Journal of Peptide Science
J. Peptide Sci. **6**: 186–199 (2000)

Predominant Torsional Forms Adopted by Dipeptide Conformers in Solution: Parameters for Molecular Recognition

BARRY M. GRAIL and JOHN W. PAYNE*

School of Biological Sciences, University of Wales Bangor, Gwynedd, UK

Received 6 December 1999
Accepted 7 January 2000

Abstract: The present paper describes the predominant conformational forms adopted by dipeptides in aqueous solution. More than 50 dipeptides were subjected to conformational analysis using SYBYL Random Search. The resultant collections of conformers for individual dipeptides, for small groups with related side chain residues and for large groups of about 50 dipeptides were visualized graphically and analysed using a novel three-dimensional pseudo-Ramachandran plot. The distribution of conformers, weighted according to the percentage of each in the total conformer pool, was found to be restricted to nine main combinations of backbone psi (ψ) and phi (ϕ) torsion angles. The preferred ψ values were in sectors A7 ($+150°$ to $\pm 180°$), A10 ($+60°$ to $+90°$) and A4 ($-60°$ to $-90°$), and these were combined with preferred ϕ values in sectors B12 ($-150°$ to $\pm 180°$), B9 ($-60°$ to $-90°$) and B2 ($+30°$ to $+60°$). These combinations of ψ and ϕ values are distinct from those found in common secondary structures of proteins. These results show that although dipeptides can each adopt many conformations in solution, each possesses a profile of common conformers that is quantifiable. A similarly weighted distribution of dipeptide conformers according to distance between amino-terminal nitrogen and carboxyl-terminal carbon shows how the preferred combinations of backbone torsional angles result in particular *N–C* geometries for the conformers. This approach gives insight into the important conformational parameters of dipeptides that provide the basis for their molecular recognition as substrates by widely distributed peptide transporters. It offers a basis for the rational design of peptide-based bioactive compounds able to exploit these transporters for targeting and delivery. Copyright © 2000 European Peptide Society and John Wiley & Sons, Ltd.

Keywords: conformational analysis; conformers; molecular recognition; random search; rational drug design; substrate recognition parameters; structures of dipeptides; torsion angles

INTRODUCTION

Attempts to understand the molecular basis of recognition and binding of small peptides by proteins has long been the subject of intense study. Small peptides (2–5 amino acid residues) derived from the enzymic hydrolysis of proteins occur universally and are recognized and bound by a variety of enzymes, transporters and receptors [1–4]. These interactions can vary widely in their specificity. For example, peptide transporters possess broad specificity for small peptides, which is largely independent of their sequence, whereas the charged terminal amino and carboxylate groups play an important role [3–8]. The molecular recognition templates (MRTs) [9] of the substrates of such transporters must encompass structural features that are common to all components of the natural peptide pool, and this requirement will have exerted a strong selective pressure upon the evolution of these transporters [5–13]. An overall description of the peptide conformations present in solution would provide valuable insight into the MRT likely to be recognized by a particular enzyme, transporter or receptor. However, the conformational flexibility

* Correspondence to: School of Biological Sciences, University of Wales Bangor, Bangor, Gwynedd LL57 2UW, UK; e-mail: j.w.payne@bangor.ac.uk

of peptides in solution has deterred the general use of computational methods in the study of these compounds. Furthermore, when small peptides have been modelled in solution they have almost invariably been considered as portions of protein structure, in which their terminal charges have been removed, typically through *N*-acylation and *C*-amidation, and a dielectric constant between 1 and 20 has been used to simulate their location in an internal environment of a protein [14–17].

Many therapeutic agents work by modulating the activity of key enzymes or receptors. This requires not only specific recognition and binding of the drug but also its delivery to its target site in suitable concentration and, increasingly for peptide-based therapeutics, this is being achieved by exploitation of peptide transporters [2,8,12]. The design of improved drugs or ones active against new targets is a considerable challenge [2,12,14]. Use of computer modelling as part of a rational drug design process is increasing, as advances in the predictive quality of molecular modelling simulations allied with high-throughput data processing and database searching offer the medicinal chemist insight into structure–activity relationships obtained from chemical and biological screening processes [14,18]. How well a drug is recognized by its target protein depends upon the extent to which it matches the MRT of the true substrate [9,14,18]. Because for any substrate or drug its crystal structure, NMR-observed structures and calculated single minimum energy conformation may actually have little, if any, biological relevance, consideration of a representative conformer dataset for such compounds is important. However, the ability to relate the subtleties of conformational structures to biological activity has several requirements: firstly, the generation of relevant conformer datasets; secondly, assessment of structural relationships between substrate conformers; and, finally, iteration of the correlations between theoretical, structural datasets and experimental results. To achieve this for flexible molecules, such as dipeptides, grid or systematic searches are too time consuming when carried out at the level of discrimination required, and the most appropriate way of obtaining conformer datasets is to use stochastic methods, which randomly sample conformational space [19–21]. In this paper, we generate datasets of conformers for a range of dipeptides and present a convenient, novel way to analyse such results so as to give insight into a peptide's complement of bioactive conformations.

MATERIALS AND METHODS

Computer Modelling of Peptides

These studies were carried out using SYBYL molecular modelling software (Tripos Incorporated, St Louis, MO) installed on Silicon Graphics Indigo II or Octane workstations. The Indigo II used a 5k platform with a 200 MHz processor with Irix 5.2 operating system, whereas the Octane used Irix 6.4 with a 10k platform and a 175 MHz processor.

Construction of dipeptide structures. Dipeptides were constructed from pre-formed L-amino acid fragments using the sketch facility of SYBYL 6.2 or 6.4. Dummy atoms on free terminal carboxyl and amino groups were removed and fragments joined with a peptide bond of SYBYL bond type am. All dipeptides were modelled as zwitterions with protonated *N*-terminal amino groups and ionized *C*-terminal carboxyl groups; residues such as K, D, E had their charged side chains treated analogously. SYBYL default atom types were used for force field calculations, with protonated amino groups defined as N4 and carboxylate oxygens as Oco2. Completed structures with a *trans* peptide bond were subjected to an initial minimization to provide reasonable starting structures for conformational analysis, and were saved in mol2 format in a molecular database.

Conformational analysis using Random Search. Random Search is a stochastic type conformational analysis tool within the SYBYL molecular modelling package. It combines a random torsional perturbation with energy calculation, minimization and conformer comparison in an iterative process to output sets of unique conformations. Implementations in SYBYL versions 6.2 and 6.4 were used. Minimized starting structures were submitted to Random Search with all rotatable backbone and side chain bonds included in the searches. An absolute energy cut-off of 70 kcal mol^{-1} (SYBYL 6.2) or a relative cut-off of 7.0 kcal mol^{-1} (SYBYL 6.4) were used. To ensure only unique conformers with correct chirality were collected, an r.m.s. threshold of 0.2 Å was set with chirality checking.

The energy of each conformer was calculated using Tripos force field with electrostatics included in the calculation. Charge on the molecules was calculated as implemented in SYBYL, based upon procedures proposed by Pullman and by Berthod, which combines two quantum methods to calculate the sum of atomic charges on the σ and π components. Solvation was implied by using a first level

approximation of a distance-based dielectric function with a dielectric constant value of 80. Explicit solvation, although being a more realistic simulation, proved to be too demanding of computational time in this context and did not give significantly different results when tested with AA [Marshall NJ, Grail BM, Payne JW. Unpublished results]. Each new conformation generated was subjected to an energy minimization within Random Search, using the Powell method with gradient termination. Minimization was set for 100 iterations with a gradient termination limit of 0.05. Unique conformers generated by Random Search were output in the form of a database of molecules in mol2 file format. These molecules were automatically read into SYBYL molecular spreadsheet (MSS), where their energy and count were displayed. Calculations on a conformer set were carried out within this spreadsheet using the autofill functions or user-defined scripts written in SYBYL programming language (SPL). Conformers were sorted according to increasing energy, and a Boltzmann distribution calculated using the minimum energy value. The Boltzmann distribution ($N_i/N_o = e^{-\Delta E/RT}$) relates the abundance of each conformer to that of the minimum energy conformer dependant upon the difference in energy between the two. Summation of the Boltzmann values allowed the contribution of each conformer to be expressed as a percentage of the total.

The backbone torsion angles and the distance (*N*–*C*) between the *N*-terminal amino nitrogen and the *C*-terminal carboxylate carbon were measured for all conformers and entered in the molecular spreadsheet.

N–C distance measurements of conformers. The overall length of a peptide, as determined by its *N*–*C* distance, is a structural feature that is often considered when trying to relate structure to activity. However, calculation of the mean *N*–*C* distance does not take into account the weighted contribution of conformers within a population. Using the Boltzmann distribution to calculate a percentage contribution for each conformer allows weighting of the mean, which gives a better indication of the distribution of conformers based upon length. A distance range of 2.5–10 Å was divided up into 0.05 Å bins and conformers present in each bin were evaluated. Summation of the percentage contributions of individual conformers gave the total percentage contribution for each distance bin. The percentage contribution for each *N*–*C* distance bin was aggregated for all conformers for 50 dipeptides to give an accumulated percentage contribution. Plotting the accumulated percentage contribution against the *N*–*C* distance bins allows visualization of the different length groups within a mixed population of dipeptides. Similar calculations were performed to determine how *N*–*C* distance varied between certain sets of conformers that had specific combinations of ψ and ϕ torsion angles.

Three-dimensional pseudo-Ramachandran (3DPR) plots. A conventional Ramachandran plot is used to represent protein secondary structure graphically. It plots ψ and ϕ torsion angles of adjacent residues in two dimensions to produce a scatter graph of torsion value occurrence. Within the four quadrants, areas characteristic of regular secondary structure features, such as a right- or left-handed α-helix, β-sheet, etc., can be identified. Areas frequented by particular amino acid residues, such as glycine and proline, as well as conformationally unavailable areas can be defined. A Ramachandran plot has no means of weighting for extent of occurrence of conformational forms or of displaying data accumulated from a number of different substrate compounds. A potential energy surface allows torsion angles to be plotted against the energy of a conformation to display this relationship graphically as a contour map. Although there is an element of weighting in terms of energy it is difficult to display data from several substrates in a single, meaningful plot. Dipeptides have a backbone consisting of ψ, ω and ϕ torsion angles. Since the peptide bond is essentially a planar structure, the orientation of the ψ and ϕ bonds about this describe the spatial orientation of the *N*-terminal amino nitrogen and the *C*-terminal carboxylate carbon, in effect defining the peptide unit. Combining the torsional angle components of such a peptide unit, analogous to a Ramachandran plot, and the weighted contour mesh of the potential energy surface effectively results in a 3DPR. This plot gives a weighted distribution of conformers that is easily interpreted and can be extended to the accumulated data from several compounds.

To provide an improved visual presentation of conformer distribution and to identify prevalent features, the conformational space around ψ and ϕ was divided up into 36 10° sectors. ω was classified as either *cis* (0°($\pm$5°)) or *trans* ($\pm$180°($\pm$5°)), which were treated separately. Conformers existing in any combination of ψ–ϕ sectors were identified and their individual percentage contributions accumulated to give an overall value for each ψ–ϕ

combination. The 10° ψ–ϕ sectors were labelled across their complete range of $-180°$ to $+180°$, and this was used as a grid to produce a three-dimensional plot of ψ versus ϕ versus accumulated percentage conformer contribution for individual peptides. Data were plotted using the three-dimensional mesh plot of SigmaPlot 4.0 for Windows (SPSS Incorporated, Chicago, IL); for simplicity, axes were labelled at 30° intervals, using A1–A12 for ψ and B1–B12 for ϕ angles. Percentage contribution data were calculated in SYBYL MSS for each ψ–ϕ combination and output to a file in XYZ triplet format ready for importation into a SigmaPlot worksheet. Data for individual peptides were added to the worksheet as separate columns and summations made using the transform option. 3DPR were plotted directly from these data or after interpolation using a 36×36 interval with a weighting of 3. For groups of dipeptides, the percentage contribution data for every peptide were aggregated and the overall percentage contribution summed. These data, plotted as a 3DPR, provide a visual representation of the total conformer distribution for that set of peptides. As an initial approximation to the complete substrate pool of 400 dipeptides, conformers for 50 different ones were combined and plotted on a single 3DPR.

RESULTS

Output from Random Search

Conformational analysis of over 50 (L,L) dipeptides was carried out using Random Search to establish a dataset. These were selected as being representative of the wide range of side chain chemistries found amongst the 400 possible dipeptides present in the substrate pool derived from protein hydrolysis [22], and also because they were available for experimentation. The resultant conformers from each search were sorted by increasing energy and the percentage contribution of each was calculated using a Boltzmann distribution. These percentage values are not 'absolute', but sufficient iterations were performed for them to provide a valid measure of the distribution of conformers for any individual peptide, and general conclusions about predominant torsional types were strengthened by combining outputs for collections of dipeptides. Details of the minimum energy conformer found by each search are summarized in Table 1. Values for the minimum energy conformers were generally around 2–6 kcal mol^{-1}, although peptides with aromatic residues covered a wide range, e.g. from FA which was low, to LW and WL, which had values >15 kcal mol^{-1} (Table 1). Electrostatic and van der Waals energy terms are important stabilizing factors in the energy calculations (Table 2). The interaction of aromatic and aliphatic side chains, e.g. in AF, FA and LM, with the planar peptide bond gave substantial stabilization, which is reflected in the low van der Waals energy contribution and the presence of stable, low-energy conformers (Table 2). The minimum energy conformer of KK, although having a negative van der Waals energy term, has positive energy contributions from bond stretching, angle bending and 1–4 van der Waals that puts its overall energy in the 2–6 kcal mol^{-1} range. DF has a significant electrostatic stabilization but this is counteracted by positive bond stretch, angle bend and 1–4 van der Waals contributions, giving an energy value comparable with that for AA (Table 2). Inspection of minimum energy conformations indicates the predominant conformer in solution, but these generally account for only about 10% of the total conformers. Only peptides with partially constrained structures, such as *C*-terminal proline, or ones with particular stabilization due to charge or hydrophobic interactions, have minima that may comprise up to about 20–30% of the total (Table 1). The number of unique conformers found in a particular search depends mainly upon the conformational chi-space available to the side chain residues of the peptide. AA and GG produced around 20 conformers, whereas peptides containing residues with extended chi-space, such as KK, produced several hundred (Table 1).

Backbone torsion angles, psi (ψ), omega (ω) and phi (ϕ) for each minimum energy conformer found are given in Table 1. The minimum energy conformers have ψ values of around $+160°$, $-60°$ or $+60°$ with no obvious correlation with the *N*-terminal residue. Peptides with *N*-terminal glycyl had ψ angles that were just outside these values. The ω bond was invariably *trans*, except for dipeptides with a *C*-terminal proline. ϕ values fall into three groups with values around $+60°$, $-60°$ and $-160°$. Again, there is no obvious correlation with the *C*-terminal amino acid residue.

The overall length (*N*–*C*) of the minimum energy conformers as measured from amino terminal nitrogen to carboxylate terminal carbon varies from 4.38 Å in AP, with a *cis* ω bond, to more than 6 Å for fully extended dipeptides.

Table 1 Properties of Minimum Energy Conformers for 50 Dipeptides

Peptide	Energy (kcal mol^{-1})	Number of conformers	ψ (°)	ω (°)	ϕ (°)	*N–C* distance (Å)	% of total conformers
AA	2.40	18	165.0	178.4	−65.1	5.30	23
AD	2.72	73	165.3	178.0	53.5	5.08	9
AE	2.13	196	163.6	−177.4	−166.1	6.13	9
AF	0.89	80	164.0	−176.6	−149.8	6.09	23
AG	2.16	35	164.9	178.5	−65.5	5.32	21
AI	3.19	184	165.0	−178.7	−161.3	6.32	8
AK	3.39	704	165.9	178.3	56.6	5.10	3
AL	2.72	190	165.7	178.3	56.0	5.10	13
AP	9.43	20	147.4	−1.2	−77.7	4.38	30
AQ	1.37	416	163.2	−178.0	−166.6	6.13	11
AT	2.37	86	165.2	177.8	49.7	5.05	10
AT[a]	2.37	86	165.2	−178.9	−159.5	6.12	10
AV	2.44	64	164.6	−179.1	−143.8	6.06	10
AY	1.62	76	164.0	−176.6	−149.7	6.09	23
DA	2.51	79	164.4	178.7	−65.2	5.30	10
DE	2.34	480	163.9	−177.8	−166.1	6.13	4
DF	0.83	221	63.7	−177.8	−152.9	5.38	9
DK	3.07	847	164.3	176.5	53.6	5.08	4
DL	2.46	458	164.4	177.6	54.5	5.08	10
DQ	1.49	696	163.5	−178.4	−166.6	6.13	7
EA	2.77	200	−66.4	−179.6	−68.4	4.66	6
EE	2.46	702	145.9	−176.9	−166.3	6.05	5
FA	0.86	81	155.5	177.0	−153.2	6.09	23
GA	2.89	25	−179.4	179.3	−65.2	5.24	13
GF	1.48	76	168.5	−176.0	−150.6	6.10	16
GG	2.61	26	−179.3	179.3	−65.1	5.25	12
GG[a]	2.61	26	179.3	−179.3	65.1	5.25	12
GL	3.42	187	−76.2	178.2	55.7	5.31	7
GP	9.03	15	171.7	−0.3	−78.6	4.64	21
GP[a]	9.06	15	171.2	−179.7	−75.7	5.41	20
GY	2.19	72	168.4	−176.0	−150.6	6.10	16
KA	3.63	714	−64.4	−177.5	−160.5	5.15	4
KD	3.38	836	166.3	173.6	54.2	5.13	5
KK	3.86	922	−58.5	−176.4	−159.5	5.11	25
LA	3.24	234	−68.1	−178.7	−159.9	5.14	12
LG	3.03	261	−65.6	−178.1	68.5	5.32	13
LL	3.48	718	−65.5	178.6	−92.3	4.67	10
LM	1.25	858	−59.9	−176.0	−160.0	5.13	10
LW	15.92	506	−69.3	−178.6	−87.8	4.73	34
ML	1.22	817	163.4	176.0	52.8	5.07	8
PA	11.00	69	177.5	177.7	−64.5	5.23	13
PG	10.78	78	177.4	177.9	−65.0	5.25	12
QA	2.02	418	−66.2	−179.9	−65.5	4.65	7
QD	2.27	672	170.6	179.7	−146.7	6.06	4
QQ	1.06	901	170.5	−179.7	−164.3	6.12	17
SA	2.36	81	173.1	178.2	−65.8	5.27	11
SS	2.73	239	173.2	178.5	−63.4	5.25	4
TS	2.56	247	171.7	177.7	−67.9	5.28	6
VA	3.17	64	148.4	178.9	−60.9	5.31	14
VV	3.12	160	−59.0	−177.7	−158.8	5.05	8
WL	15.44	466	155.8	178.5	−150.7	6.07	19
YA	1.55	81	155.4	176.9	−153.2	6.09	22
YG	1.56	91	152.0	177.8	−79.1	5.51	23

[a] Peptide exists with two distinct conformers of equivalent minimum energy.

Table 2 Energy Contributions to Minimum Energy Conformers of Dipeptides

	Dipeptides					
	AA	AF	KK	DA	FA	LM
Bond stretching energy	0.285	0.375	0.948	0.309	0.384	0.734
Angle bending energy	0.591	0.705	2.545	0.669	0.657	1.816
Torsional energy	2.180	2.568	2.494	3.443	2.829	2.377
Out of plane bending energy	0.006	0.052	0.029	0.005	0.022	0.032
1–4 van der Waals energy	1.345	2.574	3.697	1.305	2.669	1.562
van der Waals energy	−1.855	−5.171	−5.897	−2.885	−5.519	−5.089
1–4 Electrostatic energy	0.194	0.155	0.216	0.101	0.180	0.151
Electrostatic energy	−0.346	−0.368	−0.174	−0.433	−0.361	−0.331
Total energy (kcal mol^{-1})	2.402	0.891	3.857	2.514	0.861	1.252

Three-dimensional Pseudo Ramachandran Plots of Individual Peptides

To take account of all the conformations produced by a search and their relative abundance, conformer distribution is conveniently visualized using a 3DPR. For an individual peptide, the 3DPR is related to an inverted potential energy surface and shows where in ψ–ϕ conformational space the conformers are located. To facilitate description of the weighted conformer distribution, the ψ and ϕ torsions were divided into 12 30° sectors. The ψ sectors were labelled A1–A12, with A1–A6 covering the range from $-180°$ to 0° and A7–A12 from $+180°$ to 0°; ϕ sectors were labelled B1–B12, with B1–B6 covering the range from 0° to $+180°$ and B7–B12 from 0° to $-180°$.

Different dipeptides, although broadly showing the same main conformational types, vary in the distribution of conformers between these types. AA has a tight distribution of conformers, with three main ψ categories: A7, A10 and A4, and three main ϕ categories: B9, B12 and B2. (Figure 1(a)). The main contribution of conformers being distributed within A7, with those from A10 and A4 combinations being slightly lower. With LL, which has aliphatic side chains similar to AA, a very different weighted distribution of conformers occurs (Figure 1(b)). The conformer contributions are more dispersed, the main type being A4, with very little from A10 and the A7 conformer contribution being broadly spread. DF and AF show similar distributions (Figure 1(c) and (d)), arising from a stabilizing interaction between the aromatic phenylalanine side chain and the peptide bond. DF has contributions from A7, A4 and A10 but the ϕ torsion contribution is mainly concentrated around B12. AF has a predominance of A7B12, the shift to A7 presumably arising from the influence of the *N*-terminal alanine (*cf*. Figure 1(a)).

Three-dimensional Pseudo Ramachandran Plots of Groups of Related Peptides

For even quite similar dipeptides, the differences in their distribution of conformers in ψ and ϕ space makes analysis of their combined conformers far from trivial. However, by considering such peptides in groups, e.g. ones with similar side chains, and by aggregating percentages of related conformers with similar torsional angles before visualizing in a 3DPR, one can reveal how the main conformational types are distributed and highlight structural features that predominate.

Figure 2(a) shows a 3DPR plot for the combined data from 12 dipeptides with an *N*-terminal alanyl residue. The main ψ and ϕ combinations are clear: A7, A4 and A10 for ψ and B9, B12 and B2 for ϕ. The distribution is similar to that obtained for AA alone (Figure 1(a)) but the peaks have broadened slightly as more dipeptides have been included. *N*-terminal glycyl has a marked effect upon the weighted distribution of conformers (Figure 2(b)). The main conformer-containing ψ sectors are A1/A7, A10/A11 and A4 but these are offset by about 10° from those seen with AA and other dipeptides. This shift in distribution is specific to glycyl peptides and arises from its unique flexibility and lack of a β-carbon. The distribution of conformers throughout ϕ-space is largely unaffected, with B9, B12 and B2 predominating. Dipeptides with *C*-terminal proline residues contain a high proportion of conformers with *cis* ω bonds, e.g. AP was computed to contain 46% *cis* forms; uniquely, they are also constrained to adopt

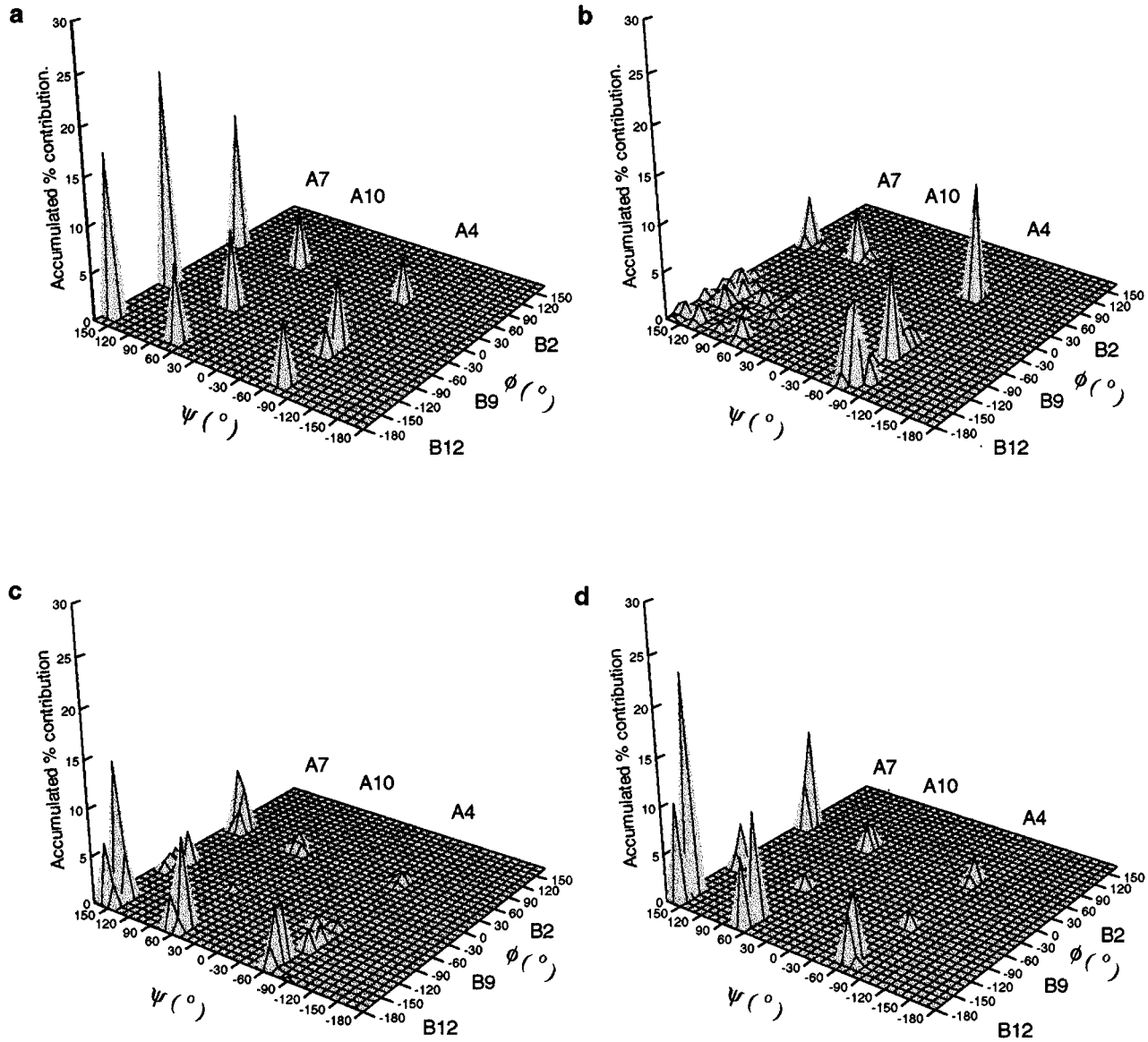

Figure 1 3DPR plots for dipeptides. The accumulated percentage contributions of conformers with particular ψ and ϕ torsional angles are plotted against the ψ and ϕ angles. (a) AlaAla, (b) LeuLeu, (c) AspPhe, (d) AlaPhe.

only B9 conformations. Together, these features contribute to the poor recognition of such peptides by peptide-binding proteins of broad specificity, such as transporters [5–7] and to the evolution of specific proteins and enzymes that capitalize upon recognition of their unique features. Dipeptides with anionic *N*-terminal amino acid residues (D or E) are a group with a wide distribution of conformers (Figure 2(c)). The main weighted distribution of conformers in ψ-space is around A7, the main peak being A7B2; smaller, broader peaks occur at A7B12 and A7B9. The small amount of A10 is mainly split between A10B12 and A10B2. Unusually, A4 conformations extend into A5 and A6; A4 conformers are distributed between A4B9, A4B12 and A4B2 with the peaks being very broad. Dipeptides containing aromatic amino acid residues tend to favour stable, low energy forms, which account for a large proportion of their conformers in solution. Combining ten dipeptides with aromatic residues at either the *N*- or *C*-terminus reveals a situation in which A7 predominates (Figure 2(d)). The bulk of conformers adopt A7B12 conformations with others at A7B9 and A10B12; smaller, broad peaks occur at

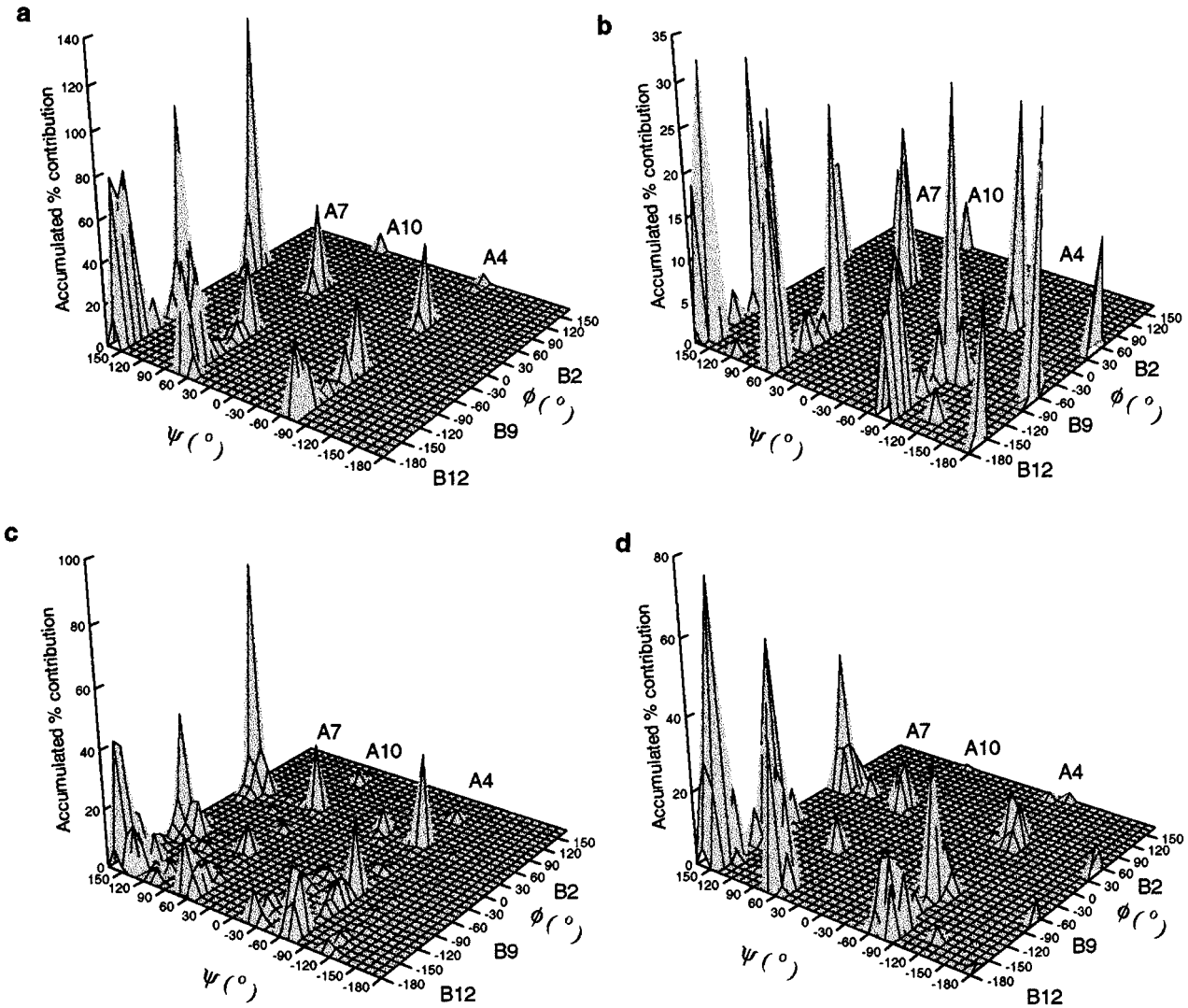

Figure 2 3DPR plots for groups of related dipeptides. The accumulated percentage contributions of conformers with particular ψ and ϕ torsional angles are plotted against the ψ and ϕ angles. (a) Twelve dipeptides with *N*-terminal Ala, (b) six dipeptides with *N*-terminal Gly, (c) nine dipeptides with *N*-terminal Asp or Glu, (d) ten dipeptides containing aromatic residues.

A4B9/B12 with very little A10B9. The spread of conformers into A1 arises from inclusion of the glycyl peptides, GF and GY.

Three-dimensional Pseudo Ramachandran Plot of 50 Dipeptides

As has been shown, plotting the weighted conformer distribution for any dipeptide gives the specific conformer profile for that dipeptide. Combined plots for structurally related peptides show their similarities and highlight any conformational idiosyncrasies of particular dipeptides, e.g. those with Gly or Pro. When plots are done for large collections of dipeptides the result is dominated by their similarities, and any features that are particular for a peptide group, although still present, are largely overshadowed. Thus, to approximate to the 400 natural members of the dipeptide pool that are available as substrates for peptide transporters and peptidases, data for 50 different dipeptides were aggregated. Plate 1 shows the resultant 3DPR, with the main conformational combinations being (A7B9, A7B12, A7B2), (A10B9, A10B12, A10B2) and (A4B9, A4B12, A4B2). The distribution of conformers amongst these nine common forms is apparent, although inclusion of many dipeptides containing

G/P residues leads to the peaks becoming broadened somewhat. These torsional preferences are illustrated in the sector diagram in Figure 3. For these calculations (see the Methods section) torsional space was divided into 30° sectors, so designated regions, e.g. A7, etc., are those that best cover the preferred torsions, although these may actually extend somewhat into adjacent sectors and/or be tightly focussed within a 30° sector. Conformers corresponding to certain of these various torsional combinations are illustrated for AA in Plate 2.

N–C Distance

A plot of the *N–C* distance against accumulated percentage contributions for all conformers for 50 dipeptides shows that they distribute into discrete groups with different lengths (Figure 4(a)). The main groups have conformer lengths between 4.5 and 6.3 Å; the smallest group, around 3.5 Å, has conformers with a *cis* ω bond, arising from XP peptides. This grouping of *N–C* distances is a consequence of the favoured conformations described above, with particular combinations of backbone torsional angles imposing particular *N–C* geometries on the conformers. This effect is illustrated by plotting *N–C* distances for specific torsional combinations, e.g. A7(B9,B12) (Figure 4(b)), A4(B9,B12) (Figure 4(c)) and A10(B9,B12) (Figure 4(d)). We focus upon these combinations (excluding B2) because they prove to be the only ones that are relevant biologically in dipeptide recognition (see below). These latter *N–C* plots illustrate the several conformational symmetries that occur with dipeptides. For example, with respect to ψ torsions, A4 and A10 form an approximate symmetric pair (see Figure 3), which, in combination with any favoured ϕ angle (B2, B9, B12), produce shorter *N–C* distances than found for A7 with comparable ϕ angles. Analogous symmetries occur with the preferred ϕ angles (see Figure 3), e.g. when combined with A7, B2 and B9 form a comparable symmetric pair, giving rise to a shorter *N–C* distance than for B12.

Side Chains and Chi-space

Conformers that may appear similar on the basis of backbone torsional angles and *N–C* distances may nevertheless be distinguishable by their chi_1-space (χ_1) values (for amino acid residues other than G and A). We examined chi_1-space for an extensive collection of computed dipeptide conformers and found that they overwhelmingly adopted the common forms of *gauche*(+) (g^+), *gauche*(−) (g^-) and *trans* (t), which are found for residues in proteins [23,24]. However, for dipeptides with charged *N*- and *C*-termini in water, certain side chains can produce chi-space effects that do not arise within protein chains. For example, with *N*-terminal anionic residues, variation in chi-space can markedly

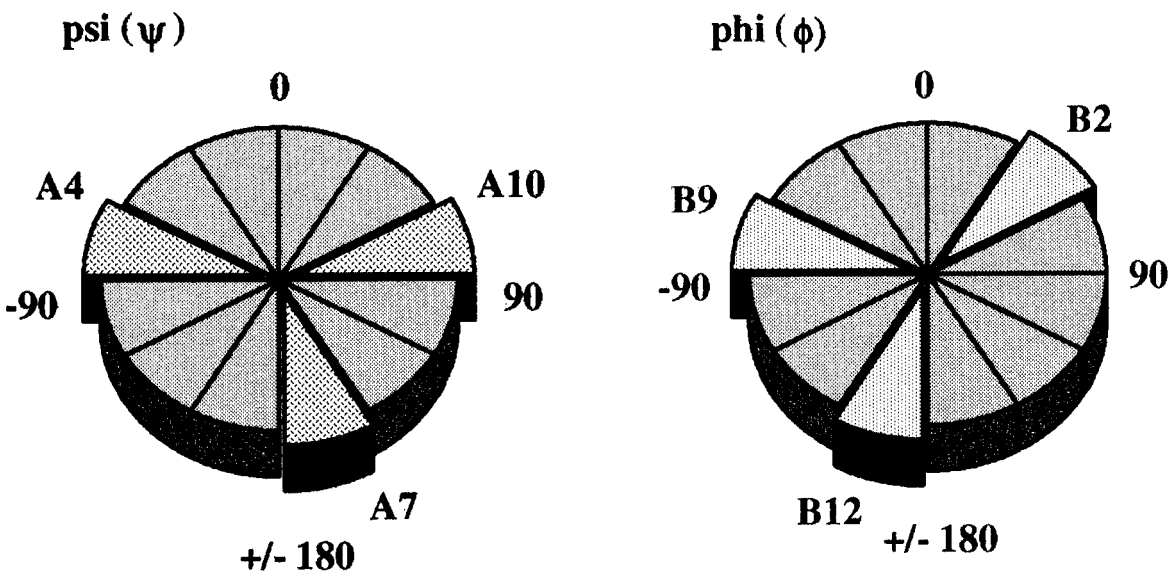

Figure 3 Pie chart illustrating the preferred psi (ψ) and phi (ϕ) torsion angles adopted by dipeptides. Conformational space for ψ and ϕ was divided up into 30° sectors covering 0° to ±180°. Preferred ψ angles are labelled A7 (+150° to ±180°), A10 (+60° to +90°) and A4 (−60° to −90°), and ϕ values are labelled B12 (−150° to ±180°), B9 (−60° to −90°) and B2 (+30° to +60°).

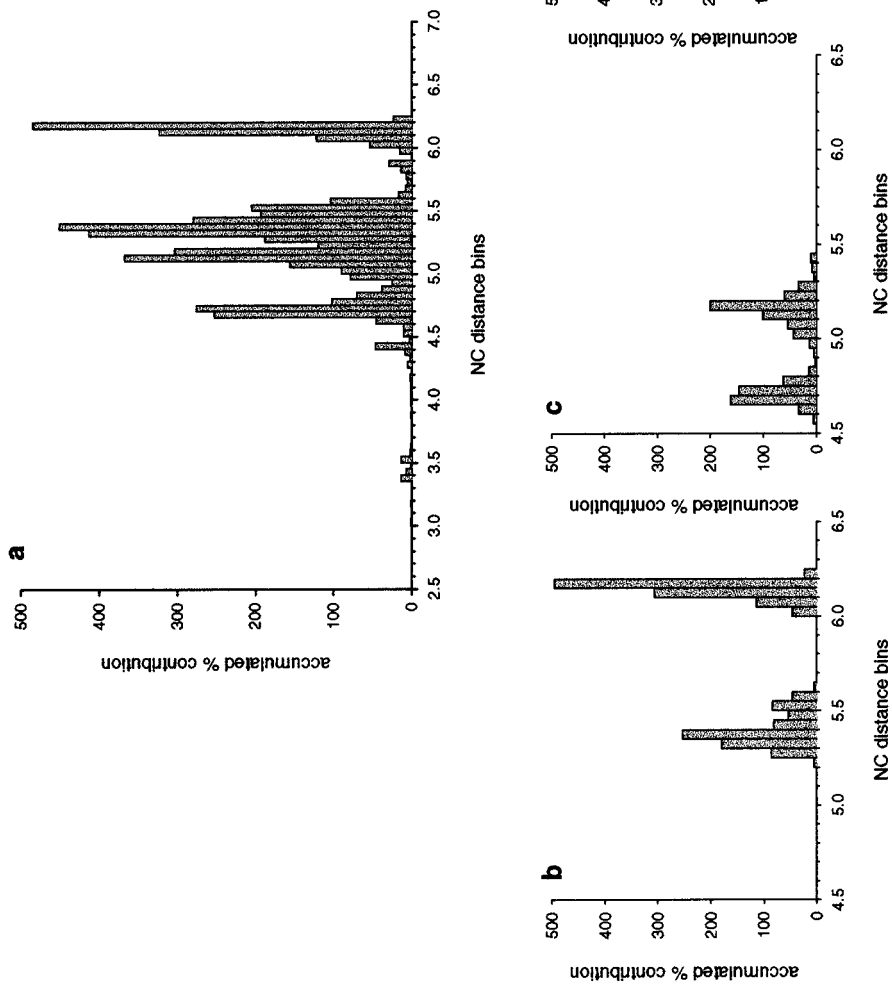

Figure 4 Relationships between *N–C* distance and ψ and ϕ torsional angles. *N–C* distance was divided into 0.05 Å and the accumulated percentages of conformers within each 0.05 Å bin were calculated. *N–C* distances versus percentages for all conformers for the 50 dipeptides listed in Table 1 were then plotted: (a) all conformers, (b) all A7B9 ($\approx$ 5.4 Å) and A7B12 ($\approx$ 6.2 Å) conformers, (c) all A4B9 ($\approx$ 4.7 Å) and A4B12 ($\approx$ 5.2 Å) conformers, (d) all A10(B9,B12) ($\approx$ 5.4 Å) conformers.

affect the properties of the dipeptide and its biological activity (see below). For DA, about 45% of its conformers are computed to exist as A7(B9,B12) and for those that adopt g^+ and g^- orientations stabilization can occur between the β-carboxylate group on D and the positively charged *N*-terminus; in contrast, with *trans* forms no such interaction occurs and the amino nitrogen effectively retains its positive charge. These effects are illustrated in the different distribution of potential fields around the terminal amino group (Plate 3). Analogous effects can arise with any charged residue at either the *N*- or the *C*-terminus. The importance of chi-space alone in the absence of a charged side chain is exemplified by AP, which is computed to exist with about 45% of its conformers in A7B9 (it cannot adopt B2, B12). Of these A7B9 forms, about 85% exist with the P residue in g^+ and 15% in g^-; only the latter form places the carboxylate and side chain into an orientation that can be recognized by dipeptide transporters (see below).

DISCUSSION

Predominant Backbone Conformations of Dipeptides

The results here describe the preferred backbone conformations adopted by dipeptides in aqueous solution. These largely define the repertoire of conformers that exist for all members of the naturally occurring dipeptide pool. This pool represents the complement of ligands and substrates for various proteins, e.g. the dipeptide transporters found in microorganisms, plants and animals [3–6]. Thus, the total conformational forms of this pool contain the structural information that is the basis for the specific molecular recognition of these substrates by their transporters. In other words, it is this structural information that has driven the evolution of dipeptide transporters to optimize their recognition and binding of this conformer collection [3–6,9,12].

Bioactive Conformations

Although the idea of dipeptides having specific bioactive conformation(s) is a basic premise in studies of their molecular recognition, attempts to identify them have largely been restricted to inspection of crystal structures of free and/or bound forms or consideration of results from NMR studies [1,2]. Computer-based conformational analysis of dipeptides as a class has been largely neglected, probably because of reservations concerning the conformational flexibility of these compounds [14-16]. Here, we have shown that this approach can provide a useful description of the complement of conformers present in solution for any individual or group of dipeptides. This 'snapshot' provides insight into the structural information content of the peptide(s) in solution. In practice, conformers of an individual dipeptide may undergo interconversion in solution, but a dynamic equilibrium will exist in which the repertoire of conformers remains essentially constant under fixed conditions, with each particular conformer being present at a relatively fixed concentration related to its energy. Our proposal for a 3DPR (Figures 1 and 2 and Plate 1) allows graphical visualization of this equilibrium conformer profile for any dipeptide substrate(s) under a particular set of conditions. The range of side chain chemistries and their combinations in the 400 natural dipeptides affects the distribution of backbone conformations. These effects can be studied by comparing 3DPRs for individual and groups of dipeptides.

The predominant backbone features displayed by the dipeptide pool overall can be present to varied degrees in the conformer population of an individual dipeptide. However, they may not necessarily be present in the minimum energy conformer of any particular dipeptide. Focussing upon ligand crystal structures or minimum energy conformations when trying to elucidate bioactive conformations may limit the information that is necessary to identify the molecular recognition parameters for the compound(s) acting as a ligand or substrate. This information may only be obtainable from consideration of the complete conformer profile of the substrate(s).

Molecular Recognition Templates

The parameters important for molecular recognition of dipeptides include: (i) charged *N*-terminal α-amino and *C*-terminal α-carboxylate groups, allowing electrostatic and hydrogen bond donor and acceptor interactions; (ii) combinations of torsion angles (ψ, ϕ and ω) in the backbone; (iii) stereochemistry at α-carbon chiral centres; (iv) *N*–*C* distance between terminal amino N and carboxylate C atoms; (v) chi(χ)-space torsion angles of side chains; (vi) hydrogen bond acceptor and donor properties of peptide bond O and N atoms; (vii) charge fields around *N*-terminal α-amino and *C*-terminal α-carboxylate groups. Incorporation of these parameters into a definition of the molecular recognition template(s) (MRT) of a dipeptide substrate(s) for

particular proteins, e.g. transporters, can only be realised by comparison of the structural information of the conformer pool with experimentally determined results for bioactivity, e.g. binding and transport. These studies have been performed for peptide transporters mainly with microorganisms but also with intestinal, kidney and plant transporters [6,7,9,12]. For the archetypal bacterial transporters of dipeptides, the di- (Dpp) and tripeptide permease (Tpp), it has been found that in addition to the common requirements for charged termini, all L-stereochemistry and a *trans* ω bond, their specificities are distinguished by different preferences for ψ and ϕ torsional angles and associated distinct *N*–*C* distances [5,7,9]. Thus, Dpp recognizes members of the conformer pool with A7(B9,B12) torsions (see Figure 3), whereas Tpp has specificity for conformers with A4,A10(B9,B12) torsions (Figure 3) [9; Marshall NJ, Grail, BM, Payne JW. Unpublished results]. Previous attempts to use computer modelling to determine the structural requirements for peptide transporters have not led to general descriptions of the type described here [25,26]. Determination of these distinctive MRTs was only possible by comparison of the transport properties for a range of substrates with the amounts of each conformational form present in their computed conformer pools [9]. The overall distribution of conformers for a dipeptide(s) is well described by its 3DPR, however, when identifying features that help to define an MRT one needs to consider not only the peak height but also its contour shape and the area under the peak [9].

N–*C* Distance

The distance between *N*- and *C*-terminal positive and negative charges in dipeptides is important as a molecular recognition parameter, and the variations that occur contribute to the different specificities of dipeptide transporters [9]. Thus, A7(B9,B12) conformers recognized by Dpp generally have a longer *N*–*C* distance (Figure 4(b)) than do Tpp substrate conformers with A4,A10(B9,B12) torsions (Figure 4(c) and (d)). Both Dpp and Tpp also recognize tripeptides, albeit less well than dipeptides [3–9,12]; various peptidases can also recognize both peptide classes. Then the question arises as to the nature of the structural parameters shared by conformers in the combined pools of di- and tripeptides that permit this extended recognition. Initial analysis of the pooled conformers for a collection of tripeptides [9] on the basis of their *N*–*C* distances combined with torsional preferences, has led to the identification of subsets of folded tripeptides that match the MRTs of Dpp and Tpp sufficiently well for them to act as transport substrates. Thus, it may only be possible to define precisely the MRT for various peptide-binding proteins by reference to the combined conformer pool of di- and tripeptides.

Side Chain Conformational Space and Molecular Recognition Parameters

For systems such as peptide transporters, which have evolved to optimize binding of the complete profile of amino acid residues, a mechanism that can accommodate variation in size and chi-space is required, and this has been achieved using large, water-filled pockets [27–31]. Nevertheless, even such accommodating side chain pockets cannot accept the chi-space distribution of all conformers; a feature that needs to be allowed for when evaluating the extent to which members of any pool of substrate conformers may be recognizable substrates. On the other hand, systems that are more specific in their ligand binding, e.g. receptor for thyrotropin release hormone, may be quite restricted in the type and orientation of the side chains that they will accept, which would be reflected in precise side chain binding pockets permitting only specific protein–side chain interactions; in these cases, chi-space could be a more important feature in defining the MRT. Charged side chains present additional features that may compromise recognition and binding of conformers that otherwise match an MRT on the basis of their backbone torsion angles. *N*-terminal positive and *C*-terminal negative charges of dipeptides are frequently critical in stabilizing ligand–protein complexes, as in peptide transporters [1,5,7,27–31]. For dipeptides lacking charged side chains, the distribution of their terminal charges can be described by rather symmetric potential fields around the *N*- and *C*-charged groups (see Plate 3). For dipeptides having *N*-terminal residues with negatively charged side chains, i.e. Asp or Glu, many of the conformers formed are in part stabilized by interaction of this side chain charge with the *N*-terminal positive charge. Consequently, such conformers lack the correct charge distribution around the *N*-terminus, which is a critical feature of the MRT, effectively removing them from the pool of substrate conformers. Analogous situations may arise with either positively or negatively charged side chains at either the *N*- or *C*-termini, leading to distorted charge fields and

making such dipeptides particularly poor substrates for general peptide transporters [9].

Tri- and Oligopeptides

We have carried out analogous Random Searches for a selection of about 50 oligopeptides containing 3–5 amino acid residues [9; Marshall NJ, Grail BM, Payne JW. Unpublished results]. Comparable analysis of the resultant conformer pools has identified analogous predominant torsional forms. As mentioned, certain folded tripeptide conformers have been shown to possess a sufficient match with the dipeptide-based MRTs of Dpp and Tpp for them to be recognized as substrates by these transporters. In addition, particular elongated conformers of tri- and oligopeptides have been identified whose structures match the substrate specificity of the oligopeptide-binding protein OppA, which is the recognition protein for bacterial oligopeptide transporters [1,5,7,9,29–31] and [Marshall NJ, Grail BM, Payne JW. Unpublished results].

Acknowledgements

The authors thank N.J. Marshall and G.M. Payne for their many helpful contributions during these studies. This work was supported in part by the Biotechnology and Biological Sciences Research Council (grant 5/PAC02709) and the Research and Development Committee of the North West Wales NHS Trust.

REFERENCES

1. Wilkinson AJ. Accommodating structurally diverse peptides in proteins. *Chem. Biol.* 1996; **3**: 519–524.
2. Taylor MD, Amidon GL (eds). *Peptide-Based Drug Design: Controlling Transport and Metabolism*. ACS: Washington, DC, 1995.
3. Matthews DM, Payne JW. Transmembrane transport of small peptides. *Curr. Top. Memb. Transp.* 1980; **14**: 331–425.
4. Matthews DM. *Protein Absorption: Development and Present State of the Subject*. Wiley–Liss: New York, 1991.
5. Payne JW, Smith MW. Peptide transport by microorganisms. *Adv. Microbial Physiol.* 1994; **36**: 1–80.
6. Fei YJ, Ganapathy V, Leibach FH. Molecular and structural features of the proton-coupled oligopeptide transporter superfamily. *Prog. Nucl. Acid Res. Mol. Biol.* 1998; **58**: 239–261.
7. Smith MW, Tyreman DR, Payne GM, Marshall NJ, Payne JW. Substrate specificity of the periplasmic dipeptide binding protein from *Escherichia coli*: experimental basis for the design of peptide prodrugs. *Microbiology* 1999; **145**: 2891–2901.
8. Payne JW. Bacterial peptide permeases as a drug delivery target. In *Peptide-Based Drug Design: Controlling Transport and Metabolism*, Taylor MD, Amidon GL (eds). ACS: Washington, DC, 1995; 341–367.
9. Payne JW, Grail BM, Marshall NJ. Molecular recognition templates of peptides: driving force for molecular evolution of peptide transporters. *Biochem. Biophys. Res. Commun.* 2000; **267**: 283–289.
10. Abouhamad WN, Manson M, Gibson MM, Higgins CF. Peptide transport and chemotaxis in *Escherichia coli* and *Salmonella typhimurium*: characterisation of the dipeptide permease (Dpp) and the dipeptide-binding protein. *Mol. Microbiol.* 1991; **5**: 1035–1047.
11. Hiles ID, Gallagher MP, Jamieson DJ, Higgins CF. Molecular characterization of the oligopeptide permease of *Salmonella typhimurium*. *J. Mol. Biol.* 1987; **195**: 125–142.
12. Yang CY, Dantzig AH, Pidgeon C. Intestinal peptide transport systems and oral drug availability. *Pharm. Res.* 1999; **16**: 1331–1343.
13. Tyreman DR, Smith MW, Marshall NJ, Payne GM, Schuster CM, Grail BM, Payne JW. Peptides as prodrugs: the smugglin concept. In *Peptides in Mammalian Protein Metabolism: Tissue Utilization and Clinical Targeting*, Grimble GK, Backwell FRC (eds). Portland Press: London, 1998; 141–157.
14. Nikiforovich GV. Computational molecular modelling in peptide drug design. *Int. J. Pep. Prot. Res.* 1994; **44**: 513–531.
15. Cornell WD, Caldwell JW, Kollman PA. Calculation of the ϕ–ψ maps for alanyl and glycyl dipeptides with different additive and non-additive molecular mechanical models. *J. Chim. Phys.* 1997; **94**: 1417–1435.
16. Marshall GR, Beusen DD, Nikiforovich GV. Peptide modelling: an overview. In *Peptides: Chemistry, Structure and Biology. Proceedings of the 13th American Peptide Symposium*, Hodges RS, Smith JA (eds). ESCOM Science Publishers: Netherlands, 1993; 1105–1117.
17. Knapp-Mohammady M, Jalkanen KJ, Nardi F, Wade RC, Suhai S. L-Alanyl-L-alanine in the zwitterionic state: structures determined in the presence of explicit water molecules and with continuum models using density functional theory. *Chem. Phys.* 1999; **240**: 63–77.
18. Willett P. Searching for pharmacophoric patterns in databases of three-dimensional chemical structures. *J. Mol. Recog.* 1995; **8**: 290–303.
19. Ghose RS, Jaeger EP, Kowalczyk PJ, Peterson ML, Treasurywala AM. Conformational searching methods for small molecules. 1. Study of the SYBYL Search method. *J. Comput. Chem.* 1993; **14**: 1050–1065.

20. Judson RS, Jaeger EP, Treasurywala AM, Peterson ML. Conformational searching methods for small molecules. 2. Genetic algorithm approach. *J. Comput. Chem.* 1993; **14**: 1407–1414.
21. Treasurywala AM, Jaeger EP, Peterson ML. Conformational searching methods for small molecules. 3. Study of stochastic methods available in SYBYL and MACROMODEL. *J. Comput. Chem.* 1996; **17**: 1171–1182.
22. Doi H, Kitajima M, Watanabe I, Kikuchi Y, Matsuzawa F, Aikawa S, Takiguchi K, Ohno S. Diverse incidences of individual oligopeptides (dipeptidic to hexapeptidic) in proteins of human, bakers yeast, and *Escherichia coli* origin registered in the Swiss-Prot database. *Proc. Natl. Acad. Sci. USA* 1995; **92**: 2879–2883.
23. Dunbrack RL, Karplus M. Conformational analysis of the backbone-dependent rotamer preferences of protein side chains. *Nature Struct. Biol.* 1994; **1**: 334–340.
24. Hruby VJ, Li G, Haskell-Luevana C, Shenderovich M. Design of peptides, proteins and peptidomimetics in chi space. *Biopolymers* 1997; **43**: 219–266.
25. Li J, Hidalgo IJ. Molecular modelling study of structural requirements for the oligopeptide transporter. *J. Drug Target.* 1996; **4**: 9–17.
26. Swaan PW, Tukker JJ. Molecular determinants of recognition for the intestinal peptide carrier. *J. Pharm. Sci.* 1997; **86**: 596–602.
27. Dunten PW, Mowbray SL. Crystal structure of the dipeptide-binding protein from *Escherichia coli* involved in active transport and chemotaxis. *Prot. Sci.* 1995; **4**: 2327–2334.
28. Nickitenko AV, Trakhanov S, Quiocho FA. 2 Å resolution structure of DppA, a periplasmic dipeptide transport/chemosensory receptor. *Biochemistry* 1995; **34**: 16 585–16 595.
29. Tame JRH, Dodson EJ, Murshudov GN, Higgins CF, Wilkinson AJ. The crystal structures of the oligopeptide-binding protein OppA complexed with tripeptide and tetrapeptide ligands. *Structure* 1995; **3**: 1395–1406.
30. Sleigh SH, Tame JRH, Dodson EJ, Wilkinson AJ. Peptide binding in OppA, the crystal structures of the periplasmic oligopeptide binding protein in the unliganded form and in complex with lysyllysine. *Biochemistry* 1997; **36**: 9747–9758.
31. Davies GD, Hubbard RE, Tame JRH. Relating structure to thermodynamics: the crystal structures and binding affinity of eight OppA–peptide complexes. *Prot. Sci.* 1999; **8**: 1–13.

