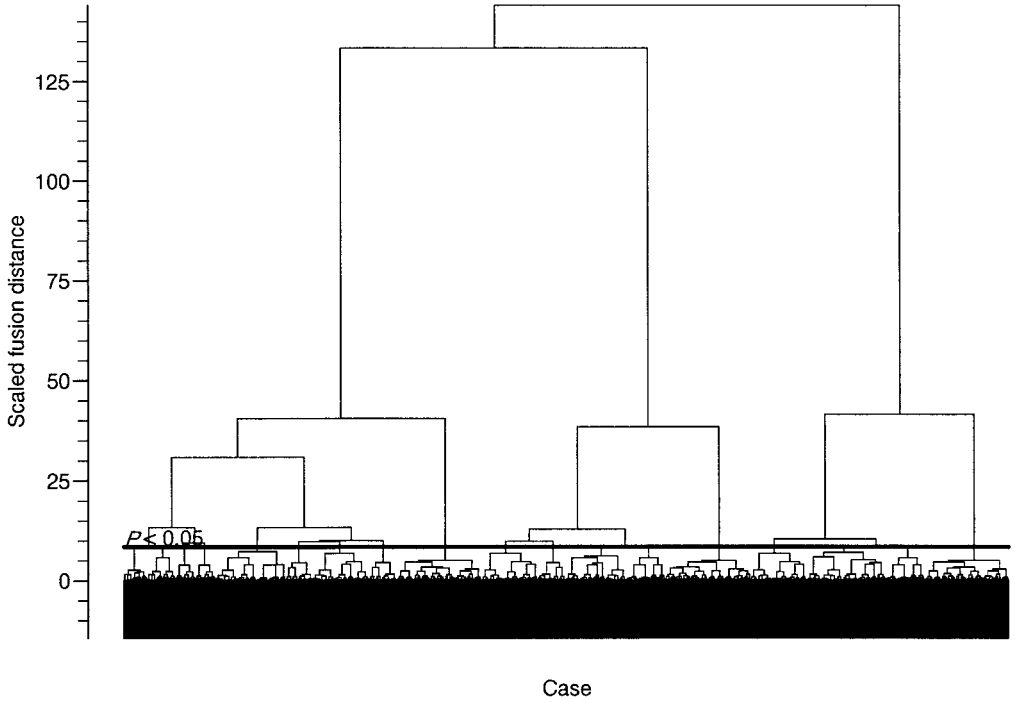

Figure 6. Dendrogram for the cluster analysis of the superpositions of the vector triplets.

Equivalency removal

The presence of such patterns in the class membership of the data provides a means for removing the equivalencies in the data. As with the 2-vector problem, there are additional data elements present that are not necessary from a drug design point-of-view. A molecule designed to mimic any particular fragment will also serve as a mimetic for the five alternative orientations by changing the way in which the core is substituted upon synthesis.

The removal of the redundancy requires the optimal selection of one of the orientations for each of the various classes observed in the data. With examples such as the I-123 fragment, where all orientations are present in the same class, the same procedure is employed as with the 2-vector problem. The first occurrence of the fragment is retained with the vectors numbered in ascending order and all future examples (III-123, etc.) have the six rms deviations between each orientation and the initial representative examined, the one with the lowest deviation being kept. The removal of equivalencies where there is only a partial symmetry to the original classification of the fragments is far more complex. In essence, the problem is to identify the arrangement that correctly maps one fragment onto another that will ultimately be in the same class. This problem does not occur with elements that show no such symmetry. If a fragment is placed in six different classes, one for each orientation, then there will only be one correct arrangement. For instance, the fragment II-123 exhibits the colour-coded pattern pale-red-dark blue-blue-light green-red-magenta. Fragment V-123 shows exactly the same pattern so the selection process is simple; the same relative orientation of the two turn fragments should be retained. The fragment II-231 is taken as representative and so the V-231 fragment is also selected preserving relative orientation. In this way it is possible to classify all of the fragments that display the exact same colour-code pattern.

There are additional cases that must be placed into the same class, however, this includes the I-124 fragment. Here the colour-code pattern is light green-dark blue-red-magenta-light red-blue. Whilst the colour-code is not the same as the II-123 pattern, it is clearly related since the six colours are the same but are specified in a different order. In fact it represents a case where the two fragments have the same classification arrangement but need to be re-oriented to provide the best fit onto one another. Selection of the correct orientation is simple; the orientation that is classified

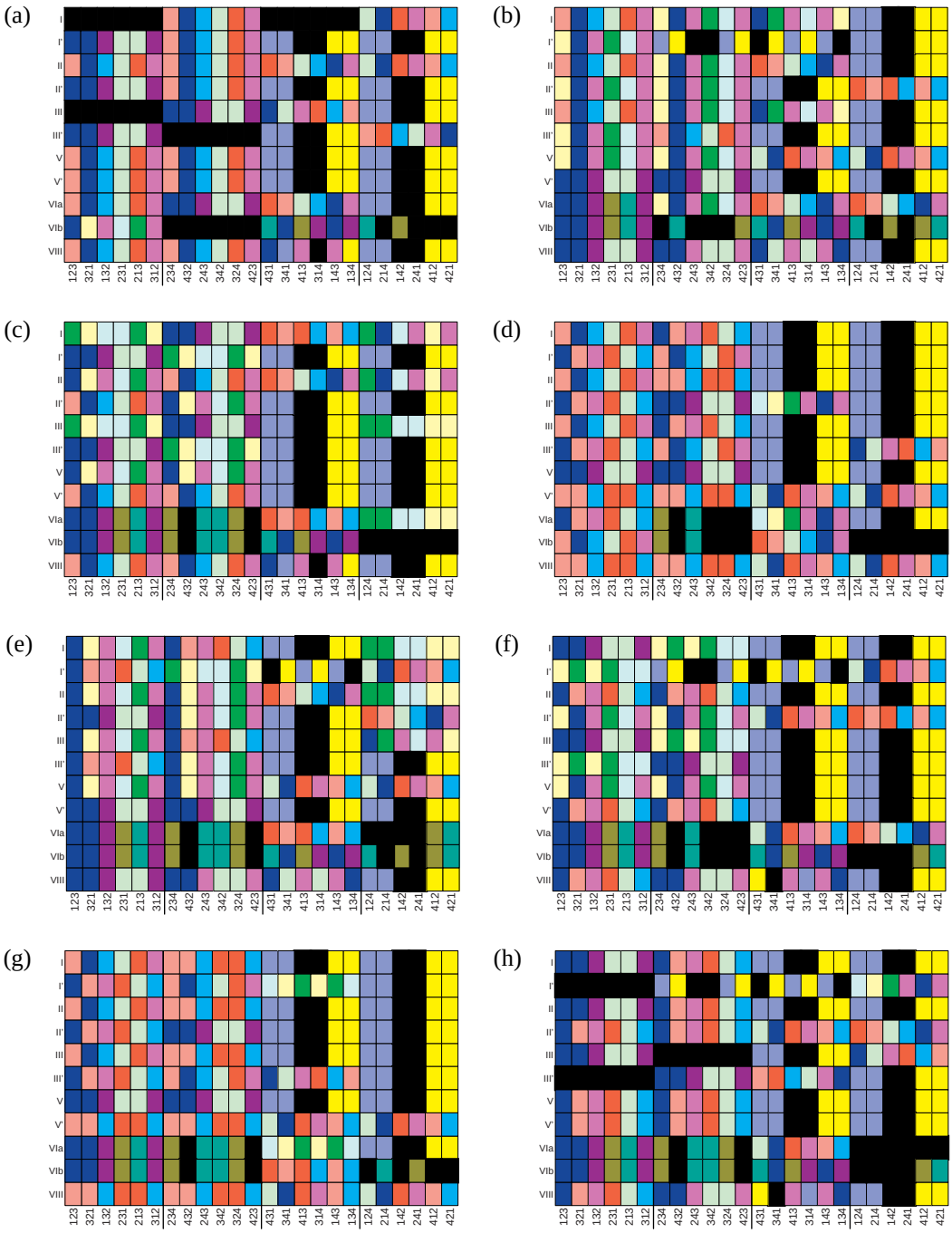

Figure 7. Classification of (a) LLL, (b) LLD/DLL/LDL/LDL/LLD/DLL, (c) LDL/LDbL/LLD/DLL/DLL/LLD, (d) DLL/LLD/DLL/LLD/LDL/LDL, (e) LDD/DDL/LDD/LLD/LDL/LDL, (f) DLD/DLD/DDL/LDD/LDD/DDL, (g) DDL/LDD/DLD/DLD/DDL/LDD and (h) DDD vector triplets.

Table 4. Checking the classification of turn fragments related to one another via a single orientation expressed as all possible permutations of the vectors

Type I fragment	Colour code	Type II fragment	Colour code
124	Light green	231	Light green
214	Dark blue	321	Dark blue
142	Red	213	Red
241	Magenta	312	Magenta
412	Light red	123	Light red
421	Blue	132	Blue

along with the representative fragment is taken. The representative for this equivalency-removed class was II-231 and was colour-coded light green. Thus the I-124 fragment is retained since it too is colour-coded light green. Such classification must not be carried out blind, however. It is possible that the six elements of colour presented in the code are the same between a pair of turn fragments but there is no correct superposition arrangement between the two. With the previous example, it is important to check that with the superposition of the I-124 and II-231 fragments, we produce the correct classification of the alternative orientations in the correct order. To do this, the superposition is expressed formally. The two turn types are superposed with vector 1 of the first being fitted to vector 2 of the second ($1 \rightarrow 2$), vector 2 to vector 3 ($2 \rightarrow 3$) and vector 4 to vector 1 ($4 \rightarrow 1$). Thus, the appearance of an alternative orientation may be calculated from the first using the mappings expressed (Table 4). For the I-214 orientation (colour-coded dark blue), the equivalent orientation for the type II fragment is II-321, also colour-coded dark blue. All such arrangements are explicitly checked to ensure the relationship is valid.

For those cases where there is partial symmetry, a combination of the two classification procedures is carried out. For example, an equivalency-removed class may be formed from those elements coloured violet-violet-black-black-yellow-yellow in the original LLL classification chart (Figure 7a). Here, however, there will always be two possible superpositions for a given pair of fragments. Taking the I′-124 fragment as representative and examining the relationship with the I′-431 fragment illustrates the problem. The simple view is that there are two orientations of the I′-431 fragment (I′-431 and I′-341) that share the same violet colour-code as the representative element. Checking the superposition according to the procedure described above shows both to be valid. In these cases, the rms deviations between the representative fragment and the alternative orientations of the query fragment are examined. The orientation with the lower rms deviation is retained and the other discarded. Thus, the I′-341 fragment is retained and colour-coded light red, along with the I′-124 fragment, and the I′-431 fragment is removed. Following these various procedures provides the equivalency-removed classification of all 2,112 3-vector fragments (Figure 8). Now, however, the data falls into just 12 classes, down from the 16 classes found after the initial cluster analysis rather than the increase (from 8 to 10) observed with the 2-vector problem.

Nature of the superpositions

Comparison of the mean rms deviations obtained in the 3-vector problem (Table 5) with the 2-vector problem (Table 3) shows a drop in the degree of fit, with the mean rms deviation between the member vector triplets when optimally superposed ranging between 0.50 and 1.71 Å (Table 5). This result should be anticipated, since it is clearly more difficult to place three vectors accurately onto one another than just two.

The nature of the superpositions of the C_α–C_β bond vectors is such that they form haystack-like arrangements (Figure 9). The inter-midpoint distances between vector triplets within the classes are fairly well conserved but the angular variation in the position of the side-chain about those points is much greater than with the 2-vector problem.

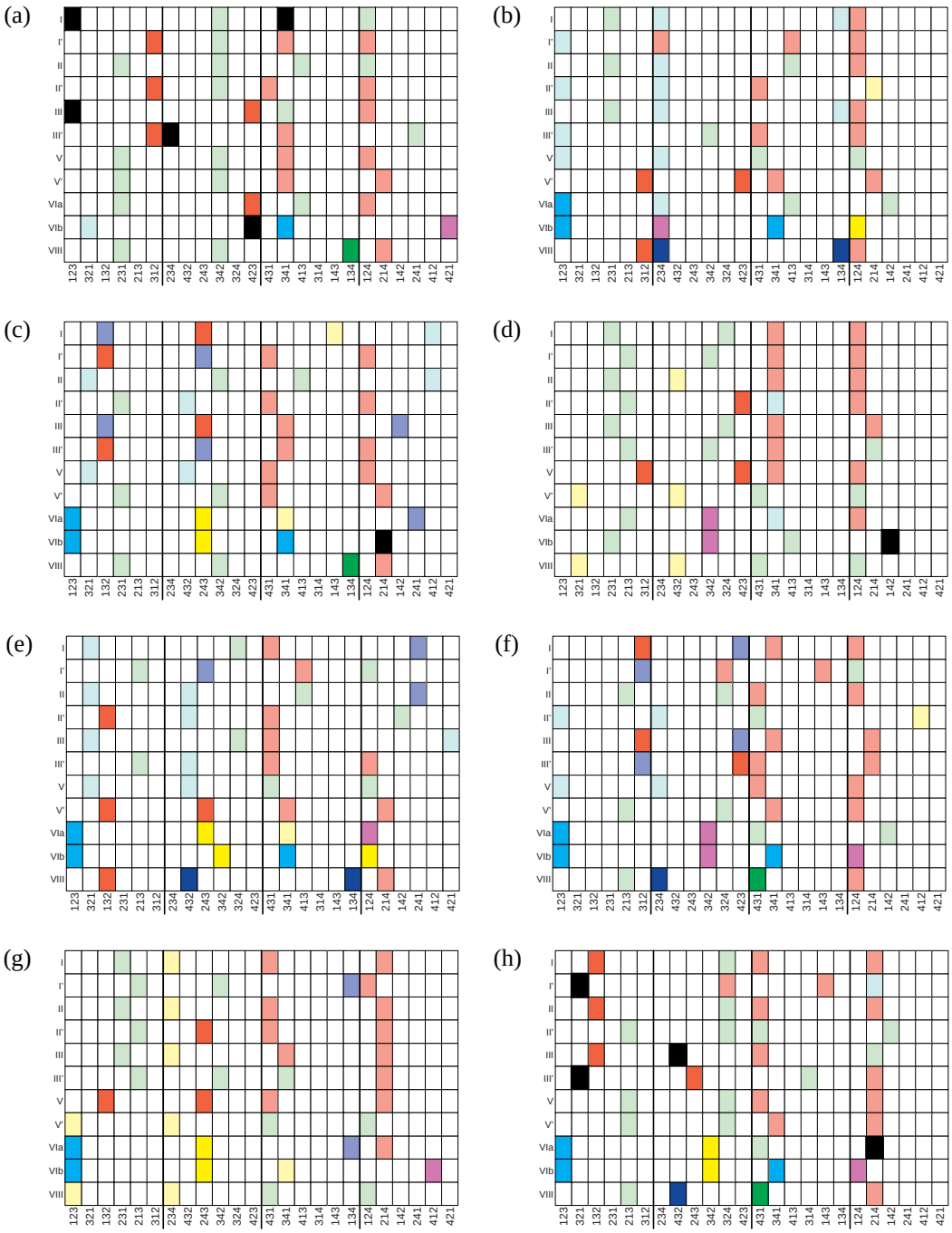

Figure 8. Classification of (a) LLL, (b) LLD/DLL/LDL/LDL/LLD/DLL, (c) LDL/LDbL/LLD/DLL/DLL/LLD, (d) DLL/LLD/DLL/LLD/LDL/LDL, (e) LDD/DDL/LDD/LLD/LDL/LDL, (f) DLD/DLD/DDL/LDD/LDD/DDL, (g) DDL/LDD/DLD/DLD/DDL/LDD and (h) DDD vector triplets after equivalency removal.

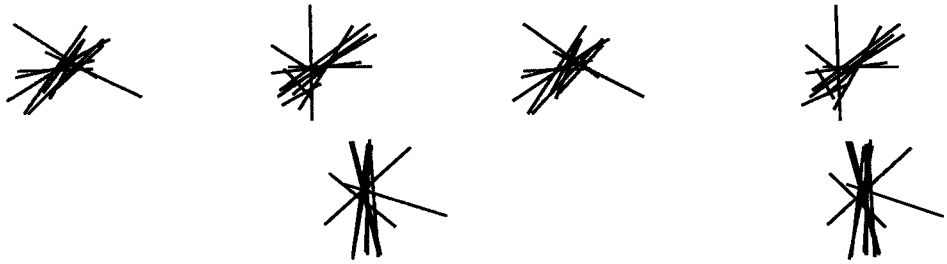

Figure 9. Example superposition of β-turn fragments according to 3-vector classification.

Table 5. Mean and standard deviation of the rms fit for the superposition of all vector triplets within each class measured in Å

Class number	Mean and standard deviation
1	0.50 ± 0.57
2	1.17 ± 0.69
3	1.25 ± 0.72
4	1.28 ± 0.45
5	1.16 ± 0.51
6	1.15 ± 0.45
7	1.38 ± 0.33
8	1.39 ± 0.64
9	1.71 ± 0.75
10	1.58 ± 0.78
11	1.44 ± 0.57
12	1.50 ± 0.86

Discussion

2-Vector problem

The degree of variation in the fits obtained between the various vector doublets may seem relatively poor since the rms deviation is measured between only four atoms. In our experience, most known peptidomimetics fit to within approximately 1.0 Å of the target constraints and this appears to be a reasonable threshold for peptidomimetic drug design. This yields a moderate fit to the target structure and a molecule with two side-chains placed to within that distance of the bioactive conformation often appears to exhibit activity. Using this criterion, it can be seen that six of the ten classes are within this arbitrary limit and two more (8 and 10) are only a small amount above it. This leaves two classes (5 and 7) that are, perhaps, not as well defined as would seem desirable.

It should be borne in mind, however, that the true bioactive conformation is rarely going to be known to an equivalent degree of accuracy. Even if a crystal structure of the native ligand bound to its putative receptor were available (possibly the ideal circumstances for drug design), the resolution of the structure is seldom this good. Also, the amino-acid side-chains often extend beyond the C_β atom and have additional flexibility. It is possible for the side-chains to move a fair amount in response to the environment of the receptor and to take maximal advantage of potential binding interactions (the so-called induced fit).

Setting the number of classes is a balancing act between the number of structures that may be synthesized and the compactness of each cluster. Changing the significance level used in the clustering process can readily provide more compact clusters with smaller rms deviations between the members.

To do so, however, requires an increase in the number of classes and consequently an increase in the number of mimetic structures that would need to be synthesized to span the conformational space available to the β-turn. At a significance level of $p < 0.05$ and following equivalency removal just ten classes are found. These structures efficiently span the available conformational space at the upper limit of what is thought necessary for recognition at the receptor. Given these considerations, it is perhaps sufficient that the mimetics are designed such that the two side-chains are projected towards the receptor in approximately the correct direction for interaction. For this design problem, the vector doublet classes determined are the ideal design constraints. They perform the desired function, providing coordinates for the position of amino-acid side-chains such that they have a good chance of interacting at the receptor if the bioactive conformation of the peptide is indeed one of the member β-turn fragments. The vector doublet classes also perform this function in such a way that the conformational space available to all peptide fragments is encompassed as efficiently as possible. The original 528 fragments of the β-turns have been characterized

by a mere ten vector doublet classes, which serves to decrease substantially the potential number of syntheses required to develop libraries for two side-chain β-turn peptidomimetic structures of this sort.

3-Vector problem

Whilst the bond vectors in the 3-vector analysis are not as well superposed as with the 2-vector analysis, the classes identified are optimal for the design of a small set of structures that may best mimic the 2112 identified β-turn triplet fragments. It is also worth noting that whilst the side-chains may not be placed as precisely as with the 2-vector problem, they may not need to be. The placement of three side-chains in approximately the correct position in space has the potential to form a stronger binding interaction than with just two side-chains. Once again, should a more precise grouping of side-chains be required, the significance level could be changed in the clustering process, although this would necessitate the synthesis of a greater number of scaffolds to span the conformational space available to triplet fragments of the β-turn.

Non-idealized β-turns

The analysis carried out thus far has been restricted to idealized β-turns. The definition of the β-turn allows for up to a 30° variation in as many as two of the torsion angles [21]. This variation has been ignored through necessity. The number of fragments that would need to be analysed rapidly becomes too large due to the combinatoric nature of the problem. Examining fragments with changes in conformation taken in 10° intervals, for instance, requires that seven different conformations ($-30° \rightarrow +30°$) be considered for each angle. With two angles allowed to vary, this gives 7^2 (49) different conformations. There are also six different ways in which two angles may be selected from four. Thus, for the 2-vector problem, the number of fragments that need to be considered is $49 \times 6 \times 528$ (155,232) and for the 3-vector problem it is $49 \times 6 \times 2112$ (620,928). Creation of the distance matrix requires $n(n-1)$ superposition measurements (2.4×10^{10} and 3.8×10^{11} for the 2-vector and 3-vector problems, respectively). This number of calculations cannot currently be completed within a realistic period of time. There would also be great difficulties storing the amount of data derived and then performing the cluster analysis. Thus, it has proved necessary to ignore the possibility for variation in backbone angle and to consider only the idealized β-turn types.

Conclusions

The analysis of C_α–C_β bond vector placement for 2- and 3-residue fragments of the β-turn has considerable advantages from a peptidomimetic drug design standpoint. It is extremely difficult to devise structures that correctly place all four of the side-chains about the turn. Given that it appears rare for all side-chains projecting from the turn to be involved in the binding process, it is highly desirable to be able to design molecules that mimic only the relevant portions of the β-turn structure. This dramatically reduces the constraints and makes possible the design of many thousands of potential peptidomimetic scaffolds. However, losing either one or two of the side-chains about the turn results in a very large number of β-turn fragments. The design and synthesis of so many scaffolds is extremely undesirable. The loss of target side-chain vectors not only serves to create a combinatoric explosion in the number of fragments to be mimicked but also tends to make many of the turn fragments highly similar to one another.

By measuring the rms deviation when β-turn fragments are optimally superposed using the C_α–C_β bond vectors of the target side-chains, it is possible to quantify the similarity between a given pair of β-turn fragments. This distance matrix may be clustered to determine the relationships in the similarity between the fragments. The procedure produces a small number of classes of vector doublets and triplets that may be used as design constraints. Whilst the mean rms deviation between fragments that cluster together at a significance level of $p < 0.05$ is quite large, the side-chains are likely to be positioned with sufficient accuracy for an interaction with the receptor to be formed. Thus, the structures should yield positive biological data if the conformation on which they are based is indeed the correct one. This makes the reduced set of bond vectors ideal for the design of libraries of structures for high-throughput screening based on peptides in the β-turn conformation. The design of combinatorial libraries with all possible amino-acids at each substitution point on the 2- and 3-side-chain structures provides a means by which to search for biological activity efficiently. The molecules are designed from constraints that optimally span the conformational space available to the β-turn within a very small number of classes. This maximizes the chances of success whilst keeping the number of structures that need to be synthesized as small as possible. Coordinate data for the β-turn types super-

posed according to the relations described above may be downloaded from the Drug Design Group website http://www.phar.cam.ac.uk/DDG/beta_turn.html.

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. Beck-Sickinger, A.G., Wieland, H.A., Wittneben, H., Willim, K.D., Rudolf, K. and Jung, G., Eur. J. Biochem., 225 (1994) 947.
4. Peeters, T.L., Macielag, M.J., Depoortere, I., Konteatis, Z., Vantrappen, G., Lessor, R. and Florance, J., Regul. Pept., 40 (1992) 226.
5. Tam, J.P., Liu, W., Zhang, J.-W., Galantino, M., Bertolero, F., Christiani, C., Vaghi, F. and de Castiglione, R., Peptides, 15 (1994) 703.
6. Kahn, M., Synlett., 11 (1993) 821.
7. Giannis, A. and Kolter, T., Angew. Chem. Int. Ed. Engl., 32 (1993) 1244.
8. Gante, J., Angew. Chem. Int. Ed. Engl., 33 (1994) 1699.
9. Damewood, J.R., Rev. Comp. Chem., 9 (1996) 1.
10. Goodman, M. and Ro, S., In Wolff, M.E. (Ed.) Burger's Medicinal Chemistry and Drug Discovery, Vol. 1, 1995, p. 803.
11. Wüthrich, K., NMR of Proteins and Nucleic Acids, John Wiley, New York, NY, 1991.
12. Fesik, S.W., Neri, P., Meadows, R., Olejniczak, E.T. and Gemmecker, G., J. Am. Chem. Soc., 114 (1992) 3165.
13. Weber, C., Weber, G., von Freyberg, B., Traber, R., Braun, W., Widmer, H. and Wüthrich, K., Biochemistry, 30 (1991) 6563.
14. Spitzfaden, C., Weber, H.P., Braun, W., Kallen, J., Wider, G., Widmer, H., Walkinshaw, M.D. and Wüthrich, K., FEBS Lett., 300 (1992) 291.
15. 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.
16. Kopple, K.D., In Rich, D.H. and Gross, E. (Eds.) Peptides: Synthesis, Structure, Function, Pierce Chemical Co., Rockford, IL, 1981, p. 295.
17. Momany, F.A., J. Am. Chem. Soc., 98 (1976) 2990.
18. Walter, R., Fed. Proc., 36 (1977) 1872.
19. Fox, J.W., Vavarek, R.J., Tu, A.T. and Stewart, J.M., Peptides, 1 (1982) 193.
20. Bradbury, A.F., Smythe, D.G. and Snell, C.R., Nature, 260 (1976) 165.
21. Rose, G.D., Gierasch, L.M. and Smith, J.A., Adv. Protein Chem., 37 (1985) 1.
22. Wilmot, C.M. and Thornton, J.M., J. Mol. Biol., 203 (1988) 221.
23. Wilmot, C.M. and Thornton, J.M., Protein Eng., 3 (1990) 479.
24. Ball, J.B., Hughes, R.A., Alewood, P.F. and Andrews, P.R., Tetrahedron, 49 (1993) 3467.
25. Farmer, P.S., In Ariëns (Ed.) Drug Design, Vol. 10, Academic Press, New York, NY, 1980, pp. 119–143.
26. Garland, S.L. and Dean, P.M., J. Comput.-Aided Mol. Design, 13 (1999) 469.
27. SYBYL 6.4, 1997, Tripos Associates Inc., St. Louis, MO.
28. McLachlan, A.D., J. Mol. Biol., 128 (1979) 49.
29. Heisterberg, D.J., Ohio Supercomputer Center, Unpublished, 1990.
30. Ward, J.H., J. Am. Stat. Assoc., 58 (1963) 236.
31. Mojena, R., Comp. J., 20 (1977) 359.