Journal of Peptide Science
J. Peptide Sci. **7**: 175–189 (2001)
DOI: 10.1002/psc.312

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

NEIL J. MARSHALL, BARRY M. GRAIL and JOHN W. PAYNE*

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

Received 20 September 2000
Accepted 8 November 2000

Abstract: In this paper, we describe the predominant conformational forms adopted by tripeptides and higher oligopeptides in aqueous solution. About 50 tripeptides and almost 20 higher oligopeptides (4–6 residues) were subjected to conformational analysis using SYBYL Random Search. As with dipeptides (Grail BM, Payne JW. *J. Peptide Sci.* 2000; **6:** 186–199), both tripeptides and higher oligopeptides were found to occupy relatively few combinations of psi–phi space that were distinct from those associated with predominant protein secondary structures (e.g. helices and β-sheets). Again, the preferred psi (ψ) values for the first residue ($i-1$) were in sectors encompassed by the ranges from $+150°$ to $\pm 180°$, $+60°$ to $+90°$ and $-60°$ to $-90°$, which were combined with preferred phi (ϕ) values for the second residue (i) in sectors with ranges from $-150°$ to $\pm 180°$, $-60°$ to $-90°$ and $+30°$ to $+60°$. It was notable that tripeptides and, to a greater extent, higher oligopeptides adopted an initial psi (ψ) (Tor2) from $+150°$ to $\pm 180°$. For tripeptides, their *N*–*C* distances (distance between *N*-terminal nitrogen and *C*-terminal carbon atoms) distribute about 6.5 Å to give shorter, 'folded' conformers that are similar in length to dipeptides, and longer, 'extended' conformers that are distinct. Furthermore, for higher oligopeptides, their *N*–*C* distances did not increment in relation to their increasing number of residues and short, 'folded' conformers were still present. These findings have a bearing upon the recognition of these molecules as substrates for widely distributed peptidases and peptide transporters. Copyright © 2001 European Peptide Society and John Wiley & Sons, Ltd.

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

INTRODUCTION

We have previously described the use of molecular modelling to determine the solution conformations of dipeptides [1], and have shown how distinct conformational forms of dipeptides are recognized by the dipeptide permease (Dpp) and the tripeptide permease (Tpp) of bacteria such as *Escherichia coli* [2,3]. By logical extension, the analogous transporters in the mammalian intestine and kidney, PepT1 and PepT2, respectively, recognize and transport these same conformational forms, the so-called 'Molecular Recognition Templates' (MRTs). Tripeptides are transported by Dpp and Tpp in *E. coli* [4], and also by a third peptide permease, the oligopeptide permease (Opp) [5,6]. Given that Dpp and Tpp recognize distinct conformational forms of dipeptides, it is likely that Dpp, Tpp and Opp will also recognize different conformational forms of tripeptides, even though their mode of binding by ionic interactions, H-bonds, and hydrophobic interactions may be similar [7–9]. As only Opp has the ability to recognize and to transport peptides larger than tripeptides [5,10,11], it may be anticipated that Opp will recognize tripeptide conformers with an extended backbone [10,12,13], whereas Dpp and Tpp will be restricted to shorter tripeptide

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

conformers that approximate an *N–C* distance found for their preferred dipeptide substrates [2,3].

Thus, we describe in this paper molecular modelling studies of a collection (about 50) tripeptides and representative higher oligopeptides and report the predominant torsional forms they adopt in solution. Subsequently, these data can be used to describe the precise torsional forms of tripeptides recognized by the three peptide permeases (Dpp, Tpp and Opp), and to explain why only Opp can recognize and transport the higher oligopeptides. Our approach offers two immediate advantages. Firstly, the data obtained for the *E. coli* peptide transporters Dpp and Tpp are directly applicable to their clinically important counterparts in the mammalian intestine, PepT1, and kidney, PepT2, respectively [14,15]. Second, the full description of the MRTs recognized by Dpp-type (PepT1 and PepT2) and Tpp-type transporters will allow molecular modelling to be used as a tool to rationally design *in silico* therapeutic agents with improved uptake by these transporters. The advantage is that bioavailability can be 'designed in' early on in a drug development process before any chemical syntheses are attempted, a sentiment endorsed by Rademann and Jung [16]. The modelling of tripeptides is particularly relevant to the further characterization of PepT1 (and PepT2), as it is responsible for the uptake of orally active β-lactams, which are tripeptide mimetics. Tripeptides are substrates for all three *E. coli* peptide transporters and molecular modelling can help reveal why this is so. It is unlikely that approaches such as X-ray crystallography will be able to address this question in the foreseeable future.

MATERIALS AND METHODS

Construction of Peptide Models

Molecules were constructed from first principles as described previously using either SYBYL 6.2 running on a Silicon Graphics Indigo 2 workstation (R5K platform, 200 MHz processor) or SYBYL 6.4 running on a Silicon Graphics Octane workstation (R10K platform, 175 MHz processor) [1]. As with dipeptides, we considered that the physiologically relevant state of oligopeptides would be as zwitterions with charged termini. Thus, the amino terminus was protonated and the carboxyl terminus was modelled as a carboxylate. The side-chains for Lys, Orn, Asp and Glu were treated analogously. The default atom type for the nitrogen atoms of protonated amino groups was changed to a tetrahedral N4, the carbon of dissociated carboxylates was changed to trigonal planar C2 and the two carboxylate oxygens were changed to Oco2. Because of the inherent difficulties in determining the protonation states and/or sites of the side-chains of His and Arg, these two residues were purposefully omitted from this study. This omission is justified because His and Arg also occur less commonly as residues within proteins than many other amino acids [17].

Choice of Solvation Method

As we were interested in the conformations adopted by peptides in solution, it was paramount that we simulated their behaviour in aqueous solution. Also, conformational analysis of AlaAla (charged-form) using explicit solvation [18] is extremely computationally demanding [3], and consequently, we have opted for an approximation using a distance-based dielectric function of 80, which is less demanding computationally but still gives realistic results [19]. Furthermore, distance-based dielectric functions of 4 and below results in conformers with cis peptide bonds, a situation that is unlikely to occur with most peptides in solution [19].

Choice of Conformational Analysis Procedure

The minimum number of relevant rotatable backbone torsion angles for a tripeptide will be four (for GlyGlyGly and AlaAlaAla), making the use of grid search conformational analysis even more impractical than with dipeptides [1], and the case for using a random search procedure more clear cut. However, a particular concern was whether conformational space was adequately searched and, to ensure this, we generally performed several searches and also increased the number of search iterations (using between 1000 and 50000). The random search procedure was performed as described previously [1,3,19], using versions 6.2 and 6.4 of the SYBYL software package. The search algorithm used in SYBYL 6.4 incorporated a 'Bumps' checking procedure to eliminate high energy conformers from consideration early on, and also used a relative energy cut-off value of 7 kcal/mol rather than an absolute energy cut-off (default value of 70 kcal/mol in SYBYL 6.2). Where several random searches were performed for a peptide, the resulting sets of conformers were combined and referenced to the lowest energy member of the combined set.

Definition of Backbone Torsion Angle Segments

The conformational preferences of oligopeptides as a group were analysed separately from that of dipeptides [1], although the backbone torsion angles were still divided into $12 \times 30°$ sectors. The sector numbers were prefixed by letters to continue the series giving Tor2 (A1–A12 sectors), Tor4 (B1–B12 sectors), Tor6 (C1–C12 sectors), Tor8 (D1–D12 sectors), Tor10 (E1–E12 sectors), Tor12 (F1–F12 sectors) and so on (see Figure 1).

N–C distance measurement of conformers

The overall length of a peptide, its *N–C* distance, was defined and manipulated as previously described [1].

RESULTS

Choice of Peptides to Model

In this study, we restrict our consideration of oligopeptides to peptides composed of 3–6 residues. Even so, this gives a possible set of 8000 tripeptides, 160000 tetrapeptides etc., that could be chosen to model, an impractical proposition that was first noted when we modelled dipeptides (400 possible structures). The largest potential source of naturally occurring peptides are those derived from the hydrolysis of proteins and these will comprise a common pool of peptides available to peptide transporters and peptidases in all organisms [3]. Doi *et al.* [17] have examined the distribution of individual oligopeptides derived from the amino acid sequence of human, *Saccharomyces cerevisiae* and *E. coli* proteins and have determined that, essentially, all possible tripeptide species are present, although the complement of tetra- to hexapeptide species diminished to approximately 1.5% of the theoretical possible. Furthermore, the copy numbers of these peptides was phylum-specific and not correlated with the distribution of individual amino acids within proteins contained within the database at the time of this analysis. Interestingly, homopeptides were particularly prevalent, although the significance of this is unknown. Thus, although the choice of which peptides to model is not immediately obvious from this study, it would appear that modelling peptides so as to include the full range of side-chain chemistries (aliphatic, aromatic, polar, and charged) would be a suitable compromise.

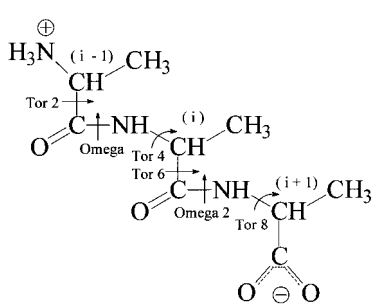

Figure 1 Schematic of tripeptide AlaAlaAla showing residues ($i-1$, i, and $i+1$), electrostatic charges and torsion angles. Tor2 is ψ_{i-1}, Tor4 is ϕ_i, Tor6 is ψ_i, and Tor8 is ϕ_{i+1}.

Computational Demands and 'Completeness' of the Random Search Procedure when Applied to Flexible Molecules

The random search procedure is designed to be less computationally demanding than a full, systematic grid search using small step sizes ($\approx 10°$) [20]. However, a specific concern for conformational analysis, particularly when random conformational searches are used, is ensuring that most, if not all, conformational space is searched rigorously. Previously, we have modelled dipeptides [1] and found that random searches using the default setting of 1000 search cycles searched conformational space adequately for most dipeptides, LysLys being a possible exception. Tri- and higher oligopeptides, however, have a more complex conformational space because of their increased number of backbone torsion angles and side-chains and their associated conformational flexibility.

As a specific test of the completeness of the random search procedure implemented, a series of random searches was carried out for the two extremes of tripeptide conformational dimensionality represented by AlaAlaAla (6 backbone) and LysLysLys (6 backbone + 12 side-chain torsions). The implicit assumption to be tested was that if conformational space had been sampled sufficiently well, then increasing the number of search iterations was unlikely to find conformers with significantly lower energies (i.e. new energy minima). Thus, these two peptides were subjected to random searches for between 1000 and 50000 search iterations. For AlaAlaAla (Figure 2(a)), increasing the number of search iterations from 1000 to 20000 did not result in finding conformers with lower energies, although the number of unique conformers found did increase. The minimum energy conformer found for

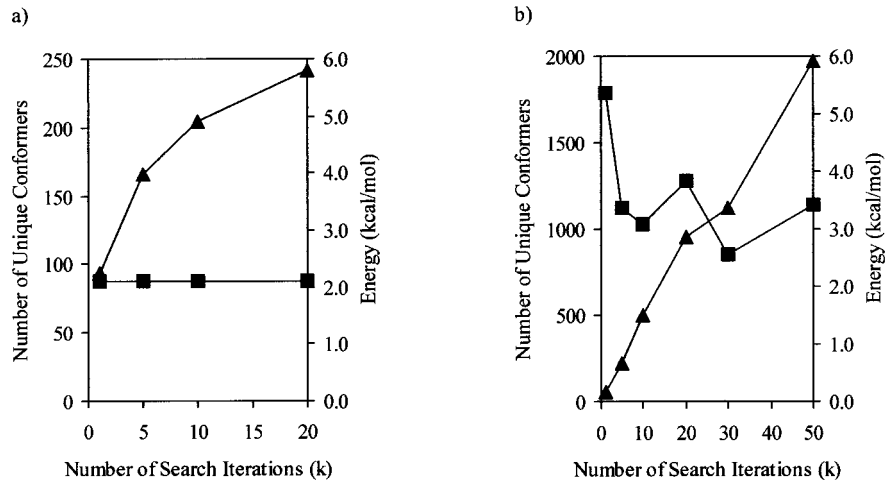

Figure 2 Variation of energy (kcal/mol) of minimum energy conformer (■) and number of unique conformations found (▲) for (a) AlaAlaAla and (b) LysLysLys for increasing search iterations used in random search.

each of these searches was essentially the same for each search with values similar to those reported in Table 1 (data not shown). The additional, unique conformers found with the increased search iterations can be regarded as close variants of previously found conformers. With LysLysLys (Figure 2(b)), the position is different in that the energy of the minimum energy conformer found does vary with the number of search iterations, generally decreasing with increased search iterations. Similarly, the number of unique conformers found increases approximately linearly with number of search iterations performed. This result was not entirely unexpected since Lys has a longer side-chain than Ala (and most other amino acids). Therefore, although the degree of 'completeness' of the conformational searching for LysLysLys may be lower owing to its more extensive conformational space, this particular tripeptide should be considered as a computationally challenging case for random search conformational analysis rather than a prohibitive limitation in its successful application. In addition, it must be stressed that the data derived are not 'absolute' [1], and that variations between searches are to be expected given their stochastic nature. Also, the computational times (Total CPU) taken for the random searches of these two molecules differed; the AlaAlaAla searches (1000–20000 iterations) took between 0.8 and 19.9 h, the LysLysLys searches (1000–50000 iterations) taking between 1.1 and 176.8 h using a Silicon Graphics Octane workstation (R10K platform, 175 MHz processor) running SYBYL 6.4. Thus, although the searches for LysLysLys take considerably longer than those for AlaAlaAla, most other peptides will fall somewhere between these two extremes, generally towards shorter computational times.

By its nature, random search may find the same conformer several times, and it has been suggested that the number of times the search finds each unique conformer (recorded as 'counts' in SYBYL) can be used as a measure of the 'completeness' [21]. In this case, the probability that a search is complete is given by $(1-(\frac{1}{2})^n)$, where n is the least number of times each unique conformer has been found (i.e. the conformer with the lowest 'counts'). Thus, if each unique conformer has been found at least six times, then the probability that the search has found all possible conformers is $(1-(\frac{1}{2})^6) = 0.98$, and this criterion is used to stop the SYBYL random search before all search iterations have been performed. With the peptide set modelled here, none was terminated before exhausting all the permitted search iterations, a finding echoed by Treasurywala *et al.* [20] in their study, although the minimum energy conformer of each search was often one of the most commonly found conformations (data not shown). That these random searches were not terminated because of this 'counts' criterion should not be taken as an indication that the random search procedure has sampled conformational space poorly for these compounds, nor should termination because of the 'counts' criterion be taken as absolute proof that conformational space has been searched well. Occasionally, particularly in the case of constrained structures, we have found that

random search terminates well before the search iterations have been exhausted because one low energy conformer has been found six times in rapid succession (BM Grail, S Gupta, NJ Marshall, GM Payne and JW Payne, unpublished data). In these cases, it is apparent that the random search has found conformers that, when energy minimized, always result in the production of the same low energy conformation without finding any other low energy conformations. Thus, total reliance upon the 'counts' for assessing the 'completeness' and 'quality' of the random search as implemented in SYBYL may be misleading.

Combining of Multiple Random Searches for a Single Peptide

Given that it may be difficult to assess reliably the degree of conformational space searched by the SYBYL random search procedure and that the computational time associated with increased search iterations (>10000 cycles) may become prohibitive, we have implemented a procedure to circumvent such limitations. In this approach, more than one random search is performed for each peptide, often using different starting conformations and numbers of search iterations. The resulting conformations from the different searches are combined to form a single set of conformations which are referenced to the lowest energy conformer found within all searches. Duplicate conformers (those with identical energies, backbone torsions and *N–C* distances) are removed from this combined set and the Boltzmann distributions determined using the lowest energy conformer of the combined searches. Because the computational demands imposed by higher numbers of search iterations do not increase linearly, it is frequently quicker to perform several smaller searches (e.g. 1000–10000 cycles) than a larger, single search (e.g. 50000 cycles). The stochastic nature of random search ensures that performing several smaller searches instead of one larger search does not bias the finding of particular sets of conformers repeatedly without visiting other regions of conformational space. Much of the data reported here is the result of combining two or more searches in this way (see Table 1).

Conformational Preferences of Tripeptides

Table 1 presents a summary of the tripeptides modelled in this study and shows data for the minimum energy conformer found for each peptide. Forty-eight tripeptides were modelled using 90 independent random searches (the number of searches for each tripeptide modelled varied from one (e.g. AlaGlyGly) to seven (LysLysLys)). A total of 490626 search iterations were performed to locate 36353 unique conformers of the 48 tripeptides modelled (the number of unique conformers found varied from 42 (LysTrpLys) to 3482 (LysLysLys)). The only residues not considered in this analysis were Arg and His for reasons discussed earlier. The energies of the minimum energy conformers varied from circa -5 kcal/mol (for PhePhePhe) to $+24$ kcal/mol (ProProPro). A similar situation was found when dipeptides were modelled in that those with aromatic residues (e.g. PhePhe, TyrTyr), had lower energy minimum energy conformers than those containing Pro residues (e.g. ProAla, GlyPro). The stabilizing factors pertinent to peptides and involved in the energy calculations have been discussed in some detail [1]. The minimum energy conformer generally only accounts for a small proportion of the total conformers (20%), unless there are particular energetic stabilizations operating (e.g. PhePhePhe, TyrTyrTyr). The number of unique conformations found is much greater than that for dipeptides (as expected) [1], and is mainly related to the chi torsions of the side-chain, although SerSerSer (a relatively simple side-chain) produces an unexpectedly large number of unique conformations when compared with the structurally related homotripeptides AlaAlaAla and ThrThrThr. The *N–C* distance of the minimum energy conformers varies from 3.78 Å (GlyGlyPro) (that contained one *cis* peptide bond) to 8.87 Å (TrpGlyGly) (where both peptide bonds were *trans*), the vast majority falling in the range of 5–7 Å.

To examine the conformational preferences of tripeptides (residues $i-1$, i and $i+1$), we compared pairs of torsion angles along the peptide backbone. Thus, we compared the two peptide units encompassing the pairs Tor2/Tor4 (ψ_{i-1}/ϕ_i) and Tor6/Tor8 (ψ_i/ϕ_{i+1}) (pseudo-Ramachandran analyses [1–3]) and also the torsion angle pair Tor4/Tor6 (ϕ_i/ψ_i) (Ramachandran analysis) for the central residue (i) (Figure 1). To facilitate these analyses, dedicated macros (written in SYBYL Programming Language (SPL)) were written to categorize and to extract the required data from the SYBYL spreadsheets. First, these were used to categorize the two peptide bonds (omega and omega_2) as either *cis* ($0° \pm 90°$) or *trans* ($\pm 180° \pm 90°$). When the two peptide bonds were examined in detail for all these tripeptides, it was found that the spread of torsion angle values about the 'ideal' *trans* value of $\pm 180°$ was

Table 1 Summary of Tripeptide Random Searches and Properties of Minimum Energy Conformers

Peptide[a]	Iterations (k) (searches)[b]	Unique conformers[c]	Energy[d] (kcal/mol)	Percentage contribution[e]	Tor2	Omega	Tor4 (°)	Tor6	Omega_2	Tor8	*N*–*C* distance (Å)
AAA	36 (4)	666	2.10	2.86	166	176	−58	−47	177	−67	5.44
AAG	11 (2)	498	2.05	4.35	166	177	−58	−48	177	−68	5.48
AAK	21 (3)	2061	2.13	2.44	172	179	54	61	−177	−158	6.96
AFA	11 (2)	571	0.69	4.54	167	176	−55	−41	178	−70	5.37
AFG	11 (2)	741	1.10	1.89	163	−179	52	52	−176	70	4.90
AGA	11 (2)	643	2.58	4.10	166	176	−58	−48	177	−66	5.47
AGG	1 (1)	270	2.49	7.54	166	177	−58	−50	177	−68	5.53
AMA	11 (2)	1140	0.60	2.71	166	176	−59	−52	179	−65	5.54
CCC	11 (2)	1276	−0.60	3.24	164	175	−64	−49	−178	−144	6.25
DAA	11 (2)	771	1.77	3.10	159	178	−56	−49	177	−67	5.57
DDD	11 (2)	1362	2.24	2.18	159	178	−55	−43	178	−73	5.55
EAA	11 (2)	1032	2.34	2.26	131	177	−70	−52	±180	−156	7.13
EEE	16 (3)	1470	2.71	4.54	164	175	52	58	−176	−162	6.68
FFF	16 (3)	438	−4.60	26.59	166	173	−59	−42	177	−72	5.37
GAA	11 (2)	447	2.76	3.28	−80	175	−59	−49	175	−72	4.34
GGA	1 (1)	310	3.22	4.99	−80	175	−60	−53	174	−71	4.33
GGF	1 (1)	510	1.49	5.21	−167	−179	−77	78	−179	−150	6.45
GGG	6 (2)	776	3.09	1.67	−80	175	−60	−54	175	−73	4.36
GGG					80	−175	60	54	−175	73	4.35
GGL	1 (1)	604	3.01	8.95	−176	178	−63	−60	177	53	7.34
GGP	1 (1)	184	9.04	8.25	143	±180	56	59	−1	−83	3.78
GGV	1 (1)	487	2.75	4.78	175	−179	57	53	−176	−160	6.77
GLY	1 (1)	363	1.34	29.77	−79	176	−56	−96	177	−113	4.20
III	11 (2)	385	4.34	12.61	162	175	−70	−54	−178	−162	6.80
KAA	1 (1)	281	2.10	31.67	163	179	−60	−46	−179	−75	5.80
KKK	118 (7)	3482	2.55	7.43	149	−179	−55	−57	−176	−162	6.91
KWK	1 (1)	42	16.03	17.58	147	−179	−50	−46	179	−72	5.68
LGG	1 (1)	398	2.31	23.26	−66	−179	65	64	−178	79	6.64
LLL	6 (2)	369	3.01	18.56	114	177	−59	−54	−179	−161	7.06
MAS	11 (2)	1375	0.30	5.41	163	±180	−58	−46	±180	−77	5.71
MGM	11 (2)	1537	−0.48	3.27	148	−177	−63	−61	−179	60	7.98
MLG	11 (2)	1163	1.38	4.38	163	176	−93	70	−175	74	6.35
MMM	16 (3)	1313	−2.51	5.34	97	175	−60	−48	179	−162	6.88
NNN	7 (3)	895	1.11	4.56	170	177	52	54	−175	−158	6.63
OOO	1 (1)	179	3.76	56.62	163	174	−70	−51	−179	−165	6.77
PAA	15 (2)	108	10.51	18.64	179	176	−58	−49	176	−67	5.21
PFK	11 (2)	312	9.95	14.27	179	175	−55	−49	−179	−157	5.87
PGG	1 (1)	321	10.82	8.27	±180	178	−56	−44	179	−61	5.52
PPP	1 (1)	146	23.67	12.19	163	−1	−78	164	−2	−78	5.25

Table 1 (continued)

Peptide[a]	Iterations (k) (searches)[b]	Unique conformers[c]	Energy[d] (kcal/mol)	Percentage contribution[e]	Tor2	Omega	Tor4 (°)	Tor6	Omega_2	Tor8	*N*–*C* distance (Å)
QAA	1 (1)	408	1.86	10.36	132	177	−70	−53	±180	−157	7.13
QQQ	1 (1)	446	2.09	9.04	162	174	−78	−54	−179	−163	7.05
SSS	16 (3)	2421	2.05	1.51	168	176	−57	−49	178	−64	5.49
TTT	11 (2)	1200	1.27	3.06	168	173	−70	−54	±180	−141	6.25
VGG	1 (1)	435	2.77	8.78	147	177	−67	−69	176	−73	6.18
VVV	12 (2)	1034	1.97	5.77	161	173	−65	−49	±180	−157	6.36
VYV	1 (1)	393	2.11	6.32	159	175	−63	−43	179	−155	6.25
WGG	1 (1)	473	15.40	11.28	155	±180	−160	−173	±180	69	8.87
YGG	1 (1)	389	0.94	10.00	173	±180	−58	96	−178	77	5.43
YYY	11 (2)	228	−3.22	29.40	143	−176	−160	59	179	91	8.53
Totals	491 (90)	36 353	160.93	478.80	–	–	–	–	–	–	–
Mean	10 (1.88)	757	3.35	9.98	–	–	–	–	–	–	–
S.D.	17 (1.06)	653	4.94	10.50	–	–	–	–	–	–	–
Maximum	118 (7)	3482	23.67	56.62	–	–	–	–	–	–	–
Minimum	1 (1)	42	−4.06	1.51	–	–	–	–	–	–	–

[a] Peptides are described by their standard one letter code and homotripeptides are shown in *italics*. The non-standard abbreviation O stands for ornithine (Orn).
[b] Number of search iterations performed in random search (number of independent searches performed).
[c] Number of unique conformers found.
[d] Energy of the lowest energy conformer found.
[e] Percentage contribution of minimum energy conformer.

asymmetric and about ±6° (data not shown), indicating that the above classification would effectively separate *cis* from *trans* conformations. Second, the remaining four backbone torsions (Tor2, 4, 6 and 8, respectively [1–3]) were divided into 12 × 30° sectors (A1–12, B1–12, C1–12 and D1–12, respectively). In this way, all torsion angles had been categorized, and conformers containing particular combinations of torsion angles, e.g. Tor2 = A7, omega = trans, Tor4 = B9, could be identified and selected. Finally, conformations belonging to each of the 12 × 12 = 144 combinations of Tor2 and Tor4 sectors (with a trans omega) were identified, their percentage contributions aggregated and these data exported to a separate file. This was repeated for all 48 tripeptides modelled and the final totals summed. A similar procedure was followed for the Tor6/Tor8 comparison (with a trans omega_2); for the Tor4/Tor6 comparison, only conformers with both omega and omega_2 at trans were considered.

These analyses are shown in Plate 1(a)–(c) and are, conceptually, the mirror image of a potential energy surface scaled by the Boltzmann distribution. Hence, the peaks on these plots actually represent the energy minima on the potential energy surface and are, in effect, an 'average' for all tripeptides. Most, if not all, tripeptides will have a significant proportion of their conformations within these peaks although particular tripeptides may occupy unique areas of conformational space or reside just outside the major volumes of these peaks. This was also found to be the case for dipeptides where those containing Gly and Leu residues frequently lay towards or just outside of these major peaks [1]. Thus, the effect of including peptides containing Gly and Leu residues, for example, in these analyses is to broaden these peaks. The graphs were plotted as 3-D Mesh Plots using SigmaPlot 2000 (SPSS Inc., Chicago, IL) using a 36 × 36 interpolation grid to generate intervening data applying the inverse distance smoother with default sampling proportions and exponents of 0.5 and 1, respectively.

For the Tor2/Tor4 analysis (Plate 1(a)), it is apparent that tripeptides, as with dipeptides [1], occupy the Tor2 sectors A4 (from −90° to −60°), A7 (from 150° to ±180°) (+A8 (from 120° to 150°)) and A10 (from 60° to 90°), a relatively greater proportion residing within A7 than within A4 and A10. For Tor4, the preferred sectors are B9 (from −60° to −90°), B8 (from −30° to −60°), B2 (from 30° to 60°) and B12 (from −150° to ±180°) (+B11 (from −120° to −150°). The B8 + B9 sectors are relatively more occupied than any of the other sectors. Thus, tripeptides have a great tendency to adopt the A7–A8/B8–B9 conformation at their *N*-terminus, and more so than was seen with dipeptides.

Analysis of the torsion angles about the central residue (i), Tor4 (ϕ) and Tor6 (ψ), a 'true' Ramachandran plot (Plate 1(b)), showed that the most populated sectors were the B8 + B9 sectors with C5 (from −60° to −30°), B2 with C11 (from 30° to 60°), and B8 + B9 sectors with C10 (from 60° to 90°).

With the Tor6/Tor8 analysis (Plate 1(c)), tripeptides populate the Tor6 sectors C5, C11 + C10, and C7; sectors D9 (from −90° to −60°), D12 (from −150° to ±180°), and D2 + D3 (30° to 90°) being predominant for Tor8. Thus, for the second ψ angle (Tor6, C sectors), a less extended torsion angle sector (C5) is chosen, although the extended torsion of C7 is still a major sector. For Tor8, there has been little change in the sectors adopted, i.e. the preferred sectors are still D9, D12, and D2 + D3.

That such a small proportion of conformational space is occupied by tripeptides is, perhaps, surprising. This was also shown for dipeptides [1], where only nine psi–phi combinations were occupied.

Although these pairwise comparisons are informative and allow a complete description of backbone conformational space for dipeptides, in the case of larger peptides it would be desirable to be able to view relevant information for all torsion angles and percentage contributions on a single plot. The application of some variant of parallel coordinate plots to visualize such complex, multidimensional data could prove valuable here [22].

N–C Distance Data for Tripeptides

The *N–C* distance data for all peptide conformers were divided into 0.5 Å bins from 2.5 to 10.0 Å and the percentage contributions of all conformers falling within the bins summed. These data are shown in Figure 3, and show a broadly symmetric distribution about the 6.5 Å bin. A significant proportion of the conformers have *N–C* distances comparable with those of dipeptides, despite containing an extra residue [1,2]. This situation has relevance to the differential recognition of tripeptide conformers by peptide transporters and peptidases and must, therefore, be considered when describing their potential substrates. Particular *N–C* distance bins encompass specific conformational subtypes (e.g. those with common Tor2, 4, 6 and 8 values), because the *N–C* distance is a consequence of the backbone torsion angles. Thus, although, in

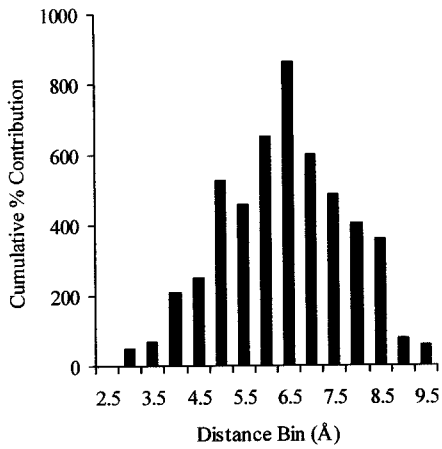

Figure 3 *N–C* distance data for the 48 tripeptides listed in Table 1. The *N–C* distances have been divided into 0.5 Å bins such that the bin labelled 2.5 Å represents conformers with an *N–C* distance of >2.5 Å and ≤ 3.0 Å.

principle, it is possible to determine which torsional forms are present within each *N–C* distance bin, in practice, this is far from trivial. Indeed, distances are frequently used by medicinal chemists to describe particular features of pharmacophores (e.g. distances between H-bond donors and acceptors) for the very reason that it is simpler. However, despite the inherent difficulties in determining which torsional subtypes occupy particular *N–C* distance bins, it is safe to assume that more elongated peptide conformers will have trans peptide bonds and mostly extended torsion angle values ($\approx \pm 180°$).

Conformational Preferences of Higher Oligopeptides

Table 2 summarizes the higher oligopeptides modelled in this study and shows relevant data for the minimum energy conformers. Higher oligopeptides take considerably longer than most tripeptides to model and, consequently, fewer have been modelled, although we have tried to be selective and model mainly those which have been studied as ligands for OppA (e.g. ValLysProGly), or homo-oligopeptides of varied side-chain chemistries (e.g. AspAspAspAsp).

Fifteen higher oligopeptides were modelled using 28 independent random searches in this study, the number of searches per molecule varying from one (e.g. AlaAlaAlaAlaAlaAla) to three (e.g. AlaAlaAlaAla), with most having two. A total of 496808 search iterations were performed to find 16668 unique higher oligopeptide conformers, the number of unique conformers per molecule ranging from 88 (for ValLysProGly) to 2905 (for AlaAlaAlaAla). The energies of the minimum energy conformers varied from circa -10 kcal/mol (TyrTyrTyrTyrTyr) to $+24$ kcal/mol (ProLeuTrpAla). As with the tripeptides (Table 1), homopeptides comprising aromatic residues such as Tyr and Phe appear to gain very favourable energetic stabilizations brought about by interactions between aromatic rings and peptide bonds. The presence of internally strained prolyl residues in the tetrapeptides ProLeuGlyGly and ProLeuTrpAla results in noticeably higher energies for these molecules. With the notable exception of TyrTyrTyrTyrTyr, the percentage contributions of the minimum energy conformer generally accounts for a small proportion of the total conformers (ranging from 3 to 30%). As would be expected with the greater conformational space of higher oligopeptides, the number of unique conformers found has increased greatly, ranging from 88 (for ValLysProGly) to 2905 (for AlaAlaAlaAla), homopeptides of Ser giving relatively more conformers than would be expected from the fewer search iterations performed for these peptides when compared with the Ala-series of homo-oligopeptides. Perhaps surprisingly, the *N–C* distance of the minimum energy conformers varies from 3.8 Å (e.g. AlaAlaAlaAla) to 10.2 Å (e.g. LysAlaAlaAla), with most falling in the range 6–9 Å and not being significantly longer than many tripeptides (Table 1).

The conformational preferences of higher oligopeptides were investigated in the same manner as tripeptides, although only the torsion angles common to all peptides (Tor2–Tor12) were investigated. Again, dedicated macros were written to define the torsional sectors (A1–A12 (Tor2), B1–B12 (Tor4), C1–C12 (Tor6), D1–D12 (Tor8), E1–E12 (Tor10), and F1–F12 (Tor12)) and extract the required data. Because of the limitations of 3D pseudo-Ramachandran plots and currently available graphical display methods discussed earlier, these are shown in Plate 1(d) only for Tor2 and Tor4 (A and B sectors). It is apparent that higher oligopeptides occupy similarly tight torsional sector ranges as dipeptides and tripeptides, although there is a far greater tendency to populate the Tor2 sector A7. Such a preponderance of higher oligopeptide conformers in this particular torsional form at the critical *N*-terminus of a peptide is likely to exert a strong evolutionary pressure on oligopeptide transporters (e.g. *E. coli* Opp), such that they are likely to recognize this *N*-terminal conformation to the exclusion of all others. For Tor4, the sectors B8 and B9 are the most common, B2 and B12 being relatively less important. As with tripeptides, the most common Tor6 sector adopted is C5, with sector C7 being less popular. For Tor8, the

Table 2 Summary of Oligopeptide Random Searches and Properties of Minimum Energy Conformers

Peptide[a]	Iterations (k) (searches)[b]	Unique conformers[c]	Minimum energy (kcal/mol)[d]	Percentage contribution[e]	Tor2	Om_1	Tor4	Tor6	Om_2	Tor8	Tor10 (°)	Om_3	Tor12	Tor14	Om_4	Tor16	*N–C* distance (Å)
AAAA	60 (3)	2905	0.88	3.21	163	±180	−59	−45	±180	−55	−48	−179	−162	–	–	–	3.86
AAAAA	30 (2)	1375	−0.38	23.68	165	178	−57	−41	179	−52	−47	±180	−61	−46	−179	−62	7.62
AAAAAA	50 (1)	1215	−1.82	30.53	166	177	−56	−40	179	−53	−48	−179	−60	−47	±180	−55	9.25
DDDD	30 (2)	1898	1.19	11.04	162	177	−96	66	±180	53	55	−179	−158	–	–	–	6.94
DDDDD	30 (2)	1095	1.38	22.12	159	−179	−57	−52	−177	−153	49	179	−150	58	−175	−163	6.61
KKKA	25 (2)	447	1.24	19.99	124	177	−72	−56	−174	−168	167	176	−71	–	–	–	10.23
PLGG	30 (1)	304	9.74	18.20	178	179	53	54	−179	63	52	−179	81	–	–	–	3.81
PLWA	30 (1)	236	24.39	13.17	179	175	−64	−47	−179	−160	160	177	−151	–	–	–	8.77
SSSS	30 (2)	2565	0.58	7.68	172	179	−53	−38	177	−71	−59	±180	−87	–	–	–	3.74
SSSSS	37 (2)	1985	1.02	7.11	166	175	−80	−60	177	−78	170	177	47	54	−176	−162	6.69
VKPG	25 (2)	88	11.06	12.06	134	178	51	67	−2	−85	176	178	−55	–	–	–	3.96
VVVV	30 (2)	962	1.62	8.80	162	173	−64	−49	179	−157	148	178	−84	–	–	–	8.70
VVVVV	30 (2)	898	0.84	10.91	164	177	−73	−55	−176	−88	80	176	−68	−54	178	−98	10.21
YYYY	30 (2)	565	−6.58	20.04	167	173	−57	−40	176	−78	167	177	−149	–	–	–	7.46
YYYYY	30 (2)	130	−10.29	80.58	154	−179	−150	179	172	−53	−39	172	−84	156	177	−172	8.53
Totals	497 (28)	16 668	34.86	289.14	–	–	–	–	–	–	–	–	–	–	–	–	–
Mean	33 (1.87)	1111	2.32	19.28	–	–	–	–	–	–	–	–	–	–	–	–	–
S.D.	9 (0.51)	888	7.99	18.50	–	–	–	–	–	–	–	–	–	–	–	–	–
Maximum	60 (3)	2905	24.39	80.58	–	–	–	–	–	–	–	–	–	–	–	–	–
Minimum	25 (1)	88	−10.29	3.21	–	–	–	–	–	–	–	–	–	–	–	–	–

[a] Peptides are described by their standard one letter code and homo-oligopeptides are shown in *italics*.
[b] Number of search iterations performed (number of independent searches done).
[c] Number of unique conformers found for combined searches.
[d] Energy of the lowest energy conformer found.
[e] Percentage contribution of minimum energy conformer.
For hexa-alanine, the three remaining torsion angle values for the minimum energy conformer are Tor18 −47°, Omega_5 ±180°, and Tor20 −161°.

sectors D8 and D9 are still the predominant torsional forms (data not shown).

N–C Distance Data for Higher Oligopeptides

The *N–C* distance data for higher oligopeptides were considered separately from that of tripeptides, and, within this group, the tetrapeptides (nine in total), pentapeptides (five in total), and hexapeptide were treated as individual sets and plotted accordingly (Figure 4). As with tripeptides earlier, a significant proportion of these higher oligopeptide conformers have adopted a 'folded' conformation indicated by short *N–C* distances (< 4 Å). Thus, although it is to be expected that folded higher oligopeptide conformers exist in aqueous solution together with their more elongated forms, this may be partly attributable to the inclusion of ValLysProGly in the sample which has conformers with *cis* peptide bonds and, consequently, shorter *N–C* distances. It is also interesting to see that the *N–C* distances of tetrapeptide and pentapeptide conformers are not significantly different from each other, or the single hexapeptide modelled, and not correspondingly longer than those of tripeptide conformers. Although, intuitively, one might expect *N–C* distance to increment proportionally with chain length, that it does not may help to explain the evolutionary pressure on oligopeptide transporters such as *E. coli* Opp to recognize and to transport peptides composed of at least two to six residues because of their similarity in *N–C* distance [5]. There is also evidence that oligopeptide transporters can recognize peptides considerably longer than six residues [23,24] and this also has bearing upon the sequence of molecular recognition events occurring in such transporters.

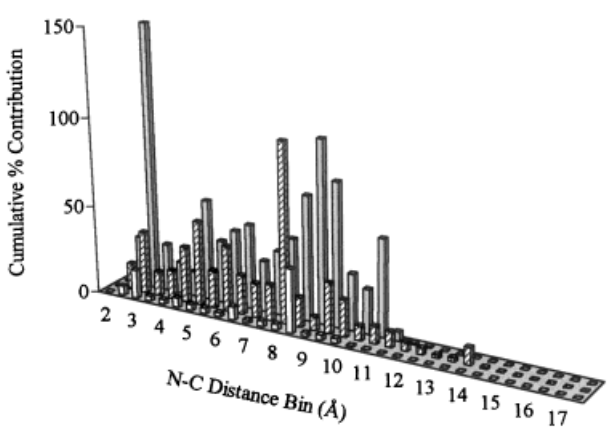

Figure 4 *N–C* distance data for the 15 oligopeptides listed in Table 2. The data for nine tetrapeptides (grey), five pentapeptides (hatched) and one hexapeptide (clear) have been combined and plotted separately. The *N–C* distances have been divided into 0.5 Å bins such that the bin labelled 3 Å represents conformers with an *N–C* distance of > 3 Å and ≤ 3.5 Å.

DISCUSSION

Predominant Backbone Conformations of Tripeptides and Higher Oligopeptides

The results presented here extend those previously reported for dipeptides [1], and provide a representative analysis of the combined di- and tripeptide pool using computer-based conformational analysis. It is apparent that tripeptides also adopt a limited number of backbone conformations, as was so evident with dipeptides; for Tor2 and Tor4, these sectors are the same, although the Tor2 sector A7 is much more occupied with tripeptides than with dipeptides, with A4 and A10 being relatively less populated in comparison. For Tor6, sectors C5, C10 and C7 are preferred; C7 results in a more 'extended' backbone conformation, whereas C5 and C10 result in a more 'folded' conformation. The Tor8 sectors of D9, D12 and D2 are again the predominant torsion angles adopted. With higher oligopeptides, the conformational preferences mimic those of tripeptides, although there is an even greater tendency to adopt an extended Tor2 sector of A7. Clearly, proteins that have evolved to recognize and to bind tri- and higher oligopeptides (e.g. Opp of *E. coli*) would be expected to recognize this *N*-terminal conformation over all others. This expectation is confirmed by the crystal structures of the oligopeptide binding protein, OppA, in complex with a variety of peptide ligands (comprising 2–4 amino acid residues) [9–11]; all bound ligands adopt an 'extended' A7-type *N*-terminal torsion. Plate 2 shows, in spacefill representation, a modelled A7 'extended' AlaAlaAla conformer and a LysAlaLys conformer extracted from the crystal structure of an OppA-LysAlaLys complex [11].

Bioactive Conformations

We have previously discussed the concept of bioactive conformations of dipeptides [1–3] and the same principles will hold for tri- and higher oligopeptides. Again, the modelling results presented here are a 'snapshot' of the dynamic equilibrium situation likely to be found in solution and do not consider interconversion of conformational species. However,

in the context of molecular recognition, this interconversion is largely irrelevant, as this dynamic equilibrium will result in an essentially fixed proportion of each conformational species at any concentration. The graphical representation of dipeptide conformer profiles using three-dimensional pseudo-Ramachandran plots (3DPR) [1–3] has also been used here. However, as useful as these have proved to be for displaying conformational information for dipeptides, their limitations in their current form are highlighted with tripeptides in that at least two plots are needed to adequately display the necessary information. The need to be able to present this complex, conformational data on a single plot has been partly addressed by Becker using parallel plots [22], although there is still no easy way of including conformer weighting in the form of percentage contribution. These limitations will need to be overcome to more easily understand and graphically display oligopeptide conformations. As with dipeptides, the predominant backbone conformations adopted by the overall tripeptide pool (approximated here with about 50 tripeptides), may not represent the minimum energy conformer of any particular tripeptide. Thus, it is paramount to consider the entire conformer profile when attempting to identify the bioactive conformations and not just those matching a crystal structure (if available) or the minimum energy conformer. Furthermore, as shown in Tables 1 and 2, the minimum energy conformer frequently only represents quite a small percentage of the total and one must be careful not to over-interpret its significance.

Molecular Recognition Templates

Determining the molecular recognition templates (MRTs) for tripeptides is a challenging prospect in that they are substrates not only for the di- and tripeptide transporters, but also for oligopeptide transporters (e.g., Opp of *E. coli*) [2,3,7–13,23–28]. Although these three classes share many of their recognition features, in order to show some degree of specificity, they must also recognize particular features that are likely to be possessed only by specific conformers. For instance, dipeptides are high-affinity substrates for both Dpp and Tpp, which require that their dipeptide substrates have positively-charged amino termini and negatively-charged carboxy termini; the two systems discriminate between dipeptide conformers by recognizing different combinations of torsion angles (ψ_{i-1} and ϕ_i) [2,3]. Thus, it is likely that these same two systems recognize and transport tripeptide conformers that are able to match sufficiently well the appropriate dipeptide MRT. Therefore, one would expect a group of tripeptide conformers to possess the correct *N*-terminal ψ torsion, the correct *N*–*C* distances and positively- and negatively-charged termini, and a set of 'folded' conformations does match sufficiently well the MRTs of dipeptides to be transported by Dpp and Tpp [3]. Similarly, it would be expected that there is a corresponding set of conformers that violate the MRTs for dipeptides to such an extent that they cannot be substrates for Dpp and Tpp but are available to Opp [3]. In this respect, it is relatively simple to sort tripeptide conformers into appropriate *N*–*C* distances that can be recognized only by Tpp, Dpp or Opp; thus Tpp is restricted to folded conformers of 4.0–5.5 Å, Dpp to folded conformers of 4.5–6.0 Å, and Opp to extended conformers of >6.5 Å [2,3]. In contrast with Dpp and Tpp, Opp has no absolute requirement for a charged carboxyl terminus [5], a feature that helps to explain its broader specificity and ability to bind peptides longer than five residues [23–25]. Plate 3 shows a superimposition between a typical A7 'extended' AlaAlaAla conformer and that of the crystal structure of the OppA ligand LysAlaLys [11]. The superimposition is very good over the first nine heavy atoms (the *N*-terminal *N* to the second peptide bond carbonyl C), although less so beyond this point. However, results from an extended range of biochemical and biophysical assays indicate that the *N*-terminus of a peptide is more important to the recognition process and modifications here are less well tolerated than elsewhere and may completely abolish transport [5,6,23,25,26]. In contrast, modifications to the *C*-terminus of a peptide are relatively better accepted [5,6,23] and Opp-type transporters can bind peptides composed of at least 20 residues [23,24]. Thus, although OppA may use features of the *C*-terminus of a peptide for molecular recognition, these are of secondary importance to those of the *N*-terminus and reinforce these initial recognition events. Unfortunately, no crystal structures are available for OppA in complex with peptide ligands derivatized at the *C*-terminus to determine whether it is orientated in the same manner within the binding site although unable to form the stabilizing salt-bridge. It is interesting to speculate whether following initial docking of the *N*-terminal region there may be some conformational rearrangement of ligand such that OppA is able to accommodate a wider variety of torsional forms at

the *C*-terminus of a peptide than the crystal structures so far would suggest. Thus, the solution conformer of AlaAlaAla in Plate 3 would be well recognized and bound by OppA because of its suitable *N*-terminal conformation and could be rearranged in the process to induce a better fit at the *C*-terminus. Alternatively, during the binding of such conformers, the OppA-peptide ligand complex could adopt slightly different conformations in order to accommodate several common *C*-terminal torsional forms without exercising strict specificity. Although the crystal structures solved to date do not support this suggestion, the majority of the ligands crystallized are charged, lower affinity Lys-X-Lys peptides that may not be ideal choices from which to draw general conclusions about the entire tripeptide pool. In support of these possibilities, both OppA and DppA can adopt different conformational forms in solution depending upon the bound substrate [25,4].

Because the *N*-terminus of a peptide is such a critical determinant of molecular recognition by peptide transporters, we have analysed this feature in more detail for OppA. The two *N*-terminal torsion angles, Tor2 and Tor4, were examined about the Tor2 sectors A7 and A8, and the Tor4 sectors B8 and B9. These two pairs of 30° sectors were split into $6 \times 10°$ sectors covering from $+120°$ to $\pm 180°$ for the Tor2 sectors, and from $-30°$ to $-90°$ for the Tor4 sectors, and the percentage contributions of tripeptide conformers falling into these narrower sectors were summed. This showed that the three 10° sectors for each 30° sector were not populated equally and that for Tor2 the angular range from $+140°$ to $\pm 180°$ contained the majority of conformers and for Tor4, the equivalent angular range was from $-50°$ to $-90°$, which gave a 'weighted' mean value for an A7B9 tripeptide conformer of $+160°$ and $-65°$, respectively. This structure compares very closely with the Tor2 and Tor4 values of the ligands of some 18 crystallized OppA-Lys-X-Lys complexes [11] (Figure 5). Thus, the predictions from modelling of potential tripeptide substrates are corroborated by these experimental data.

Although Dpp and Tpp have the ability to transport some tripeptide conformers, they are completely unable to transport tetra- and higher oligopeptides, thus rendering these peptides as substrates only for the remaining transporter Opp. Given that tri- and higher oligopeptides have a greater tendency to adopt the 'extended' A7 Tor2 sector, it is unsurprising that crystal structures of

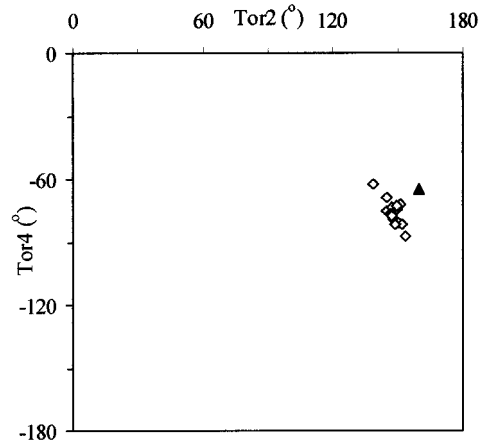

Figure 5 Pseudo-Ramachandran plot of the Tor2 and Tor4 values of 18 complexes of OppA crystallized with a series of Lys-X-Lys ligands [11]. The ligands (in one letter code) comprising this analysis are (with PDB code in parentheses): KAK (1JET), KCK (1B05), KDK (1B4Z), KEK (1JEU), KFK (1B40), KGK (1B3L), KHK (1B3F), KIK (1B3G), KKK (2OLB), KLK (1B9J), KMK (1B32), KNK (1B5I), KPK (1B46), KQK (1B5J), KSK (1B51), KTK (1B52), KWK (1JEV), and KYK (1B58), indicated with ◇. Also indicated on the plot is the weighted mean value of an A7B9 conformer derived from the modelling studies (▲).

OppA liganded with various oligopeptide ligands have shown that these ligands all adopt an A7 ψ_{i-1} torsion with an overall length of >6.5 Å [10,11]. Thus, Opp has evolved to transport the more extended tri- and higher oligopeptide conformers, whereas Dpp and Tpp complement this by transporting the folded, shorter conformers.

N–C Distance

Analysis of the *N–C* distances of tripeptide and higher oligopeptide conformers reveals why a variety of transporters and peptidases are required to handle this repertoire of substrates. Whereas Dpp and Tpp can transport most dipeptide conformers and 'folded' tripeptide conformers, they could not also recognize 'extended' tripeptide conformers and higher oligopeptides [1–3] and complementing this, Opp has evolved to recognize 'extended' conformations, at the expense of ignoring shorter, 'folded' conformers. The *N–C* distance data for tripeptides shows that there are significant amounts of 'folded', as well as 'extended' conformations. There are higher oligopeptide conformers that have *N–C* distances that approximate to 'folded' tripeptides, which are, presumably, not available as substrates to any of the three main classes of peptide

transporters. The maximum *N*–*C* distance of higher oligopeptide conformers does not increase in proportion to their number of residues, i.e. tetrapeptides are not twice as long as dipeptides (when the percentage contributions are considered). This feature may help to explain why Opp has no strict requirement for a negatively-charged *C*-terminus; firstly it can then bind peptides of various residue length without anchoring the *C*-terminal carboxylate and, second, even for peptides of the same number of residues, it can be less discriminating about the *N*–*C* distance range it will initially recognize.

Side-chain Conformational Space and Molecular Recognition Parameters

Peptide transporters are able to transport essentially any small peptide regardless of sequence and have evolved water-filled cavities to accommodate the variously-sized peptide ligand side-chains [27,28]. Even so, not all side-chain conformations can be accepted, particularly in the case of bulky, hydrophobic residues, e.g. Phe, or long-chain, polar residues, e.g. Lys. The chi-space distribution of tri- and higher oligopeptide conformers has not been investigated exhaustively, but they follow dipeptides in generally adopting the common χ_1 values of gauche($-$) (g^-), gauche ($+$) (g^+), and trans (t) which are typically seen in protein structures [1,29,30]. This generalization does not always hold for peptides containing residues with charged side-chains at their termini, e.g. Asp, Glu, Lys, because ionic interactions may result between the residue side-chain and the charged termini leading to different χ_1 values. Charged side-chains may also lead to distorted charge fields about the termini of the peptide conformers, a feature that may seriously compromise their ability to be recognized by peptide transporters as a charged *N*-terminus, for example, is a critical feature of the MRT of all three peptide transporters [3]. An additional facet of chi-space that requires consideration is the ability of Dpp and Tpp to transport certain 'folded' tripeptide conformers as well as dipeptides [2,3]. In these cases, it can be envisaged that the second peptide bond and the *C*-terminal side-chain form a structural unit that mimics a dipeptide *C*-terminal side-chain sufficiently well that it can be accommodated in the appropriate side-chain pocket [2]. Whilst this scenario has not been verified by crystallizing appropriate DppA-tripeptide complexes, it explains why 'folded' higher oligopeptide conformers with the correct *N*–*C* distance cannot be recognized by either Dpp or Tpp.

Modelling of Other Small, Bioactive Molecules

A basic interest here has been to determine the structural and conformational features of peptides recognized by the generic peptide transporters widespread in Nature, and to use this information to improve the oral bioavailability of therapeutic agents such as β-lactams and ACE inhibitors. Consequently, we have been using the same approach to model a variety of therapeutic agents such as ACE inhibitors and analogues designed *in silico* to identify modifications that would result in better recognition by mammalian peptide transporters so as to enhance their oral bioavailability (S Gupta, BM Grail, NJ Marshall, GM Payne and JW Payne, unpublished results). Together with studies upon varied peptidomimetics, this approach should be a useful addition to the drug discovery and design process.

Acknowledgements

This work was funded by the Biotechnology and Biological Sciences Research Council (Grant P12847). We gratefully acknowledge use of the facilities of the Protein Data Bank (PDB) held at the Research Collaboratory for Structural Bioinformatics [31] (http://www.rcsb.org/pdb/) (last accessed on 6 September 2000) for access to the crystal structures of OppA-ligand complexes.

REFERENCES

1. Grail BM, Payne JW. Predominant torsional forms adopted by dipeptide conformers in solution: parameters for molecular recognition. *J. Peptide Sci.* 2000; **6**: 186–199.
2. 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.
3. Payne JW, Grail BM, Gupta S, Ladbury JE, Marshall NJ, O'Brien R, Payne GM. Structural basis for recognition of dipeptides by peptide transporters. *Arch. Biochem. Biophys.* 2000; **384**: 9–23.
4. 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.

5. Payne JW, Smith MW. Peptide transport by microorganisms. *Adv. Microbial. Physiol.* 1994; **36**: 1–80.
6. Payne JW. Bacterial peptide permeases as a drug delivery target. In *Peptide Based Drug Design: Controlling Transport and Metabolism*, Taylor MD, Amidon GL (eds). American Chemical Society: Washington DC, 1995; 341–367.
7. Dunten P, Mowbray SL. Crystal structure of the dipeptide binding protein from *Escherichia coli* involved in active transport and chemotaxis. *Prot. Sci.* 1995; **4**: 2327–2334.
8. Nickitenko AV, Trakhanov S, Quiocho FA. 2 angstrom resolution structure of DppA, a periplasmic dipeptide transport chemosensory receptor. *Biochemistry* 1995; **34**: 16585–16595.
9. 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.
10. 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.
11. Sleigh SH, Seavers PR, Wilkinson AJ, Ladbury JE, Tame JRH. Crystallographic and calorimetric analysis of peptide binding to OppA protein. *J. Mol. Biol.* 1999; **291**: 393–415.
12. Tame JRH, Murshudov GN, Dodson EJ, Neil TK, Dodson GG, Higgins CF, Wilkinson AJ. The structural basis of sequence-independent peptide binding by OppA protein. *Science* 1994; **264**: 1578–1581.
13. Davies TG, Hubbard RE, Tame JRH. Relating structure to thermodynamics: The crystal structures and binding affinity of eight OppA-peptide complexes. *Prot. Sci.* 1999; **8**: 1432–1444.
14. Adibi SA. The oligopeptide transporter (PepT-1) in human intestine: biology and function. *Gastroenterology* 1997; **113**: 332–340.
15. Yang CY, Dantzig AH, Pidgeon C. Intestinal peptide transport systems and oral drug availability. *Pharm. Res.* 1999; **16**: 1331–1343.
16. Rademann J, Jung G. Drug discovery — integrating combinatorial synthesis and bioassays. *Science* 2000; **287**: 1947–1948.
17. 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 SwissProt database. *Proc. Natl. Acad. Sci. USA* 1995; **92**: 2879–2883.
18. KnappMohammady 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.
19. Marshall NJ, Payne JW. The importance of electrostatic charge and dielectric constant in conformational analysis of biologically active dipeptides. *J. Mol. Model.* 2001; in press.
20. 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.
21. Saunders M. Stochastic exploration of molecular mechanics energy surfaces — hunting for the global minimum. *J. Am. Chem. Soc.* 1987; **109**: 3150–3152.
22. Becker OM. Representing protein and peptide structures with parallel-coordinates. *J. Comput. Chem.* 1997; **18**: 1893–1902.
23. 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.
24. Lanfermeijer FC, Detmers FJM, Konings WN, Poolman B. On the binding mechanism of the peptide receptor of the oligopeptide transport system of *Lactococcus lactis*. *EMBO J.* 2000; **19**: 3649–3656.
25. Tyreman DR, Smith MW, Payne GM, Payne JW. Exploitation of peptide transport systems in the design of antimicrobial agents. In *Molecular Aspects of Chemotherapy*, Shugar D, Rode W, Borowski E (eds). Springer-Verlag: Berlin, 1992; 127–142.
26. Rostom AA, Tame JRH, Ladbury JE, Robinson CV. Specificity and interactions of the protein OppA: Partitioning solvent binding effects using mass spectrometry. *J. Mol. Biol.* 2000; **296**: 269–279.
27. Tame JRH, Sleigh SH, Wilkinson AJ, Ladbury JE. The role of water in sequence-independent ligand binding by an oligopeptide transporter protein. *Nature. Struct. Biol.* 1996; **3**: 998–1001.
28. Wilkinson AJ. Accommodating structurally diverse peptides in proteins. *Chem. Biol.* 1996; **3**: 519–524.
29. Dunbrock RL, Karplus M. Conformational analysis of the backbone-dependent rotamer preferences of protein side-chains. *Nature. Struct. Biol.* 1994; **1**: 334–340.
30. Hruby VJ, Li GG, Haskell-Luevano C, Shenderovich M. Design of peptides, proteins, and peptidomimetics in chi space. *Biopolymers* 1997; **43**: 219–266.
31. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE. The Protein Data Bank. *Nucl. Acids. Res.* 2000; **28**: 235–242.

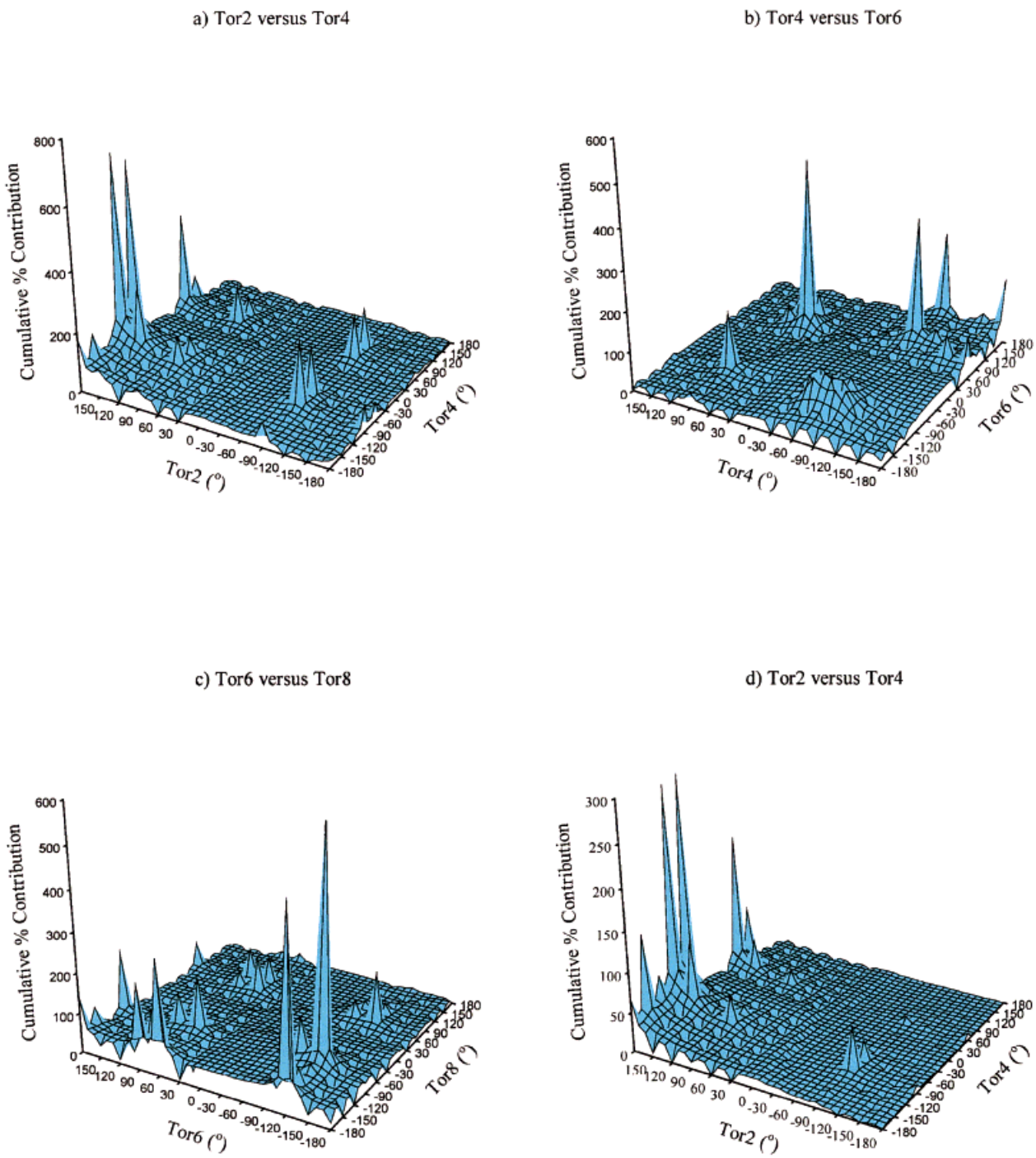

Plate 1 3D-Pseudo Ramachandran plots and 3D Ramachandran plots for the 48 tripeptides listed in Table 1 and 15 oligopeptides listed in Table 2. The cumulative percentages of conformers with particular combinations of torsion angles (ψ and ϕ) are plotted against those torsion angles. Parts (a)–(c) are for tripeptides and represent the torsion angle combinations of Tor2 with Tor4 (ψ_{i-1} and ϕ_i), Tor4 with Tor6 (ϕ_i and ψ_i), and Tor6 with Tor8 (ψ_i and ϕ_{i+1}), respectively. Part (d) relates to the oligopeptides and is for the Tor2 with Tor4 combination (ψ_{i-1} and ϕ_i).

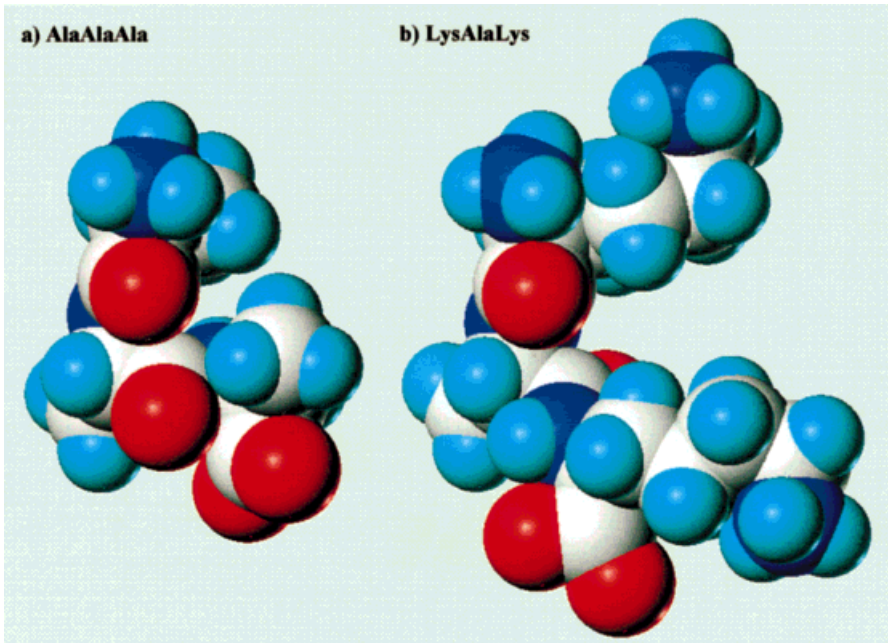

Plate 2 Comparison of a typical A7B9 'elongated' AlaAlaAla conformer with the crystal structure of the OppA ligand LysAlaLys in spacefill representation (PDB ID code 1JET).

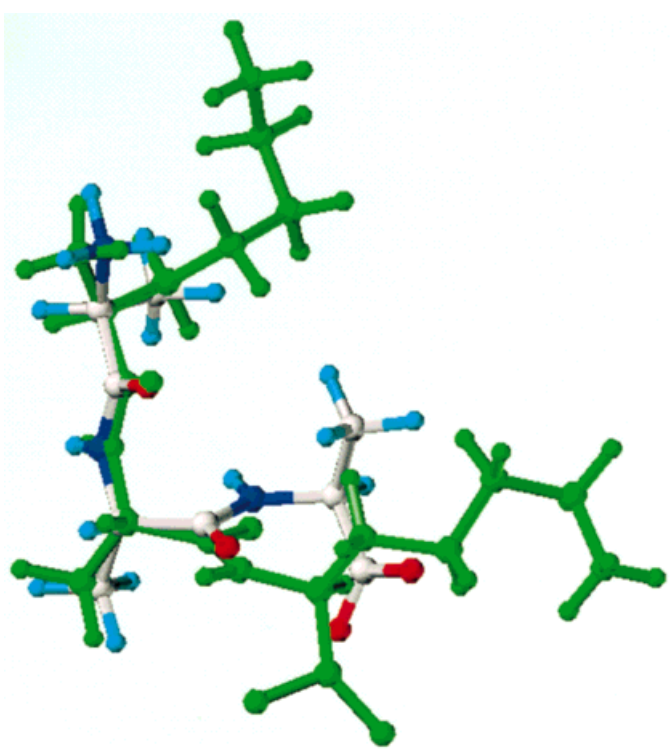

Plate 3 Superimposition of the typical A7B9 'elongated' AlaAlaAla conformer (in CPK colours) with the crystal structure of LysAlaLys (green) in ball-and-stick representation. The superimposition was performed using the nine heavy atoms between the *N*-terminal nitrogen and the carbonyl carbon of the second peptide bond. The RMS value for these two conformers over these nine atoms was 0.281 Å.