Refined models for the preferential interactions of tryptophan with phosphocholines†

John M. Sanderson

Received 17th May 2007, Accepted 16th August 2007
First published as an Advance Article on the web 31st August 2007
DOI: 10.1039/b707502b

A series of molecular models of the adducts formed between *N*-acetyl-L-tryptophan ethylamide and diacetyl-*sn*-glycero-3-phosphocholine have been generated. Using rOesy data that enabled us to place restrictions on the proximity of a number of key protons in the amino acid/phosphocholine pairs, a series of structures were generated following molecular dynamics and mechanics experiments using the CHARMM27 force field. These structures were then subjected to a series of clustering algorithms in order to classify the tight binding interactions between a single tryptophan and a phosphocholine. From these analyses, it is evident that: (i) binding is characterised by hydrogen bonding between the indole NH as donor and phosphate oxygen as acceptor, cation–carbonyl interactions between the choline ammonium and amide carbonyl groups and cation–π interactions; (ii) cation–π interactions are not always observed, particularly when their formation is at the expense of cation–carbonyl and hydrogen bonding interactions; (iii) on the basis of amino acid torsional parameters, it is possible to predict whether the phosphocholine headgroup will bind in a compact or elongated conformation. Extension of the procedures to characterise 2 : 1 Trp–PC binding revealed that the same intermolecular interactions are predominant; however, combinations of all three intermolecular interactions within the same adduct occur much more frequently due to the availability of donor/acceptor groups from both tryptophans in the 2 : 1 system.

Introduction

It is becoming increasingly clear that some amino acids have specific roles in controlling the behaviour of peptides and proteins in lipid bilayers.[1] Of the amino acids most studied, tryptophan has received particular attention because of its occurrence in transmembrane peptides and proteins at an increased frequency in interfacial regions associated with the lipid headgroups.[2,3] Studies on model peptides have identified interactions between tryptophan side chains and lipids through hydrogen bonding between the NH of the tryptophan indole and the lipid carbonyl groups.[4] Whilst these experiments are useful for determining the preferential conformations in such peptides, due to the constrained nature of the system they do not necessarily indicate the preferred conformation of the tryptophan residue in the absence of transmembrane constraints, such as during peripheral peptide binding. It is, therefore, of interest to characterise the preferred interactions between tryptophan and phosphocholine groups in the absence of competing constraints in a non-competing solvent with a dielectric constant intermediate between that of bulk water and the hydrocarbon core of the membrane. Although low dielectric solvents can exaggerate polar interactions, with the appropriate control experiments, significant factors that contribute to binding can be deduced.

In previous work,[5] the development of a system for characterising these interactions in chloroform was described. It was demonstrated that tryptophan and tyrosine have significant interactions with phosphocholines, particularly when compared with amino acids with apolar side chains, such as valine. In the particular case of tryptophan, it was demonstrated that a 2 : 1 Trp–PC complex was formed, and intermolecular rOesy cross peaks were obtained in order to generate distance restraints for molecular modelling exercises. This paper reports a refinement of the modelling approach and an analysis of the resulting structures.

Materials and methods‡

Molecular modelling

All calculations were performed with the Tinker software package[6] using the CHARMM27 force field implementation.[7] The solvent was modelled implicitly in all experiments by setting the dielectric constant to that of chloroform (4.8).[8]

Non-bonding interactions were truncated at a cut-off distance of 9 Å. The upper distance limits for protons exhibiting intermolecular cross peaks in 2D rOesy spectra were restrained, as described previously,[5] according to the magnitude of the cross peak: strong ≤4.2 Å; medium ≤4.7 Å; weak ≤5.3 Å. Seven pairs of distance restraints were included: atoms 50–17/99, 50–20/102, 50–25/107, 53–20/102, 53–25/107, 61–20/102 and 61–25/107 (see Fig. 1 for

Centre for Bioactive Chemistry, Department of Chemistry, University Science Laboratories, South Road, Durham, UK. E-mail: j.m.sanderson@dur.ac.uk; Fax: +44 (0)191 384 4737; Tel: +44 (0)191 334 2107

† Electronic supplementary information (ESI) available: Archive of structures from Table 2 (PDB format); computed energies for the entries in Table 2; full version of Table 2; tryptophan databases used for this work. See DOI: 10.1039/b707502b

‡ The author thanks the Royal Society (UK) for an equipment grant that supported part of this work (RSRG 24281). IUPAC naming conventions have been adhered to. Therefore, a torsion of +60° is classified as +*sc*.

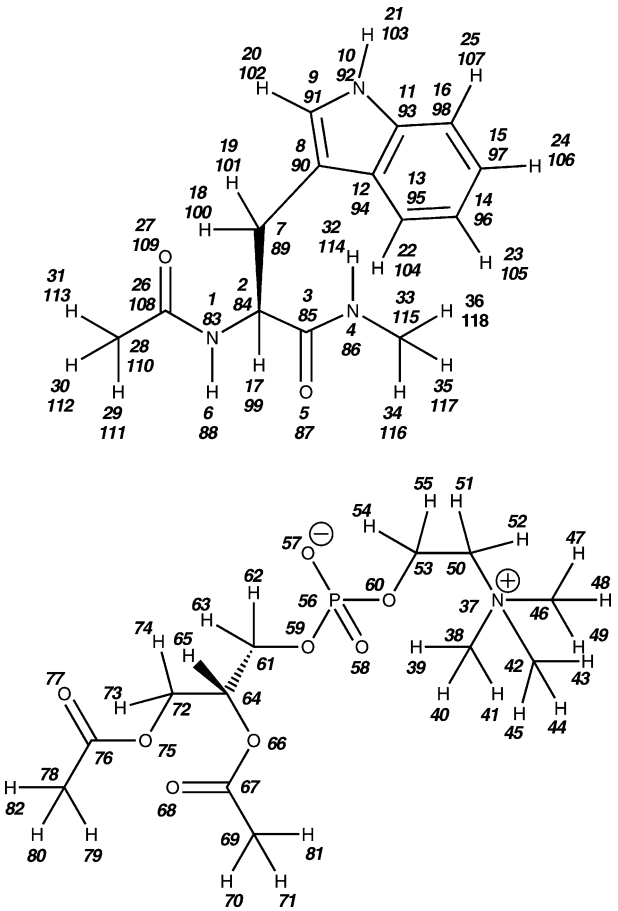

Fig. 1 Structures of *N*-acetyl-L-tryptophan ethylamide and diacetyl-*sn*-glycero-3-phosphocholine molecules described in this study and the corresponding atom numbering scheme.

numbering). Intermolecular distances were restrained using a flat-welled harmonic potential with a spring constant of 830 kJ mol^{-1}. In order to simplify the calculations and avoid complications arising from the application of distance restraints to the diastereotopic protons of methylene groups, distance restraints were applied to the carbon atom of the methylene group and relaxed by 1.12 Å (the aliphatic C–H bond length). The NH–Cα coupling constant was also extrapolated from the titration data to generate restrictions for the φ bond angles of the amino acids during the initial optimisation of the structure. These were both restrained in the range 215–265° using a flat-welled harmonic potential with a force constant of 900 kJ degree^{-1}. An initial low-energy structure was generated using a variable-metric minimisation algorithm[9] with an RMS gradient cut-off of 0.04 kJ mol^{-1} Å^{-1}. This was then subjected to repeated simulated annealing protocols between 1000 and 500 K over 10 ps, with a 1 fs integration time and a 1 ps equilibration period, using a sigmoidal cooling regime coupled to a Berendsen-type external bath.[10] During these calculations, distance and angle restrains were maintained in accordance with the earlier rOesy data. The final structure at the end of each annealing run was used for subsequent calculations, generating an initial set of 30 structures. Each of these structures was optimised using the protocol described above and then subjected to a 1 ns molecular dynamics run with a 1 fs step integration time. The temperature in these calculations was maintained at 298 K *via* coupling to a Berendsen-type external bath. These calculations were performed without torsional restraints, but with the distance restraints described above maintained. Structures were saved to the output trajectory at 1 ps intervals, with those obtained in the final 500 ps of each run used for subsequent work. Each of these structures was optimised without restraints using the protocol described above to generate a final dataset of 15 000 structures for analysis.

Data analysis

Dihedral and interatomic distance data were generated for each structure using purpose-written software. All statistical manipulations were performed on these raw data using the R programming environment (version 2.2.1),[11] unless otherwise stated.

Sampling of conformational space. In order to verify that the dynamics procedures had sufficiently sampled the available conformational space in the system, the probability distributions for several key torsions were calculated over the range −180 to +180° with a bin width of 3° for histograms and 1° for probability density plots. In the latter case, the probability densities at the bin mid-points were smoothed by applying a running lines smoothing algorithm using leave-one-out cross-validation methods.

Classification of 1 : 1 tight binding conformations. The initial group of 15 000 structures was analysed to select those in which one of the two tryptophan molecules was more tightly bound to the phosphocholine residue than the other. This was done for each structure through analysis of the indole NH-P (Fig. 1, atom numbers 21–56 and 103–56) and aryl–N$^+$ interatomic distances (Fig. 1, atom numbers 16–37 and 98–37), with tight binding being assigned when both of these distances were shorter for one of the two tryptophan residues. This produced a filtered dataset of tight-binding tryptophan residues, which was subjected to *k*-means partitioning according to tryptophan φ, ψ, χ_1 and χ_2 conformational parameters using the CLARA medoid-based algorithm for large datasets.[12] Clustering was performed using 10 subsets of the dataset containing $n/10$ randomly selected entries, where n = the total number of entries in the dataset. Dissimilarity was calculated according to rmsd. Clustering was initially performed by iterating the number of medoids k used by the algorithm (*i.e.* the number of clusters generated) over a range of values between 2 and 30. The quality of the clustering for each value of k was analysed by assessing the silhouette width[13] of the clusters generated, with the value of k producing the highest average width selected for further processing. Of the k clusters generated, those with a silhouette width less than or close to the mean width were subjected to a repeated iterative clustering procedure over a range of k values to generate subclusters. Clustering was repeated until all clusters and subclusters had a silhouette width greater than 0.6. In order to characterise the output structures, a nomenclature was adopted based on a combination of standard secondary structure and conformational naming conventions. The φ/ψ combination was labelled as α or β according broadly to the regions of the Ramachandran plot that correspond to these structures, χ_1 and χ_2 rotamers were assigned as ±*sc* (*syn*-clinal; ±60°), *ap* (*anti*-periplanar; ±180°), *sp* (*syn*-periplanar; 0°) and ±*ac* (*anti*-clinal; ±120°).[14] Additional assignments for χ_2 were ±*c* (clinal; ±90°), ±*s* (*syn*; ±30°) and

$\pm a$ (*anti*; ±150°). For phosphocholine dihedral angles, standard nomenclature was applied:[15] α_1 (64–61–59–56), α_2 (61–59–56–60), α_3 (59–56–60–53), α_4 (56–60–53–50), α_5 (60–53–50–37), α_6 (53–50–37–46) and θ_1 (72–64–61–59); torsions were assigned in the same manner as described above.

Classification of 2 : 1 binding conformations. For structures where both tryptophan molecules had close contacts with the phosphocholine molecule, hierarchical agglomerative clustering was performed using the program Xcluster (Version 1.7, Schrödinger Inc., 2004).[16] A distance matrix was generated using atom sets composed of all nitrogen atoms and the phosphorous atom from each structure. As the primary goal was to identify conformations that were statistically more abundant than would otherwise be expected, the clustering level at which the reordering entropy was at its minimum value was selected, and clusters that occurred with a relative frequency ≥0.02 identified from the resulting dataset.

Tryptophan conformations from proteins of known structure. Summaries of the backbone and side chain dihedral angles (φ, ψ, χ_1, χ_2) were generated for each tryptophan in the PDBSelect25 database.[17] An additional membrane protein-only dataset was generated. The latter comprised 1107 tryptophan entries from the 'membrane proteins of known structure' database[18] after filtering to remove homologous sequences. Correlations were formed between each of the dihedral angles in these datasets using the same *k*-means clustering procedure described above for analysis of the modelling results.

Results

Conformational properties of the model systems

In order to verify that sufficient conformational space had been sampled during the dynamics calculations to enable the determination of feasible binding conformations, torsional distributions were determined for several key bonds. Plots of the probability density for selected torsions are shown in Fig. 2. The distribution of the glycerol torsion θ_1 displayed maxima of equal probability density separated by 120° (Fig. 2A). A similar trend was observed for the α_6 torsion of the choline group (Fig. 2B). The conformations around the choline ethylene carbons exhibited maximum probabilities for the α_5 torsions of ±60° (Fig. 2C). The absence of the 180° torsion is consistent with the literature.[15,19] Both the φ (Fig. 2D) and ψ (Fig. 2E) torsions produced data consistent with Ramachandran-allowed conformations (Fig. 2G and H, respectively). Probability trends for χ_2 of the amino acids produced maxima at −90 and 90° (Fig. 2F), with lower probability conformations at torsional values close to 0°. These conformations are expected for a torsion between sp^2 and sp^3 bonded carbon atoms and are consistent with those observed for tryptophan in the PDB (Fig. 2I). Selected data also match previous reports; χ_1 conformations for all helical conformations in the PDBSelect25 dataset reflect those described previously,[20,21] with the *ap* conformation the most prevalent and $\pm sc$ conformations found with approximately equal probability. Overall, the sampling of conformational space was satisfactory for subsequent analyses to be performed with confidence.

Conformational analysis of tight-binding adducts

Tryptophan conformations. The initial intention was to probe tryptophan conformations that were implicated in binding to phosphocholine headgroups. Modelling procedures were performed with two tryptophan residues per phosphocholine group, in accordance with the binding stoichiometry determined in earlier work.[5] However, it was considered that it would be useful to characterise the instances from the dataset where the binding of one of the tryptophan residues was much tighter than the other, as this would reflect binding interactions present in the 1 : 1 complex that is a prerequisite for the formation of the 2 : 1 adduct. The interatomic indole–phosphate and indole–choline distances were determined for each tryptophan in the 15 000 structure dataset; identification of structures in which both distances were shorter for one of the two tryptophans led to a filtered dataset containing 8576 tryptophan molecules involved in close association with the phosphocholine. The filtered dataset was subjected to medoid-based clustering analysis of amino acid dihedral combinations for all values of *k* between 2 and 30, in order to optimise the process and generate clusters with the highest possible mean silhouette width. This led to the selection of an optimal value for *k* of 7, which produced clusters with a mean silhouette width of 0.66. Of these clusters, 5 were subjected to further *k*-means clustering, yielding a final total of 19 clusters that were distinct in conformational space and were present with a relative frequency ≥0.02 (Table 1). Clustering was additionally assessed by the production of Ramchandran plots for all of the amino acid dihedral combinations in each cluster (Fig. 3). The same clustering process was repeated for all of the tryptophan residues present in the PDBselect25 and membrane datasets. In some cases, two or more distinct clusters occurred with φ/ψ combinations in the same α- or β-structure region of the Ramachandran plot; these have been entered separately in Table 1, but may be distinguished by their ID in column 1.

From the data in Table 1, it is apparent that the majority of the clusters have angle distributions that are represented by tryptophan residues in the PDBselect25 and membrane datasets. However, it is notable that the $\beta\ -sc\ -c$ conformations (entries **1a** and **1b**) have a significantly higher occurrence in the model dataset, suggesting that this is a significant binding conformation. Also of note is high relative abundance of the $\alpha\ -sc\ +c$ conformer in the membrane protein dataset. Although this may be explained in part due to the higher proportion of helical proteins in this more restricted dataset, it is still significantly raised in comparison to other helical conformations when compared with the PDBSelect25 dataset and is well represented in the model dataset.

Phosphocholine conformations. For each cluster in Table 1, the dihedral angle distribution of key torsions of each of the phosphocholine molecules was assessed. Visual examination of the structures from the model dataset revealed that the majority of the interactions between tryptophan and PC molecules occurred in the region of the choline group. Clustering analysis was therefore performed using data for the α_3, α_4 and α_5 angles. Data for the final clusters are presented in Table 2.

Examination of these data indicates that the predominant conformations observed across all clusters for α_5 ($\pm sc$), α_4 ($\pm ac$ and *ap*) and α_3 ($\pm sc$) are broadly consistent with those observed

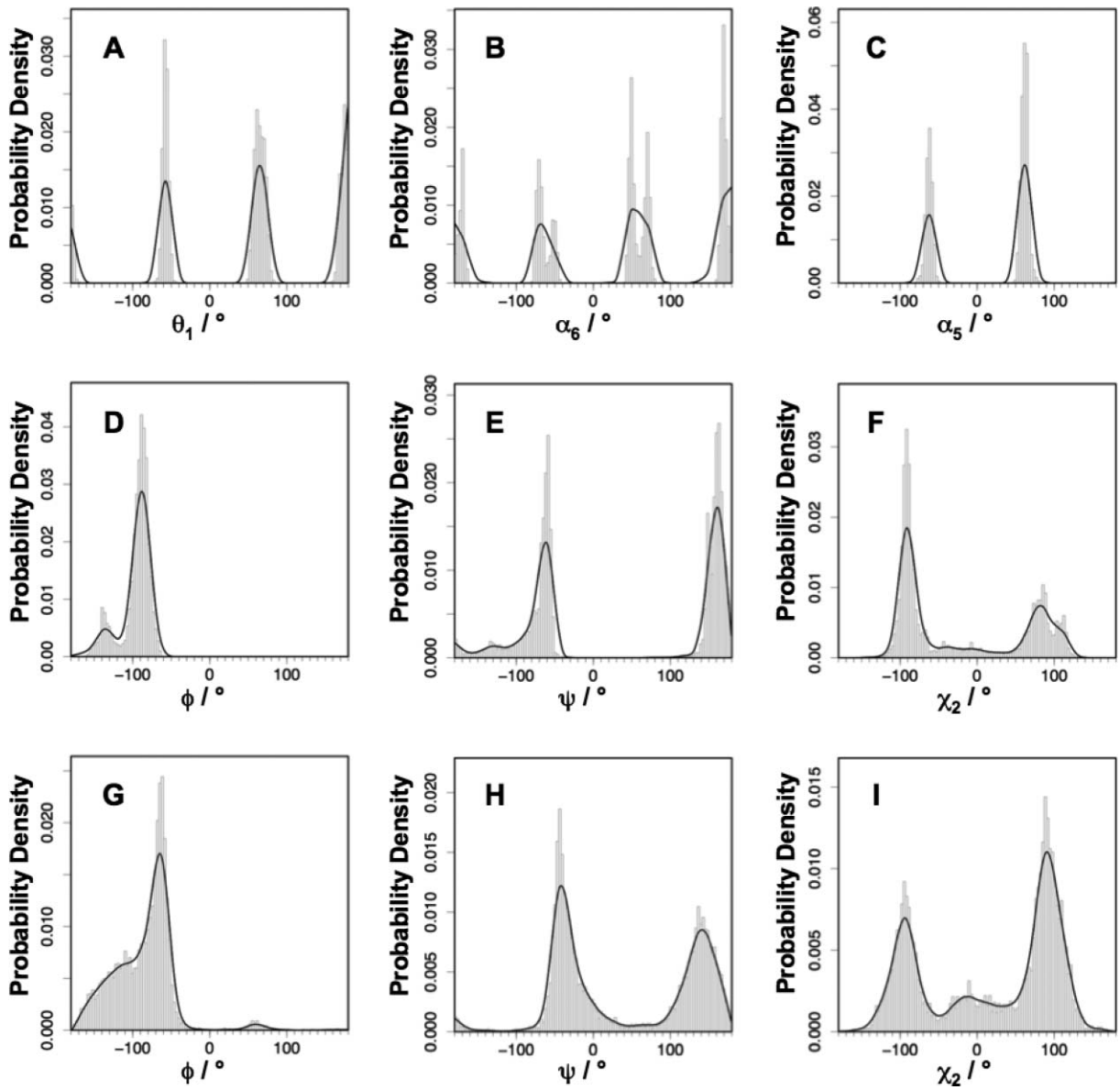

Fig. 2 Probability distributions for selected torsions from the model dataset and the PDB. A: θ_1; B: α_6; C: α_5; D: φ (combined data for 108–83–84–85 and 26–1–2–3); E: ψ (combined data for 83–84–85–86 and 1–2–3–4); F: χ_2 (combined data for 84–89–90–91 and 2–7–8–9); G: φ (all Trp in PDBSelect25); H: ψ (all Trp in PDBSelect25); I: χ_2 (all Trp in PDBSelect25). Histograms from which the probability densities were calculated are shown in grey.

in protein-bound lipids found in the PDB.[19] In contrast with the latter however, for α_5 and α_3, *ap* conformers are observed relatively infrequently.

Conformational analysis of 2 : 1 complexes

Having characterised complexes in which one of the tryptophan residues was more closely bound to the phosphocholine than the other, the remainder of the complexes that involved close binding contributions from both residues were considered. Hierarchical agglomerative clustering approaches led to the identification of a large number of bound structures, as anticipated. Nevertheless, it was possible to identify 12 structures occurring with a relative frequency $\geq$0.02, covering 34% of the structures in the dataset (Table 3). The same preferences for tryptophan and PC conformations observed in the 1 : 1 dataset are observed here, although it is notable that there is only one entry in which both tryptophans have the same conformation (α *ap* −*c*; entry **26**).

Conclusions

Tryptophan–lipid binding conformations

1 : 1 Binding of β-structures. The intention of this work was to investigate the conformational properties of tryptophan and phosphocholine molecules that delineate the binding observed in previous experiments and identify the interactions that contribute to the formation of stable adducts. Inspection of the data in Table 1 and 2 reveals that close association between a single tryptophan residue and a PC molecule in these models includes contributions

Table 1 Classification and angle distribution data for tryptophan residues in the model, PDBselect25 and membrane protein datasets

		Relative frequency[b]			Dihedral angle (°)[c]			
ID	Classification[a]	Models	PDB25	Memb.	φ	ψ	χ_1	χ_2
1a	β *−sc −c*	0.14	0.02	0.03	−135 ± 13	167 ± 17	−69 ± 37	−85 ± 16
1b	β *−sc −c*	0.17	0.02	<0.01	−94 ± 16	160 ± 21	−68 ± 32	−91 ± 13
2	β *−sc −s*	0.04	<0.01	<0.01	−104 ± 27	164 ± 17	−65 ± 24	−25 ± 30
3a	β *−sc +c*	0.12	0.16	0.10	−98 ± 29	158 ± 18	−64 ± 14	95 ± 42
3b	β *−sc +c*	—	0.04	0.02	−100 ± 33	1 ± 41	−65 ± 29	93 ± −37
4	β *−sc sp*	—	0.05	<0.01	−91 ± 46	140 ± 38	−69 ± 23	−5 ± 38
5a	β *+sc −c*	0.09	0.04	0.03	−94 ± 26	159 ± 15	51 ± 7	−89 ± 11
5b	β *+sc −c*	—	0.07	<0.01	−76 ± 71	−27 ± 67	62 ± 31	−89 ± 31
6	β *+sc +c*	—	0.03	0.03	−152 ± 22	161 ± 21	64 ± 20	89 ± 26
7	β *ap −c*	—	0.07	0.04	−86 ± 53	126 ± 38	−176 ± 31	−102 ± 37
8	β *ap +c*	<0.01	0.03	0.03	−78 ± 55	131 ± 37	−167 ± 28	77 ± 45
9	β *ap +s*	<0.01	0.02	0.03	−127 ± 41	114 ±31	−177 ± 17	24 ± 30
10a	α *−sc −c*	0.03	0.02	—	−80 ± 13	−60 ± 11	−61 ± 10	−89 ± 14
10b	α *−sc −c*	0.03	—	—	−96 ± 16	−91 ± 21	−70 ± 8	−92 ± 11
10c	α *−sc −c*	0.02	—	—	−129 ± 30	−138 ± 57	−64 ± 11	−80 ± 30
11	α *−sc +c*	0.09	0.09	0.22	−87 ± 24	−64 ± 18	−63 ± 34	90 ± 45
12	α *−sc −s*	<0.01	0.03	0.06	−65 ± 20	−38 ± 21	−67 ± 24	−24 ± 26
13	α *ap −c*	0.10	0.07	0.12	−93 ± 31	−68 ± 39	−170 ± 9	−96 ± 30
14	α *ap +c*	0.10	0.12	0.16	−88 ± 16	−62 ± 30	−168 ± 12	78 ± 29
15	α *+sc −c*	<0.01	0.07	0.05	−68 ± 47	−28 ± 26	64 ± 25	−87 ± 20
16	α *+sc +c*	—	<0.02	0.02	−75 ± 28	−15 ± 29	59 ± 30	84 ± 37

[a] For details of classification, see main text. [b] Data are only shown for instances where a particular conformation was present with a relative frequency ≥0.02 in any of the datasets. [c] Dihedral angle data are reported from the model dataset for conformations with a probability density ≥0.02, or from the PDBselect25 dataset otherwise.

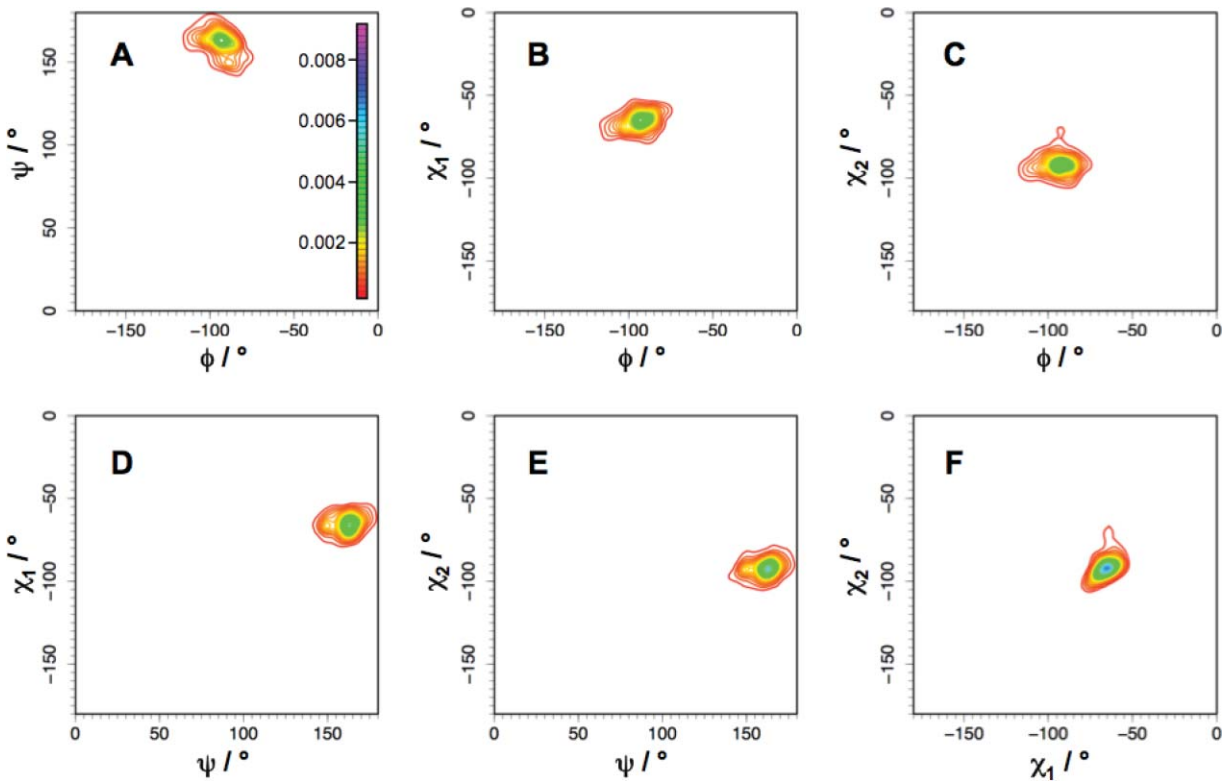

Fig. 3 Contour plots showing the probability distributions of φ, ψ, χ_1 and χ_2 torsions for the cluster with the highest relative frequency (entry **1a**, Table 1). The inset in A shows the colour scale used for the *z*-axis. Contour plots for pairwise combinations of dihedrals were generated by a two-dimensional kernel density estimation with 181 grid points in each direction and a bandwidth of 10°.

Table 2 Classification, distance and angle distribution data for phosphocholine molecules in the clustered datasets†

ID[a]			Dihedral angle (°)[c]			Selected distances/Å[d]		
1°	2°	RF[b]	α_5	α_4	α_3	NH(21)–O(x)	Ar–choline(x)	O(x)–choline
1a	**A**	0.10	69 ± 92	−167 ± 50	−83 ± 80	2.73 ± 1.00 (60)	4.30 ± 1.57 (Me)	2.58 ± 1.13 (27)
	B	0.02	63 ± 7	92 ± 93	66 ± 79	2.72 ± 1.14 (59)	4.14 ± 1.47 (52)	2.47 ± 0.45 (27)
	C	0.02	−60 ± 9	152 ± 35	66 ± 55	2.81 ± 1.26 (59)	4.44 ± 1.78 (54)	3.07 ± 1.56 (27)
1b	**A**	0.07	64 ± 66	−166 ± 60	−70 ± 54	3.13 ± 1.04 (60)	4.16 ± 2.34 (Me)	2.73 ± 1.27 (27)
	B	0.05	−64 ± 8	139 ± 26	63 ± 44	2.83 ± 1.31 (59)	4.14 ± 1.93 (Me)	2.75 ± 1.30 (27)
	C	0.03	64 ± 9	76 ± 35	65 ± 99	3.00 ± 1.36 (59)	3.65 ± 2.25 (Me)	2.42 ± 0.66 (27)
	D	0.02	−63 ± 10	−107 ± 83	−90 ± 139	3.28 ± 1.94 (59)	4.57 ± 2.92 (Me)	2.49 ± 0.79 (27)
2	**A**	0.02	93 ± 162	−143 ± 35	−80 ± 149	3.24 ± 1.42 (59)	4.20 ± 3.26 (Me)	5.58 ± 4.57 (27)
3a	**A**	0.05	65 ± 28	−143 ± 29	−60 ± 37	3.48 ± 1.60 (59)	3.57 ± 1.32 (Me)	5.10 ± 5.61 (27)
	B	0.04	−63 ± 26	−80 ± 41	−51 ± 35	2.72 ± 0.66 (59)	3.13 ± 0.45 (Me)	2.43 ± 0.76 (27)
5a	**A**	0.05	72 ± 98	−154 ± 36	−58 ± 26	2.84 ± 1.58 (57)	4.58 ± 0.74 (52)	2.37 ± 0.19 (5)
	B	0.03	−98 ± 170	149 ± 54	60 ± 140	2.89 ± 0.75 (59)	4.56 ± 2.33 (54)	3.17 ± 2.52 (5)
11	**A**	0.05	65 ± 47	−153 ± 53	−69 ± 97	3.60 ± 1.65 (59)	4.15 ± 2.09 (51)	5.07 ± 5.09 (27)
13	**A**	0.06	61 ± 10	−165 ± 46	−63 ± 38	3.66 ± 1.67 (59)	4.43 ± 2.12 (52)	2.45 ± 0.28 (27)
	B	0.03	65 ± 6	75 ± 18	52 ± 65	3.41 ± 1.49 (59)	4.26 ± 1.97 (Me)	2.84 ± 1.52 (27)
14	**A**	0.05	−64 ± 33	−178 ± 65	77 ± 119	2.83 ± 1.58 (59)	6.36 ± 1.28 (52)	2.45 ± 0.47 (27)
	B	0.04	68 ± 58	−139 ± 30	−53 ± 43	2.84 ± 1.00 (59)	3.95 ± 2.72 (55)	6.14 ± 2.58 (5)

[a] For details of classification, see main text. Data are only shown for instances where the probability density was ≥0.02. The 1° ID corresponds to the cluster ID in Table 1. Individual clusters are referred to as **1aA**, **1aB**, *etc*. [b] RF = relative frequency. [c] For dihedral angle definitions, see main text. [d] Values in parentheses indicate the atom number to which the distance was measured. Where multiple choices were available, 'x' represents the atom number that gave the shortest measured distance. 'Me' represents a hydrogen atom on one of the choline methyl groups. Distances were measured to the centroid of the 6-membered (Ar) ring of the indole.

Table 3 Classification of structures for which both tryptophans have significant close interactions with the PC molecule[a]

		Torsional classification				
ID	RF	Trp1	Trp2	α_5	α_4	α_3
17	0.05	β *−sc −c*	α *ap −c*	*+sc*	*−a*	*−sc*
18	0.02	β *−sc −s*	α *−sc +c*	*+sc*	*−ac*	*−sc*
19	0.05	β *+sc −c*	β *−sc −c*	*−sc*	*+ac*	*+sc*
20	0.02	β *+sc −c*	β *−sc −c*	*−sc*	*c*	*−sc*
21	0.04	β *+sc −c*	α *−sc −c*	*−sc*	*ap*	*+sc*
22	0.03	α *−sc +c*	β *−sc −c*	*+sc*	*ap*	*−sc*
23	0.02	α *−sc +c*	β *−sc +c*	*−sc*	*−c*	*sp*
24	0.02	α *+sc −c*	α *−sc −c*	*−sc*	*+a*	*+sc*
25	0.02	α *ap −sc*	α *−sc +c*	*+sc*	*ap*	*−ac*
26	0.03	α *ap −c*	α *ap −c*	*+sc*	*−a*	*−sc*
27	0.02	α *ap −c*	β *−sc −c*	*+sc*	*+a*	*−sc*
28	0.02	α *ap −c*	β *−sc sp*	*+sc*	*−a*	*−sc*

[a] Data are only shown for instances where the probability density was ≥0.02.

from hydrogen bonding between the indole NH atom as donor, usually to a phosphate oxygen atom as acceptor, and cation–carbonyl interactions between an amide carbonyl group and the choline ammonium group. These interactions occur readily when the indole N–H and amide C=O bonds are approximately aligned, as is the case in the β *−sc −c*, β *+sc −c*, α *−sc −c* and α *−ap +c* conformations (Fig. 4). These tryptophan conformations allow good interactions with both the ammonium and phosphate groups of the phosphocholine when the latter is in an extended conformation. This explains why the more compact headgroup conformations present in membrane lipids, with the ammonium and phosphate groups forming a salt bridge, are present less frequently in these models, reflecting a similar trend for lipid molecules found in the PDB.[19] Of the 1 : 1 structures, the most prevalent (entry **1aA**, Table 2; Fig. 5A) typifies this combination of amino acid and phosphocholine parameters, with a β *−sc −c* conformation for the former and an $\alpha_5/\alpha_4/\alpha_3$ combination of *+sc/ap/−sc* for the latter. Cation–π interactions in entry **1aA** are not present; indeed, they are generally observed only when the amino acid torsions place the indole–NH and both of the amide carbonyl groups at less than optimal angles for interactions with the phosphocholine. For example, increasing the cation–π overlap observed in entry **1aA** requires concomitant adjustments in the φ backbone torsion in order to minimise steric clashing between the carbonyl of the *N*-acetyl group and the choline ammonium group, producing the structure observed for entry **1bA** (Table 2; Fig. 5B); this is the reason for the separation of the β *−sc −c* conformation in to 2 subsets (entries **1a** and **1b**, Table 1). Conformations in which the amide C=O and indole N–H are not aligned allow binding of the phosphocholine in a more compact conformation, as found in entries **3aA** and **3aB** (Table 2; Fig. 5C), in which the phosphocholine adopts *+sc/−ac/−sc* and *−sc/−sc/−sc* $\alpha_5/\alpha_4/\alpha_3$ conformations respectively. In these cases, there is extensive cation–π overlap, again at the expense of interactions between the *N*-acetyl carbonyl group and the choline ammonium group, with only one of the carbonyl lone pairs suitably oriented for interaction. With a β *+sc −c* tryptophan conformation, cation–carbonyl interactions become possible with the α-carbonyl group, which is now aligned with the indole N–H bond. In these conformations, the phosphocholine is able to bind in an extended configuration, allowing hydrogen bonding interactions between the indole NH group and a phosphate oxygen (entry **5aA**, Table 2; Fig. 5D). In contrast with the β *−sc* conformers, which all interact *via* the *N*-acetyl group, some cation–π overlap is able to occur, although the extent of this overlap is limited by partial occlusion of the 5-membered ring by the backbone amide groups and the conformational restrictions required for the hydrogen bonding interactions of the indole NH group.

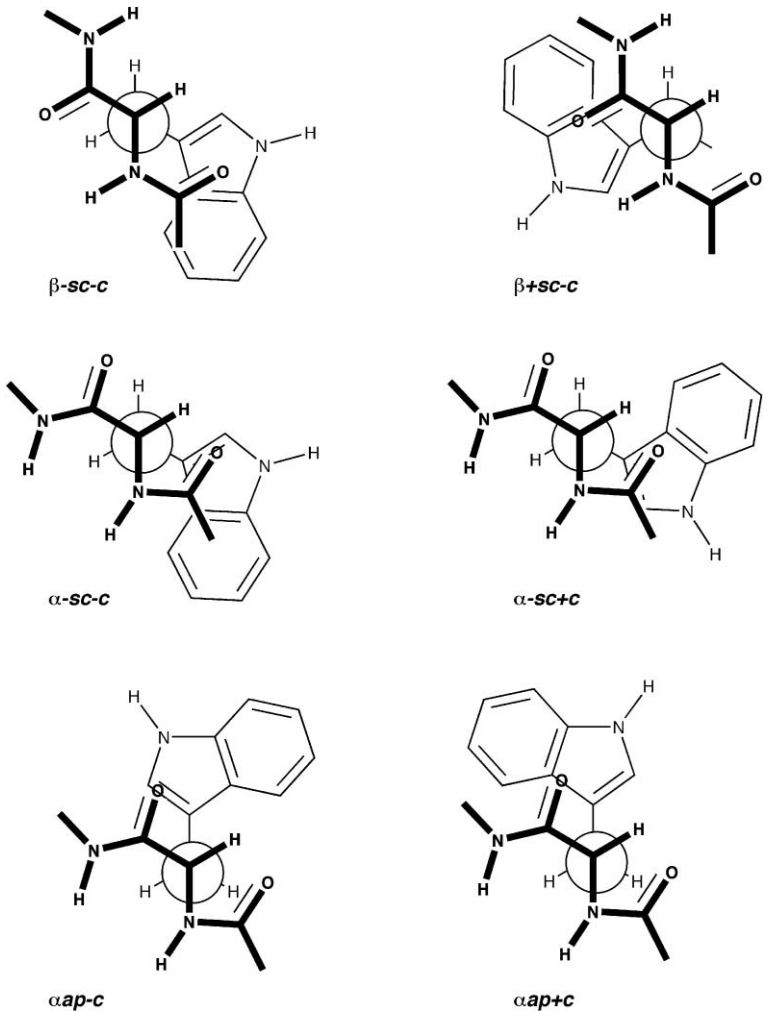

Fig. 4 Newman projections showing schematic representations of binding conformations. Each structure is viewed along the α–β bond (χ_1).

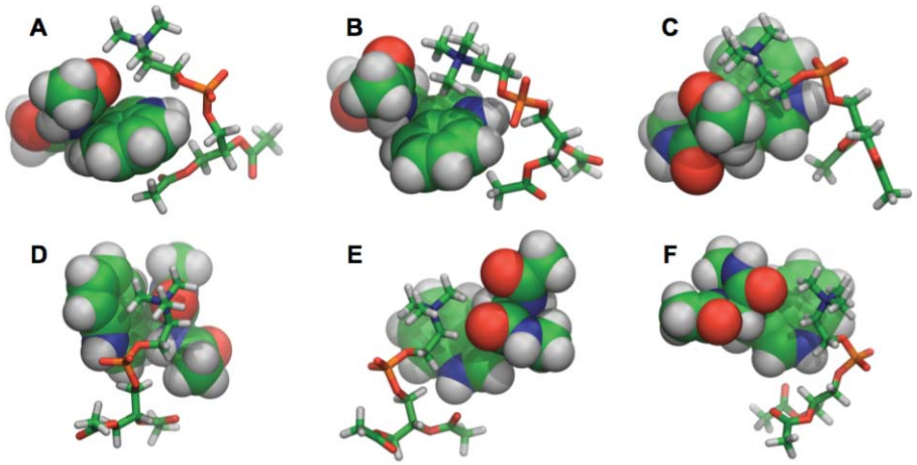

Fig. 5 Structures for selected entries in Table 2. In each case, the median structure is displayed. A, **1aA**; B, **1bA**; C, **3aB**; D, **5aA**; E, **13A**; F, **14A**.

1 : 1 Binding of α-structures. The binding characteristics observed for β-structures are also observed for helical structures. In the α *ap* −*c* conformation for example, the backbone carbonyls and the indole NH are not aligned. As a consequence, this conformation is able to exhibit significant overlap of the cationic parts of the choline group with the aromatic ring and cation–carbonyl interactions, in this case with both of the carbonyl groups. These interactions are detrimental to the hydrogen bonding interactions of the indole NH group to the extent that this now interacts with the carbonyl group of the acyl chain as acceptor

(Fig. 5E). Switching χ_2 to the $+c$ conformation aligns the α-carbonyl and indole NH groups, permitting hydrogen bonding interactions with the phosphate group at the expense of one of the cation–carbonyl interactions (Fig. 5F).

2 : 1 Binding. Tryptophan–phosphocholine association in these adducts is generally characterised by the inclusion of all of the binding interactions described above. All of the complexes display some degree of overlap of the cationic parts of the choline group with one of the indole rings, hydrogen bonding interactions between one or both of the indole NH groups and the phosphate oxygens and multiple interactions between the amide carbonyl groups and the choline ammonium group (Fig. 6). Some tryptophan conformations occur that are present in the PDBselect25 dataset, but not present in the tight-binding modelling subset, such as β *−sc sp* in entry **26** (Table 3), but otherwise the predominant conformations are similar. For most of these structures, the conformational restrictions required for tight binding to a single tryptophan are relaxed, as hydrogen bonding, cation–carbonyl and cation–π interactions are able to occur with different tryptophan molecules. All of the structures in Fig. 6 display this tendency to some extent. Nevertheless, in some cases, adducts are formed in which there is little cation–π overlap, such as entry **17** (Table 3; Fig. 6E). This may be due in part to 2° interactions—in this case hydrogen bonding between the NH of the α-amido group and the phosphate, the former of which is in the *cis* configuration. Some of the structures exhibit Trp–Trp interactions through backbone hydrogen bonding that are also likely to contribute to adduct formation; *cis*-amide conformations occasionally occur in these adducts, particularly where they are able to permit closer intermolecular contacts.

Significance with respect to peptide–lipid interactions

Phosphocholine binding conformations. The model dataset allows for analysis of feasible binding conformations between tryptophan and a phosphocholine in a non-competing solvent. Considered as a whole, the structures are consistent with the complexation-induced chemical shift changes seen during NMR titrations and cross-peaks seen in 2D rOesy experiments. Three main types of non-covalent interaction contribute to the formation of the observed adducts: indole–phosphate hydrogen bonds, carbonyl–cation interactions between backbone amides and the choline group and cation–π interactions between the choline group and the indole system. Considering the prevalence of each of these types of interaction in the model dataset, it seems their relative importance runs in the order indole–phosphate hydrogen bonding > carbonyl–cation > cation–π. Although the distribution of conformations in the model dataset is different to that in the PDBSelect25 and membrane datasets, all of the conformations adopted may be considered as allowed, as all occur to some degree in the latter datasets. In enthalpic terms, all of the complexes exhibit similar intermolecular interaction energies (in the range −50 to −70 kJ mol^{-1}), although there is a tendency for these potential energy terms to be more favourable for helical conformations. Although the effects are likely to be small,[22] they nevertheless potentially contribute to the observed differences in the association constants for Trp–PC complex formation (19 ± 2 M^{-1} for K_1 and 179 ± 21 M^{-1} for K_2). Furthermore, it has been suggested from *ab initio* calculations that polarisation effects contribute significantly to the interaction energy for cation–π interactions, being responsible for more than 50% of the binding energy in favourable cases (with electrostatic interactions responsible for the remainder).[23] The effects in these examples are further complicated by the tendency (for steric reasons) of cation–π interactions to occur with the 6-membered phenyl ring instead of the pyrrole ring, which is the usual site of interaction in globular proteins. There is, therefore, some scope for obtaining data that enable optimised polarisation effects to be included in future calculations. The lack of accounting for inductive and polarisation effects with the CHARMM27 force field is a limitation in this work that may influence the distributions of conformers observed.

Time-resolved fluorescence studies on model peptides have demonstrated that, at least in some cases, all of the three available χ_1 tryptophan rotamers are sampled in random coil

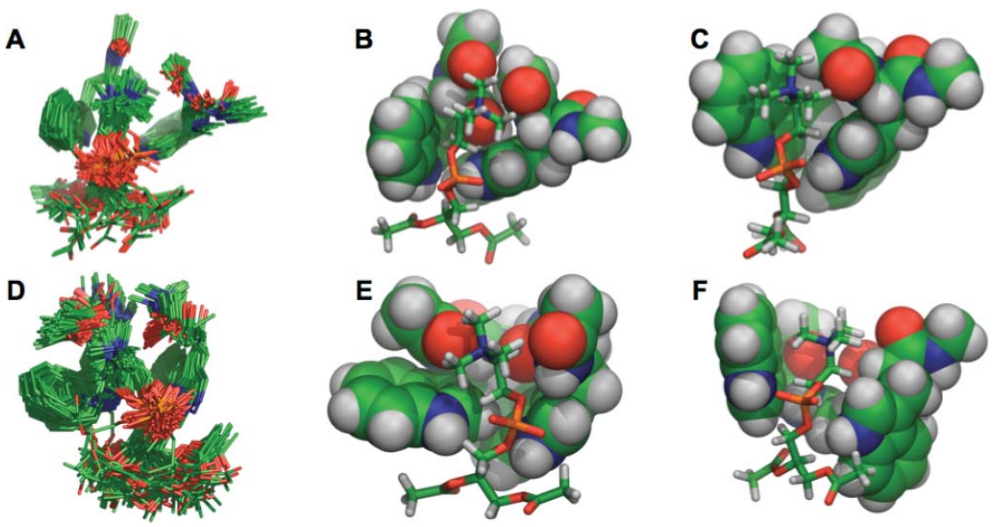

Fig. 6 Structures for selected entries in Table 3. All are representative structures unless otherwise stated A, 15 (all structures); B, **15**; C, **16**; D, **17** (all structures); E, **17**; F, **19**.

conformations in solution; in the same studies, peptide binding to membranes was characterised by the adoption of helical conformations and a reduction of the number of χ_1 rotamers sampled to two, reflecting the χ_1 distributions observed for ideal α-helices.[24] These were assigned as the *ap* and *−sc* conformers on this basis, which reflects exactly the trend seen in our model dataset when all helical conformations are considered. As the model dataset was generated without any restraints with regard to helicity and without the packing restraints of an ideal α-helix, this would suggest that the α-conformations observed match those determined experimentally during peripheral binding of helical peptides. The predominant conformations however, are those in which indole–phosphate and carbonyl–cation interactions are able to occur optimally, *i.e.* those in which amide carbonyl and indole NH bonds are aligned, which favours binding to a phosphocholine in an extended conformation. The torsional angle distributions for the choline headgroup are consistent with those observed experimentally for lipids, although there is a tendency for extended headgroup conformations to occur more frequently than observed for membrane lipids. The significance of this with respect to membrane binding is a likely disruption to lipid packing within one leaflet of the membrane. It is interesting to speculate that this might be sufficient to expose hydrophobic regions of the membrane to facilitate further insertion, or facilitate the release of membrane contents.

Peptide–lipid binding. There is a general consensus that for integral (intrinsic) membrane proteins, tryptophan has a distinctive role in anchoring the protein within the lipid bilayer. This may be in part a consequence of the amphipathic nature of indole, allowing hydrogen bonding interactions of the polar NH group and concomitant partitioning of the 6-membered aromatic ring into more hydrophobic regions of the fatty acyl chains. In addition, a favourable alignment of the indole dipole with the electrostatic field of the membrane may contribute.[25] As a consequence, the indole tends to be oriented with the NH group directed towards the exterior surface of the bilayer. The situation for peripheral (extrinsic) proteins and peptides is different. The role of tryptophan in the membrane binding properties of a number of proteins and peptides has been addressed, including examples that remain bound peripherally, undergo insertion, or are translocated across the bilayer. In general, binding to the membrane is favoured by complementary electrostatic charge profiles between the peptide or protein and the membrane, with the former tending to be positively charged and the latter negatively charged. Binding is further dependent on the hydrophobicity of the protein, with most being appreciably hydrophobic or amphipathic in nature. However, in all of these cases, examples exist where the presence of tryptophan is essential for activity.

Even for peptides in which membrane binding is driven largely by charge complementarity, tryptophan residues still impact upon membrane behaviour. For example, although partitioning of penetratins into the membrane is promoted principally by charge complementarity,[26] substitution of the N-terminal tryptophan in the peptide RQIKIWFQNRRMKWKK by phenylalanine disrupts translocation without affecting partitioning,[27] suggesting that factors other than simple charge matching between the membrane and the peptide contribute to the process. Similar cases exist for peripheral membrane proteins, such as annexin V, in which mutation of a single tryptophan to alanine removes the ability of the protein to bind to mixed PC/PS membranes.[28] As there is no evidence for the involvement of this surface-exposed tryptophan in direct interactions with PS molecules, the explanation for these observations presumably lies in the subtle thermodynamics of the membrane association process. In contrast, mutation of a valine residue on the membrane-binding surface of human phospholipase A_2 to tryptophan increased the activity of this enzyme towards PC membranes by a factor of 100.[29] This enhanced activity was attributed to enhanced interfacial binding. Interestingly, membrane binding was reduced in high salt buffers, suggesting the involvement of electrostatic factors in the binding interaction.

All of the above observations implicate tryptophan as having a significant role in controlling the behaviour of peptides and proteins following peripheral interaction with membranes. In most cases, although structures of membrane-active peptides may be characterised in solution or following interaction with membranes or micelles, the role of critical tryptophan residues during the initial interactions with the bilayer or during the insertion process itself are largely uncharacterised. In contrast to the hydrogen bonding interactions with acyl carbonyl groups characterised for integral proteins, hydrogen bonding interactions in the model systems occur with phosphate oxygen atoms as acceptor, with 40% of the examples exhibiting an H–O contact distance of less than 3.0 Å. This is not entirely unexpected, as this is a particularly favourable donor–acceptor pairing.[30] Furthermore, in contrast to integral proteins, entropic and dipolar effects that favour the partitioning of the 6-membered aromatic ring into the lower dielectric regions of the bilayer are not present in the model system. This also holds true for peripheral binding to the membrane, so it is not unreasonable *a priori* that phosphate–indole hydrogen bonding could occur in these systems. Another common feature of the models is the presence of carbonyl–cation interactions, with 52% of the structures having O–choline contact distances of less than 3 Å. In the majority of cases, the availability of backbone carbonyl groups in integral proteins is expected to be limited due to their participation in the formation of 2° and 3° structure. However, this is not the case for peptides, where higher conformational flexibility or the presence of unusual folds and loops stabilised by disulfide bridges frequently occur. A number of membrane-active peptides have been characterised in which the presence of several tryptophan residues, frequently clustered in a localised structural fold, are essential for membrane binding and insertion.[31,32] Two related peptides from wheat endosperm, puroindoline-a (PIN-a) and puroindoline-b (PIN-b) exhibit strong binding to phosphocholine membranes, with reported dissociation constants of the order of 10^{-7} and 10^{-8} M, respectively.[33] PIN-a has been demonstrated to form cation-selective channels, both in *Xenopus* oocytes and planar lipid bilayers.[34] Both peptides contain tryptophan-rich sequences that are implicated in membrane activity (WRWWKWWK in the case of PIN-a and WPTWWK in the case of PIN-b). Despite being cationic peptides, both display higher binding affinity to neutral phospholipid membranes than negatively-charged membranes containing wheat glycolipids, which contrasts with the typical behaviour of less tryptophan-rich peptides. Carbonyl–cation interactions may therefore be significant in more rigid peptides and a transient feature of more dynamic ones. The third non-covalent feature seen in many of the

models, cation–π interaction, tends to occur wherever geometric features allow its formation, although for many of the structures, a compromise is reached between carbonyl–cation and cation–π interactions. Accordingly, a modest 35% of the models have close contacts of less than 4.2 Å between the choline group and the centroid of the indole aromatic ring. These interactions have been well described for membrane proteins, and for some synthetic membrane receptors,[35] but the extent to which they might contribute to the binding interactions of peptides is difficult to ascertain with certainty, although their presence has been speculated from studies on the binding of indole derivatives to membranes.[25,36,37]

There is no doubt that charge neutralisation contributes significantly to the enthalpy of binding of cationic peptides to negatively charged membranes, as there are several examples where binding is significantly diminished or removed completely by disruption of the charge complementarity. In many cases, substitution of tryptophan in a cationic peptide does not prevent membrane binding, although the behaviour following binding may be modified. These cases serve to illustrate that electrostatic interactions between tryptophan and phosphocholines are generally insufficient to produce strong binding with membranes formed from mixtures of anionic and neutral lipids, but are able to exert subtle effects on processes subsequent to binding. The exception to this occurs when several tryptophan residues are clustered together in a compact region of the peptide; in these cases, binding of a cationic peptide to a neutral membrane may actually be even more pronounced than that to a negatively charged membrane. From our previous experimental data, we can estimate that the maximum free energy available for 1 : 1 and 2 : 1 binding between tryptophan and a phosphocholine molecule are of the order of −7 and −20 kJ mol^{-1}, respectively. From the modelling experiments, the total available enthalpic contribution to binding is of a similar order of magnitude (data not shown), suggesting that entropic factors impact less on the free energy of binding in this case. For binding of a peptide to a membrane, the available free energy will be offset by desolvation effects and enthalpic and entropic factors relating to membrane destabilisation, making the net free energy benefit from tryptophan–phosphocholine binding considerably less. The major difficulty in assessing the enthalpic contributions of individual amino acids to binding is accounting for contributions from other processes, such as the formation of 2° structural elements following during binding; in this regard, position-dependent effects are frequently observed, with the substitution of a single isoleucine for tryptophan in the synthetic peptide $(KIGAKI)_3$–NH_2 being a good example.[38] In a similar vein, the binding of the peptide Ac-WL_5 to neutral membranes has zero enthalpy at 25 °C, whereas the binding of the related peptide Ac-WLWLL is exothermic. This may reflect higher contributions from a 2 : 1 Trp–PC binding, but it supports the idea that clusters of tryptophan molecules promote binding to neutral membranes in peptides and illustrates the complexity of analysing the thermodynamics of binding in peptide systems.[39] Nevertheless, it is feasible that lipid binding by tryptophan is favourable on purely enthalpic grounds. It is notable that the binding of homologues of the antimicrobial peptide magainin to neutral membranes has been shown to be enthalpically driven. The binding of a tryptophan-containing homologue was enthalpically more favourable by 10–15 kJ mol^{-1} at 30 °C than homologues without tryptophan.[40] These differences were attributed to a non-classical hydrophobic effect. Although other substitutions were made in these peptides, making the contribution of tryptophan difficult to ascertain, the available free energy from tryptophan–phosphocholine binding could at least account for part of this difference. Studies on the binding of LamB-W, with a nominal net charge of +2.5 (with 0.5 of a charge unit from the N-terminus), to both negatively charged and neutral membranes indicated that the enthalpic contribution to binding was 7.6 kJ mol^{-1} greater in the case of negatively charged membranes, or 3 kJ mol^{-1} per charge unit.[41,42] A lower value of ∼1.2 kJ mol^{-1} per charge unit was found for the binding of peptide PGLa to negatively-charged and neutral membranes at 30 °C.[43] However, the latter study considered the effects of local surface concentrations in the calculation of binding isotherms, leading to a conclusion that the difference in binding of cationic peptides to negatively charged and neutral membranes may be largely due to localised concentration effects on the surface of the membrane. This is consistent with the higher enthalpic contribution per unit charge observed for the binding of LamB-W to neutral membranes. If the contribution to binding from charge matching is indeed unrelated to enthalpic effects, the enthalpic contributions of tryptophan–phosphocholine interactions may be more significant than expected.

References

1 J. M. Sanderson, *Org. Biomol. Chem.*, 2005, **3**, 201.
2 C. Landolt-Marticorena, K. A. Williams, C. M. Deber and R. A. Reithmeier, *J. Mol. Biol.*, 1993, **229**, 602.
3 M. B. Ulmschneider and M. S. Sansom, *Biochim. Biophys. Acta*, 2001, **1512**, 1.
4 S. Ozdirekcan, D. T. S. Rijkers, R. M. J. Liskamp and J. A. Killian, *Biochemistry*, 2005, **44**, 1004.
5 J. M. Sanderson and E. J. Whelan, *Phys. Chem. Chem. Phys.*, 2004, **6**, 1012.
6 R. V. Pappu, R. K. Hart and J. W. Ponder, *J. Phys. Chem. B*, 1998, **102**, 9725.
7 B. R. Brooks, R. E. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathan and M. Karplus, *J. Comput. Chem.*, 1983, **4**, 187.
8 N. S. Isaacs, *Physical Organic Chemistry*, Harlow: Longman, 2nd edn, 1995.
9 W. C. Davidon, *Math. Program.*, 1975, **9**, 1.
10 H. J. C. Berendsen, J. P. M. Postma, W. F. Vangunsteren, A. Dinola and J. R. Haak, *J. Chem. Phys.*, 1984, **81**, 3684.
11 *R programming environment (version 2.2.1): A language and environment for statistical computing*, 2005, R Development Core Team.
12 A. Struyf, M. Hubert and P. J. Rousseeuw, *Comput. Stat. Data Anal.*, 1997, **26**, 17.
13 P. J. Rousseeuw, *J. Comput. Appl. Math.*, 1987, **20**, 53.
14 M. Nic, J. Jirat and B. Kosata, *IUPAC Compendium of Chemical Terminology*, electronic version: http://goldbook.iupac.org/T06406.html.
15 D. Marsh, *Protein Sci.*, 2003, **12**, 2109.
16 P. S. Shenkin and D. Q. McDonald, *J. Comput. Chem.*, 1994, **15**, 899.
17 U. Hobohm and C. Sander, *Protein Sci.*, 1994, **3**, 522.
18 S. H. White, *Membrane proteins of known 3D structure*: http://blanco.biomol.uci.edu/Membrane_Proteins_xtal.html.
19 D. Marsh and T. Pali, *Biochim. Biophys. Acta*, 2004, **1666**, 118.
20 M. J. Mcgregor, S. A. Islam and M. J. E. Sternberg, *J. Mol. Biol.*, 1987, **198**, 295.
21 H. Schrauber, F. Eisenhaber and P. Argos, *J. Mol. Biol.*, 1993, **230**, 592.
22 A. P. Bisson, C. A. Hunter, J. C. Morales and K. Young, *Chem.–Eur. J.*, 1998, **4**, 845.
23 E. Cubero, F. J. Luque and M. Orozco, *Proc. Natl. Acad. Sci. U. S. A.*, 1998, **95**, 5976.
24 A. H. Clayton and W. H. Sawyer, *Biophys. J.*, 1999, **76**, 3235.

25 H. C. Gaede, W.-M. Yau and K. Gawrisch, *J. Phys. Chem. B*, 2005, **109**, 13014.
26 B. Christiaens, S. Symoens, S. Verheyden, Y. Engelborghs, A. Joliot, A. Prochiantz, J. Vandekerckhove, M. Rosseneu and B. Vanloo, *Eur. J. Biochem.*, 2002, **269**, 2918.
27 G. Dom, C. Shaw-Jackson, C. Matis, O. Bouffioux, J. J. Picard, A. Prochiantz, M.-P. Mingeot-Leclercq, R. Brasseur and R. Rezsohazy, *Nucleic Acids Res.*, 2003, **31**, 556.
28 B. Campos, Y. D. Mo, T. R. Mealy, C. W. Li, M. A. Swairjo, C. Balch, J. F. Head, G. Retzinger, J. R. Dedman and B. A. Seaton, *Biochemistry*, 1998, **37**, 8004.
29 S. F. Baker, R. Othman and D. C. Wilton, *Biochemistry*, 1998, **37**, 13203.
30 C. A. Hunter, *Angew. Chem., Int. Ed.*, 2004, **43**, 5310.
31 D. J. Schibli, R. F. Epand, H. J. Vogel and R. M. Epand, *Biochem. Cell Biol.*, 2002, **80**, 667.
32 G. Fimland, V. G. H. Eijsink and J. Nissen-Meyer, *Biochemistry*, 2002, **41**, 9508.
33 L. Dubreil, J.-P. Compoint and D. Marion, *J. Agric. Food Chem.*, 1997, **45**, 108.
34 P. Charnet, G. Molle, D. Marion, M. Rousset and V. Lullien-Pellerin, *Biophys. J.*, 2003, **84**, 2416.
35 M. E. Weber, E. K. Elliott and G. W. Gokel, *Org. Biomol. Chem.*, 2006, **4**, 83.
36 K. E. Norman and H. Nymeyer, *Biophys. J.*, 2006, **91**, 2046.
37 F. N. R. Petersen, M. O. Jensen and C. H. Nielsen, *Biophys. J.*, 2005, **89**, 3985.
38 Y. Jin, H. Mozsolits, J. Hammer, E. Zmuda, F. Zhu, Y. Zhang, M. I. Aguilar and J. Blazyk, *Biochemistry*, 2003, **42**, 9395.
39 W. C. Wimley and S. H. White, *J. Mol. Biol.*, 2004, **342**, 703.
40 T. Wieprecht, M. Beyermann and J. Seelig, *Biochemistry*, 1999, **38**, 10377.
41 J. D. Jones and L. M. Gierasch, *Biophys. J.*, 1994, **67**, 1534.
42 J. D. Jones and L. M. Gierasch, *Biophys. J.*, 1994, **67**, 1546.
43 T. Wieprecht, O. Apostolov, M. Beyermann and J. Seelig, *Biochemistry*, 2000, **39**, 442.

