Peptide - lipid interactions analyzed via Tryptophan derivatives: experimental and computational results

Georg Blaser

Peptide – Lipid Interactions Analyzed *Via* Tryptophan Derivatives: Experimental and Computational Results

Georg Blaser

Department of Chemistry

Durham University

November 2007

Submitted in partial fulfilment of the requirements for the degree of Doctor of Philosophy.

Declaration

The work in this thesis is based on research carried out at Durham University. The material contained in this thesis has not previously been submitted for any other degree at Durham University or any other institution. The research within this thesis has been conducted by the author unless indicated otherwise.

Georg Blaser

November 2007

Abstract

Tryptophan analogues have been shown to play an important role in the binding and anchoring of peptides in lipid membranes. Mono-substituted tryptophan derivatives were synthesized to study the importance of the relative contribution of electrostatic, quadrupolar and induced dipolar interactions to binding. Binding constants for the adduct formation of amino acid derivatives, lipid molecules and water with themselves and each other were analysed on a molecular level both experimentally, using NMR host-guest titrations and isothermal titration calorimetry (ITC), and computationally *via* molecular dynamics simulations. Free-energy calculations support the data presented and allowed a quantitative comparison.

When free energies of the association of 5-monosubstituted tryptophan analogues (Figure 0.1) with DMPC lipid molecules were plotted against a measure of electron density of the indole ring (the Hammett parameter), an n-shaped pattern was observed. Binding was found to increase for substituted amino acid derivatives with maximal responses for the most electron withdrawing (5-nitro) and the most electron donating (5-methoxy) compounds. This trend was seen both in the titration-data and in the molecular dynamics simulations. The latter allowed the sampling of preferred binding conformations and a determination of the importance of hydrogen bonding and cation-π interactions for and in adduct formation.

Additionally, amino acid analogues were incorporated into a model peptide (AcWLWLL) to study the electrostatic effects of the substituent's on the peptide-lipid interactions in the biologically important lipid bilayer environment.

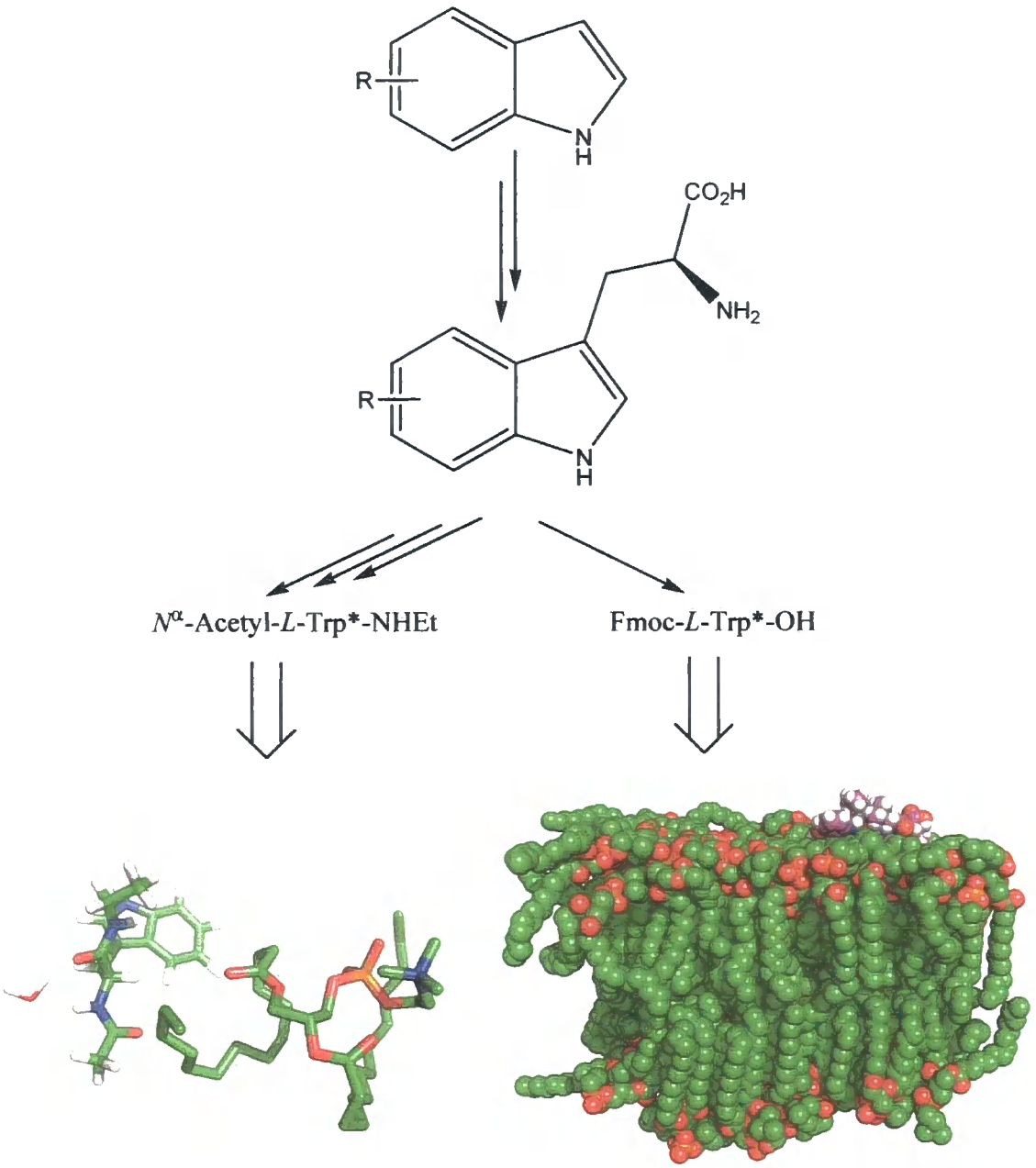


Figure 0.1: Schematic representation of the research described in this thesis. Trp* stands for 5-monosubstituted tryptophan compounds, with the substituent R being H, F, Cl, Br, I, Me, MeO or NO_2.

Acknowledgements

Thanks go to my academic supervisors Dr John Sanderson and Dr Mark Wilson, Dr Lars-Olof Pålsson and Dr Andy Beeby who helped me with fluorescence spectroscopy, and Dr Andrei Batsanov for x-ray crystallography. Special thanks go to Dr Elizabeth Grayson for her continuous support and the technical staff for their analytical services.

I am grateful for the financial sponsoring I have received from Durham University and the EPSRC Research Council.

The following persons have made my work possible and I would like to acknowledge them by name: My family, always supporting and strengthening me; Lucinda, who gave me inspiration and motivation particularly in my final year and days where it was required the most; my friends in England who persuaded me to pursue my study and interests with conviction and dedication, especially noteworthy are Bader Obeidat, Liz Ellis, Zubi Al'Zubi, and various others who came and left throughout the years. They have made my time in Durham memorable and exciting.

Thanks also to my German friend Hans, who has been a steadfast link to Germany.

Table of Contents

List of Figures

List of Tables

Abbreviations Used

$[\alpha]_d^{20}$:	optical rotation at 20 °C using sodium light
°C:	grad Celsius
AA:	amino acid
Ac:	acetyl (functional group CH_3CO-)
Ac_2O:	acetic anhydride
AcOH:	acetic acid
Boc:	t-butoxycarbonyl
calcd.:	calculated
CD:	circular dichroism
$CDCl_3$:	deuterated chloroform
c.f.:	Latin: *confer* – compare
$CHCl_3$:	chloroform
COSY:	correlation spectroscopy
D:	dextrorotatory enantiomer (with respect to glyceraldehyde)
d:	doublet
DAPC:	1,2-diacetyl-*sn*-3-phosphocholine
DCM:	dichloromethane
dd:	doublet of doublet
DIC:	diisopropylcarbodiimide
DIPEA:	*N*-*N*-diisopropylethylamine
DMF:	dimethylformamide
DMPC:	1,2-dimyristoyl-*sn*-glycero-3-phosphocholine
DOPC:	1,2-dioleoyl-*sn*-glycero-3-phosphocholine
DOPE:	1,2-dioleoyl-*sn*-glycero-3-phosphoethanolamine
DPPC:	1,2- palmitoyl-*sn*-glycero-3-phosphocholine
dq:	doublet of quartet
dt:	doublet of triplet
DTPC:	ditetradecylphosphatidylcholine
ε:	dielectric constant
e.g.:	Latin: *exampli gratia* – example given
EI:	electrospray ionisation
Et:	ethyl (functional group CH_3CH_2-)
EtOAc:	ethyl acetate
EtOH:	ethanol
Fmoc:	9-fluorenylmethyloxycarbonyl succinimide
FT-MS:	flight of time – mass spectrometry
FTIR:	Fourier transform infrared spectroscopy
GROMACS:	Groningen machine for chemical simulations
HCl:	hydrochloric acid
HOBt:	1-hydroxybenzotriazole hydrate
HPLC:	high pressure liquid chromatography
HRMS:	high-resolution mass spectrometry
HW:	hydrogen atom in the SPC water representation
i.e.:	Latin: *id est* – that is
ITC:	isothermal titration calorimetry
K:	binding constant
L:	levorotatory enantiomer (with respect to glyceraldehyde)

L, Leu:	*L*-leucine
LJ:	Lennard-Jones parameter
M:	molar (concentration in mol dm^{-3})
m:	multiplet
MD:	molecular dynamics
Me:	methyl (functional group CH_3-)
MeOH:	methanol
min:	minutes
m*M*:	millimolar
MolS:	molecular sieves
mp.:	melting point
MS:	mass spectrometry
nm:	nanometre
NMR:	nuclear magnetic resonance
NOESY:	nuclear Overhauser enhancement spectroscopy
NpT:	canonical ensemble with constant number of particles N, pressure p and temperature T
ns:	nanosecond
NVT:	canonical ensemble with constant number of particles N, volume V and temperature T
NγT:	canonical ensemble with constant number of particles N, surface tension γ and temperature T
OPLS-AA:	optimised parameters for liquid simulations – all atom
OW:	oxygen atom in the SPC water representation
PDB:	protein data bank (http://www.pdb.org)
PME:	particle mesh Ewald summation
POPC:	1-palmitoyl,2-oleoyl-*sn*-glycero-3-phosphocholine
PyBOP:	benzotriazol-1-yl-oxytripyrrolidinophosphonium hexafluorophosphate
R_f:	retention factor
RF:	reaction field
s:	singlet
SAS:	solvent accessible surface area
SPC:	single point charge (water representation)
t:	triplet
TFA:	trifluoracetic acid
TIP4P:	transferable intermolecular potential four point water model
TIS:	triisopropylsilane
TLC:	thin layer chromatography
TMS:	tetramethylsilane
Trp:	N^{α}-Acetyl-*L*-tryptophan ethyl amide derivate
W:	*L*-tryptophan
ΔG:	Gibbs free energy
µl:	microliter

Chapter 1: Introduction

1.1 Overview

This chapter provides the reader with an introduction to the area of peptide-lipid interactions (§1.2), their importance in biological systems (§1.2.2.a), and how they have been investigated in the past (§1.2.2.b). It also outlines how the research presented in this thesis aims to contribute to the field (§1.2.3). An introduction to and background information about the key concepts and models important in peptide-lipid research is described in the following sections. Membrane structure and properties and the underlying models that describe them are presented (§1.3), including information about the structure and properties of single lipid molecules, and in particular of the lipid used in this study (DMPC) and a short introduction to membrane proteins. Details, specific to the synthesis and literature preparation of tryptophan derivatives (Figure 1.1) are discussed in §1.4. Subsequently, the model peptide AcWLWLL and its characteristics are presented (§1.5). §1.6 of this chapter provides information about computational approaches to simulate peptide lipid interactions. A detailed breakdown of factors that are important for peptide lipid interactions and their importance for tryptophan derivative-lipid interactions (as viewed in the literature) are given thereafter (§1.7).

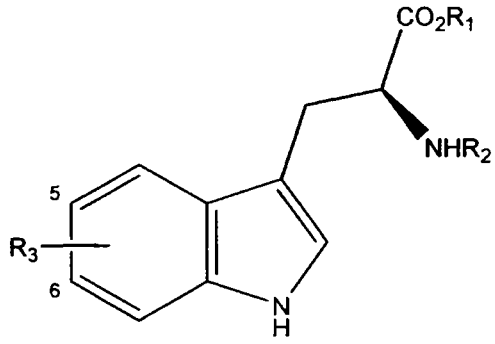


Figure 1.1: Common tryptophan derivatives. Tryptophan derivatives commonly carry substitutions on the carboxylic acid and the amine functional groups (R_1 and R_2 respectively) and can exhibit substitutions (R_3) on the indole ring, which is highlighted here. The tryptophan derivatives synthesised and studied for this thesis are: R_3 = F, Cl, Br, I, Me, MeO, or NO_2, on the indole 5 or 6 position of the indole ring, with R_1 = OH and R_2 = Fmoc, and R_1 = NHEt and R_2 = Ac respectively.

1.2 Introduction to Peptide-Lipid Interactions

1.2.1 Definition of Peptide-Lipid Interactions

Peptide-lipids interactions (and protein-lipid interactions respectively) have received much attention since proteins have been discovered to play a role in the structure and functioning of cell membranes in 1972.[1] The exact nature and strength of these interactions varies between proteins and lipids of different types but is generally influenced by electrostatic and non-polar (*e.g.* hydrophobic) effects. Each amino acid and lipid type may exhibit a different relative combination of these forces and may be influenced differently. For tryptophan residues, many favourable non-covalent interactions between the aromatic amino acid and lipids are feasible. Besides the amphiphilic character of the indole side chain and intrinsic polarity, there is evidence for favourable dipolar interactions between the aromatic amino acid side chain and lipids[2], and hydrogen bonding with the headgroups or carbonyl groups of the lipids or with water molecules in the headgroup region.[3] Some of the possible favourable interactions (divided into non-polar and polar forces) are discussed briefly here. More detail can be found in §1.7.

1.2.1.a Non-polar forces

Non-polar forces can arise from two effects:

The **hydrophobic effect** (§1.7.1.a), which favours exclusion of hydrophobic moieties (*e.g.* phenyl ring of the indole) from aqueous solvent. Tryptophan has the largest surface area of any natural amino acid, which will favour such an exclusion.

The **lipophobic effect**[4] (§1.7.1.b), describes the solvation of uncharged solutes by the bilayer lipids. It is the analogue of the hydrophobic effect. The geometry of tryptophan, being rigidly planar, may favour localization due to compatibility with the liquid-crystalline-like order of the lipid bilayer, which varies strongly with depth in the hydrocarbon core.

1.2.1.b Polar Forces

Polar forces can be subdivided into:

Hydrogen bonding (§1.7.2.a) between the tryptophan residue and lipids. An example of such hydrogen bonding includes the N1 proton of indole (Figure 1.1) and the lipid carbonyl groups located in the glycerol region of the lipid bilayer.

Cation-π interactions (§1.7.2.b) between the aromatic indole ring and positively charged choline groups in phosphatidylcholine lipids.

Electrostatic interactions (§1.7.2.c) include charge-charge interactions between proteins and bilayers and electrostatic interactions between the indole dipole moment and the electric field of the bilayer. Some lipids carry charged head groups attracting proteins and peptides that carry a charge of opposing sign by coulombic interactions. Additionally, phosphocholine bilayers normally have a strong positive electrostatic potential in the hydrocarbon core, which rapidly drops off in the interfacial region.[5] This strong electrostatic gradient ($2 < \varepsilon < 80$, where ε is the dielectric constant) could trap and hold molecules, such as indole, which have a large molecular dipole ($D \sim 1.7$, where D is the dipole moment).

Other nonspecific electrostatic interactions between indole and the lipid and water molecules of the interfacial region. For example, simulations indicate that the interfacial region possesses a large dielectric (*i.e.* 80)[6] but a lower surface tension than bulk water.[7] Molecules with a large surface area might prefer to move into the interfacial region to lower the free energy of nonpolar solvation, but they might be too polar to move into the low dielectric hydrocarbon core region. They are effectively being trapped in the interfacial region.

1.2.2 The Study of Peptide-Lipid Interactions

1.2.2.a Motivation

The study of peptide-lipid interactions is of great scientific value, as it provides a simplified model system to aid the understanding of the complex relationship between proteins, their folding, association, and integration with lipid membranes. Water-soluble peptides or proteins binding to or inserting into lipid bilayers profoundly alter the properties of the bilayer. Examples include the activation of some cytosolic enzymes upon post-translational insertion of proteins into membranes.[8,9] Other peptides act as antibiotics upon partition into a bacterial membrane.[10] Knowledge of the precise nature of peptide-lipid interactions will help to gain insights to molecular mechanisms such as the action of antibiotic peptides, the association of proteins used in cell signalling, and membrane fusion. This in turn will aid the development of new drugs (such as anti-inflammatories), drug delivery methodology (efficient crossing of the drug through the cell membrane or using carrier molecules/particle

for crossing), and potentially new cancer therapies (binding of proteins and drug molecules alters the properties of the bilayer and might lead to its disruption and subsequent apoptosis).[11]

1.2.2.b Techniques

Several groups have used different experimental and computational techniques to elucidate peptide-lipid interactions in detail. The experimental studies range from site-directed spin labelling[12], isothermal titration calorimetry (ITC)[13], infrared spectroscopy[14], fluorescence quenching[15], x-ray[16] and neutron scattering[17], to nuclear magnetic resonance (NMR) techniques, in the solid-state[18] and *via* host-guest titrations[19]. Other techniques that work well on water-soluble proteins (*e.g.* nuclear magnetic resonance (NMR) and crystallography) are more difficult to employ due to the hydrophobic nature of the proteins and their tendency not to crystallize well in the absence of lipids. Nevertheless, there are a number of protein structures published in the protein databank (PDB) that contain one or more bound lipids of varying type.

There have been numerous approaches to simulate peptide lipid interactions varying in detail and size of the analysed system as well as in technique. These include coarse-grained computational analysis of membrane proteins and peptides[20], molecular dynamics and stochastical approaches (for a recent review see Ref. 21).

1.2.3 Scope of this Thesis

This thesis aims to investigate the specific influence of tryptophan residues on peptide-lipid interactions. In particular, electrostatic interactions of the indole ring of the aromatic amino acid with lipids were analysed. This was achieved by synthesis of a homologous series of tryptophan derivatives with monosubstitutions on the 5 and 6-position of the indole ring, altering its electron density to access electron rich compounds (such as derivatives carrying a methoxy group) and electron poor compounds (such as derivatives carrying a nitro functionality) (see §2.2). The synthesis demonstrates an improved version of the original Snyder *et al.*[22] synthesis, building on the expertise from Yokoyama *et al.*[23] The tryptophan derivatives (N^{α}-acetyl-*L*-tryptophan ethyl amides) were designed to mimic the non-terminal tryptophan residue in proteins.

NMR host-guest titrations and isothermal titration calorimetry (ITC) were employed as experimental tools to determine binding constants of tryptophan-lipid adducts in the non-competitive solvent chloroform (see §2.3 and §2.4). Additionally, these interactions were modelled using molecular dynamics (MD) simulations helping to visualise and study interactions on a time-scale inaccessible by experiment and to gain insight into the binding conformation of key adducts (see §2.5). Furthermore, computations allowed the calculation of free energies of association by using thermodynamic cycles (see §2.6). Tryptophan derivatives were also incorporated into a model peptide (AcWLWLL) and its binding to a DMPC bilayer analysed (§2.7).

The obtained data allowed a detailed analysis of the electrostatic influence of the π-electron system of the tryptophan derivatives and the importance of hydrogen bonding in various complexes (see §Chapter 3). The information gained, also helps to understand the underlying principles of tryptophan-lipid interactions. As tryptophan has been shown to play an important role in the anchoring, localisation and orientation of proteins in lipid bilayers and membranes (§1.4.2), this research has implications for the wider field of peptide-lipid binding.

1.3 Introduction to Membranes

1.3.1 The Structure of the Membrane

All cells are surrounded by a membrane, which separates cellular contents from the external environment and forms special compartments within the cytoplasm (the interior of a cell) that separate various cellular processes from the remainder of the cell (*e.g.* respiratory mitochondria). The membrane itself consists of a semi-permeable lipid bilayer, which is a complex, dynamic system with unique structural and mechanical properties which are critical to membrane organization and function.[24] The types of the lipids present in the bilayer and their relative ratios affect the partitioning of peptides and proteins into membranes[24b] (Figure 1.2) and hence the proper functioning of membrane receptors, channels and enzymes depends on the composition, structure and mechanical properties of the bilayer.[25]

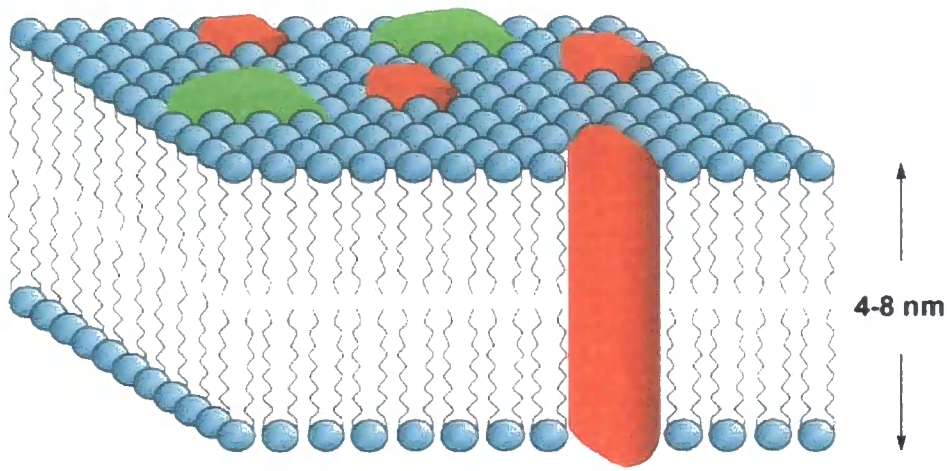


Figure 1.2: Model representation of a membrane. Blue circles represent the head group of a lipid to which two fatty acyl chains are attached. Red barrels represent trans-membrane proteins, while green barrels depict surface proteins. (Figure adapted from Ref. 19)

The membrane itself is formed by a two molecule thick bilayer of lipids (between 4 - 8 nm), which are mainly phospho- and sphingolipids (Figure 1.2).[26] A schematic picture of a general lipid molecule (Figure 1.4) and a description of its properties are given in §1.3.4. The essential structure of the bilayer was suggested by Gorter and Grendel (1925) who discovered that a cell membrane consisted of "a layer of fatty substances that is two molecules thick".[27] Later, in 1935, Danielli and Davson[28] proposed a model for the membrane that suggested a lipid bilayer core that is coated with protein on either side. This model was subsequently refined to the model that is used today: the **fluid-mosaic model** (§1.3.2), devised by Singer and Nicolson in 1972.[1]

The stability and form of the membrane is mainly determined by the aggregation behaviour of the lipid to form a bilayer; length and degree of saturation of the lipid acyl chains determine the thickness and ordering of the hydrophobic region of the membrane.

1.3.2 The Fluid-Mosaic Model[1]

The fluid-mosaic model explains the structure of biological membranes including those of animals, plants, and microorganisms. Since its introduction, no modifications of the model have been necessary. The fluid-mosaic model assumes that the amphipathic lipid molecules form a bilayer with a tail-to-tail arrangement (Figure 1.2). The driving force for this association arises from a favourable like-with-like arrangement that keeps the hydrophobic fatty acyl chains, the tails of the lipid molecules, segregated from the surrounding aqueous

fatty acyl chains, the tails of the lipid molecules, segregated from the surrounding aqueous environment. This also allows the polar heads of the lipid molecules (hydrophilic groups such as choline) to interact favourably with the water they are exposed to.

The model states that the lipid molecules as well as the proteins embedded and associated with the membrane are mobile and diffuse freely, both retaining their vertical orientation within the bilayer during this process. The lipid molecules diffuse about 100 times faster than an average membrane associated protein ($D_{protein}$ = 10 pm cm^{-2}, $D_{phospholipid}$ = 1000 pm cm^{-2}; where D is the diffusion constant). The lipids in the membrane can exist either in a gel-like (solid) state or in a liquid-like (liquid crystalline) state, which gives the lipid molecules more mobility. In living cells, the membrane seems to be in a transition between the two states, depending on physical conditions and the lipids and proteins that are present in the membrane layer.

The fluid-mosaic model was confirmed experimentally by fusing two cell membranes together that each carried distinct, labelled proteins. These proteins were found equally distributed over the newly created cell membrane, confirming the diffusion of the proteins.[29] Similarly the diffusion of lipid molecules may be demonstrated.

1.3.3 Properties

The bilayer itself is constituted of amphipathic lipid molecules which are comprised of a polar, hydrophilic head group and non-polar, hydrophobic fatty acid chains. This gives rise to a complex environment of hydrophobicity (the bilayer interior), hydrophilicity (aqueous environment on either side of the membrane) and a complex water-lipid interface of polar lipid head groups and interfacial water molecules (see Figure 1.3, **A**). There is a steep polarity gradient across the interfacial region from highly non-polar ($\varepsilon = 2$) near the hydrocarbon core of the bilayer to highly polar (ε = 80) near the aqueous outer sides of the membrane, where ε is the dielectric constant.[30] The bilayer essentially functions as an electrical isolator separating aqueous environments (Figure 1.3, **B**). Only small, hydrophobic particles can diffuse through the bilayer.

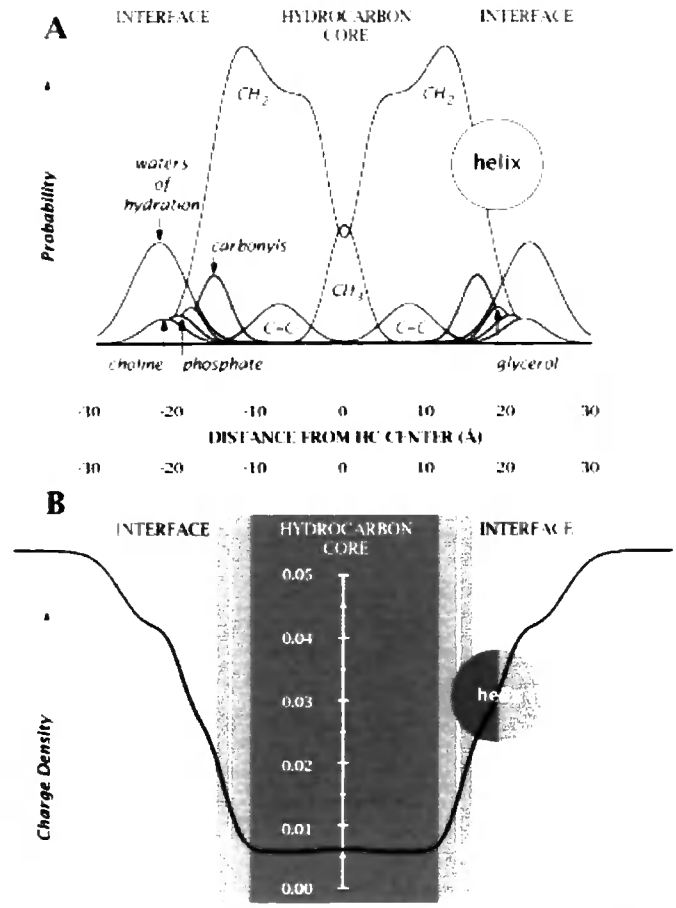


Figure 1.3: Hydrophobicity profile of a bilayer. **A** represents the structure of a fluid DOPC bilayer. Its polarity profile is shown in **B**. (Adapted from Ref. 31)

The lipid-water interfacial region accounts for 50% of the thermal thickness of the bilayer[32] and highly influences and promotes the binding and folding of proteins and peptides and determines their precise orientation in the membrane.[33,34]

1.3.4 Lipids

1.3.4.a General Structure and Overview

A lipid molecule is comprised of a polar head group, a linkage region and one or two fatty acyl chains (see Figure 1.4, **A**). Lipid molecules can be classed according to the differences in the various regions such as varying head groups and linkage regions (sphingolipids and phospholipids (Figure 1.4, **B**), fatty acyl chain length and degree of saturation.

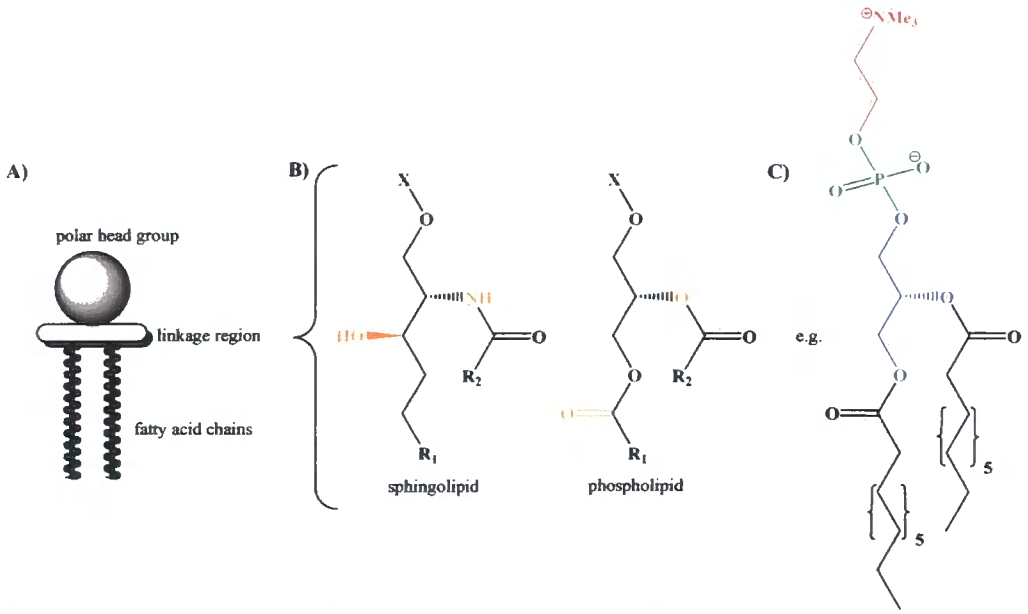


Figure 1.4: A schematic picture of a typical lipid molecule is given in **A**. **B** shows the chemical structure of the most common lipid types in cell membranes (sphingo- and phospholipids) with their major differences highlighted in orange. **C** depicts DMPC, as a typical example of a choline phospholipid; the polar choline head group (red), the phosphate group (green), the glycerol linker (blue) and the two fatty acid chains (black) are highlighted.

The hydrocarbon chains differ in their chain lengths from C_{14} to C_{26} including significant, but not predominating chains of odd carbon numbers.[35] In mammalian cerebrosides, typically sphingosine is linked to fatty acids of 16–24 carbons, which may be α-hydroxylated.[36] In yeast, longer chain fatty acids of 26 carbons predominate, the majority of which are hydroxylated at the α-position.[37] In plants saturated C_{16}, C_{20}, C_{22} and C_{24} α-hydroxylated fatty acids are the most abundant.[36]

1.3.4.b 1,2-Dimyristoyl-sn-glycero-phosphocholine (DMPC)

The lipid used in this study (1,2-dimyristoyl-*sn*-glycero-phosphocholine, DMPC) is a neutral, zwitterionic, fully saturated lipid with two C_{14} hydrocarbon tails (Figure 1.4, C). Depending on the temperature and other, external factors (such as the degree of hydration), aqueous dispersions of DMPC can exist under at least three lamellar phases.[38] In both the gel phase, $L_{\beta'}$, and the 'rippled' phase, $P_{\beta'}$, in which the membrane adopts a wave-like interface, the phospholipids are tilted with respect to the lamellar plane. In the liquid crystalline lamellar phase L_{α}, the molecules align themselves perpendicular to the lamellar plane. The main transition temperature from the gel to fluid phase for pure DMPC was found to be at 24 °C.[39]

In apolar solvents, such as chloroform, DMPC forms inverse micelles in which the hydrophobic lipid tails are directed towards the solvent, burying the polar head groups in a central sphere to avoid enthalpically unfavourable solvent contacts of the latter (Figure 1.5).

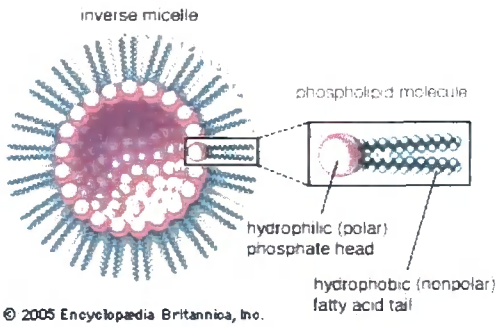


Figure 1.5: Schematic of an inverse micelle, which is formed by certain lipid molecules (such as DMPC) in non-polar solvents (*e.g.* chloroform). Figure was adapted from Ref. 40.

Haque *et al.*[41] using NMR studies of DMPC in chloroform, suggested that there are three lipids per "micelle". Hydration studies on the closely related 1,2-dipalmitoyl-*sn*-glycero-phosphocholine (DPPC) lipid in chloroform indicated that, contrary to the aggregated DPPC, the monomers of DPPC do not appear to be hydrated.[42] Evidence for this was given through the ^{1}H chemical shift and spin-lattice relaxation time of water and the ^{31}P chemical shift of the phosphatidyl group of DPPC.

1.3.5 Membrane Proteins

Proteins which are interacting with the bilayer can be distinguished into two categories: **Integral** (membrane-spanning) proteins and **peripheral** proteins, which are bound to the surface and are inserted in the bilayer but only traverse one side of it. It is estimated that approximately 30% of genes may encode membrane proteins.[43,44]

It was shown that the folding of membrane proteins into the bilayer can occur spontaneously, as is the case for pore-forming toxins[45], or through a series of protein-catalysed steps.[46] The folding appears to be governed by three steps[7b]: Binding of the protein to the membrane interface, formation of secondary structure, and insertion of secondary-structure elements into the membrane. However, other mechanisms and order might occur. The precise nature of the favourable interactions and the origin of the free energy gain for

proteins upon interaction with lipid bilayer is still being investigated by various groups (see §1.2.2.b).

1.4 An Introduction to Tryptophan – Importance, Synthesis and Uses

1.4.1 The Amino Acid Tryptophan – An Overview

Tryptophan is synthesised by bacteria and plants only and, as an essential amino acid, is a required ingredient in the diet of all animals. It is the second least abundant of the common amino acids, generally constituting 1% or less of the average protein mass.[47] Some proteins contain no tryptophan residues at all. For example, two of the tryptophan biosynthetic enzymes in *Escherichia coli*, anthranilate synthase (520 residues) and the α subunit of tryptophan synthase (268 residues) do not contain any tryptophan residues.[48,49] However, proteins with an unusually high content of tryptophan such as β-galactosidase (38 of 1021 residues, 3.7%) are also known.[50]

In bioenergetic terms, tryptophan is the most expensive amino acid to produce: each mole of tryptophan requires 78 moles ATP.[51] This energy cost is 20% higher than that for the biosynthesis of the second most costly amino acid, phenylalanine, and approximately double the biosynthetic energy of the average amino acid. The complex biosynthesis of tryptophan is split in two parts. The Shikimate pathway describes seven enzyme-catalyzed steps in the synthesis of chorismate, the last common precursor of the aromatic amino acids. The elaboration into tryptophan requires five more enzyme-catalyzed reactions.

Tryptophan has the largest nonpolar surface of all of the naturally abundant amino acids. However, its side chain is also capable of forming hydrogen bonds through its NH group. This gives rise to an ambivalent behaviour of tryptophan, and its hydrophobicity is ranked very differently on different hydrophobicity scales, which are generally based on the partitioning of single tryptophan molecules between polar and nonpolar solvents.[52,53]

1.4.2 The Role of Tryptophan

In 1989 the first high-resolution, 3D structure of a membrane protein revealed a statistical anomalous over-representation of tryptophan at the putative interface between the polar aqueous environment and the more apolar interior of the lipid membrane.[54] Since then this asymmetric pattern has been seen in an increasing number of membrane peptides, not only

for tryptophan, but also for tyrosine and phenylalanine.[55,56] For example, crystal structures of membrane proteins such as the bacterial KcsA potassium channel,[57] maltoporin[58], bacteriorhodopsin[59], and others show aromatic residues preferentially clustered in a saddle-like "aromatic belt" around the membrane interfacial region. Statistical studies[60] of sequence databases (such as the Protein Data Bank (PDB)) and available crystal structure of integral membrane proteins gave further support to this preferred localisation of tryptophan residues in the regions where the protein intersects the membrane interface region (Figure 1.6).[149]

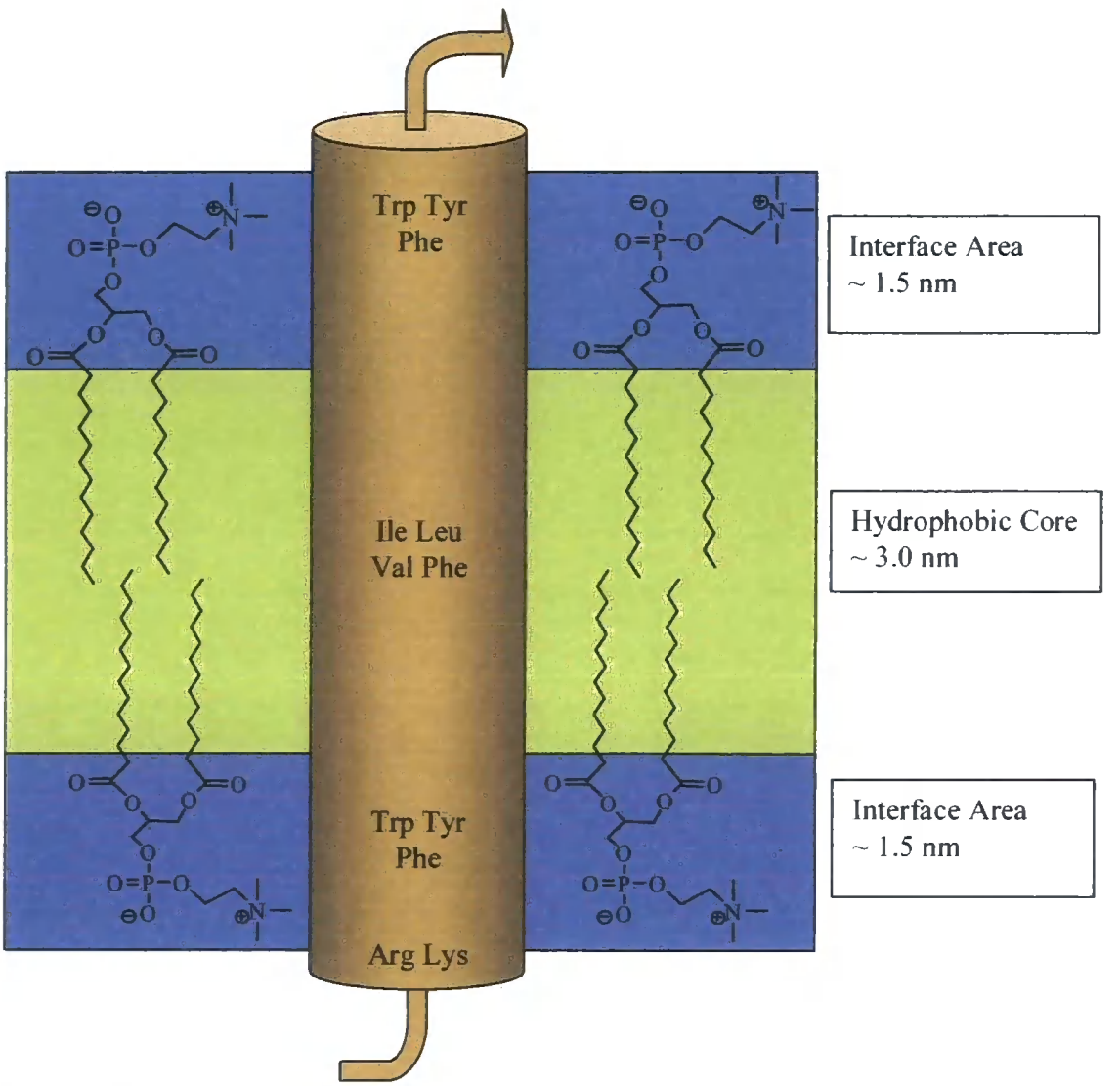


Figure 1.6: Schematic representation of an α-helix transmembrane segment inserted in a lipid bilayer. The interfacial areas and hydrocarbon core of the bilayer, consisting of DMPC phospholipid molecules, are highlighted. The preferred locations of a number of amino acids are illustrated. (Figure adapted and modified from Ref. 61)

It is believed that the aromatic amino acids can assist in stabilizing the protein in the membrane[62,63,56] and that they are essential for the proper function of membrane proteins.[64,65] It was further proposed that tryptophans are involved in the ability of trans-membrane α-helical peptides to induce non-bilayer structures in model membrane systems under conditions of hydrophobic mismatch.[66,67] For membrane-active water-soluble proteins, tryptophans are likely to play an important role in membrane interaction. An example of the functional importance of tryptophan residues in membrane proteins is the gramicidin A channel. Here, tryptophan residues have been shown not only to be important for the anchoring of the protein in the membrane[68], but also to increase the ion permeability through the channel.[69] This was demonstrated by substitutions of tryptophan residues with a more polar tryptophan derivative (5-fluoro-tryptophan), which increased the conductance of the channel. Another example is the signalling helix of the aspartate receptor of bacterial chemotaxis, where the position of aromatic residues in relation to the membrane interface is critical for its activity, as revealed by mutagenesis studies.[70]

These examples illustrate the preference of tryptophan residues to reside in the water-lipid interface of the membrane. This is further corroborated by the observation that tryptophan residues, when placed into the centre positions of hydrophobic membrane-spanning helices, can pull these helices out of transbilayer orientations and into interfacial orientations showing a strong disfavour for the hydrocarbon core of the membrane.[71] Another study by Shank *et al.*[72] concluded that the introduction of aromatic residues in synthetic channel-forming sequences optimized membrane insertion, orientation, and channel activity as a result of the efficient anchoring of the sequence at the membrane interface.

Recent studies have tried to elucidate and analyse the nonuniform distribution of tryptophan and its strong preference for the interfacial region in different model systems. The most commonly used model system employs glycerophospholipids.[24b,73]

1.4.3 Synthesis of Tryptophan Derivatives

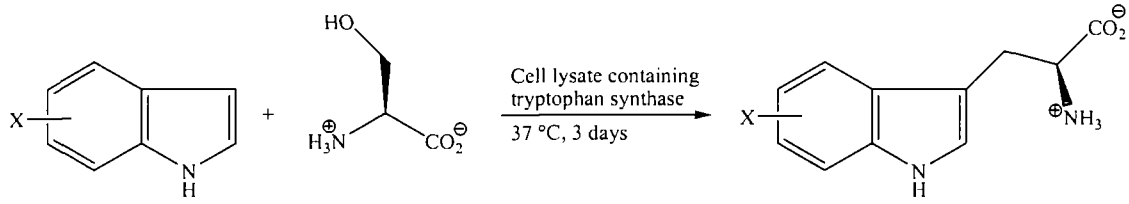
There are two basic routes to produce tryptophan derivatives: One is based on the enzyme tryptophan synthase and uses biological conditions to express various tryptophan derivatives mostly employing simple indole derivatives as substrate. The other methodologies are all based on a chemical synthesis. Here, there is a trend to exploit the same general idea of coupling indole derivatives to serine, amino acrylate or nitroacetate/paraformaldehyde

analogues. Alternatively, tryptophan analogues have been synthesised using metal catalysed reactions. Examples of such approaches are highlighted and their advantages/disadvantages discussed in the following sections.

1.4.3.a Biological Approaches

All enzymatic approaches to synthesise tryptophan derivatives are based on tryptophan synthase, which can be extracted from various sources. Tryptophan synthase (*L*-serine hydrolase, EC 4.2.1.20) was first recognized by Tatum and Bonner (1943-1944) who demonstrated the ability of extracts of *Neurospora crassa* to catalyze a condensation between indole and *L*-serine to yield *L*-tryptophan. The enzyme also catalyses the final step in tryptophan synthesis (the replacement of the triosephosphate side-chain of indole-3-glycerol-phosphate by *L*-serine). Wilcox[74] used this enzyme to produce tryptophan analogues with moderate yields (43-85%).

In the most recent synthesis, tryptophan synthase was extracted from a pellet of *E. coli* pSTB7 by sonification.[75] The crude cell lysate was then used as a reagent for the conversion of monosubstituted indoles to *L*-tryptophan derivatives using *L*-serine as substrate (Scheme 1.1).



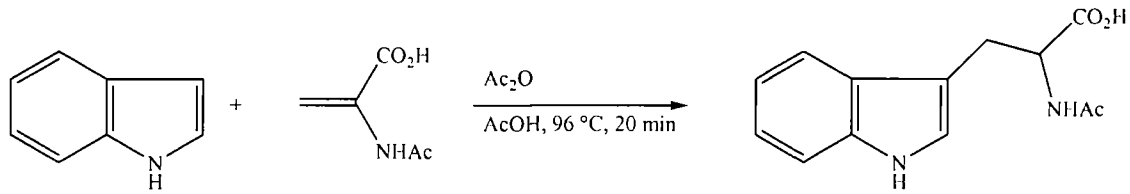
Scheme 1.1: *L*-tryptophan derivatives were synthesized using mono-substituted indole derivatives (X = F, Cl, Br in positions 4, 5, 6 and 7), *L*-serine and tryptophan synthase in buffer solution.

The advantage of the methodology is the direct, one-pot synthesis of enantiomerically pure free amino acid derivatives. However, the scope of the reaction is limited by the substrate specificity of the tryptophan synthase. Sterically hindered indole derivatives are only converted very poorly or not at all (*e.g.* 4-bromoindole and 7-chloroindole lead to yields of < 10%). Other indole derivatives proved more suitable with yields of up to 82% after one cycle (6-fluoroindole). It was found that substrates which have nitrogen for carbon replacements adjacent to the position (3-) where condensation occurs are either completely unreactive (pyrimidazole) or convert only very slowly (2-azaindole).[74] The author (Wilcox) suggested that the electronegativity of the nitrogen leads to a withdrawal of electrons and

hence a deactivation of the reactive 3-position. Methyl substituted indoles are not susceptible to conversion either, having reported yields ranging from 4% (5-methylindole) to 24% (7-methylindole).[76] Another limiting factor of the enzymatic methodology is the breakdown of serine, either by deamination[77] or in a reaction catalysed by some other activity present in the enzyme preparation. Even though the yield of the enzymatic reactions can be improved by recycling unreacted starting material and submitting it to the reaction more often, this methodology does not seem suitable for the large scale production of a varied series of tryptophan derivatives due to the restrictions the enzyme places.

1.4.3.b Chemical Approaches

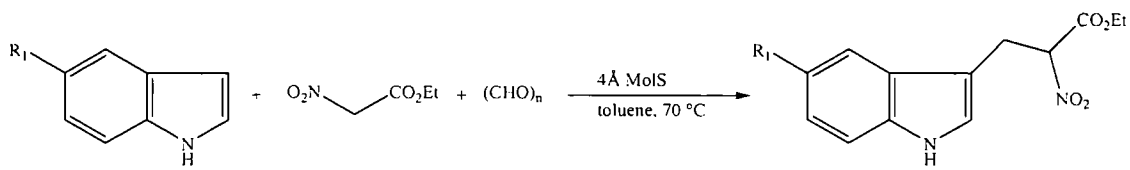
A successful chemical synthesis of N^{α}-acetylated tryptophan was introduced by Snyder *et al.* in 1955.[22] The authors reported the reaction of indole with α-aminoacrylic acid to yield the desired product (Scheme 1.2).



Scheme 1.2: α-Amino acrylic acid can be reacted with indole to form N^{α}-acetyl-tryptophan in moderate yield.

Even though the yield was moderate (58%), others[23,78] improved the reaction further, also demonstrating that substituted indole derivatives could be used in this methodology.

Recently, Sui *et al.*[78] offered a variation of the approach in which indole derivatives are reacted with nitroacetate and paraformaldehyde to produce ethyl 3-(1H-indol-3-yl)-2-nitro-propanoate derivatives (Scheme 1.3).

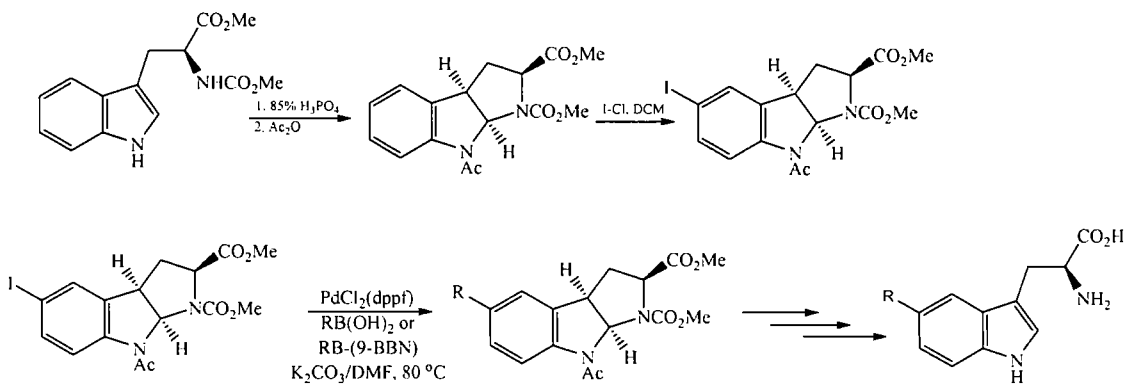


Scheme 1.3: Schematic reaction scheme of indole reacting with nitroacetate and paraformaldehyde, where MolS stands for molecular sieves and R_1 = H, Br, OH, OMe, CO_2Me.

Molecular sieves had to be added to improve the yield of the reaction by trapping the generated water which is liberated in the nitroaldol condensation. The reaction has the

drawback that the nitro group needs to be reduced to the amine in order to access free amino acid tryptophan analogues. It also requires at least one further step of optical resolution if one is interested in enantiomerically pure products.

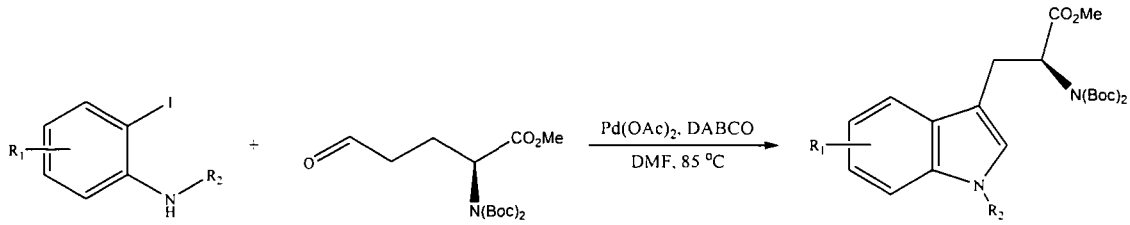
Zembower and Ames[79] demonstrated the use of cyclic tryptophan tautomers in the derivatisation of tryptophan compounds (Scheme 1.4). The tautomers were treated with iodine monochloride resulting in a 5-iodo substituted tautomer. Then, cross-coupling with arylboronic acids or *B*-alkyl-9-bora-bicyclo[3.3.1]nonane (9-BBN), catalyzed by 3 mol% [bis(1,1'-di-phenylphosphino)ferrocene]palladium(II) chloride ($PdCl_2$ (dppf)), afforded 5-aryl and 5-alkyl cyclic tryptophan tautomers that were decyclised and deprotected to yield *L*-tryptophan derivatives in 33% overall yield.



Scheme 1.4: Synthesis of 5-substituted tryptophan analogues using tryptophan tautomers and palladium cross-coupling of boronic acids.

Even though this synthesis provides an elegant solution to produce a variety of 5-substituted tryptophan derivatives, it is limited to this position and with the need of decyclisation and deprotection additional steps that lowered the overall yield noticeably.

Recently, a palladium-catalysed direct annulation reaction of substituted *ortho*-haloanilines and aldehydes was shown to be a viable route for the synthesis for tryptophan compounds (Scheme 1.5).[80]



Scheme 1.5: Palladium-catalysed direct annulation of substituted *ortho*-haloanilines and aldehydes to produce N^{α}-Boc$_2$-*L*-tryptophan methyl esters. DABCO stands for 1,4-diazabicyclo[2,2,2]octane; R_1 = 4-, 5- or 6-MeO, 5- or 6-NO_2, or 6-Cl; R_2 = H or Me.

The authors could synthesise enantiomerically pure N^{α}-Boc-protected *L*-tryptophan derivatives in good to excellent yield (51-85%) using enantiomerically pure methyl (*S*)-2-*N*,*N*-di-*tert*butoxycarbonyl-5-oxopentanoate. The deprotection of the carboxylic acid and the amine functionalities and potentially the indole amine then introduces additional steps for the access of free *L*-tryptophan derivatives using this route.

1.4.4 Uses of Tryptophan Derivatives

Apart from its synthetic use as protein and peptide building block, tryptophan provides a biosynthetic precursor for many naturally occurring alkaloids.[81] Furthermore, tryptophan analogues have been used extensively to probe biological systems: α-Tryptophan was reported to inhibit the growth of *Bacterium coli*[82] and can be used as IDO inhibitor.[83] Methyl-substituted indoles inhibit the biosynthesis of tryptophan and are employed to elucidate the mechanism and detail of tryptophan biosynthesis.[76] 6-Chlorotryptophan has been identified as a non-nutritive sweetener.[84]

The physical properties of tryptophan were also exploited extensively: Tryptophan molecules absorb strongly in the region between 275 and 280 nm and make up a large contribution to the ultraviolet absorption of protein molecules. These luminescence properties can be employed to study membrane-interactions (tryptophan is arguably the most frequently used tool for examining the interactions of proteins and peptides with vesicular unilamellar model membranes).[85] The fluorescence lifetime and quantum yield, are environmentally dependent, and can be used to determine its membrane/water partition coefficient.[86,87] In addition, spectral shifts undergone by the emission spectra of tryptophan residues, as well as changes in accessibility to aqueous quenchers can be exploited to characterize peptide membrane binding. The peptides either naturally contain one or more

tryptophan residues or are engineered (*e.g.* mutated) accordingly. Tryptophan has also been found to be an ideal optical probe for hydration dynamics and protein-water interactions.[88]

1.5 The Model Peptide AcWLWLL

Biomembranes present a complex environment in terms of the diversity and size of their components. It is therefore difficult to obtain detailed and unambiguous information about individual interactions of peptides with the lipid bilayer. Therefore, simplified model systems need to be established, where structural parameters (such as the lipid size, type and composition) can be systematically altered allowing the various effects of these alterations to be studied in detail. Analogously, the interacting peptides can be adjusted and the resulting effects analysed at a molecular level. A simplified membrane system should consist of a model lipid bilayer consisting of only one or two lipid species. Model peptides should be small enough to avoid the formation of complicated tertiary structures or aggregates and ideally interact with the membrane in a well-defined manner. The interactions between a peptide and the lipid interface will be easier to study if there are no side effects such as partitioning and/or aggregation of the peptide on the membrane which will have significant effects on the binding and interaction with the membrane and would complicate any direct measurements. The peptides may also include reporter amino acid residues (such as luminescent tryptophan) or tags to allow a detailed experimental analysis.

The pentapeptide Ac-Trp-Leu-Trp-Leu-Leu-OH (AcWLWLL, Figure 1.7) fulfills all these requirements.

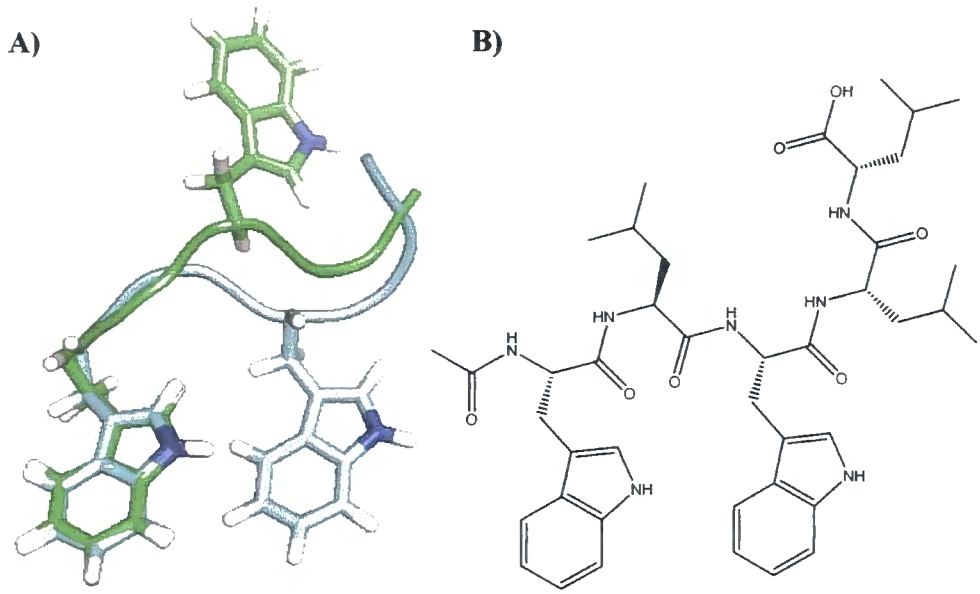


Figure 1.7: Model peptide AcWLWLL. **A** depicts two major confomers of the AcWLWLL as determined by molecular dynamics simulations of the pentapeptide in SPC water (§5.1.4.c). **B** shows the chemical structure of AcWLWLL.

The lack of secondary structure of this peptide was determined by White and Wimley.[89] The authors showed AcWLWLL to be strictly monomeric and random coil in aqueous phases using CD spectroscopy, which showed minima at ~200 nm and intensities of about -20,000 deg $dmol^{-1}$ cm^{-2}, both strongly characteristic of random coil peptides. Isothermal titration calorimetry indicated that there is no significant aggregation of the peptide even in a saturated solution as no detectable heat change was observed when the concentrated solutions were injected into buffer solution.[90] Further support was given through fluorescence spectroscopy and quenching studies.[90]

It was shown previously that AcWLWLL partitions into a lipid bilayer and the partitioning is independent of the peptide concentration.[90] The partition coefficient is around 0.025 at room temperature. It was also demonstrated that the partitioning of AcWLWLL depends only weakly on temperature and decreases with increasing temperature with the partition coefficient changing only from ~0.038 at 5 °C to ~0.018 at 65 °C.[89] White and Wimley could calculate the partitioning free energy of AcWLWLL from water into octanol ($\Delta G_{water\text{-}octanol}$ = 2.96±0.01 kcal mol^{-1}) and into a POPC bilayer phase ($\Delta G_{water\text{-}POPC}$ = -1.85±0.06 kcal mol^{-1}) by comparison to the free energy of the partitioning of other pentapeptides of the AcWL-X-LL family, where X stands for any of the 20 amino acid

residues.[91] Contents leakage of lipid membranes that can be induced by some pore-forming peptides was shown to be very inefficient and low for AcWLWLL.[92]

The peptide under investigation (AcWLWLL) is very similar in sequence, length and hydrophobicity to the hexapeptide $AcWL_5$, which was extensively analyzed by the researchers mentioned above.[90,93] While $AcWL_5$ assembles cooperatively into β-sheets upon partitioning into a lipid bilayer from an aqueous phase where the peptide is strictly monomeric and random coil, the AcWLWLL pentapeptide does not aggregate nor form secondary structure.[90]

1.6 Computational Studies of Peptide-Lipid Interactions

1.6.1 Introduction

The advantage of computational studies is the ability to describe both detailed thermodynamic and fast kinetic aspects of binding interactions. Binding interactions between pairs of atoms, molecules and entire complexes can be simulated and analysed with relative ease. Molecular dynamics simulations also give a unique insight into binding conformations of complexes and their dynamics on a time scale from pico- to sub microseconds. The main disadvantage is the computational cost associated with the simulation of larger systems (such as wide arrays of lipid bilayers). This often leads to a compromise between a) the level of atomic detail of the lipid representation (e. g. all atom *vs.* united atom or coarse grained approaches), and b) the computed internal degrees of freedom for proteins and the number of simulated particles. In this section, the main parameters used in computational studies are introduced and discussed. A detailed break-down of the parameters used in this thesis are given in the appropriate sections (§5.1 and following).

Here, molecular dynamics simulations were used to simulate i) the binding of tryptophan derivatives to DMPC lipid molecules in chloroform at varying concentrations, ii) the importance of water on these interactions, iii) the self-association of tryptophan compounds in the absence of lipid, and iv) to simulate peptide-lipid interactions of model peptides with a DMPC bilayer in SPC water. Details of set-ups and results can be found in (§5.2 and §2.5 respectively).

1.6.2 Molecular Dynamics (MD) Simulations

1.6.2.a General Overview

In molecular dynamics simulations, all atoms are treated as classical particles with their elemental mass. The interactions between atoms are expressed as non-bonded interactions, for any pair of atoms within a pre-defined cut-off radius, and bonded interactions between atoms connected by a chemical bond. Each atom gets assigned a partial charge (usually derived from *ab initio* calculations of the molecule) and parameters for repulsion and attraction (such as Lennard-Jones parameters) to compute non-bonded interactions. Bonded interactions are usually described by harmonic oscillators for bonds and angles and by a suitable cosine expansion for dihedral angles. The parameters and equations used in the simulations of this thesis are described in detail in §5.1. The following assumptions are commonly employed:

1) *Only pair-additive interactions are taken into account.* This means that non-bonded interactions involving three or more atoms (multi-body interactions) are neglected.
2) *Atoms are represented as partial charges.* This assumption neglects the effects of electronic polarizability which could be introduced using dummy atoms that are coupled to electronegative atoms by a harmonic oscillator and represent a partial charge in close proximity, approximating dipolar effects.
3) *Simple harmonics are used represent bonds and angles.* Other models (such as a Morse potential[94] or other anharmonic descriptions) give better descriptions of bond lengths and angles at an increased computational cost.

The potential functions are combined in the so-called force field (§1.6.3) and then used to solve Newton's equations of motions for all atoms in the system by calculating the forces on all atoms and integrating with respect to time. Such treatment, in principle, allows simulating the dynamics of any system that can be described in terms of a simple interaction potential.[95] The output (trajectory) of a molecular dynamics simulation contains the xyz coordinates of all atoms and their respective velocities as a function of the number of integration steps (*i.e.* time).

1.6.2.b Limitations

Parameterisation

The limitations of molecular dynamics simulations originate from the need to parameterize. The parameters used to describe, for example the Lennard-Jones potential, are mainly derived from experimental values. A parameter set for a given system will be adjusted to reproduce macroscopic, experimentally detectable quantities including density, heat of vaporisation, and in the case of lipid bilayers area per lipid and lipid order parameters. These empirically adjusted parameters are therefore able to reproduce the averaged global factors of a system, but it is difficult to estimate and represent local effects. This holds especially true for assumption 2) (§1.6.2.a); the assignment of partial charges and their parameterisation leads to the effect that atomic polarizability is not accounted for properly and even though the average effects of polarizability are retained in the system, detailed effects are not properly represented. Furthermore, the height of barriers in the dihedral potentials and exact values for the representation of van der Waals forces are difficult to determine exactly. This makes extensive testing of the parameters necessary. In this study, the parameters used were obtained from the literature and have been tested extensively elsewhere (as described in §1.6.3).

Time Step

Another limitation in molecular dynamics simulations is the integration of the equations of motions. If the time step with which an integration is performed is chosen to be too large, the system becomes unstable. A typical value is 2 fs (as chosen in this study). However, the use of constraints can potentially increase the time step to up to 5 fs. In any case, a reasonably long (10-100 ns) molecular dynamics simulation will still require hundreds of thousands of computationally expensive integration steps, limiting the length and system size of the calculation. Typically, MD simulations do not exceed 100 000 simulated particles and are restricted to about 100 ns.

Classical Treatment

The approximation of bonded interactions to be described by harmonic oscillators makes the study of chemical reactions by pure molecular dynamics simulations impossible: The harmonic description prevents the breaking of bonds (dissociation of atoms). This problem

might be overcome by using more appropriate descriptions of bonded interaction at a computational expense or by describing at least part of the system quantum mechanically.

1.6.3 Force Fields

Atomistic simulations require expressions for the potential energies between all atoms in the system as input parameters, including the energies between bonded pairs of atoms and non-bonded pairs of atoms (Equation 1.1). Generally, harmonic potentials model chemical bonds (V_b) and angles (V_a) whereas non-bonded interactions are approximated using coulombic (V_C) and Lennard-Jones potentials (V_{LJ}), which model interactions due to polarization effects between atomic electron clouds. Dihedral functions (proper V_{pd} and improper dihedral potentials V_{id} respectively) model interactions between next-nearest neighbour atoms on the same molecule. The potential function V_{pot} of a system can be described as

$$V_{pot} = V_b + V_a + V_{pd} + V_{id} + V_C + V_{LJ} . \qquad (1.1)$$

The set of functions, and the parameters characterizing the strengths of the various interactions, is commonly referred to as a 'force field'. Most common force fields share common assumptions such as a simplified treatment of electrostatic interactions as point charges centred on atoms, which probably poses the most serious simplification as it ignores higher multiple interactions and the details of electronic polarizability, limiting the accuracy of properties such as the relative orientation of aromatic residues, interactions with ions, and likely the solvation free energy of amino acid side chains in solvents of different dielectric constants.

Simulation packages such as AMBER[96], CHARMM[97], GROMOS[98], and special force fields have been introduced to analyse different kind of structures such as lipid bilayers and are generally developed independently from each other and tested against experimental data. There are only two phospholipid force fields in common use today: The official all-atom CHARMM distribution and a parameter set that was developed by Berger *et al.*[99] using parameters taken from united-atom versions of OPLS[100] and AMBER. Both force fields simulate lipids that reproduce experimental information on the structure and dynamics of phospholipid bilayers reasonably well, particularly for phosphatidylcholine bilayers. None of the force fields seem to perform substantially better or worse than the other in terms of reproducing benchmark parameters such as the lipid order parameter or the area per lipid.[101]

There are also differences in the detail of the atomistic descriptions: An all-atom description has been used extensively[102] as well as a unified description of CH_n groups, which may be further separated into the united atom[103] or the anisotropic united atom[104] model.

1.6.4 The Canonical Ensemble

Three alternative canonical ensembles have been used in simulations of lipid bilayers so far: constant number of particles at constant volume and constant temperature (*NVT*); constant isotropic pressure (*NPT*); and constant surface tension which is equivalent to a constant anisotropic pressure (*NγT*). Different methods of implementation of the ensembles have been used by Feller *et al.*[102d] and Zhang *et al.*[105] As reviewed by Berger *et al.*[99] and Tieleman and Berendsen[95] the most appropriate method seems to be an *NPT* ensemble as detailed briefly below.

In the *NVT* ensemble constant volume requires that the dimensions of the simulation box are constant. Although this is the standard condition for simulations of a protein in a crystal lattice,[99] it is less appropriate for a lipid bilayer, because the dimensions of the box are determined by the area and length per lipid, which are not well known. Moreover, the bilayer can undergo phase transitions which involve changes in dimensions that would not be represented accurately within such an ensemble.

In the constant pressure ensemble (*NPT*) one has to consider that the pressure might be better represented as being anisotropic. The symmetry of the surface and the condition of mechanical equilibrium leads to a diagonal pressure tensor with a constant normal pressure P_N that is equal to the external pressure and a pressure P_T in the two transverse directions, which may depend upon the depth of the bilayer.

The *NγT* as used by Chiu *et al.*[106] has the disadvantage that it may not accurately represent the fact that lipid bilayers are free to adjust their surface area to attain equilibrium with the surroundings. Surface tension γ is defined by the work $\delta W = \gamma dA$ that is required to change the surface area by dA. The area will adjust so that the free energy becomes minimal. For systems with a positive surface tension (like oil and water), the free energy increases monotonously with area. *I.e.* the system tries to minimize the contact area (at constant volume). If the surface tension is negative the two fluids mix and there is no phase separation. Lipid bilayers exhibit a behaviour that can not be modelled accurately by either of these situations. The hydrophobic effect, as for oil and water, will give rise to an increase in

free energy with surface area. The bilayer would minimize its area at all temperatures and go into the ordered gel phase. However, in reality there are other contributions to the free energy that increase with decreasing surface area. A small surface area forces the lipid chains into a more ordered state and thus reduces the entropy of the system. Therefore, the free energy no longer increases monotonously with area but has a minimum. The bilayer adjusts its area to attain this minimum in thermodynamic equilibration, which implies by definition that the surface tension is zero and the average transverse pressure is equal to the external pressure, as claimed by Jähnig[99] in response to Chiu *et al.*[106]

1.6.5 Lennard-Jones (LJ) Parameters

The interaction between two atoms is governed by a long-ranged attractive van der Waals force and a short ranged repulsive coulombic force which arises when the electronic clouds of these atoms start to overlap. The latter is also referred to as Pauli repulsion (from the Pauli Exclusion Principle). The Lennard-Jones potential, as shown schematically in Figure 1.8 describes the strength of these two forces with respect to the distance between them and thus the separation of them. The Lennard-Jones potential is commonly also referred to as the 6-12 potential due to its mathematical expression:

$$V(r) = \sum_{i>j}^{N} 4\varepsilon \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^{6} \right], \qquad (1.2)$$

where epsilon (ε) is the well depth and sigma (σ) the hard sphere diameter, and i and j denote two particles. These two parameters (ε and σ) are known as the Lennard-Jones (*L-J*) parameters. The first term in Equation 2 describes the repulsive and the second term the attractive force.

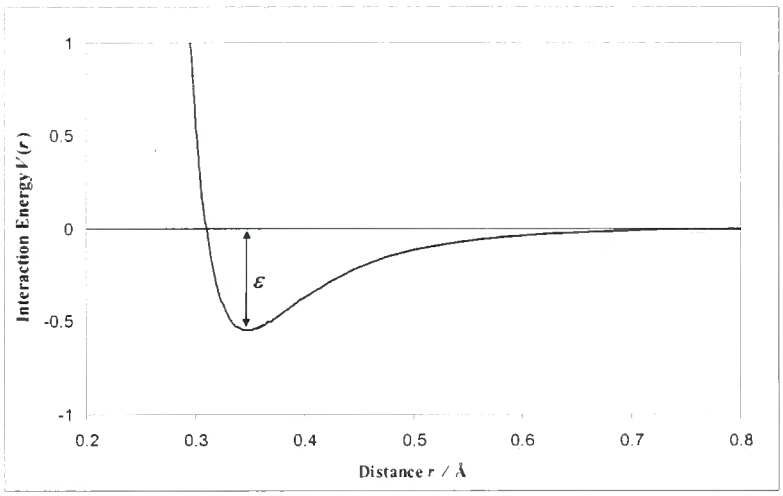


Figure 1.8: Schematic drawing of the Lennard-Jones potential between two atoms or molecules. The distance r and the interaction energy $V(r)$ are in arbitrary units. The depth of the well is defined by the parameter ε.

The Lennard-Jones potential is a good approximation of the interaction between two particles and is used extensively due to its simplicity. It reflects the real potential very well at short and long distances. Small deviations arise due to an incorrect treatment of the long-range part of the repulsion term.

The Lennard-Jones parameters are fitted to experimental data or deducted from quantum chemistry calculations. In computational chemistry these parameters are also used not only to describe the interactions between two atoms, but also between groups of atoms. An example is the united atom model, where heavy atoms and attached hydrogens can be grouped together and treated as one unit rather than individually. Although the interaction potential can be approximated to a higher accuracy by extending the Lennard-Jones model (*e.g.* Stockmayer equation, Møller-Plesset perturbation theory, and coupled cluster approach), the computational demand increases so much that the Lennard-Jones model remains the most widely used description of inter-atomic potentials.

Different sets of Lennard-Jones parameters have been used in simulations. Tieleman and Berendsen[107] studied the influence of the Lennard-Jones interactions of the CH_2 and CH_3 groups of the acyl chains with the oxygen water atom. At constant sigma value (σ) smaller parameters for epsilon (ε) resulted in a more defined and smaller interface region in lipid bilayer/water systems and a smaller area per lipid head group. The order of the tails along the interface is less than found in experimental data for higher values of epsilon. The authors suggested the use of $\varepsilon = 0.529$ kJ/mol for the CH_2-Ow, $\varepsilon = 0.637$ kJ/mol for the CH_3-Ow and,

$\sigma = 0.310$ nm for both interactions. This corresponds to 53% of the original values used in the simulation of a decane/water interface.[108]

1.6.6 Treatment of Electrostatic Interactions

Electrostatic interactions between two charged particles are described by a coulombic force between them. The charges are generally obtained as partial atomic charges *via ab initio* computer calculations and quantum physical considerations. Especially in the united atom models groups of atoms are combined and assigned such a charge. Typical groupings might include methylene (CH_2) and methyl (CH_3) units as well as chemical groups such as phosphate or ammonium. Alternatively, in simulations, groups of atoms can be combined in charge groups as part of an electrostatic treatment (*e.g.* by the particle mesh Ewald summation (PME)) and then in turn interact with each other.

Electrostatic interactions play an important role in the interaction (*e.g.* solvation) of solutes with solvent. This parameter has not only great influence on the dimensions and characteristics of the bilayer/water interface region but also on the bulk properties of the solvent itself. A careful assignment of partial charges is crucial for an accurate calculation of solvation free energies and binding constants.

In current MD force fields, electronic polarizability is not treated explicitly;[109] rather polarizability is included implicitly in the form of partial atomic (electrostatic) charges that typically overestimate molecular dipoles.[110] Although the polarizability of the π-electron distribution is not accounted for explicitly, current pairwise additive electrostatic (Coulomb) forces capture the basic physics of the cation-π interaction.[111,112]

The area per lipid is strongly influenced by the methodology used in the treatment of electrostatics. Lipid areas obtained with PME are significantly higher than those obtained with an abrupt truncation of the electrostatic interactions.[113] The difference between cut-off method and PME becomes larger as the cut-off radius increases. For example, in a model system of a DPPC bilayer in water, a small cut-off of 1.4 nm produced the largest average area, which dropped by 4% when the cut-off radius was increased to 1.8 nm and a further 2.5% when increased to 2.4 nm.[113] The total volume, volume per lipid, and especially the area per lipid decrease as the cut-off is increased. Further evidence that cut-off treatments of the electrostatic interactions should be avoided when simulating bilayer/water systems was given by Patra and Karttunen[114] in 2004 in their review of electrostatic methodologies used in

bilayer simulations. The authors reported that the lateral diffusion coefficients obtained by PME and cut-off truncation at 1.8 nm, for example, differed by a factor of 10, while the PME results are consistent with experimental values.

Different choices of charge group sizes and the charge they are representing have a strong influence on the simulation. It has been shown[113], that a change from larger to smaller charge groups result in an increase of the area per lipid by 2% with a Coulomb cut-off treatment and in a decrease by 2% with PME. The effect seen in PME can be attributed to the effect of changing the effective cut-off for the Lennard-Jones interactions because the electrostatic cut-off in PME serves only as a numerical device to separate direct- and reciprocal-space sums.

1.6.7 Long-range electrostatics

Coulomb cut-off methods often show an artificial ordering of water in bulk solution resulting in a higher viscosity.[115] As an artefact, the water dipoles are anti-correlated slightly below the cut-off and correlated above the cut-off value. The long-range dipole interactions decrease significantly when using the reaction field (RF) and especially when using the particle mesh Ewald (PME) methodology. Short- and intermediate-range correlations are still stronger with RF than with PME. This trend is also seen in interfacial water/lipid systems.[102a,115b] However, the PME method may induce artificial effects due to the infinite periodicity implied by this technique. Such periodicity artefacts were found to be particularly important in systems involving a solvent of low dielectric permittivity, a small unit cell, or a solute with a net charge or dipole. Ewald-enhanced stability of peptides, for instance, has been reported in some simulations.[116,117]

In conclusion, none of the methods mentioned above seem to be very well suited for membrane calculations: simple cut-off treatments induce artificial ordering; the RF method ignores the heterogeneous nature of the membrane and dielectric constants for the latter are difficult to justify; and the PME technique enhances periodicity. In a recent comparison of electrostatic methodology in the simulation of a water/lipid interface Anézo *et al.*[113] concluded that PME and RF techniques would both be suitable methods for such simulations.

1.6.8 Solvent Models

1.6.8.a Chloroform Model

In 1994, Tironi and van Gunsteren[118] compared three different models used for chloroform simulations. The authors concluded that an all-atom representation as used by Dietz and Heinzinger[119] reproduced a variety of experimental properties such as isothermal compressibility, constant volume heat capacity, and dielectric constant of chloroform the best.

1.6.8.b Water Models

Tieleman *et al.*[107,120] compared the three most widely used water models in molecular dynamics simulation. These are the single point charge (SPC)[121] model, the single-point extended charge (SPC/E)[122,123] model, and the transferable intermolecular potential model with 4 points (TIP4P)[124] (Figure 1.9).

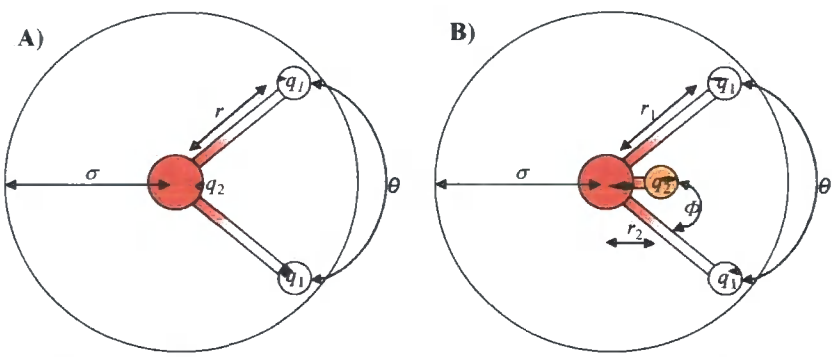


Figure 1.9: Commonly used water models. **A** depicts an SPC type water model; **B** represents a TIP4P type water representation. Here, q_1 and q_2 are the partial charge associated with the hydrogen atoms, and the oxygen atom respectively. The latter is represented either on the atom itself (**A**) or on a dummy particle a distance r_2 away from the oxygen atom (**B**); r_1 is the distance between the oxygen and hydrogen atoms, the latter being at an angle θ; Φ is the angle between the oxygen-hydrogen bond and the oxygen-dummy particle bond, σ stands for the Lennard-Jones parameter.

Although the differences in partial charges and geometry for a single molecule between these models are small, they accumulate to give noticeable effects on the total system. SPC/E was originally developed because previous models did not take the self-energy due to polarization into account.[122] This also limits the model as the thermodynamic potential for SPC/E is only correct if the proper polarization self-energy correction is applied.

In simulations where water is in equilibrium in different environments such corrections cannot be applied and the effective free energy of SPC/E is too low (-27.6 kJ/mol), compared to -24.3 kJ/mol for SPC, the latter being close to the experimental value.[123,125] This means that the liquid state is thermodynamically too favoured for SPC/E. Since solubility is important at interfaces, the SPC water model seems preferable. Furthermore, SPC leads to a slightly broader (*i.e.* wider) interface, which can be rationalised by the increased partial charge of the water oxygen atom in the SPC/E model. In experimental work the area per lipid is typically larger than in simulations. This favours the SPC model since it results in larger values for the area per lipid than SPC/E as has been reported for a DPPC/water interface by Tieleman and Berendsen.[107] Although SPC/E results in a better density, radial distribution function, self-diffusion constant and dielectric constant than SPC in bulk solution,[126] it was concluded that SPC water performs generally better than the SPC/E model for interface simulations where these trends are not seen as such.[107]

The TIP4P model uses a 4 point model to simulate water (Figure 1.9, **B**). Three atoms represent the oxygen and two hydrogen atoms in the water molecule and a fourth atom is introduced that does not carry mass but partial charge. This so-called dummy atom is intended to reproduce the lone pair on the water oxygen. Although bulk quantities are simulated more accurately with the TIP4P model than with either SPC or SPC/E, simulations are computationally more demanding because of the additional site. In bilayer simulations the SPC model is therefore preferred.[99]

In calculations of hydration free energies of amino acid side chain analogues using different force fields and conditions, Tieleman and MacCallum[120] reported that both water models (SPC and TIP4P) give comparable results.

1.7 Factors important in Peptide-Lipid Interactions

1.7.1 Non-polar Effects

1.7.1.a Hydrophobic Effect and Mismatching

The hydrophobicity of amino acid residues in proteins and peptides is suggested to be one of the driving forces responsible for their association and insertion in lipid bilayers and membranes. Model peptides of known length, 3-dimensional structure and hydro-phobicity

can be used to assess the importance and strength of the hydrophobic effect by analysing their interactions with model membranes of various thickness and composition.

The hydrophobic lengths and mismatch effects of model peptides (*e.g.* the WALP and KALP polypeptide series) have been studied extensively both experimentally and computationally in the past.[127] The family of the WALP α-helical transmembrane peptides (Table 1.1) contains two flanking tryptophan residues (W) on either side of an alternating arrangement of alanine (A) and leucine (L) residues. The peptide inserts into the bilayers as shown schematically in Figure 1.10.

Peptide Family	General Sequence	Hydrophobic length[a] / nm
WALP	Ac-GWW$(LA)_{5\text{-}10.5}$WWA-Etn	1.5-3.75
KALP	Ac-GKKL$A_{8.5\text{-}12.5}$KKA-amide	2.55-3.75

Table 1.1: Amino acid sequences of the WALP and KALP peptides and calculated length of the leucine-alanine stretch when present as an α-Helix. [a]It is assumed that each amino acid has a length of 0.15 nm.

Figure 1.10: Schematic figure of a WALP23 α-helix orientation in a lipid bilayer. The bilayer-water interface is indicated by a dashed line. (Adapted from Ref. 127a)

The bilayer could adjust to mismatched peptides in various ways (as reviewed in Ref. 128), but the main response was found to be the exclusion of peptide material from the planar bilayer.[127a] The authors showed by circular dichroism, Fourier transform infrared (FTIR), and fluorescence spectroscopy that peptides are less likely to insert themselves into the bilayer, when they have a positive or negative hydrophobic mismatch. Depending on the hydrophobic mismatch of the length of the protein with respect to the depth and length of the lipid, non-bilayer structures were induced, whose nature depended on the precise extent of the mismatch. De Planque *et al.*[127d] determined by fluorescence measurements on a set of

WALP/phospholipid combinations that the inserted peptides adopt an orientation, in which the tryptophan side chains are located at the lipid-water interface, characterized by a broad maximum around 337 nm, typical for a relatively deep interfacial partitioning of the indole rings. This was further supported by a combination of mass spectrometry and hydrogen/deuterium exchange which showed that the tryptophan residues in WALP peptides in DMPC vesicles though buried in the bilayer region were still accessible for deuterium exchange over time.[129] It was concluded, that the tryptophan side chains preferred to reside in a rather narrow positional range at the membrane-water interface and resisted displacement in either direction away from that position. Energetically favourable interactions were thought to be more important than hydrophobic matching for this resistance.

1.7.1.b Lipophobic Effect

The lipophobic effect is the analogue of the hydrophobic effect. It is a purely empirical property of amino acid side chains which, according to their lipophilicity, will either try to avoid unfavourable contacts with lipid molecules or to be exposed to them.[130] A contribution to the lipophobic effect arises from an unfavourable loss of entropy of lipid molecules interacting with the peptide, instead of freely diffusing in the bilayer.

Tryptophan side chains, due to their rigidly planar geometry, may exhibit favourable localisation of the indole ring in the lipid bilayer. Such orientations were detected for tryptophan residues in WALP model peptides (see §1.7.1.a, Figure 1.10) by solid state NMR experiments[131], where it was shown that the indole moieties of the tryptophan residues resided approximately at the same level as the glycerol group and below the choline moiety and close to the carbonyl region of the DMPC phospholipid. The authors further report that the tryptophan side chains are oriented roughly parallel to the bilayer surface instead of pointing toward the aqueous phase (as Seelig *et al.* suggested[132]). Preferred localisations and orientations were further supported by the studies of small water-soluble mimics of the tryptophan side chain that were found to be positioned close to the lipid carbonyls using fluorescence quenching methods[73a] and lipid-directed and analogue-directed NMR studies[2,73b]. Recently however, the importance of the lipid carbonyls in anchoring tryptophan side chains in the bilayer was challenged by van der Wel *et al.*[133] The authors studied the interactions of WALP peptides with the ether-linked DMPC analogue ditetradecylphosphatidylcholine (DTPC), which misses the lipid carbonyls using solid-state

NMR. The very similar dynamics of the aromatic amino acid side chain in both, DMPC and DTPC bilayers suggested that the tryptophan side chains adopt similar orientations and that the lipid carbonyls therefore play at most a minor role in the anchoring of tryptophan side chains in the interfacial region.

1.7.2 Electrostatic Effects

1.7.2.a Hydrogen bonds

Even though the backbone of amino acids is often involved in the formation of secondary protein structure (helix and sheet), exposed segments in turns and loops could participate in hydrogen bonding interactions. In a typical lipid/water interface, possible candidates with which tryptophan residues can form hydrogen bonds are water molecules, the phosphate of the lipid head group and the carbonyl groups that link the acyl chains to the glycerol backbone of the ester lipids. Previously, measurements of the free energy of transfer of tryptophan analogues from water into cyclohexane suggested that the NH group in tryptophan was considerably less polar than expected which implied that hydrogen bond formation of the tryptophan indole NH to water was particularly weak.[134] In lipid bilayers, the importance and involvement of hydrogen bonding of the indole NH was previously investigated by comparing the binding behaviour of free NH analogues to both ether and ester lipid bilayers with that of analogues where the NH group was masked by an alkyl chain.[2] If hydrogen bonding to lipid carbonyls is important for the ordering of solubilised tryptophan analogues, a large decrease in the ordering of tryptophan and no appreciable effects on *N*-methyl tryptophan would be expect when the ester lipids are substitute for ether lipids. However, experiments showed a similar change in the ordering of indole and *N*-methyl indole as tryptophan analogues when the lipid type was altered, suggesting that hydrogen bonding interactions are not likely to play a major role in the interfacial preference of tryptophan.[2,3] This was supported by linear dichroism spectroscopy of various different indole derivatives (indole, 3-methylindole, 1-methylindole and tryptophan octyl ester) suggesting that the adopted orientation is independent of the protein the tryptophan residue is part of and that they are directed by dipole-dipole interactions and steric constraints in the membrane hydrocarbon region with hydrogen bonding and cation-π-interactions with lipid head groups being less important.[135]

1.7.2.b Cation-π interactions

Cation-π interactions are increasingly recognized as an important non-covalent binding interaction relevant to structural biology, where they are considered to be one of the driving forces in molecular recognition.[136] The main contribution to cation-π binding is the interaction between the quadrupole moment of aromatic rings and an external charge distribution. Additionally, cation-π interactions can also be affected by induced dipoles, polarizabilities, dispersion forces, and charge transfer effects.[137] In biological systems, the role of cation-π interactions in protein and peptide conformational states has been studied by means of interactions between positively charged lysine and arginine residues (the ammonium NH_3^+ and guanadinium $C(NH_2)_3^+$ groups, respectively) and the aromatic six-membered ring centres of phenylalanine, tyrosine, and tryptophan.[136d,138] Following work by Levitt and Perutz[139] who suggested a hydrogen bond between aromatic and amino groups, Burley and Petsko[140] identified the ''amino aromatic'' interaction, in which NH-containing groups tend to be positioned near aromatic rings within proteins. It is now appreciated that the interaction of a cationic group with an aromatic ring is much more favourable than the analogous interaction involving a neutral amine.[141,142] A cation-π interaction with an aromatic can be attractive or repulsive. The indole part of tryptophan may participate in attractive cation-π binding with positively charged residues of the lipids (such as the choline head group of phosphocholine lipids).

Theoretical and experimental studies have shown that cation-π interactions can be quite strong, both in the gas phase and in aqueous media, and a number of possible applications to biological systems have already been suggested.[136] In globular proteins, a cation-π interaction is more stabilizing than an analogous salt bridge.[143] The strengths of the interactions are strongly dependent on the dielectric constant of the medium, as would be expected for interactions with large electrostatic components. Interestingly, the cation-π-interaction maintains its strength across a range of solvents. Whereas the strength of the salt bridge is reduced over 50-fold on moving from the gas phase to water, the strength of a cation-π interaction is weakened by less than a factor of 3.[143]

The strength of cation-π-interactions of tryptophan molecules exceeds that of other aromatic amino acids.[136] Its strength towards the binding of choline has been estimated to be ~2 kcal mol^{-1} by studying the effects of substitutions on tryptophans indole ring and its implications for the ligand recognition by serotonergic (5-HT3A) and nicotinic acetylcholine receptors.[144] The authors used 4-fluoro-tryptophan as an electrostatically neutral benchmark.

Ratios of relative ligand binding of the derivative with respect to the natural tryptophan compound allowed an estimation of binding free energies, which the authors correlated to cation-π interactions.

The face of the aromatic ring can also act as an acceptor in hydrogen bonding interactions with donors such as amide NH groups[142,145] with considerable free energy contributions (-4.5 ± 0.5 kJ mol^{-1}).[146] Cation-π interactions can thus provide strong, specific interactions on the surfaces of proteins. Consistent with this prediction, cation-π interactions are rarely buried within proteins found in the Protein Data Bank, and more often than not, are exposed to water. The Protein Data Bank also yields that one favourable cation-π interaction can be expected for every 77 amino acid residues. This takes into account any interactions between an aromatic side chain (Phe, Tyr, or Trp) and a cationic one (Lys or Arg). A strong bias toward Trp was found, as over one-quarter (26%) of all tryptophans in the data bank experienced at least one energetically significant cation-π interaction.[147] Computer simulations had postulated that tryptophan would be overrepresented at cation-π-sites because, in the gas phase, indole binds cations more tightly than either benzene or phenol.[136,148]

The statistical significance and abundance of tryptophan residues in the water/lipid interfacial region in lipid bilayers was reviewed in detail by Caffrey *et al.*[149] Cation-π interactions for interfacial tryptophan residues were reported to depend on both the tryptophan position relative to the bilayer-water interface and on the chemical nature of the lipid headgroup.[150] In molecular dynamics simulations, ethanolamine headgroups (POPE) were favoured over choline headgroups (POPC), whereas the hydrogen-bonding ability between tryptophan and lipid carbonyls was found to be largely independent of the chemical nature of the headgroup.[150] Cation-π interactions have been suggested to be involved in the localization of indole in the choline region,[3] but the factors primarily responsible for the preference are still unknown. Recently, computer simulations suggested that electrostatic interactions are the main reason for the preferred localization of indoles.[73e]

Classical additive force fields perform poorly when quantitatively reproducing the strength of cation-π interactions, primarily because they do not include the explicit polarizability of the aromatic molecule and its bound cation.[151] However, Norman and Nymeyer[4] successfully reproduced the localisation of indoles at a POPC/water interface using a GROMACS G45A3 force field. The authors further demonstrated (by alternation of the indole partial charges) that non-specific electrostatic effects, cation-π and dipolar

interactions between the indole and the bilayer as well as hydrogen bonding to the bilayer carbonyl groups play a role in this preferred localisation.

1.7.2.c Dipolar interactions

The lipid interface is a region with varying sources of dipoles. The phosphatidylcholine headgroup carries a dipole moment of 18.5-25 D, oriented nearly parallel to the membrane surface.[152] Additional dipoles in the order of 2.0 D arise from the lipid carbonyls; the dipole of the carbonyl on the *sn*1 chain is oriented preferentially in the plane of the bilayer, while the carbonyl on the *sn*2 chain is oriented toward the water phase.[153] Furthermore, interfacial water shows a dipole moment of 1.8 D, with the hydrogen atoms of the first water layer pointed preferentially toward the membrane. This arrangement of dipoles results in a positive electrical potential in the membrane interior.[154]

Polarizability does not contribute significantly to the total dipole moment. Persson *et al.*[2] speculated that dipolar interactions may be dominant for the molecular ordering of tryptophan analogues. However, the situation may be different for transmembrane proteins with interfacially localized tryptophans, as in these cases the indole ring may be situated closer to the hydrocarbon interior, where it is less exposed to water.

It has been suggested that the indole dipole moment and hydrogen bonding do play a role in localization, but do not by themselves result in the interfacial preference.[2, 3, 73b]

1.8 References

[1] Singer, S. J., Nicholson, G. L., *Science*, **1972**, *175*, 720&

[2] Persson, S., Killian, S. A., Lindblom, G., *Biophys. J.*, **1998**, *75*(3), 1365-1371

[3] Gaede, H. C., Yau, W.-M., Gawrisch, K., *J. Phys. Chem. B.*, **2005**, *109*(26), 13014–13023

[4] Norman, K. E., Nymeyer, H., *Biophys. J.*, **2006**, *91*, 2046-2054

[5] a) Davis, M. E., McCammon,. J. A., *Chem. Rev.*, **1990**, *90*, 509–521
b) Mulgrew-Nesbitt, A., Diraviyam, K., Wang, J. Y., Singh, S., Murray, P., Li, Z. H., Rogers, L., Mirkovic, N., Murray, D., *Biochim. Biophys. Acta*, **2006**, *1761*(8), 812-826

[6] Stern, H. A., Feller, S. E., *J. Chem. Phys.*, **2003**, *118*, 3401–3412

[7] a) Marrink, S. J., Berendsen, H. J. C., *J. Phys. Chem.*, **1996**, *100*, 16729–16738
b) Jacobs, R. E., White, S. H., *Biochemistry*, **1989**, *28*(8), 3421–3437

[8] Majerus, P. W., Connolly, T. M., Deckmyn, H., Ross, T. S., Bross, T. E., Ishii, H., Bansal, V. S., Wilson, D. B., *Science*, **1986**, *234*, 1519-1526
[9] Hannun, Y. A., Bell, R. M., *Science*, **1987**, *235*, 670-674
[10] a) Vaara, M., *Microbiol. Rev.*, **1992**, *56*, 395-411
b) Ludtke, S., He, K., Huang, H., *Biochemistry*, **1996**, *34*, 16764-16769
c) Lohner, K., Latal, A., Leher, R. I., Ganz, T., *Biochemistry*, **1997**, *36*, 1525-1531
[11] Shai, Y. in *Protein-Lipid Interactions – New Approaches and Emerging Concepts*, Ed. Mateo, C. R., Gómez, J., Villalaín, González Ros, J. M., Springer, **2006**, 177-235
[12] Marsh, D., Horvath, L. I., *Biochim Biophys Acta*, **1998**, *1376*(3),267-296
[13] a) Seelig, J., *Biochim. Biophys. Acta*, **2004**, *1666*(1-2), 40-50
b) Heerklotz, H., *J Phys. Conds. Matter*, **2004**, *16*(15), R441-R461
[14] Martin, I., Goormaghtigh, E., Ruysschaert, J. M., *Biochim. Biophys. Acta*, **2003**, *1614*(1), 97-103
[15] London, E., Ladokhin, A. S., *in* Peptide-Lipid interactions, *Current Topics in Membranes*, Elsevier Academic Press, **2002**, *52*, 89-115
[16] Salditt, T., Li, C. H., Spaar, A., *Biochim. Biophys. Acta*, **2006**, *1758*(9), 1483-1498
[17] White, S. H., in Peptide Solvation and H-Bonds, *Advances in Protein Chemistry*, Elsevier Academic Press, **2006**, *72*, pp.157
[18] Bechinger, B., *Biochim. Biophys. Acta*. **1999**, *1462*(1-2),157-183
[19] Sanderson, J. M., *Org. Biomol., Chem.*, **2005**, *3*, 201-212
[20] Bond, P. J., Holyoake, J., Ivetac, A., Khalid, S., Sansom, M. S. P., *J. Struc. Biol.*, **2007**, *157*(3), 593-605
[21] Ash, W. L., Zlomislic, M. R., Oloo, E. O., Tieleman, D. P., *Biochim. Biophs. Acta*, **2004**, *1666*(1-2), 158-189
[22] Snyder, H. R., MacDonald, J. A., *J. Am. Chem. Soc.*, **1955**, *77*, 1257-1259
[23] Yokoyama, Y., Hikawa, H., Mitsuhashi, M., Uyama, A., Hiroki, Y., Murakami, Y., *Eur. J. Org. Chem.*, **2004**, *6*, 1244-1253
[24] a) Mouritsen, O. G., *Curr. Opin. Colloid Interface Sci.*, **1998**, *3*, 78-87
b) White, S. H., Wimley, W. C., *Annu. Rev. Biophys. Struct.*, **1999**, *28*, 319-365
c) Bloom, M., Evans, E., Mourtisen, O. G., *Q. Rev. Biophys.*, **1991**, *24*, 293-397
d) Gil, T., Ipsen, J. H., Mouritsen, O. G., Sabra, M. C., Sperotto, M. M., Zuckermann, M. J., *Biochim. Biophys. Acta* ,**1998**, *1376*, 245-266
[25] a) Baldwin, P. A., Hubbell, W. L., *Biochemistry*, **1985**, *24*, 2633-2639
b) Bienvenue, A., Marie, J. S., *Curr. Top. Membr.*, **1994**, *40*, 319-354
c) Litman, B. J., Mitchell, D. C., *Lipids*, **1996**, *31*, S193-S197
d) Dumas, F., Tocanne, J.-F., Leblanc, G., Lebrun, M.-C., *Biochemistry*, **2000**, *39*, 4846-4854
[26] Edidin, M., *Nature Reviews Mol. Cell Biol.*, **2003**, *4*, 414-418
[27] Gorter, E., Grendel, F., *J. Exp. Med.*, **1925**, *41*, 439-443
[28] Danielli, J. F., Davson, H., *J. Cell. Comp. Physiol.*, **1935**, *5*, 495-508
[29] Murray, R. K., Granner, D. K., Mayes, P. A., Rodwell, V. W., in *Harper's Illustrated Biochemistry*, 27th edition, McGraw-Hill, New York, **2006**, pp376
[30] Sakurai, M., Tamagawa, H., Inoue, Y., Ariga, K, Kunitake, T., *J. Phys. Chem. B*, **1997**, *101*(24), 4810-4816
[31] Stephen H. White, Irvine Laboratories, UC Irvine, USA
[32] White, S. H., Wimley, W. C., *Curr. Opin. Struc. Biol.*, **1994**, *4*, 79-86
[33] White, S. H., Wimley, W. C., *Biochim. Biophys. Acta*, **1998**, *1376*, 339-352
[34] Hristova, K., Wimley, W. C., Mishra, V. K., Anantharamiah, G. M., Segrest, J. P., White, S. H., *J. Mol. Biol.*, **1999**, *290*, 99-117
[35] Bohn, M., Heinz, E., Lüthje, S., *Arch. Biochem. Biophys.*, **2001**, *387*, 35-40
[36] Holleran, W. M., *J. Clin. Invest.*, **1993**, *91*, 1656-1664
[37] Lester, R. L., Wells, G. B., Oxford, G., Dickson, R. C., *J. Biol. Chem.*, **1993**, 268, 845-856
[38] Faure, C., Bonakdar. L, Dufourc, E. J., *FEBS Letters*, **1997**, *405*, 263-266
[39] Cevc, G., Marsh, D., in Phospholipid Bilayers: Physical Principles and Models, John Wiley & Sons, New York, **1987**
[40] http://www.britannica.com/eb/art-73016/Phospholipid-molecules-composed-of-fatty-acid-tails-and-a-phosphate
[41] Haque, R., Tinsley, I. J., Schmedding, D., *J. Biol. Chem.*, **1972**, *247*(1), 157-161
[42] Kovac, A., Kidric, J., Hadzi, D., *J. Mol. Struct.*, **1992**, *267*, 399-404
[43] Wallin, E., von Heijne, G., *Protein Sci.*, **1998**, *7*(4), 1029-1038
[44] Jones, D. T., *FEBS Lett.*, **1998**, *423*(3), 281-285
[45] Gouaux, E., *Curr. Opin. Struct. Biol.*, **1997**, *7*(4), 566-573
[46] Do, H., Falcone, D., Lin, J. L., Andrews, D. W., Johnson, A. E., *Cell*, **1996**, *85*(3), 369-378

[47] Herrmann, K. M., Somerville, R. L., *in* Amino Acids: Biosynthesis and Genetic Regulation, Addison-Wesley Publishing Company, **1983**, Reading, MA., pp.352-378
[48] Yanosfsky, C., Drapeau, G. R., Guest, J. R., Carlton, B. C., *Proc. Nat. Acad. Sci. U.S.A.*, **1967**, *57*, 296- 298
[49] Nichols, B. P., van Cleemput, M., Yanofsky, C., *J. Mol. Biol.*, **1981**, *146*, 45-54
[50] Fowler, A. V., Zabin, I., *Proc. Nat. Acad. Sci. U.S.A.*, **1977**, *74*, 1507-1510
[51] Atkinson, D. E., *in* Cellular Energy Metabolism and Its Regulation, **1977**, Academic Press, New York
[52] Fauchere, J. L., Pliska, V., *Eur. J. Med. Chem.*, **1983**, *18*, 369 –375
[53] Radzika, A., Wolfenden, R., *Biochemistry*, **1988**, *27*, 1664 –1670
[54] Deisenhofer, J., and Michel, H., Nobel lecture, *EMBO J.* **1989**, *8*, 2149-2170
[55] Raman, P., Cherezov, V., Caffrey, M., *Cell Mol. Life Sci.*, **2005**, *63*, 36-51
[56] Schiffer, M., Chang , C.-H., Stevens. F. J., *Protein Eng.*, **1992**, *5*, 213–214
[57] Doyle, D. A., Cabral, J. M., Pfuetzner, R. A., Kuo, A., Gulbis, J. M., Cohen, S. L., Chait, B. T., MacKinnon, R., *Science*, **1998**, *280*, 1268-1277
[58] Schirmer, T., Keller, D. A., Wang, Y. F., Rosenbusch, J. P., *Science*, **1995**, *267*, 512-514
[59] Luecke, H., Schobert, B., Richter, H.T., Cartailler, J.-P., Lanyi, J. K., *Science*, **1999**, *286*, 255-261
[60] a) Reithmaier, R. A. F., *Curr. Opin. Struct. Biol.*, **1995**, *5*, 491-500
b) Adamian, L., Nanda, V., DeGrado, W. F., Liang, J., *Proteins*, **2005**, *59*, 496-509
c) Liu, W., Caffrey, M., *Biochemistry*, **2006**, 45(36), 11713-11726
[61] Raghuraman, H., Kelkar, D. A., Chattopadhyay, A., in *Reviews in Fluoresence*, Springer, New York, **2005**, Vol. 2., pp 199-222
[62] Durkin, J. T., Providence, L. L., Koeppe II, R. E., Andersen. O. S., *Biophys. J.*, **1992**, *62*, 145–159
[63] O'Connell, A. M., R. E. Koeppe II, and O. S. Andersen, *Science.*, **1990**, *250*, 1256 –1259
[64] Becker, M. D., Greathouse, D. V., Koeppe II, R. E., Andersen, O. S., *Biochemistry*, **1991**, *30*, 8830–8839
[65] Hu, W., Cross, T., *Biochemistry*, **1995**, *34*, 14147–14155
[66] Killian, J. A., Salemink, I., de Planque, M. R. R., Lindblom, G., Koeppe II, R. E., Greathouse, D. V., *Biochemistry*, **1996**, *35*, 1037–1045
[67] Morein, S., Strandberg, E., Killian, A., Persson, S., Arvidson, G., Koeppe II, R. E., Lindblom, G., *Biophys. J.*, **1997**, *73*, 3078 –3088
[68] Van der Wel, P. C. A., Reed, N. D., Greathouse, D. V., Koeppe, R. E., *Biochemistry*, **2007**, *46*(25)7514-7524
[69] Andersen, O. S., Greathouse, D. V., Providence, L. L., Becker, M. D., Koeppe, R. E., *J. Am. Chem. Soc.*, **1998**, *120*(21), 5142-5146
[70] a) Miller, A. S., Falke, J. J., Biochemistry, 2004, 43, 1763-1770
b) Draheim, R. R., Bormans, A. F., Lai, R. Z., Manson, M. D., *Biochemistry*, **2005**, *44*, 1268-1277
[71] a) Braun, P., von Heijne, G., *Biochemistry*, **1999**, *38*, 9778–9782
b) Whiles, J. A., Glover, K. J., Vold, R. R., Komives, E. A., *J. Magn. Reson.*, **2002**, *158*, 149–156
[72] Shank, L. P., Broughman, J. R., Takeguchi, W., Cook, G., Robbins, A. S., Hahn, L., Radke, G., Iwamoto, T., Schultz, B. D., Tomich, J. M., *Biophys. J.*, **2006**, *90*, 2138-2159
[73] a) Kachel, K., Ascuncion-Punzalan, E., London, E., *Biochemistry*, **1995**, *34*(47), 15475-15479
b) Yau, W.-M., Wimley, W. C., Gawrisch, K., White, S. H., *Biochemistry*, **1998**, *37*(42), 14713-14718
c) Killian, J. A., von Heijne, G., *Trends Biochem. Sci.*, **2000**, *25*, 429-434
d) de Planque, M. R., Bonev, B. B., Demmers, J. A., Greathouse, D. V., Koeppe, R. E., II, Separovic, F., Watts, A., Killian, J. A., *Biochemistry*, **2003**,*42*, 5341-5348
e) Norman, K. E., Nymeyer, H., *Biophys. J.*, **2006**, *91*, 2046-2054
[74] Wilcox, M., *Anal. Biochem.*, **1974**, *59*, 436-440
[75] Goss, R. J. M., Newill, P. L. A., *Chem. Comm.*, **2006**, 4924-4925
[76] Hall, A. N., Lea, D. J., Rydon, H. N., *Biochem. J.*, **1962**, *84*, 12-16
[77] Yanofsky, C., *in* Methods in Enzymology, Vol. II, **1955**, pp.233-268, Academic Press, New York
[78] Sui, Y., Liu, L., Zhao, J.-L., Wang, D., Chen, Y.-J., *Tet. Lett.*, **2007**, *48*, 3779-3782
[79] Zembower, D. E., Ames, M. M., *Synthesis*, **1994**, 1433-1436
[80] Zhu, J., Jia, Y., *J. Org. Chem.*, **2006**, *71*(20), 7826-7834
[81] Dalton, D. R., in *The Alkaloids*; Marcel Dekker, Inc., New York, **1979**, pp415-628
[82] Anderson, M., *Science*, **1945**, *101*, 565
[83] Van Essevldt, B. C., Van Delft, F. L., Smits, J. M., De Gelder, R., Schoemaker, H. E., Rutjes, F. P., *Adv. Synth. Catal.*, **2004**, *346*, 823
[84] Kornfeld, E. C., Sheneman, J. M., Suarez, T., German Patent 1,917,844 (**1969**); *Chem Abstr.*, **1970**, *72*, 30438c
[85] Ladokhin, A. S., Jayasinghe, S., White, S. H., *Anal. Biochem*, **2000**, *285*, 235-245
[86] Loura, L. M. S., deAlmeida, R. F. M., Coutinho, A., Prieto, M., *Chem Phys. Lipids*, **2003**, *122*, 77-96

[87] Zhang, L., Kao, Y.T., Qiu, W., Wang, L., Zhong, D., *J. Phys. Chem. B.*, **2006**, *110*, 18097-18103
[88] a) Zhong, D., Pal., S. K., Zhang, D., Chan, S. I., Zewail, A. H., *Proc. Natl. Acad. Sci. U.S.A.*, **2002**, *99*, 13
b) Pal, S. K., Zewail, A. H., *Chem. Rev.*, **2004**, *104*, 2099
[89] Wimley, W. C., Hristova, K., Ladokhin, A. S., Silvestro, L., Axelsen, P. H., White, S. H., *J. Mol. Biol.*, **1998**, *277*(5), 1091-1110
[90] Wimley, W. C., Creamer, T. P., White, S. H., *Biochemistry*, **1996**, *35*(16), 5109-5124
[91] White, S. H., Wimley, W. C., *Biochim. Biophys. Acta*, **1998**, *1376*, 339-352
[92] Ladokhin, A. S., Wimley, W. C., White, S. H., *Biophys. J.*, **1995**, *69*(5), 1964-1971
[93] Wimley, W. C., White, S. H., *Nature Struct. Biol.*, **1996**, *3*(10), 842-848
[94] Morse, P. M., *Phys. Rev.*, **1929**, *34*, 57-64
[95] Tieleman, D. P., Marrink, S, J., Berendsen, H. R. J., *Biochim. Biophys. Acta*, **1997**, *1331*, 235-270
[96] Weiner, S.J., Kollman, P.A., Case, D.A., Singh, U.C., Ghio, C., Alagona, G., Profeta, S., Weiner, P., *J. Am. Chem. Soc.*, **1993**, *106*, 765-784
[97] Brooks, B.R., Bruccoleri, R.E., Olafson, B.D., States, D.J., Swaminathan, S., Karplus, M., *J. Comp. Chem.*, **1983**, *4*, 187-217
[98] van Gunsteren, W.F., Berendsen, H.J.C., **1987**. Groningen Molecular Simulation (GROMOS) Library Manual. Biomos, Nijenborgh 4, 9747 AG Groningen, the Netherlands
[99] Berger, O., Edholm, O., Jähnig, F., *Biophysical J.*, **1997**, *72*, 2002-2013
[100] Jorgensen, W. L., Maxwell, D. S., TiradoRives, J., *J. Am. Chem. Soc.*, **1996**, *118*, 11225-11236
[101] Tieleman, D. P., MacCallum, J. L., Ash, W. L., Kandt, C., Xu, Z., Monticelli, L., *J. Phys.: Condens. Matter*, **2006**, *18*, S1221–S1234
[102] Exemplified by a) Alper, H.E., Bassolino, D., Stouch, T.R., *J. Chem. Phys.*, **1993**, *98*, 9798-9807
b) Bassolino-Klimas, D., Alper, H.E., Stouch, T.R., *Biochemistry*, **1993**, *32*, 12624-12637
c) Venable, R., Zhang Y., Hardy, B., Pastor, R.W., *Science*, **1993**, *262*, 223-226
d) Feller, S.E., Zhang Y., Pastor, R.W., *J. Chem. Phys.*, **1995**, *103*, 10267-10276
e) Tu, K., Tobias, D., Klein, M., *Biophys. J.*, **1995**, *69*, 2558-2562
[103] Exemplified a) Raghavan, M., Reddy, M.R., Berkowitz, M.L., *Langmuir*, **1992**, *8*, 233-240
b) Heller, H., Schaefer, M., Schulten, K., *J. Phys. Chem.*, **1993**, *97*, 8343-8360
c) Essex, J., Hann, M., Richards, W., *Phil. Trans. R. Soc. Lond. B.*, **1994**, *344*, 239-260
d) Zhou, F., Schulten, K., *J. Phys. Chem.*, **1995**, *99*, 2194-2207
[104] Exemplified by a) Toxvaerd, S., *J. Chem. Phys.*, **1990**, *93*, 4290-4295
b) Peters, G.H., Toxvaerd, S., Svendson, A., Olsen, H., *J. Chem. Phys.*, **1994**, *100*, 5996-6010
[105] Zhang Y., Feller, S.E., Brooks, B.R., Pastor, R.W., *J. Chem. Phys.*, **1995**, *103*, 10252-10266
[106] Chiu, S., Clark, M., Balaji, V., Subramaniam, S., Scott, H., Jakobsson, E., *Biophys. J.*, **1995**, *69*, 1230-1245
[107] Tieleman, D. P., Berendsen, H. J. C., *J. Chem. Phys.*, **1996**, *105*(11), 4871-4880
[108] Egberts, E., Marrink, S., Berendsen, H.J.C., *Eur. Biophys. J.*, **1993**, *22*, 423-436
[109] MacKerell, A. D., *J. Comput. Chem.*, **2004**, *25*, 1584–1604
[110] MacKerell, A. D., *Computational Biochemistry and Biophysics*, Marcel Dekker, New York., **2001**, 7–38
[111] Minoux, H., Chipot, C., *J. Am. Chem. Soc.*, **1999**, *121*(44), 10366–10372
[112] Woolf, T. B., Grossfield, A., Pearson, J. G., *Int. J. Quant. Chem.* **1999**, *75*, 197–206
[113] Anézo, C., de Vries, A. H., Höltje, H. D., Tieleman, D. P., Marrink, S. J., *J. Phys. Chem.*, **2003**, *107*, 9424-9433
[114] Patra, M., Karttunen, M., Hyvonen, M. T., Falck, E., Vattulainen, I., *J. Phys. Chem. B.*, **2004**, *108*(14), 4485-4494
[115] a) Hess, B., *J. Chem. Phys.*, **2002**, *116*, 209-217
b) Feller, S.E., Pastor, R.W., Rojnuckarin, A., Bogusz, S., Brooks, B.R., *J. Chem. Phys.*, **1996**, *100*, 17011-17020
c) Hünenberger, P.H., van Gunsteren, W.F., *J. Chem. Phys.*, **1998**, *108*, 6117-6134
d) Mark, P., Nilsson, L., *J. Comp. Chem.*, **2002**, *23*, 1211-1219
[116] Smith, P.E., Blatt, H.D., Pettit, B.M., *J. Chem. Phys. B.*, **1997**, *101*, 3886-3890
[117] Weber, W., Hünenberger, P.H., McCammon, J.A., *J. Phys. Chem. B.*, **2000**, *104*, 3668-3675
[118] Tironi, I.L., van Gunsteren, W.F., *Mol. Phys.*, **1994**, *83* (2), 381-403
[119] Dietz, W., Heinzinger, K., *Ber. Bunsenges. Phys. Chem.*, **1984**, *88*, 543-546
[120] MacCallum, J.L., Tieleman, D.P., *J. Comp. Chem.*, **2003**, *24*, 1930-1935
[121] Berendsen, H.J.C., Postma, F.P.M., van Gunsteren, W.F., Hermans, J., *Intermolecular Forces*, edited by B. Pullman (Reidel, Dordrecht, **1981**), pp. 331-342
[122] Berendsen, H. J. C., Grigera, J., Straatsma, T., *J. Phys. Chem.*, **1987**, *91*, 6269-6271

[123] van Buuren, A.R., Marrink, S.J., Berendsen, H.J.C., *J. Phys. Chem.*, **1993**, *97*, 9206-9212
[124] Jorgensen, W.L., *J. Am. Chem. Soc.*, **1981**, *103*, 335-340
[125] Hermans, J., Pathiaseril, A., Anderson, A., *J. Am. Chem. Soc.*, **1988**, *110*, 5982-5986
[126] Guissani, Y., Guillot, B., *J. Chem. Phys.*, **1993**, *98*, 8221-8235
[127] a) De Planque, M. R. R., Goormaghtigh, E., Greathouse, D. V., Koeppe, R. E., Kruijtzer, J. A. W., Liskamp., R. M. J., De Kruijff, B., Killian, J. A., *Biochemistry*, **2001**, *40*(16), 5000-5010
b) Strandberg, E., Morein, S., Rijkers, D. T. S., Liskamp, R. M J., van der Wel, P. C. A., Killian, J. A., *Biochemistry*, **2002**, *41*, 7190-7198
c) Nyholm, T. K. M., Özdirekcan, S., Killian, J. A., *Biochemistry*, **2007**, *46*(6), 1457-1465
d) De Planque, M. R. R., Killian, J. A., *Mol. Membr. Biol.*, **2003**, *20*, 271-284
e) Kandasamy, S. K., Larson, R. G., Biophys. J., 2006, 90, 2326-2343
f) Petrache, H. I., Zuckerman, D. M., Sachs, J. N., Killian, J. A., Koeppe, R. E., Woolf, T. B., *Langmuir*, **2002**, *18*, 1340-1351
[128] Killian, J. A., *Biochim. Biophys. Acta*, **1998**, *1376*, 401-416
[129] Demmers, J. A. A., van Duijn, E., Haverkamp, J., Greathouse, D. V., Koeppe, R. E., Heck, A. J. R., Killian, J. A., *J. Biol. Chem.*, **2001**, *276*(37), 34501-34508
[130] Adamian, L., Nanda, V., DeGrado, W. F., Liang, J., *Proteins*, **2005**, *59*, 496-509
[131] Van der Wel, P. C. A., Reed, N. D., Greathouse, D. V., Koeppe, R. E., *Biochemistry*, **2007**, *46*, 7514-7524
[132] Seelig, J., MacDonald, P. M., Scherer, P. G., *Biochemistry*, **1987**, *26*(24), 7535-7541
[133] Van der Wel, P. C. A., Reed, N. D., Greathouse, D. V., Koeppe, R. E., *Biochemistry*, **2007**, *46*(25), 7514-7524
[134] Wimley, W. C., White. S. H., *Biochemistry*, **1992**, *31*(51), 12813–12818
[135] Esbjörner, E. K., Caesar, C. E. B., Albinsson, B., Lincoln, P., Nordén, B., *Biochim. Biophys. Acta*, **2007**, *361*(3), 645-650
[136] a) Dougherty, D. A., *Science*, **1996**, *271*(5246), 163–168
b) Ma, J. C., Dougherty, D. A., *Chem. Rev.*, **1997**, *97*(5), 1303–1324
c) Scrutton, N. S., Raine, A. R. C., *Biochem. J.*, **1996**, *319*(1), 1–8
[137] Caldwell, J. W., Kollman, P. A., *J. Am. Chem. Soc.*, **1995**, *117*(14), 4177– 4178
[138] Aliste, M. P., MacCallum, J. L., Tieleman, D. P., *Biochemistry*, **2003**, *42*(30), 8976–8987
[139] a) Perutz, M. F., Fermi, G., Abraham, D. J., Poyart, C., Bursaux, E., *J. Am. Chem. Soc.*, **1986**, *108*(5), 1064–1078
b) Levitt, M., Perutz, M. F., *J. Mol. Biol.*, **1988**, *201*(4), 751–754
c) Perutz, M. F., *Philos. Trans. R. Soc. London A*, **1993**, *345*(1674), 105–112
[140] Burley, S. K., Petsko, G. A., *FEBS Lett.*, **1986**, *203*(2), 139–143
[141] Deakyne, C. A., Meot-Ner, M., *J. Am. Chem. Soc.*, **1985**, *107*(2), 474–479
[142] Rodham, D. A., Suzuki, S., Suenram, R. D., Lovas, F. J., Dasgupta, S., Goddard, W. A., III, Blake, G. A., *Nature*, **1993**, *362*, 735–737
[143] Gallivan, J.P., Dougherty, D. A., *J. Am. Chem. Soc.*, **2000**, *122*, 870-874
[144] Beene, D. L., Brandt, G. S., Zhong, W. G., Zacharias, N. M., Lester, H. A., Dougherty, D. A., *Biochemistry*, **2002**, *41*(32), 10262-1069
[145] Fong, T. M., Cascieri, M. A., Yu, H., Bansal, A., Swain, C., Strader, C. D., *Nature*, **1993**, *362*, 350
[146] Adams, H., Harris, K. D. M., Hembury, G. A., Hunter, C. A., Livingstone, D., McCabe, J. F., *Chem. Commun.*, **1996**, 2531-2532
[147] Gallivan, J. P., Dougherty, D. A., *Proc. Natl. Acad. Sci.*, **1999**, *96*, 9459–9464
[148] Mecozzi, S., West, A. P., Jr., Dougherty, D. A., *J. Am. Chem. Soc.*, **1996**, *118*, 2307–2308
[149] Liu, W., Caffrey, M., *Biochemistry*, **2006**, 45(36), 11713-11726
[150] Petersin, F. N. R., Jensen, M. Ø., Nielson, C. H., *Biophys. J.*, **2005**, *89*, 3985-3996
[151] a) Caldwell, J. W., Kollman, P. A., *J. Am. Chem. Soc.*, **1995**, *117*, 4177–4178
b) Cubero, E., Luque, F. J., Orozco, M., *Proc. Natl. Acad. Sci. USA.*, **1998**, *95*, 5976–5980
c) Woolf, T. B., Grossfield, A., Pearson, J. G., *Int. J. Quantum. Chem.*, **1999**, *75*, 197–206
[152] Seelig, J., *Biochim. Biophys. Acta*, **1978**, *515*(2), 105-140
[153] Pearson, R. H., Pascher, I., *Nature*, **1979**, *281*(5731), 499-501
[154] Gawrisch, K., Ruston, D., Zimmerberg, J., Parsegian, V. A., Rand, R. P., Fuller, N., *Biophys J.*, **1992**, *61*(5), 1213-1223

Chapter 2: Synthesis of Tryptophan Derivatives and Analysis of their Binding to DMPC

2.1 Overview

The results of the experimental and computational analysis of the binding interactions of a series of 5-monosubstituted N^{α}-acetyl tryptophan ethyl amide derivatives (**6a-g, 6n**) with DMPC are presented and discussed in this chapter. First, the synthesis of the products used will be reported (§2.2). Subsequently, the data from the experimental techniques used to quantify binding interactions will be presented, starting with NMR host-guest titrations. The binding models used in the data analysis of the titrations (§2.3.1), the influence and effects of water in the system (§2.3.2.a) and micellisation of DMPC in chloroform (§2.3.2.b) will be evaluated and discussed. This follows an analysis of NMR titration experiments of tryptophan derivatives (host) and DMPC lipid (guest) in chloroform (§2.3.3). The observations made in the titrations will be presented in §2.3.3.a, followed by an investigation into the binding stoichiometry (§2.3.3.b), data analysis and discussion (§2.3.3.c). Kinetic effects and the importance of hydrogen bonding in the amino acid-lipid interactions are investigated using deuterium exchange NMR titrations (§2.3.5).

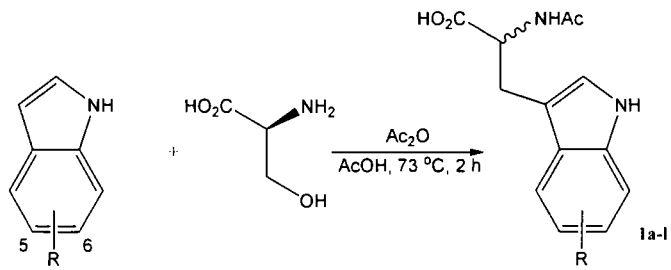
The results of the second analytical technique used in this study (isothermal titration calorimetry) are presented in §2.4. After a general introduction to the technique and the theory of the data analysis (§2.4.1), results for the titration of hydrated DMPC into chloroform (§2.4.2.a) to assess the critical micelle concentration (CMC) of DMPC in this solvent (§2.4.2.b) and for host-guest titrations of concentrated lipid solutions into tryptophan derivative containing chloroform solutions are shown (§2.4.2.c).

Finally, binding events were investigated using molecular dynamics simulations of the system (§2.5). Various adduct formation of the tryptophan derivatives (**6a-g, 6n**) with lipid molecules (§2.5.2 & §2.5.4) and water (§2.5.3) as well as the adduct formation of two amino acid derivatives (**6a**) themselves (§2.5.5) (without any lipid present) will be presented and critically analysed. Free energy calculations of some of the binding interactions are evaluated and compared to empirical results in §2.6. §2.7 analyses peptide-lipid interactions as studied *via* model peptides (of the general type AcWLWLL) and DMPC lipid bilayers in aqueous solutions by molecular dynamics simulations.

2.2 *Synthesis of Tryptophan Derivatives*

2.2.1 Synthesis of Racemic Tryptophan Derivatives (1a-l)

A series of racemic N^{α}-acetyl tryptophan derivatives was produced by reaction of the corresponding indoles with *L*-serine in a mixture of acetic acid and acetic anhydride (Scheme 2.1).



Scheme 2.1: Reaction of substituted indoles with *L*-serine to produce racemic N^{α}-acetyl-tryptophan compounds (**1a-l**).

Products were isolated in good yields (Table 2.1) and cover a broad spectrum of electron density on the indole ring, from electron rich (Table 2.1, entries **V** and **XI**; R = MeO) to electron poor (Table 2.1, entries **VI** and **XII**; R = NO_2), demonstrating the applicability of this reaction to the preparation of a broad range of analogues.

Entry	R	Pos.	[Serine][a]	Yield (%)	Entry	Pos.	Yield (%)
I	F	C5	2.0	82 (**1a**)	**VII**	C6	98 (**1g**)
II	Cl		2.0	88 (**1b**)	**VIII**		86 (**1h**)
III	Br		2.0	77 (**1c**)	**IX**		73 (**1i**)
IV	I		1.5	84 (**1d**)	**X**		73 (**1j**)
V	MeO		1.0	92 (**1e**)	**XI**		50 (**1k**)
VI	NO_2		2.5	44 (**1f**)	**XII**		52 (**1l**)

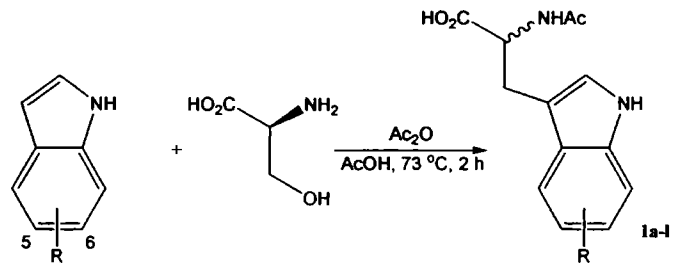
Table 2.1: Data for the reaction of monosubstituted indoles with *L*-serine. [a]molar equivalents of *L*-serine with respect to indole.

The relative success of the electrophilic substitution depends on the electron density distribution on the indole ring. The less electron rich nitro-substituted indoles were of lower reactivity and gave moderate yields. More electron rich indoles were found to be highly reactive towards electrophilic attack, leading to double-substitution at positions C2 and C3 (see also §2.2.3). Reaction yields were therefore maximised by adjusting the stoichiometry accordingly, either by increasing or decreasing the amount of indole for electron deficient and electron rich compounds respectively. The highest yields were observed for substituents

2.2 *Synthesis of Tryptophan Derivatives*

2.2.1 Synthesis of Racemic Tryptophan Derivatives (1a-l)

A series of racemic N^{α}-acetyl tryptophan derivatives was produced by reaction of the corresponding indoles with *L*-serine in a mixture of acetic acid and acetic anhydride (Scheme 2.1).



Scheme 2.1: Reaction of substituted indoles with *L*-serine to produce racemic N^{α}-acetyl-tryptophan compounds (**1a-l**).

Products were isolated in good yields (Table 2.1) and cover a broad spectrum of electron density on the indole ring, from electron rich (Table 2.1, entries **V** and **XI**; R = MeO) to electron poor (Table 2.1, entries **VI** and **XII**; R = NO_2), demonstrating the applicability of this reaction to the preparation of a broad range of analogues.

Entry	R	Pos.	[Serine][a]	Yield (%)		Entry	Pos.	Yield (%)
I	F	C5	2.0	82 (**1a**)		**VII**	C6	98 (**1g**)
II	Cl		2.0	88 (**1b**)		**VIII**		86 (**1h**)
III	Br		2.0	77 (**1c**)		**IX**		73 (**1i**)
IV	I		1.5	84 (**1d**)		**X**		73 (**1j**)
V	MeO		1.0	92 (**1e**)		**XI**		50 (**1k**)
VI	NO_2		2.5	44 (**1f**)		**XII**		52 (**1l**)

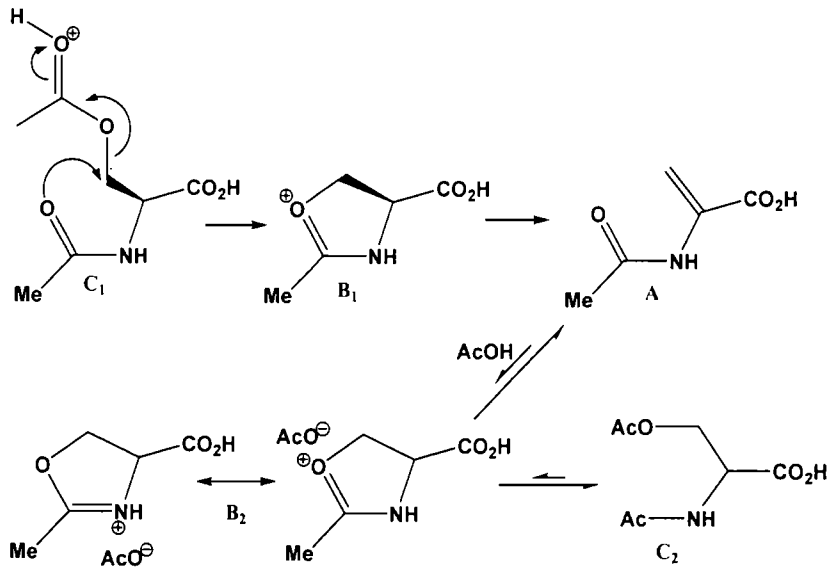
Table 2.1: Data for the reaction of monosubstituted indoles with *L*-serine. [a]molar equivalents of *L*-serine with respect to indole.

The relative success of the electrophilic substitution depends on the electron density distribution on the indole ring. The less electron rich nitro-substituted indoles were of lower reactivity and gave moderate yields. More electron rich indoles were found to be highly reactive towards electrophilic attack, leading to double-substitution at positions C2 and C3 (see also §2.2.3). Reaction yields were therefore maximised by adjusting the stoichiometry accordingly, either by increasing or decreasing the amount of indole for electron deficient and electron rich compounds respectively. The highest yields were observed for substituents

with intermediate electron releasing ability (Table 2.1, entries **II** & **III**, and **VII** & **VIII**; R = Cl, F). The total conversion of indole to product could be further improved by recycling unreacted indole recovered from the reaction.

2.2.2 Racemisation

A drawback of the substitution reaction (Scheme 2.1) is racemisation of the amino acid stereocentre. As the reaction between indole and *L*-serine is conducted under acidic conditions, it is probable that the observed racemisation precedes neither *via* oxazolone formation, nor *via* a classical direct enolisation, both of which require the presence of base. On the other hand, azlactone formation is promoted under the conditions used in this synthesis, which will undoubtedly contribute to the formation of racemates.[1] It also seems likely that racemisation can occur during the formation of reactive intermediates from serine. Interestingly, *N*-acetyl dehydroalanine (**A**, Scheme 2.2) was isolated as a by-product in reactions with less reactive indoles.



Scheme 2.2: Proposed equilibria established by the reaction of *N*-acetyl *L*-serine with acetic anhydride in acetic acid.

This is consistent with the known reaction between indole and **A** to produce tryptophan, which occurs under acidic conditions.[2,3] Formation of **A** from either *L*-serine or *D,L*-serine by heating with acetic anhydride in acetic acid has been described.[10] However, on repeating this reaction we found the major component of the mixture to be *N,O*-diacetyl *D,L*-serine (**C_2**).

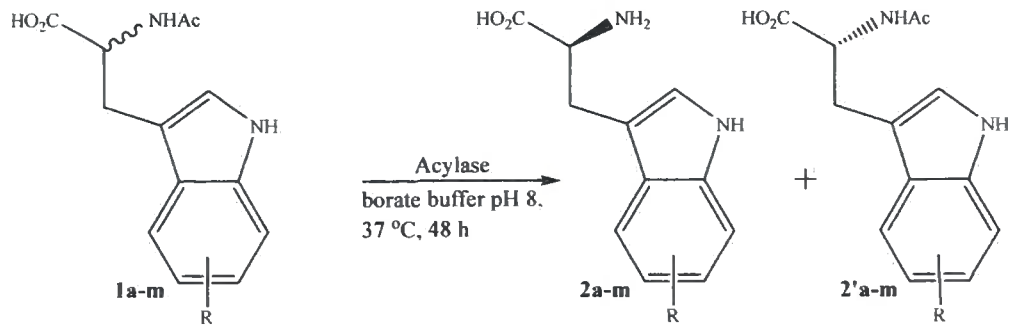
A plausible mechanism for the formation of **A** is through acid-promoted loss of acetic acid from the acetylated *L*-serine intermediate $\mathbf{C_1}$ (Scheme 2.2). This may be assisted by neighbouring group participation from the *N*-acetyl group, leading to the formation of the cationic species $\mathbf{B_1}$ as a transient intermediate, before loss of a proton to form **A**. These kinds of neighbouring group effects have been observed for serine in the gas phase[4] and are likely to persist to a lesser extent in solution. Under the conditions used, *N*-acetyl dehydroalanine would be susceptible to addition of acetic acid across the olefinic double bond to form $\mathbf{C_2}$ (the racemic counterpart of $\mathbf{C_1}$). Protonation of the olefinic bond of **A** as a precursor to addition of acetic acid will form $\mathbf{B_2}$ (the racemic counterpart of $\mathbf{B_1}$), with neighbouring group participation again helping to stabilise the cationic species. It may be anticipated that reaction of indoles with the electrophile $\mathbf{B_2}$ is the product forming step in the conversion to tryptophan analogues. Under the experimental conditions used, the steady-state concentration of the electrophile $\mathbf{B_2}$ is expected to be extremely low, but nevertheless sufficient for product formation. Direct reaction of *N*-acetyl dehydroalanine (**A**) with indoles in the presence of acetic anhydride and acetic acid tends to give significantly lower yields than the corresponding reaction with serine,[5] which suggests that the equilibrium between **A** and $\mathbf{B_2}$ lies in favour of the former. This would also suggest that a significant amount of product from the reaction arises from reaction with $\mathbf{B_1}$ and is subjected to subsequent racemisation *via* an azlactone route.

2.2.3 Multiple Electrophilic Substitutions

Electron rich indole derivatives were found to be susceptible to multiple substitutions. In addition to the 3-position, the 2-position was prone for an electrophilic attack from both *L*-serine compounds and other suitable electrophiles (such as excess acetic anhydride). Doubly substituted products were observed in particular for the reaction with methoxyindoles. These side products were preliminarily identified *via* NMR and mass spectrometry, but being only of minor interest in the scope of the study, a full characterisation was not pursued.

2.2.4 Enzymatic Resolution of Racemic Tryptophan Derivatives (2a-m, 2'a-m)

The racemic N^{α}-acetyl tryptophan derivatives were then optically resolved using an enzymatic approach (Scheme 2.3).



Scheme 2.3: Enzymatic resolution of tryptophan derivatives (**1a-m**) using acylase Amano in buffered conditions.

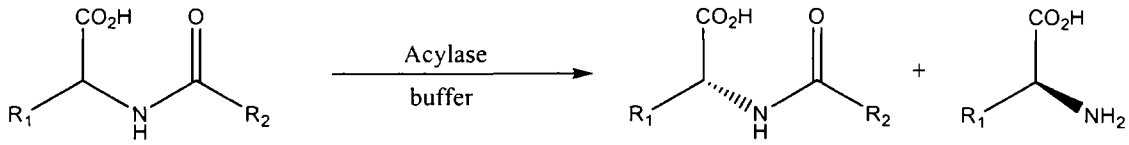
Enzymatic resolution of the racemates using Acylase "Amano" made both the free *L*-amino acid (**2**) and the corresponding N^{α}-acetyl *D*-enantiomer (**2'**) readily available (Table 2.2). The yield of **2'** exceeds the theoretical yield slightly as the crude product contained traces of starting material, as determined by optical rotation measurements. However, isolation of the N^{α}-acetyl *D*-enantiomer (**2'**) provided a means for preparing similar analogues of the naturally less abundant stereoisomer. In addition, **2'** could potentially be racemised and recycled to produce more *L*-enantiomer (**2**) following enzymatic resolution.

Entry	R	Pos.	Yield 2 (%)[a]	ee[b]	Yield 2' (%)[a]
I	F	C5	99 (**2a**)	100	76 (**2'a**)
II	Cl		80 (**2b**)	100	>100* (**2'b**)
III	Br		96 (**2c**)	100	83 (**2'c**)
IV	I		79 (**2d**)	-	93 (**2'd**)
V	Me		98 (**2e**)	89	98 (**2'e**)
VI	MeO		99 (**2f**)	100	95 (**2'f**)
VII	NO_2		68 (**2g**)	92	>100* (**2'g**)
VIII	F	C6	99 (**2h**)	91	74 (**2'h**)
IX	Cl		63 (**2i**)	91	84 (**2'i**)
X	Br		97 (**2j**)	93	71 (**2'j**)
XI	Me		94 (**2k**)	93	78 (**2'k**)
XII	MeO		95 (**2l**)	-	84 (**2'l**)
XIII	NO_2		61 (**2m**)	-	63 (**2'm**)

Table 2.2: Data for the isolation of free *L*-tryptophan analogues and N^{α}-acetyl-*D*-tryptophan derivatives following resolution. [a]Yield calculated based on the maximum theoretical recovery from resolution of a racemic mixture. [b]Enantiomeric excesses were estimated from optical rotation measurements and literature data (when available) and have an error of ±5%. *The reported yield exceeds 100% as the reaction did not proceed to the theoretical maximum that can be achieved by resolution of a racemic mixture and reflects impurities of starting material.

2.2.4.a The Enzyme Amano Acylase

The enzyme used for the enantiomeric resolution of the N^{α}-acetyl-*D*,*L*-tryptophan derivatives belongs to the class of amino acylase 1 (*N*-Acylamino acid aminohydrolase; EC 3.5.1.14). It catalyses the deacetylation of acetylated *L*-amino acids to the free amine (Scheme 2.4).



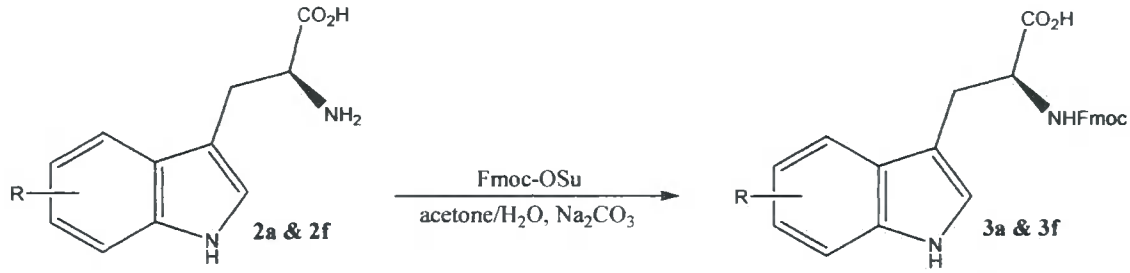
Scheme 2.4: Enzymatic cleavage of an acetyl group from an amino acid, generating the free amino acid and acetic acid, where R_1 is any *L*-amino acid side chain other than *L*-aspartic acid and R_2 an alkyl chain.

The specific acylase used in this study (acylase Amano®) was provided as a slightly yellow to yellow-brownish powder. It is an extremely stable enzyme in the dry state, in buffer solution (pH 6–10) and at high temperatures (70 °C) with optimum conditions for enzymatic activity at pH 8.0 and 50 °C. Below pH 5 the enzyme is rapidly and irreversibly inactivated. The enzyme shows a high degree of optical specificity toward its substrates and cleaves the acetyl group selectively from *L*-amino acids. For this reason, it has been used extensively for the resolution of racemic amino acids,[6,7] which are easily converted to *N*-acetyl derivatives.

The enzymatic resolution of the *N*-acetyl *D*,*L*-tryptophan derivatives (**1**) was achieved in good to excellent yields for both, the free *L*-amino acid derivatives (**2**) and the *N*-acetyl-*D*-compounds (**2'**) (Table 2.2). Enantiomeric excesses were estimated from optical rotation measurements and literature data (when available) and have an error of ±5%. A similar optical resolution of N^{α}-acetyl-*D*,*L*-tryptophan derivatives had previously been demonstrated by Yamada *et al.*[8] using *D*-amino acylase.

2.2.5 Synthesis of Fmoc-Tryptophan Analogues (3a-m)

The *L*-tryptophan derivatives were used for further functionalisation, either through introduction of an Fmoc protecting group, or conversion to the corresponding N^{α}-acetyl *N*-ethyl amides (Scheme 2.5).



Scheme 2.5: Preparation of Fmoc-*L*-tryptophan derivatives (**3a** & **3f**) using *N*-(9-Fluorenylmethoxy-carbonyloxy) succinimide (Fmoc-OSu). R stands for a 5-F or 5-MeO substituent respectively

The applicability of the protection reaction was exemplified on two model compounds (**2a** & **2f**) (Table 2.3) (Table 2.3).

ID	R	Pos.	Yield (%)
3a	F	C5	96
3f	MeO		90

Table 2.3: Fmoc protection of compounds **2a-m** and respective yields.

Purification of the desired products was achieved by HPLC chromatography. A typical trace can be seen in Figure 2.1.

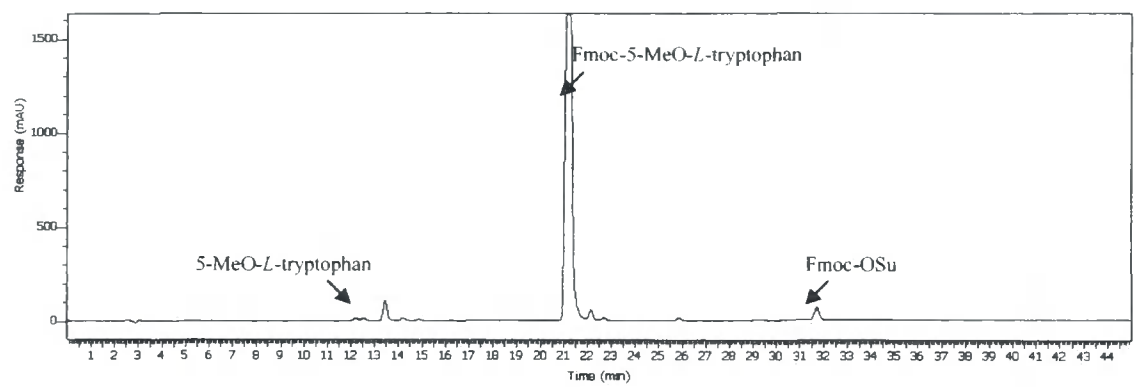


Figure 2.1: Typical HPLC trace of crude Fmoc-5-MeO-*L*-tryptophan (**3f**). A linear gradient system with 0.1% trifluoracetic acid (TFA) in water and acetonitrile respectively was used with an analytical reverse phase C18 column at a flow rate of 0.5 ml/min. The detection wavelength was 254 nm. The major peak at 21.05 min corresponds to Fmoc-5-MeO-*L*-tryptophan (**3f**).

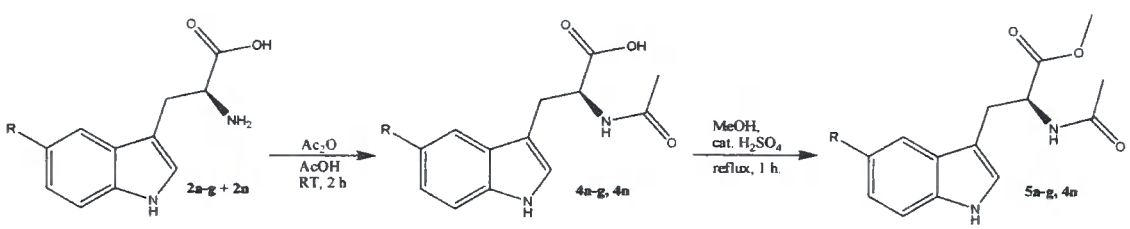
The remaining Fmoc-protected amino acids (from **2b-e** and 2**g-m**) could be synthesised in good to excellent yields, but due to problems with the HPLC system, an appropriate purification and characterisation of the products could not be achieved.

2.2.6 Conversion of Tryptophan Analogues to N^{α}-Acetyl Ethyl Amide Derivatives (6a-g, 6n)

The free amino acid analogues were first acetylated using standard conditions, providing N^{α}-acetyl-*L*-tryptophan derivatives (**4**) in excellent yields (Table 2.4, Scheme 2.6), which were further derivatised by conversion to a methyl ester on the carboxylic acid functionality *via* an acidic Fisher methyl esterification to produce **5**. The starting material for **4n** was racemic *D,L*-tryptophan (**2n**) as purchased from Sigma Aldrich.

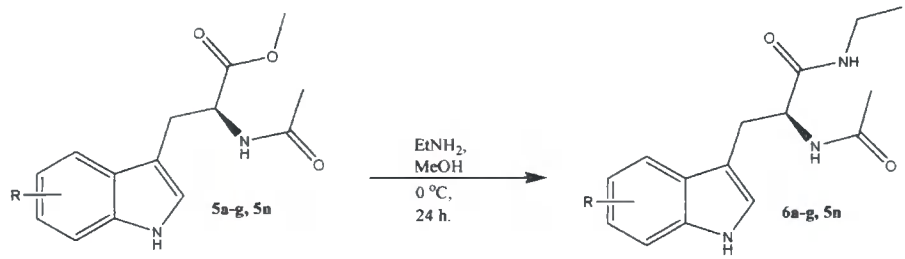
R	Pos.	Yield 4 (%)	Yield 5 (%)
F	C5	93 (**4a**)	88 (**5a**)
Cl		83 (**4b**)	93 (**5b**)
Br		87 (**4c**)	89 (**5c**)
I		85 (**4d**)	91 (**5d**)
Me		83 (**4e**)	93 (**5e**)
MeO		92 (**4f**)	92 (**5f**)
NO_2		44 (**4g**)	88 (**5g**)
H		75 (**4n**)	98 (**5n**)

Table 2.4: Yields of the tryptophan derivatives N^{α}-acetyl-*L*-tryptophan (**4a-g, 4n**) and N^{α}-acetyl-*L*-tryptophan methyl esters (**5a-g, 5n**).



Scheme 2.6: Methyl esterification of tryptophan derivatives. R stands for H, F, Cl, Br, I, Me, MeO, or NO_2.

The outcome of the esterification was evaluated by the purification and analysis of examples of the electron rich (**5f**), electron poor (**5g**), and intermediate N^{α}-acetyl-*L*-tryptophan methyl esters (**5a** and **5n**). The esterification was not expected to be influenced by the electron density on the indole ring system as the reaction was taking place on the somewhat remote carboxylate group. The remaining compounds were converted into ethyl amide derivatives (**6a-g, n**) without further purification (Scheme 2.7).



Scheme 2.7: Ethyl amidation of N^{α}-acetyl-*L*-tryptophan methyl ester derivatives (**5**).

Purification of the final derivatives by chromatography yielded compounds of good purity, as confirmed analytically by polarimetry (Table 2.5) and HPLC (Figure 2.2).

ID	Yield 6 (%)	$[\alpha]_d^{25}$	Conditions
6a	82	+29.8	c = 0.1, MeOH
6b	86	+24.3	c = 0.1, MeOH
6c	77	+19.2	c = 0.1, MeOH
6d	74	+28.6	c = 0.5, MeOH
6e	80	+28.6	c = 0.4, MeOH
6f	44	+32.8	c = 0.1, MeOH
6g	98	+8.7	c = 0.15, MeOH
6n	99	+4.1	c = 0.7, MeOH

Table 2.5: Yields and optical rotation data for N^{α}-acetyl-*L*-tryptophan ethyl amides (**6a-g, 6n**).

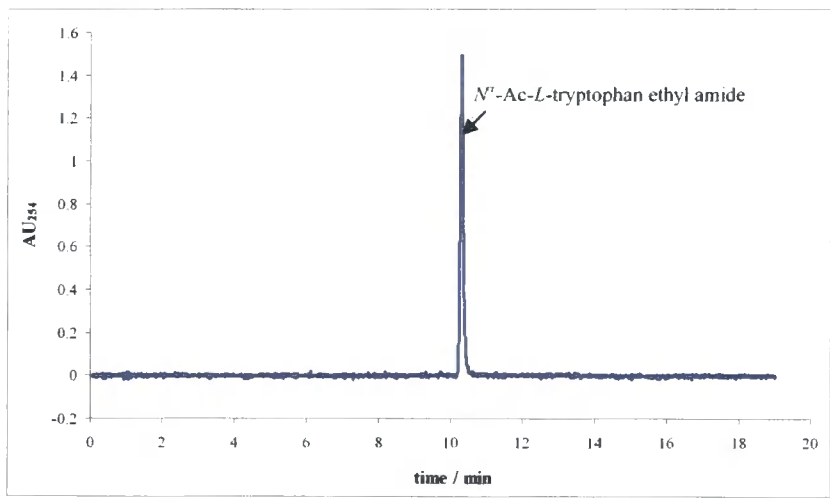


Figure 2.2: HPLC trace of crude N^{α}-acetyl-*L*-tryptophan ethyl amide (**6n**) obtained using a linear gradient system with 0.1% trifluoracetic acid (TFA) in water and acetonitrile respectively with an analytical reverse phase C18 column at a flow rate of 0.5 ml/min. The detection wavelength was 254 nm.

2.2.7 Synthesis of AcWLWLL

AcWLWLL was produced in near quantitative yield as calculated by the mass difference of the resin before and after the coupling reaction. The peptide was then cleaved from the resin and submitted to HPLC purification as described in §4.2.2. Purified peptide was obtained in 92 % yield (by mass) and in good purity (determined by LC-MS, Figure 2.3).

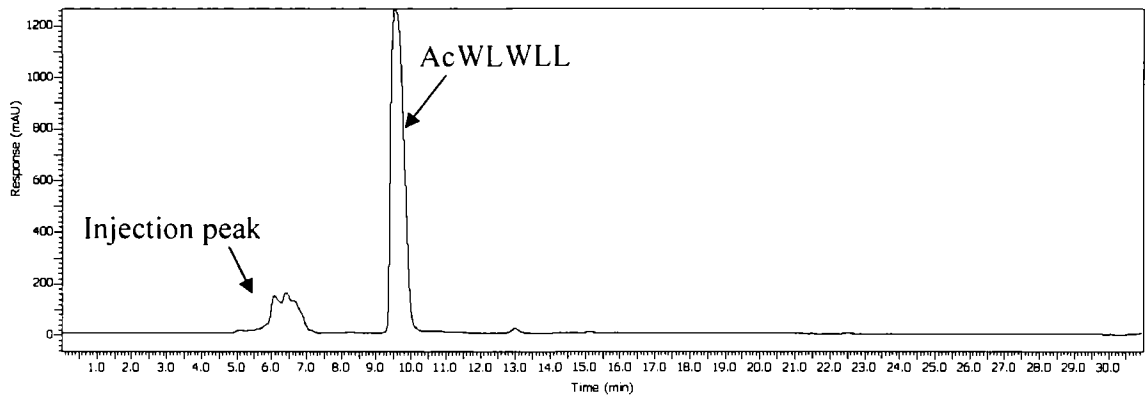


Figure 2.3: Typical HPLC trace of crude AcWLWLL after cleavage. A mixture of acetonitrile and water (both containing 0.1% TFA) was used on a C18 analytical column with a linear gradient system and the chromatogram recorded at 254 nm (see §4.2.2 for more details).

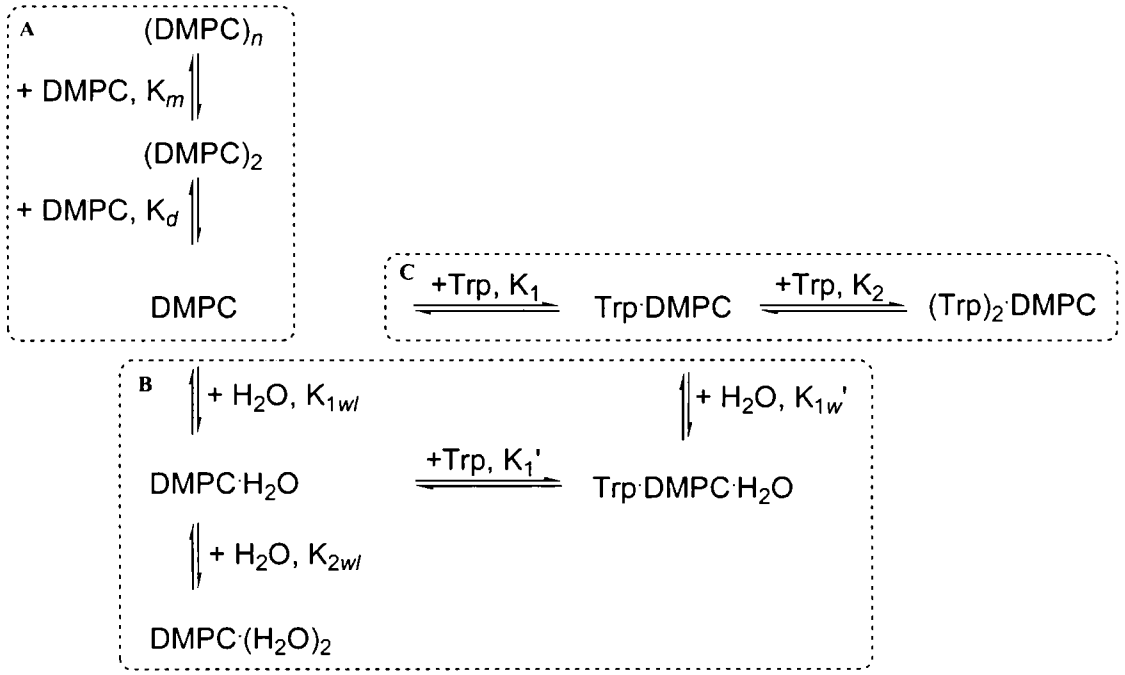
As illustrated in Figure 2.3, the crude product from the peptide cleavage contained the desired peptide. The product, eluting at about 9.5 min, was recovered in as the predominant component.

2.3 NMR Titrations

Binding interactions of N^{α}-acetyl-*L*-tryptophan ethyl amides (**6a-g, 6n**) and DMPC lipid molecules were analysed in chloroform using ^{1}H NMR host-guest titrations. The time averaged NMR signal is a combination of the chemical shifts of the unbound, freely solvated molecules, and the bound species. Using a 200 MHz spectrometer, these signals were present as a single coalesced signal. At 300 MHz, significant line broadening was observed for tryptophan, making the signal unsuitable for monitoring during the course of a titration.

2.3.1 Binding Models

Various association events need to be accounted for. A general overview of selected equilibria is given in Scheme 2.8.



Scheme 2.8: Equilibria and binding constants that were analysed. Box **A** accounts for the micellisation of DMPC molecules (binding constant K_d for dimerisation and K_m for further micellisation); box **B** encompasses the association of water to the lipid (K_{1wl} and K_{2wl}) and to a Trp:DMPC 1:1 complex (K_{1w}'). The interactions of one (K_1) or two (K_2) tryptophan molecules with a DMPC molecule is described in box **C**.

The association model needed to be accurate in the description of all important equilibria whilst keeping the number of variables at a reasonable level to enable their determination by regression analysis. The model uses an isodesmic description of DMPC self-association in chloroform. It is assumed that the binding constant for this micellisation process (K_m) remains the same for dimerisation and oligomerisation (*i.e.* $K_d = K_m$). The explicit inclusion of K_d allowed the evaluation of the accuracy of the isodesmic (non-cooperative) model, which could be compared to an anti-cooperative model ($K_d > K_m$) and a cooperative model ($K_d < K_m$). Furthermore, a previous study of DMPC molecules in chloroform suggested DMPC-trimers to be the predominant species in solution.[9] An explicit calculation of the dimer concentration was therefore a compromise between the number of fittable variables and level of detail in the model. The evaluation of the various models and

the determination of the micellisation constant K_m could be achieved by titrations of DMPC into chloroform (§2.3.2).

The experimental conditions made titrations in the presence of water unavoidable due to the working conditions, *i.e.* continuous additions of small aliquots of a solution of hydrated lipid to the host. Water concentrations determined by integration of the ^{1}H NMR signals for lipid and water (see §2.3.2.a) showed that the commercial sample of DMPC contained 2.2 water molecules per DMPC molecule.

Polar water molecules that are initially complexed to the lipid molecules will be in equilibrium with unbound water and compete with tryptophan analogues in binding interactions with DMPC molecules. Previous work on 1,2-diacetyl-*sn*-3-phosphocholine (DAPC) interacting with N^{α}-acetyl-*L*-tryptophan ethyl amide in chloroform showed that the binding constant for the amino acid-lipid association was very sensitive to the concentration of water in the sample.[10] DAPC produced lower binding constants than DMPC because of the competitive nature of the water binding. Reverse titrations of N^{α}-acetyl-*L*-tryptophan ethyl amide into DMPC containing chloroform also showed that the water signal tended to that of free water in $CDCl_3$ as the concentration of derivative increased. This could be seen as evidence that water and tryptophan derivatives bind competitively for DMPC.

The binding model used here accounts for some of these interactions but is limited to a 2:1 interaction of water with lipid and a 1:1:1 interaction of water-lipid and tryptophan to keep the number of variables reasonable. The 2:1 binding of water to DMPC was essential to represent the initial starting complexation of the lipid (as it contained approximately 2 water molecules per molecule of lipid). Although higher-order (3:1, 4:1, etc) water-lipid complexes are likely, the water concentration in the sample justifies the decision not to consider these. Of the possible interactions, the interaction of water with a 1:1 DMPC-tryptophan complex was incorporated in the model as this equilibrium had the potential to influence the observed complexation-induced chemical shifts during the course of a titration. Water-water and water-tryptophan adduct formation was not considered. In the former case, this was because of difficulties associated with the determination of water oligomerisation in chloroform. In the latter case, the low concentrations of free water (due to association with lipid) and tryptophan allowed this approximation to be made without being detrimental to the analysis. The micellisation binding constant K_m and the water to lipid binding constants $K_{1w/}$ and $K_{2w/}$ for a 1:1 and 2:1 binding respectively were determined using titrations of hydrated DMPC into chloroform. The binding constant for water to a tryptophan-lipid 1:1 complex K_{1w}' was

accounted for during the evaluation of the host-guest titrations. Job plots performed both, in the current work and in a previous study,[10] suggest a 2:1 tryptophan-lipid to be the predominant adduct of the two in chloroform. The model accounts for this species explicitly. Due to the low concentration of tryptophan derivatives used and the low percentage of their binding to DMPC the omission of higher order adducts (3:1, 4:1, etc) is justified.

2.3.2 Titrations of Hydrated DMPC into Chloroform

2.3.2.a Water Presence – Its Importance and Associated Equilibria

A blank titration of a chloroform solution of hydrated DMPC into wet chloroform allowed the amount of water present after each addition to be extrapolated by integration of the NMR signals. Four signals, belonging to the water protons, the lipid choline head group methyl and glycerol *sn*-CH_2 protons, exhibited observed complexation-induced chemical shift changes. This allowed the number of water molecules per lipid to be calculated by fitting the data to the straight line described by $[W]_0 = x[L]_0 + c$. An initial water:lipid ratio of approximately 2:1 ($x = 2.20 \pm 0.04$) was determined by this method and the concentration of free water in the $CDCl_3$ solvent was calculated to be $c = 5.4 \pm 0.3$ mM (Figure 2.4).

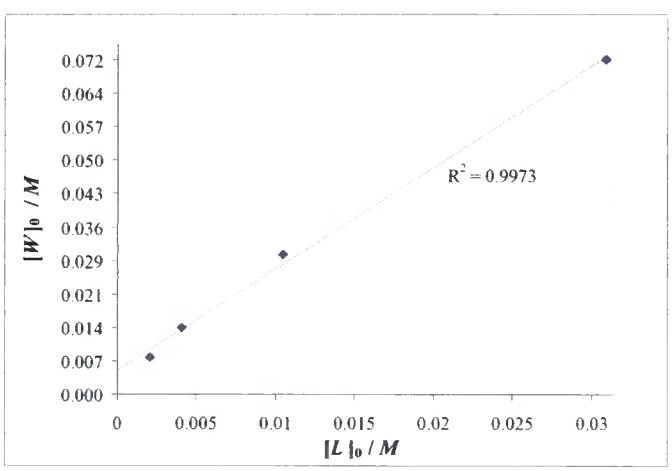


Figure 2.4: Determination of lipid hydration. Graphical analysis of water $[W]_0$ to lipid $[L]_0$ concentration ratios allowed the extrapolation of the initial water concentration in neat chloroform (the intercept of a linear regression) and the water per lipid ratio (given by the slope).

This water was found to partake in adduct formation as its chemical shift changed from the typical value of bulk water in chloroform (1.57 ppm) to approximately 4 ppm

during the course of a titration (Figure 2.5). This implied that water molecules were either released from their association site in the lipid or replaced by the tryptophan derivatives.

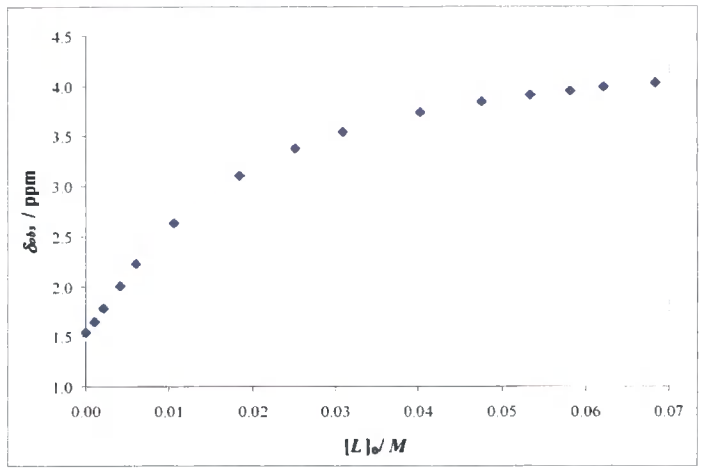


Figure 2.5: Observed chemical shift change of water protons δ_{obs} as a function of the total lipid concentration $[L]_0$. Experimental data is indicated by points on the graph, a line represents the binding isotherm of water as calculated by fitting the experimental data to Equation 2.4.

2.3.2.b Critical Micelle Concentration (CMC)

A previous study suggested that water associates with DMPC in chloroform and that the lipid is able to form inverse micelles.[9] The authors estimated a trimeric DMPC inverse "micelle" to be the most stable form. In order to mitigate against the possibility that higher-order micelles were able to form, general micellisation was accounted for during analysis of the binding constants for tryptophan adducts. The titration of hydrated DMPC into chloroform permitted the estimation of the critical micelle concentration (CMC) of the lipid in this solvent. This assumes that the lipid exists in a fully dissolved, monomeric form with a chemical shift of $\delta_{monomer}$, in equilibrium with a micellar form. At low total concentrations of DMPC, the observed chemical shift can then be approximated as arising mainly from monomers; at high total concentrations the equilibrium will have shifted mostly towards a micellar form of the lipid giving rise to a chemical shift of $\delta_{micelle}$. The observed chemical shift δ_{obs} could than be fitted against the reciprocal total lipid concentration $[L]_0$ allowing the determination of the critical micelle concentration (CMC)[9]:

$$\delta_{obs} = \delta_{micelle} + \frac{CMC}{[L]_0}(\delta_{micelle} - \delta_{monomer}) \qquad (2.1)$$

The graphical analysis of Equation 2.1 (Figure 2.6) and the introduction of two linear regressions, one at low total lipid concentrations and another at high total lipid concentrations allowed the CMC to be calculated.

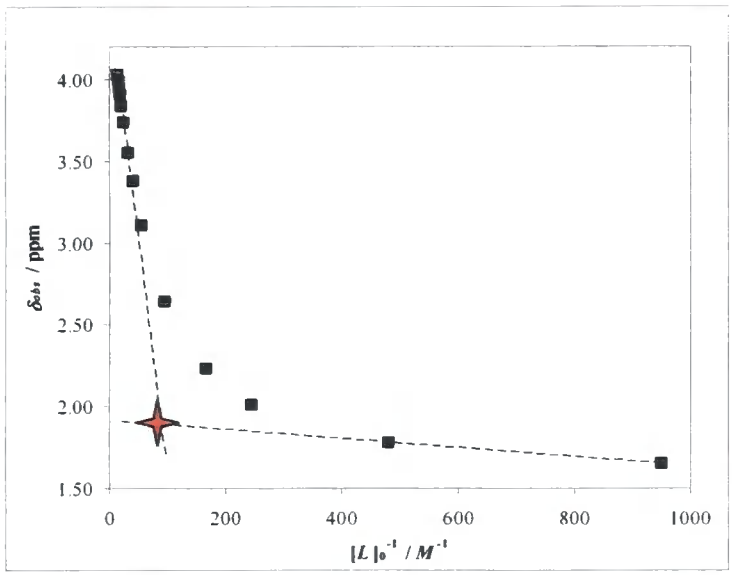


Figure 2.6: Graphical determination of the CMC of DMPC in chloroform. The graph shows the observed chemical shift change of water protons δ_{obs} as a function of the reciprocal of the total lipid concentration $[L]_0^{-1}$. Two linear regressions were added at the DMPC concentration ranges where the lipid could be assumed to be either fully in monomeric or micellar forms respectively. Their intersection, the estimated CMC, is marked with a red star.

The regression for the monomeric concentration regime was performed for the first two data points of the NMR titration when the lipid concentration was below 2 m*M* and the calculated dimer and trimer concentrations of DMPC micelles were virtually zero. The regression for the micellar region was obtained by fitting the last six points of the NMR titration, where it was assumed that most of the lipid was in the micellar phase. The calculated concentrations of micellar lipid in this concentration regime with respect to the concentration of monomeric lipid justified this assumption (Figure 2.7). The intersection of these two regression lines could be used to estimate the CMC at the temperature of the measurement (21 °C). The complexation-induced chemical shifts of several lipid protons from two independent titrations were averaged to calculate a CMC of 13.73 ± 0.17 m*M*), which is in good agreement to the previously reported value of 16 m*M* at 24 °C.[9] However,

when the complexation-induced chemical shifts of the water protons, which were more pronounced than those of the lipid, were included a slightly higher value of 20.54 m*M* for the CMC was obtained. A weight averaged value over all the signals analysed predicts the CMC at 19.71 ± 3.28 m*M* at 21 °C. The evolution of the various concentrations of free, unbound and bound species is depicted in Figure 2.7.

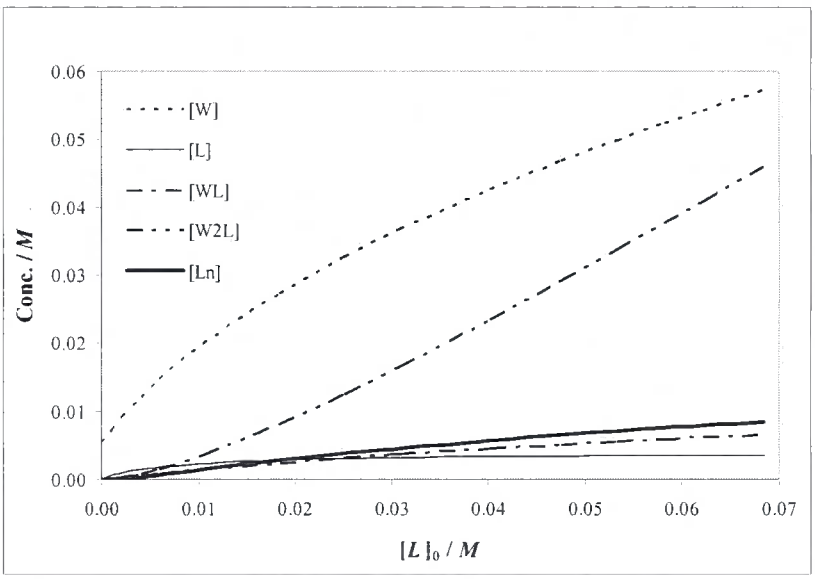


Figure 2.7: Concentration profiles. Concentrations of unbound lipid and water molecules ([*L*] and [*W*] respectively), lipid micelles ($[L]_n$) and water-lipid 1:1 and 2:1 adducts ([*WL*] and [W_2L] respectively), as function of the total lipid concentration, calculated using the equations in §7.2.

As expected, the concentration of free lipid molecules [*L*] increased at a slower rate than the concentration of micellar lipid $[L]_n$. The crossover from the calculated curves for the free lipid and the micellar lipid concentrations (at 18.43 m*M*) lies within experimental error of the approximated value for the CMC. The concentration of water-lipid adducts increases steadily with increasing concentrations of total lipid. The concentration of the 2:1 complex [W_2L] increased more significantly than the concentration of the 1:1 complex [*WL*]. All concentrations of lipid originate at 0, as the starting solution only contains chloroform and approximately 5 m*M* free water molecules, whose concentration increases steadily during the course of the titration as each added molecule of lipid introduced two molecules of water to the system.

2.3.2.c Data Analysis

Titrations of hydrated DMPC into chloroform allowed the evaluation for the binding constants of lipid-lipid self-association and water-lipid association (see Scheme 2.2). The concentrations of the bound and free species can be expressed in terms of the known total concentration of water $[W]_0$ and lipid $[L]_0$ (Equations 2.2 and 2.3):

$$[W]_0 = [W] + [WL] + 2[W_2L] \tag{2.2}$$

$$[L]_0 = [L] + [WL] + [W_2L] + 2[L]_2 + \sum_{n=3}^{n=N^0} n[L_n] \tag{2.3}$$

Here, $[W]$, $[WL]$ and $[W_2L]$ correspond to the concentrations of free water, 1:1 water:DMPC and 2:1 water:DMPC respectively. The parameter n indicates the size of the lipid micelle $[L_n]$ and assumes values between $n = 3$ and N^0, the total number of the assembly. The individual binding constants are summarised in Table 2.6:

$$K_{1w/} = \frac{[WL]}{[W][L]} \qquad K_{2w/} = \frac{[W_2L]}{[WL][W]} \qquad K_m = \frac{[L_n]}{[L_{n-1}][L]}$$

Table 2.6: Equations for water-lipid and lipid-lipid binding constants. $K_{1w/}$ and $K_{2w/}$ refer to the binding constant for the adduct formation of a 1:1 and 2:1 water:DMPC complex respectively; K_m refers to the binding constant for lipid micellisation; $[L_{n-1}]$ refers to the lipid $(n-1)$-mer. Description of the remaining constants is given in the text.

The observed complexation-induced chemical shifts were calculated using Equations 2.4 and 2.5 and values for the binding constants $K_{1w/}$, $K_{2w/}$ and K_m as well as the values for the limiting chemical shifts δ_{b1w}, δ_{b2w}, δ_m and δ_f obtained by simultaneous, multiple, non-linear fitting of the calculated to the observed data.

$$\delta_{obsW} = \left[\frac{K_{1w}[L](\delta_{b1w} - \delta_f) + 2K_{2w/}[WL](\delta_{b2w} - \delta_f)}{1 + K_{1w}[L] + 2K_{2w/}[WL]} \right] + \delta_f . \tag{2.4}$$

$$\delta_{obsL} = \frac{[WL]}{[L]_0}(\delta_{b1w} - \delta_f) + \frac{[W_2L]}{[L]_0}(\delta_{b2w} - \delta_f) + \frac{[L_n]}{[L]_0}(\delta_{bm} - \delta_f) + \delta_f , \tag{2.5}$$

Here, δ_{obsW} corresponds to the observed chemical shift change of the water signal; δ_{b1w} and δ_{b2w} are the limiting chemical shifts of the water protons in the 1:1 and 2:1 water-lipid adduct respectively, δ_f corresponds to the limiting chemical shifts of micellar lipid and δ_f the

chemical shift of free, unbound water molecules; [L] stands for the concentration of DMPC monomers, [WL] for the concentration of the water-lipid 1:1 complex; $[L]_0$ stands for the total concentration of lipid molecules in the solution, [W_2L] for the concentration of 2:1 water-lipid complex, and [L_n] for the concentration of micellar lipid molecules. The equations used to calculate the concentrations of the individual species present can be found in §7.2.

The resulting values for the binding constants and limiting complexation-induced chemical shifts are summarised in Table 2.7.

	H_2O	$-N(CH_3)_3$	$sn\text{-}CH_2$	$sn\text{-}CH_2$	$sn\text{-}CH$	
δ_f	1.468	3.402	4.294	4.501	5.256	ppm
δ_{1wl}	3.326	3.430	2.450	3.619	5.151	ppm
δ_{2wl}	5.759	3.270	3.934	4.272	5.202	ppm
δ_m	-	3.399	4.504	4.616	5.180	ppm
$\Delta\delta_{1wl}$	1.858	0.028	-1.844	-0.882	-0.105	ppm
$\Delta\delta_{2wl}$	4.291	-0.132	-0.359	-0.229	-0.054	ppm
$\Delta\delta_m$	-	-0.003	0.210	0.115	-0.076	ppm
K_{1wl}	31.73	31.73	32.05	31.73	31.73	M^{-1}
K_{2wl}	122.09	122.09	117.99	122.09	122.10	M^{-1}
K_m	143.78	143.78	146.16	143.79	143.73	M^{-1}

Table 2.7: Binding constants and complexation-induced chemical shifts for the binding of water to DMPC and DMPC self-association in $CDCl_3$. $\Delta\delta$ entries refer to chemical shift changes of the chemical shift indicated by the subscript referenced against δ_f.

The values obtained for the association constants ($K_{1wl} = 31.86 \pm 0.11\ M^{-1}$, $K_{2wl} = 121.80 \pm 1.47\ M^{-1}$, and $K_m = 144.06 \pm 0.86\ M^{-1}$) and the complexation-induced chemical shifts ($\delta_{1wl} = 3.326$ ppm, and $\delta_{2wl} = 5.759$ ppm of the water proton signal) were used as constants in the analysis of tryptophan-DMPC binding isotherms. The calculated complexation-induced chemical shifts for the water proton signals (δ_m) was found to be negligibly small in respect to the accuracy of the NMR instrument.

chemical shift of free, unbound water molecules; $[L]$ stands for the concentration of DMPC monomers, $[WL]$ for the concentration of the water-lipid 1:1 complex; $[L]_0$ stands for the total concentration of lipid molecules in the solution, $[W_2L]$ for the concentration of 2:1 water-lipid complex, and $[L_n]$ for the concentration of micellar lipid molecules. The equations used to calculate the concentrations of the individual species present can be found in §7.2.

The resulting values for the binding constants and limiting complexation-induced chemical shifts are summarised in Table 2.7.

	H_2O	$-N(CH_3)_3$	*sn*-CH_2	*sn*-CH_2	*sn*-CH	
δ_f	1.468	3.402	4.294	4.501	5.256	ppm
δ_{1wl}	3.326	3.430	2.450	3.619	5.151	ppm
δ_{2wl}	5.759	3.270	3.934	4.272	5.202	ppm
δ_m	-	3.399	4.504	4.616	5.180	ppm
$\Delta\delta_{1wl}$	1.858	0.028	-1.844	-0.882	-0.105	ppm
$\Delta\delta_{2wl}$	4.291	-0.132	-0.359	-0.229	-0.054	ppm
$\Delta\delta_m$	-	-0.003	0.210	0.115	-0.076	ppm
K_{1wl}	31.73	31.73	32.05	31.73	31.73	M^{-1}
K_{2wl}	122.09	122.09	117.99	122.09	122.10	M^{-1}
K_m	143.78	143.78	146.16	143.79	143.73	M^{-1}

Table 2.7: Binding constants and complexation-induced chemical shifts for the binding of water to DMPC and DMPC self-association in $CDCl_3$. $\Delta\delta$ entries refer to chemical shift changes of the chemical shift indicated by the subscript referenced against δ_f.

The values obtained for the association constants ($K_{1wl} = 31.86 \pm 0.11\ M^{-1}$, $K_{2wl} = 121.80 \pm 1.47\ M^{-1}$, and $K_m = 144.06 \pm 0.86\ M^{-1}$) and the complexation-induced chemical shifts ($\delta_{1wl} = 3.326$ ppm, and $\delta_{2wl} = 5.759$ ppm of the water proton signal) were used as constants in the analysis of tryptophan-DMPC binding isotherms. The calculated complexation-induced chemical shifts for the water proton signals (δ_m) was found to be negligibly small in respect to the accuracy of the NMR instrument.

2.3.3 Host-Guest Titrations

2.3.3.a Observed Complexation-Induced Chemical Shifts

NMR host-guest titrations with DMPC as guest permitted the identification of the parts of the molecules that were affected by binding interactions. In the guest molecule, the lipid DMPC, some protons in the polar head group region of the molecule showed small observed complexation-induced chemical shifts: the methyl protons of the amine shifted by about -0.08 ppm, the protons of the glycerol moiety by about -0.05 to -0.08 ppm during early stages of the titration, when the concentration of tryptophan was greater than that of the lipid. This is to be expected as most of the lipid is unbound at all concentrations and the observed chemical shifts are not limiting or maximal complexation induced changes, but arise from a coalesced signal. On the other hand, most of the protons in the guest molecules, the tryptophan derivatives, showed a change in chemical shift upon addition of the lipid. Up to seven protons exhibited a complexation-induced chemical shift change > 0.05 ppm. The chemical environment of the indole amine and backbone amide protons changed dramatically during the course of a guest-host titration resulting in complexation-induced chemical shift changes of the indole amine proton between +1.78 ppm (**6n**) to +3.31 ppm (**6g**). The two other protons that could participate in hydrogen bonding, the amide protons, shifted typically by about +1.5 ppm (EtN*H*) and +0.7 ppm (AcN*H*). The indole ring protons were shifted both positively and negatively by approximately 0.1 ppm, suggesting that the indole ring itself was partaking in binding probably through cation-π electron or π-facial interactions. The remaining protons of the tryptophan derivative were only affected slightly (<0.05 ppm) rendering them less useful for the determination of binding constant. These results are consistent with an involvement of the tryptophan analogues in binding interactions with DMPC.

2.3.3.b Adduct Stoichiometry

Job plots of N^{α}-acetyl-*L*-tryptophan ethyl amide (R = MeO (**6f**) and NO_2 (**6g**)) mixtures with DMPC were produced. Complexation-induced chemical shifts were observed for various protons of both molecules. The plots showed maxima at mole fractions, χ, of ~0.70 – 0.76 for the indole amine, ethyl amide and acetyl amide N*H* protons respectively (Figure 2.8). Complexation-induced chemical shifts changes arising from protons in the lipid (choline-

$(CH_3)_3$ and glycerol CH_2) were found to produce a maximum response at a mole fraction of 0.4.

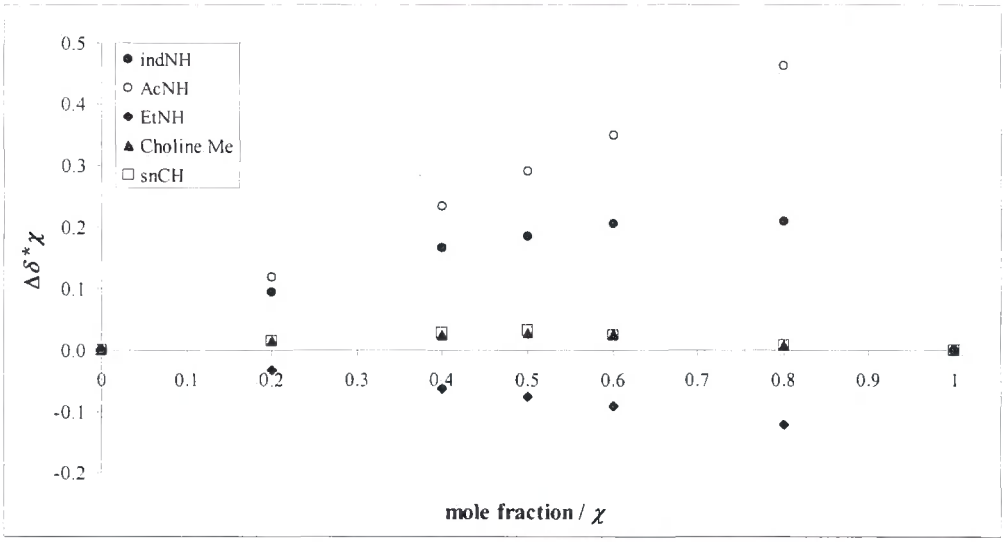


Figure 2.8: Job plot for N^{α}-acetyl-5-MeO-*L*-tryptophan ethyl amide (**6f**) with DMPC. The plots show the observed chemical shift change $\Delta\delta$ of a proton as a function of mole fraction (χ) of the species bearing the proton.

These data reflect an adduct formation that possibly exhibits some non-ideal behaviour of 2:1 association, or the formation of multimeric structures (2:2, 2:3 *etc*). The titration data produced a considerably better fit for the 2:1 adduct than for other stoichiometries however, suggesting a 2:1 adduct to be a predominant species. This observation is also in good agreement with previously published data,[10] where N^{α}-acetyl-*L*-tryptophan ethyl amide was clearly shown to form 2:1 adducts with 1,2-diacetyl-*sn*-3-phosphocholine (DAPC) in chloroform.

2.3.3.c Data Analysis

Multiple, simultaneous non-linear fitting of the NMR host-guest titration data allowed calculation of the binding parameters for the complexation of 5-monosubstituted N^{α}-acetyl-*L*-tryptophan ethyl amides to DMPC (K_1, K_2, K_1', and K_{1w}', and the limiting chemical shifts δ_{free}, δ_{bound1}, δ_{bound2}).

$$K_1 = \frac{[LT]}{[T][L]} \qquad K_2 = \frac{[LT_2]}{[LT][T]} \qquad K_{1w}' = \frac{[LTW]}{[LT][W]} \qquad K_1' = \frac{[LTW]}{[LW][T]}$$

Table 2.8: Equations for binding constants for lipid-tryptophan adducts. K_1 and K_2 refer to the binding constants for the formation of 1:1 and 2:1 Trp:DMPC adducts respectively. K_{1w}' refers to the binding constant for the association of water to the Trp:DMPC 1:1 complex and K_1' to the binding constant for the association of tryptophan to the 1:1 water-lipid complex. [*T*], [*L*], and [*W*] are the concentrations of free tryptophan, lipid and water molecules, respectively; the concentrations of lipid-amino acid complexes are given by [*LT*] and [LT_2] for the 1:1 and 1:2 adduct respectively; the concentration of the lipid-amino acid-water 1:1:1 complex is given by [*LTW*].

K_{1w}' was calculated using:

$$K_{1w}' = \frac{K_{1w} K_1'}{K_1}. \qquad (2.6)$$

The variables (K_1, K_2, K_1', and K_{1w}', and the limiting chemical shifts δ_{free}, δ_{bound1}, δ_{bound2}) were then adjusted to fit Equation 2.7 to the experimental data:

$$\delta_{obs} = \frac{\left(K_1[L](\delta_{bound1} - \delta_{free}) + K_1'[WLT](\delta_{bound1'} - \delta_{free}) + 2K_2[LT_2](\delta_{bound2} - \delta_{free})\right)}{\left(1 + K_1[L] + K_1'[WLT] + 2K_2[LT_2]\right)} + \delta_{free} \qquad (2.7)$$

Subsequently, regressions on the binding isotherm arising from each individual signal allowed refinement of the data and an estimation of error. The variance between pairs of experimental data set was typically below 8%. Full equations for the calculation of the concentrations of species present are given in §7.2.1. The parameters evaluated are given in Table 2.9. A full listing of the chemical shift data can be found in Appendix §7.3.1.

ID	K_1	K_2	K_1'	K_{1w}'
	M^{-1}	M^{-1}	M^{-1}	M^{-1}
6n	11.5 ± 0.2	210.0 ± 1.0	164.7 ± 23.6	453.6 ± 6.2
6a	17.9 ± 0.9	276.6 ± 11.6	159.8 ± 2.4	285.9 ± 15.1
6b	24.0 ± 0.3	239.5 ± 7.2	257.7 ± 3.7	336.5 ± 10.3
6c	26.3 ± 2.0	214.5 ± 19.9	257.4 ± 40.6	310.3 ± 25.2
6d	34.6 ± 2.2	210.0 ± 6.3	340.1 ± 28.6	319.8 ± 7.4
6e	35.6 ± 2.1	157.9 ± 4.4	216.8 ± 23.6	205.1 ± 18.7
6f	61.1 ± 1.3	296.1 ± 3.7	167.8 ± 3.6	87.1 ± 1.8
6g	88.2 ± 1.4	447.3 ± 8.6	381.0 ± 8.1	136.8 ± 1.4

Table 2.9: Experimentally obtained binding constants for complex formation between tryptophan and DMPC.

Binding constants were subsequently converted into binding free energies (Table 2.10) using Equation 2.8.

$$\Delta G = -RT \ln K, \qquad (2.8)$$

where R is the universal gas constant and T the temperature. Figure 2.9 shows a plot of the binding free energies for the formation of 1:1 and 2:1 complexed of the 5-substituted tryptophan analogues (**6a-g, 6n**) to DMPC *vs.* the Hammett parameter, σ_p.[11]

ID	R	σ_P	ΔG_1 kJ mol^{-1}	ΔG_2 kJ mol^{-1}	$\Delta G_1'$ kJ mol^{-1}	ΔG_{1w} kJ mol^{-1}
6g	NO_2	0.78	-11.03 ± 0.52	-15.19 ± 0.64	-14.63 ± 0.22	-12.11 ± 0.64
6d	I	0.23	-8.72 ± 0.12	-13.17 ± 0.40	-14.20 ± 0.20	-14.35 ± 0.44
6c	Br	0.23	-8.05 ± 0.62	-13.22 ± 1.22	-14.13 ± 2.23	-13.67 ± 1.11
6b	Cl	0.18	-7.83 ± 0.50	-13.49 ± 0.41	-14.33 ± 1.20	-13.67 ± 0.32
6a	F	0.06	-7.11 ± 0.13	-13.84 ± 0.12	-13.93 ± 0.32	-12.49 ± 0.17
6n	H	0.00	-6.01 ± 0.35	-11.66 ± 0.32	-15.06 ± 1.64	-12.57 ± 1.15
6e	Me	-0.17	-8.79 ± 0.18	-12.46 ± 0.15	-13.11 ± 0.28	-13.24 ± 0.27
6f	MeO	-0.27	-10.13 ± 0.16	-14.01 ± 0.25	-12.61 ± 0.27	-11.00 ± 0.20

Table 2.10: Free energies of association as determined by NMR host-guest titrations. ΔG_1 and ΔG_2 refer to the association free of Trp:lipid 1:1 and 2:1 adducts respectively, the free energy of the association of water to the *LT* complex is described by $\Delta G_1'$ and the free energy of the association of a tryptophan derivative to the W*L* complex is described by ΔG_{1w} ; σ_p refers to the Hammett parameter.

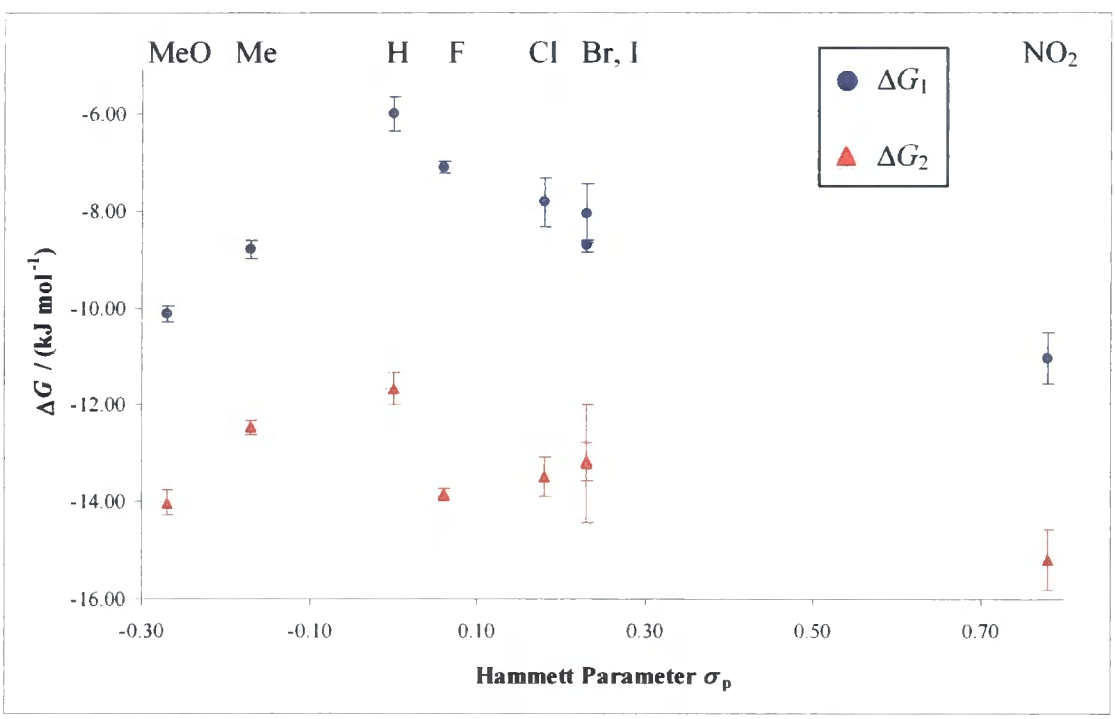


Figure 2.9: Binding free energies of 5-monosubstituted tryptophan analogues plotted against the Hammett parameter σ_p. ΔG_1 corresponds to the binding free energy for the adduct formation of Trp:DMPC 1:1; ΔG_2 is the binding free energy for the further complexation to Trp:DMPC 2:1.

When plotting the binding free energies against the Hammett parameters of the corresponding *para*-substituted compounds, a non-linear relationship with respect to the electron density of the indole ring was found (Figure 2.9). Interestingly, an n-shaped pattern was observed, indicating that not only one electronic factor affects binding free energies. All derivatives showed binding interactions with the lipid under experimental conditions, with the unsubstituted N^{α}-acetyl-*L*-tryptophan ethyl amide (**6n**) being the one interacting the weakest. The strongest influence was observed for **6g** (R = NO_2) showing a marked increase in binding compared when with N^{α}-acetyl-*L*-tryptophan ethyl amide **6n** (R = H). It is interesting to note that the free energy of binding decreased with increasing electron density from fluorine to iodine, reflecting a trend for more favourable binding for more electron-rich halogen derivatives. For electron-releasing substituents (**6e**, **6f**; R = Me, MeO), increased electron density on the indole ring system also led to more favourable binding in free-energy terms. The break in the trends either side of tryptophan (R = H) is consistent with as least two factors influencing the observed binding.

The free energies of association for Trp:lipid 2:1 adduct formation from the 1:1 complex (ΔG_2) were found to be approximately 4.91 ± 1.02 kJ mol^{-1} more favourable than a corresponding 1:1 adduct formation (ΔG_1) and comparable in size to the free energies of tryptophan:water:lipid 1:1:1 adduct formation (ΔG_{1w}) and the association of tryptophan to a lipid:water 1:1 complex ($\Delta G_1'$) (Table 2.10, Figure 2.10). These results predict a preferred Trp:lipid adduct stoichiometry of 2:1 in equilibrium, which corresponds to the preferred stoichiometry found using Job plots (§2.3.3.b).

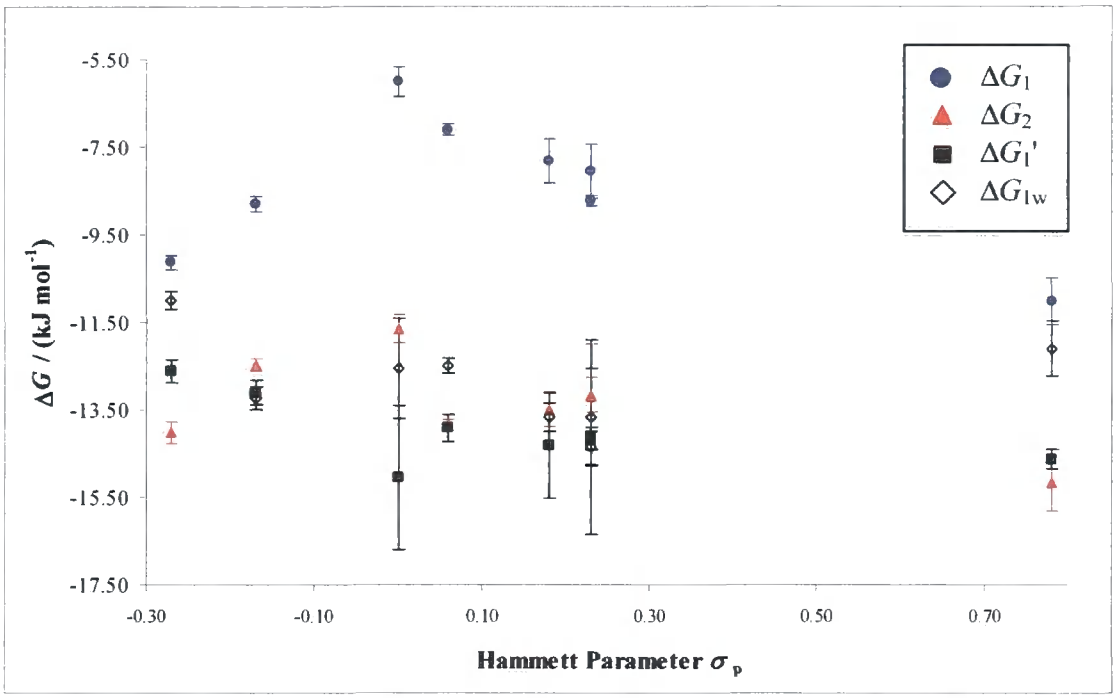


Figure 2.10: Comparison of free energies of association of Trp:DMPC and Trp:DMPC:water adduct formations as determined by NMR guest-host titrations. ΔG_1 and ΔG_2 refer to the binding free energies for the adduct formation of 1:1 and 2:1 Trp:DMPC adducts respectively; ΔG_{1w} and $\Delta G_1'$ refer to the binding free energies of the association of a tryptophan derivative to a DMPC:water 1:1 complex and the association of water to the 1:1 Trp:DMPC complex respectively.

It is noteworthy that inverse trends are to be expected for interactions involving cation-π and hydrogen bonding interactions with indole as a function of electron-donor or releasing ability of the 5-substituent. Electron withdrawing groups, such as nitro, will reduce the π-electron density of the indole ring system and produce a higher δ^+ charge on the indole NH. Electron releasing groups will have the effect of increasing the π-electron density of the ring but reducing the δ^+ charge on the indole NH. The observed trend may therefore reflect binding that is dominated by cation-π interactions for $\sigma_p < 0$, and hydrogen bonding interactions at $\sigma_p > 0$. The position of the tryptophan in the trend is such that neither cation-π nor hydrogen bonding effects will be predominant.

The macroscopical free energy of association is the sum of all free energy contributions of microscopic associations of individual molecules. Here, it can be approximated by the sum of ΔG_1 and ΔG_2 (Figure 2.11).

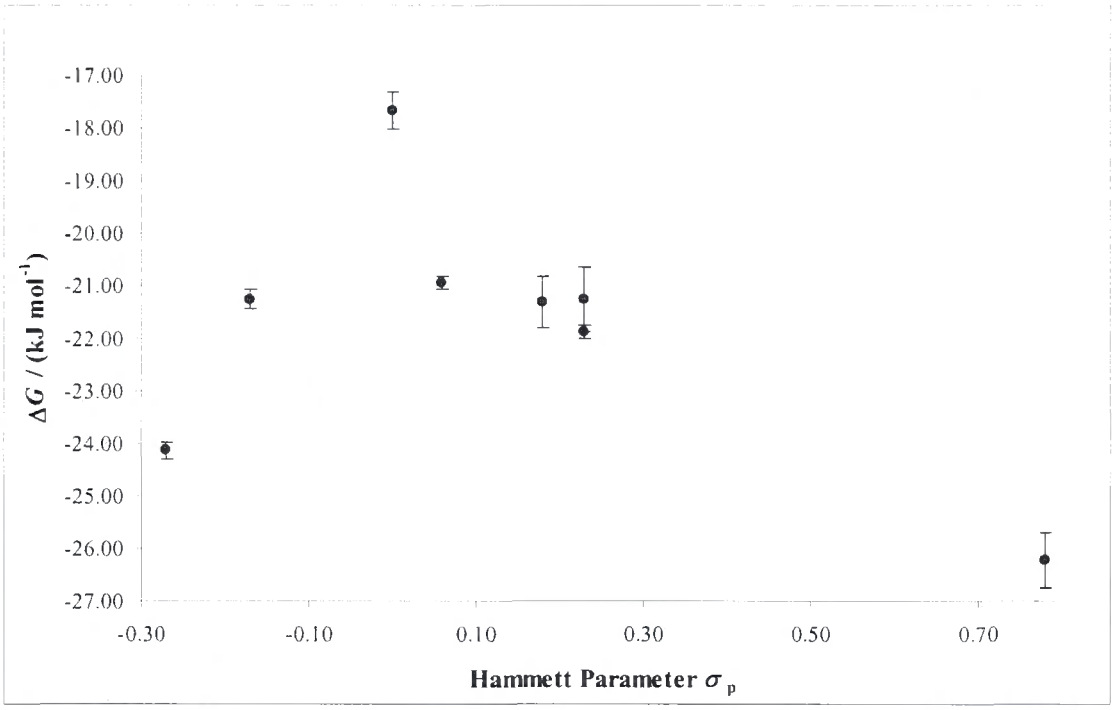


Figure 2.11: Sum of free energies of Trp:lipid 1:1 and 2:1 adduct formation.

The n-shaped pattern is mostly preserved. The association of the fluoro-substituted tryptophan analogue (**6a**) to DMPC was found to be more favourable than expected from an n-shaped pattern. A similar pattern was observed in the ITC data analysis (§2.4.2.c).

2.3.4 Water Equilibria

2.3.4.a Representation of Water Equilibria

As expected, the observed chemical shift changes for the water signal were slightly smaller in the presence of tryptophan analogues, suggesting that the model is valid for these equilibria. The observed complexation-induced chemical shift of water protons during the course of the NMR host-guest titrations could be fitted extremely well (deviations typically < 3%) using the binding constants (K_{1wl} and K_{2wl}) and limiting chemical shifts (δ_f, $\delta_{bound1w}$, $\delta_{bound2w}$) derived from the titration of hydrated DMPC into chloroform (Figure 2.12). A full listing of the chemical shift data can be found in Appendix §7.3.1.

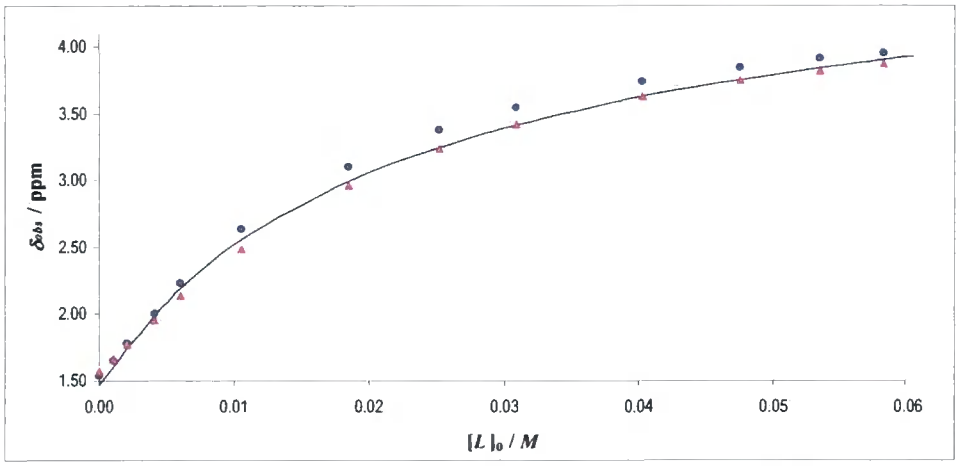
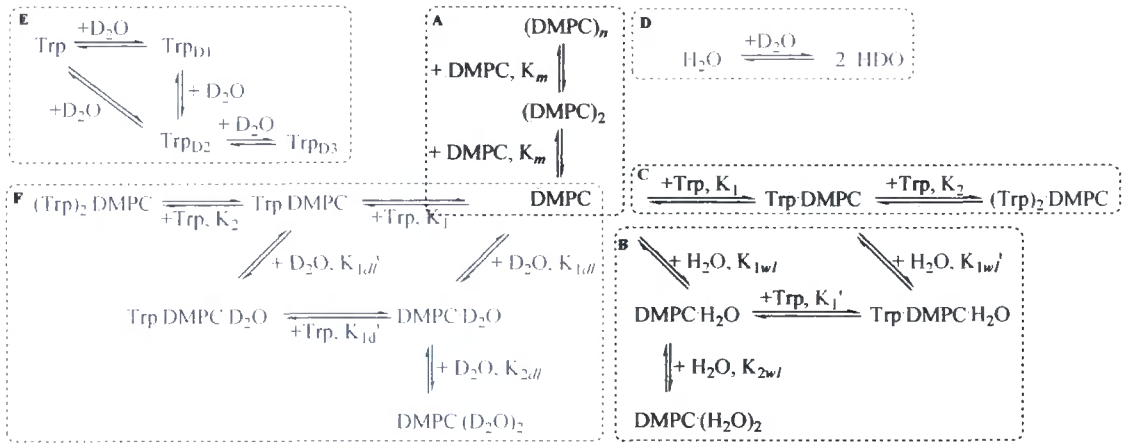


Figure 2.12: Induced chemical shifts for binding of water to DMPC in the presence (▲) and absence (●) of N^{α}-acetyl-5-fluoro-*L*-tryptophan (**6a**). A solid line shows the calculated chemical shift change for water in the titration with **6a**.

The explicit treatment of water in the modelling of the NMR titration data was found to be highly successful in reproducing the observed chemical shift changes arising from water protons. Water equilibria have been widely neglected in the analysis of NMR host-guest titrations so far[12], but are shown here to be accountable. It was shown that such water equilibria have an important influence and a considerable free energy contribution to the association equilibria of tryptophan derivatives interacting with DMPC lipids in chloroform.

2.3.5 D_2O Titrations

The importance and strength of hydrogen bonding in the Trp:DMPC adducts were evaluated by deuterium exchange kinetic studies. A sample with equimolar concentrations (55 μ*M*) of N^{α}-acetyl-*L*-tryptophan ethyl amide (**6n**) and DMPC was titrated with aliquots of deuterated water (D_2O). The addition of deuterium to the system will establish new equilibria and perturb existing ones (Scheme 2.9, *c.f.* Scheme 2.8).



Scheme 2.9: Extended equilibrium scheme. The equilibria established in Scheme 2.8 needed to be extended by the highlighted equilibria (Box **D**, **E**, and **F**) when deuterium is introduced to the system. Box **D** accounts for the deuterium exchange in water molecules. Box **F** shows equilibria of D_2O and various complexes in the system. Box **E** describes the successive deuteration of the amino acid analogue by heavy water (which is a slight simplification, as there are also three potential bisdeuterated Trp species).

The heavy water will exchange deuterium atoms with free water molecules on a very fast time scale. The exchange will affect all the remaining equilibria, according to Le Chatelier's principle (a system that is at chemical equilibrium will change in such a way as to counteract effects of external stress). Equilibria that are likely to be perturbed immediately after the addition will include all the existing equilibria that involve H_2O (*i.e.* K_{1wl}, K_{1w}', and K_{2wl}). Additionally, deuterium exchange with the three potentially labile protons of the amino acid derivatives (AcN*H*, EtN*H*, indN*H*) will be occurring. The rate of the exchange will depend on the lability and accessibility of the hydrogen atoms. If they were involved in hydrogen bonding, the strength of such an interaction could be estimated from the speed of the exchange.

The added D_2O mixed completely with the chloroform as determined by visual inspection, *i.e.* no phase separation was observed.

2.3.5.a Changes in Observed Complexation-Induced Chemical Shifts

In the experiment a proton NMR spectrum was taken immediately after each addition and subsequent agitation of the sample (Figure 2.13). This gave a response that was directly related to the fast exchange of residual water in the sample with deuterium and its effect on the existing equilibria. Chemical shift changes, $\Delta\delta$, were obtained by subtraction of the

corresponding chemical shift of the free tryptophan derivative in the absence of both lipid and D_2O (δ_{free}), from the observed chemical shifts δ_{obs}:

$$\Delta\delta = \delta_{obs} - \delta_{free}, \tag{2.9}$$

A progressive decrease in the intensity of the signals for the indole N*H* and the acetyl amide protons was observed as the quantity of D_2O added increased. Figure 2.13 shows the normalised signal intensities $\tilde{I}$ of selected protons, which were derived by dividing the individual signal intensities observed in the presence of D_2O, I_{D2O}, by the corresponding intensities of the signals in the free tryptophan derivative prior D_2O addition, I_0

$$\tilde{I} = \frac{I_{D2O}}{I_0}. \tag{2.10}$$

The indole C-H protons are unlikely to be subjected to deuterium exchange in these conditions and their intensity should therefore remain near constant throughout experiment. Fluctuations in their intensity (~12%) may be used as an indicator of the error associated with the general determination of intensities.

Within experimental error ($\Delta\delta \pm 0.005$), the chemical shift of the indole C-H proton was unaffected by the addition of deuterium, while the normalised signals of the protons typically associated with hydrogen bonding (indole N*H* and the two amide N*H* protons) decreased noticeably upon addition of D_2O (Table 2.11). A decrease in the values for the observed complexation-induced chemical shifts reflects a perturbation of the equilibria for non-deuterium labelled tryptophan analogues towards the free (non-complexed) form.

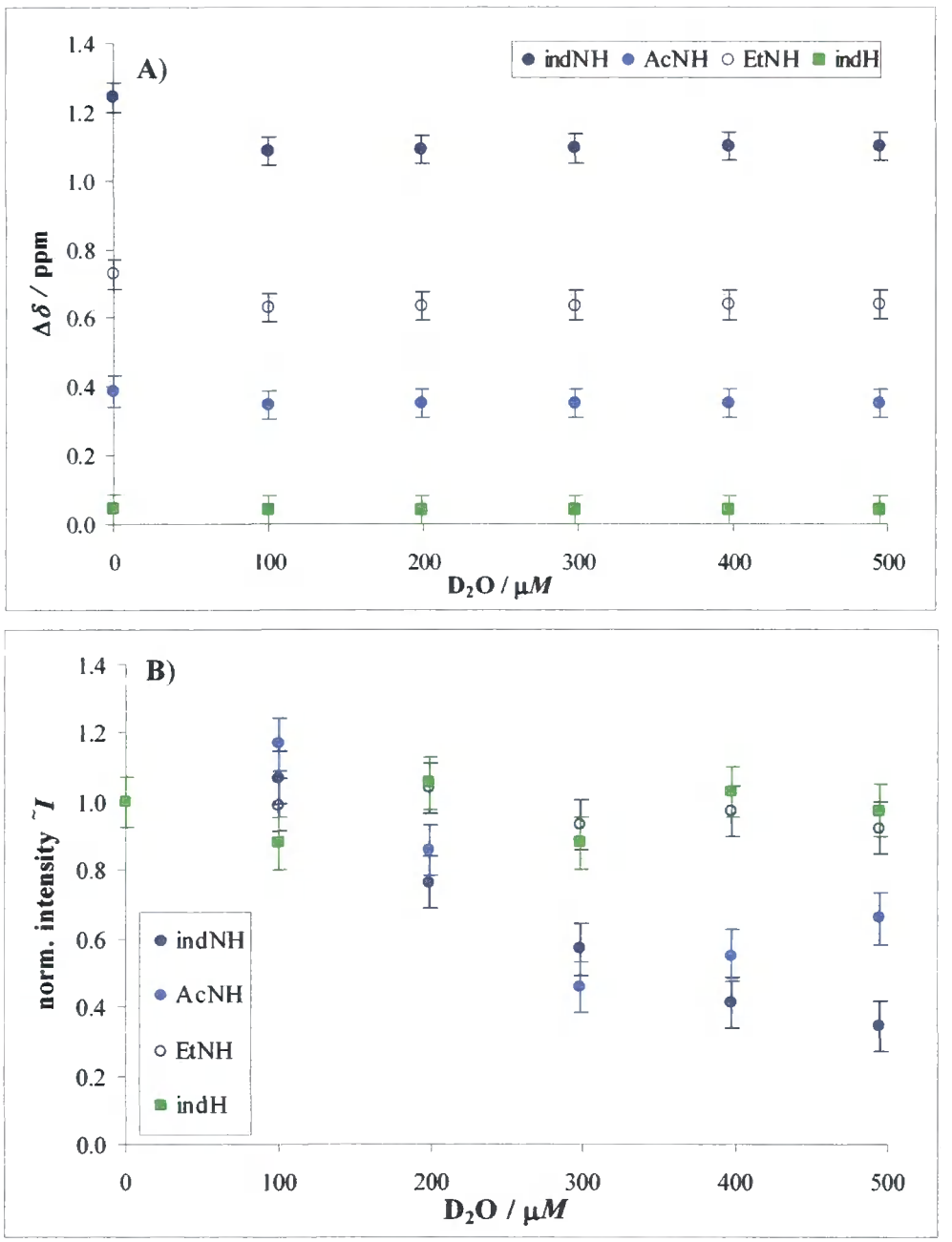


Figure 2.13: Effects of D_2O additions on tryptophan-lipid samples as a function of D_2O added. **A** shows the observed chemical shift chage $\Delta\delta$ of N^{α}-acetyl-*L*-tryptophan ethyl amide (**6n**) protons in a 1:1 solution of the tryptophan derivative and DMPC in chloroform (55 μM each) (see Equation 2.9). Measurements were taken immediately following D_2O addition. **B** shows the normalised intensities associated with the signals in **A** (see Equation 2.10).

It is noteworthy, that the observed chemical shift changes are considerably smaller than the observed complexation-induced chemical shift changes of these protons following the addition of DMPC (see §2.3.3.a).

Entry	D_2O	$\Delta\delta_{indN}$	$\Delta\delta_{AcN}$	$\Delta\delta_{EtN}$	$\Delta\delta_{ind}$
	µM	ppm	ppm	ppm	ppm
δ_{free}	-	8.402	6.600	5.846	7.686
I	0	1.243	0.388	0.728	0.044
II	50	1.087	0.347	0.631	0.041
III	100	1.092	0.35	0.632	0.041
IV	150	1.093	0.35	0.636	0.042
V	200	1.098	0.351	0.637	0.042
VI	250	1.098	0.352	0.64	0.041
II-I	-	0.145	0.036	0.088	0.003

Table 2.11: Chemical shift change $\Delta\delta$ after D_2O additions. The chemical shift change after addition of 100 µ*M* D_2O (entry **II**) with respect to the chemical shift change prior addition (entry **I**) is highlighted. The chemical shift of the signal in the absence of both DMPC and D_2O, δ_{free} is given as reference.

As expected, the intensities associated with the indole C-H signals (Figure 2.13) were not affected following the addition of deuterium. The normalised intensity of the ethyl amide proton seemed rather unaffected by the addition of D_2O, while the intensities of both, the indole amine and the acetyl amide protons decreased. This can be rationalised by a relative fast exchange of these protons with deuterium. The precise nature and rate of the exchange is detailed in §2.3.5.b

2.3.5.b Exchange Kinetics

The functional groups of N^{α}-acetyl-*L*-tryptophan ethyl amide (**6n**), the indole NH, the ethyl and acetyl amide protons, can participate in hydrogen bonding. The kinetics of these exchanges can give an indication of the presence and strength of such hydrogen bonds. Therefore, chemical shifts were recorded over time following the addition of 100 µ*M* D_2O (Figure 2.14). The kinetic effects observed were small (Figure 2.14, Table 2.12).

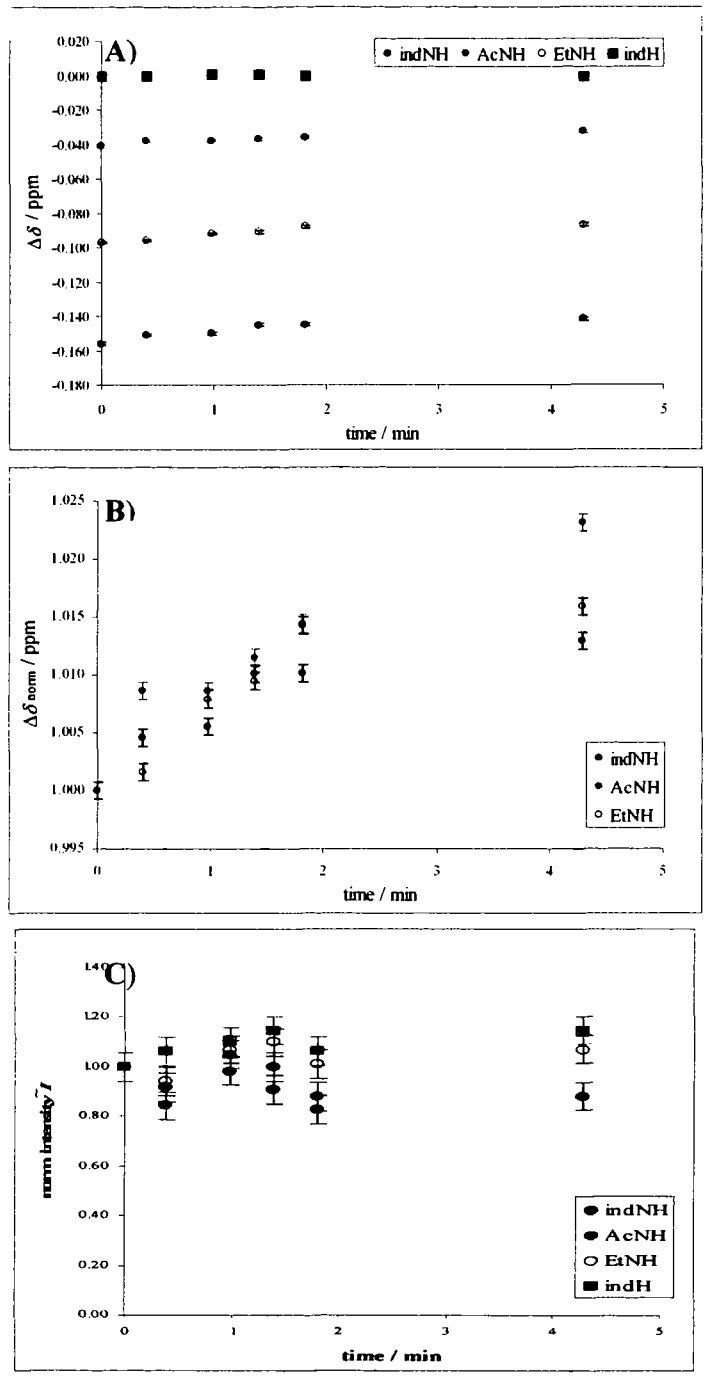


Figure 2.14: Effects of 100 μ*M* D_2O addition on a tryptophan-lipid sample as a function of time. **A** shows the chemical shift change $\Delta\delta$ of N^{α}-acetyl-*L*-tryptophan ethyl amide (**6n**) protons immediately after addition of 100 μ*M* D_2O to a 1:1 solution of **6n** and DMPC in chloroform (55 μ*M* each). **B** shows normalised values of the changing signals in **A** (see Equation 2.11). **C** shows the normalised intensities associated with the signals in **A**.

Entry	time	$\Delta\delta_{indN}$	$\Delta\delta_{AcN}$	$\Delta\delta_{EtN}$	$\Delta\delta_{ind}$
	min	ppm	ppm	ppm	ppm
I	0	-0.156	-0.041	-0.097	0.004
II	0.4	-0.151	-0.038	-0.096	0.003
III	1.0	-0.150	-0.038	-0.092	0.003
IV	1.4	-0.145	-0.037	-0.091	0.004
V	1.8	-0.145	-0.036	-0.088	0.003
VI	4.3	-0.142	-0.033	-0.087	0.003
VI-I	δ_{t0}	0.014	0.008	0.010	0.001
VII	$\Delta\delta_{shift1}$	1.243	0.388	0.728	0.044

Table 2.12: Time-evolution of the chemical shift change $\Delta\delta$ after addition of 100 μM D_2O. The difference in chemical shift change over 4.3 min is highlighted. $\Delta\delta_{shift1}$ refers to the chemical shift change from the signal upon addition of 100 μM D_2O with respect to the signal of free tryptophan derivative in the absence of both, DMPC and D_2O. δ_{t0} is the chemical shift change immediately after the addition of 100 μM D_2O.

As expected, the normalised observed chemical shift for the indole C-H proton did not change over time, as such a proton is unlikely to be exchanged by deuterium. The variation in its absolute value could therefore be used as a measure of the accuracy of the other observed chemical shift changes (± 0.001 ppm). In contrast, the proton chemical shift values for the indole amine, ethyl amide and acetyl amide protons were slowly increasing over time, shifting back towards the value of the free tryptophan derivative prior addition of D_2O, δ_{free}. This can be visualised by normalising the chemical shift change using

$$\Delta\delta_{norm} = \frac{\Delta\delta_{shift1}}{\delta_{t0} - \delta_{free}}, \qquad (2.11)$$

where $\Delta\delta_{shift1}$ refers to the chemical shift change from the signal upon addition of 100 μM D_2O with respect δ_{free}, and δ_{t0} refers to the chemical shift change immediately after the addition of D_2O. It was observed that δ_{t0} was very small given the accuracy of the spectrometer (Table 2.12, Entry **VI-I**). Also, given the large complexation-induced shift changes, this almost amounts to no change over this period.

No clear trend was observed for the time-evolution of the intensities associated to the proton signals. Within experimental error, the intensities remained constant during the analyzed time span. This suggests that no deuterium exchange occurred while the measurements were taken, as this would lead to a reduction in intensity. A deuterium

exchange with the observed hydrogens had not occurred in the time the ^{1}H NMR spectra could be taken (*i.e.* the 30 seconds sample preparation). To test for slow exchange equilibria, the amount of D_2O was increased to 495 μM and the chemical shift change observed over a longer time span (Figure 2.15). An increase in deuterium will perturb existing equilibria towards deuterated species, increasing the probability that exchangeable hydrogens will be replaced.

The observed chemical shift changes $\Delta\delta$ of the analysed protons did not change significantly ($\Delta\delta < 0.04$ ppm) over time (Table 2.13). However, the normalised intensities of the observed signals indicate a trend. Both, the signal for the acetyl amide and for the ethyl amide protons decrease over time to approximately 12 and 30% of their original value at $t = 0$ min respectively, suggesting that a deuterium exchange has taken place. The slow rate of the exchange (half-life, $t_{½}$ ~40 min and ~200 min respectively) suggests that both protons are involved in hydrogen bonding interactions.

Entry	time	$\Delta\delta_{indNH}$	$\Delta\delta_{AcNH}$	$\Delta\delta_{EtNH}$	$\Delta\delta_{indH}$
	min	ppm	ppm	ppm	ppm
I	0	-0.512	-0.064	-0.264	0.012
II	8	-0.505	-0.06	-0.256	0.011
III	20	-0.496	-0.058	-0.254	0.012
IV	32	-0.497	-0.057	-0.248	0.011
V	45	-0.498	-0.059	-0.251	0.01
VI	59	-0.494	-0.057	-0.248	0.011
VII	110	-0.521	-0.058	-0.257	0.011
VIII	243	-0.527	-0.057	-0.261	0.011
IX	301	-0.534	-0.06	-0.26	0.011
X	518	-0.485	-0.029	-0.223	0.01
X-I	-	0.027	0.035	0.041	-0.002

Table 2.13: Time-evolution of the chemical shift change $\Delta\delta$ after addition of 495 μM D_2O. The difference in chemical shift change after 518 min with respect to Entry **I** is highlighted.

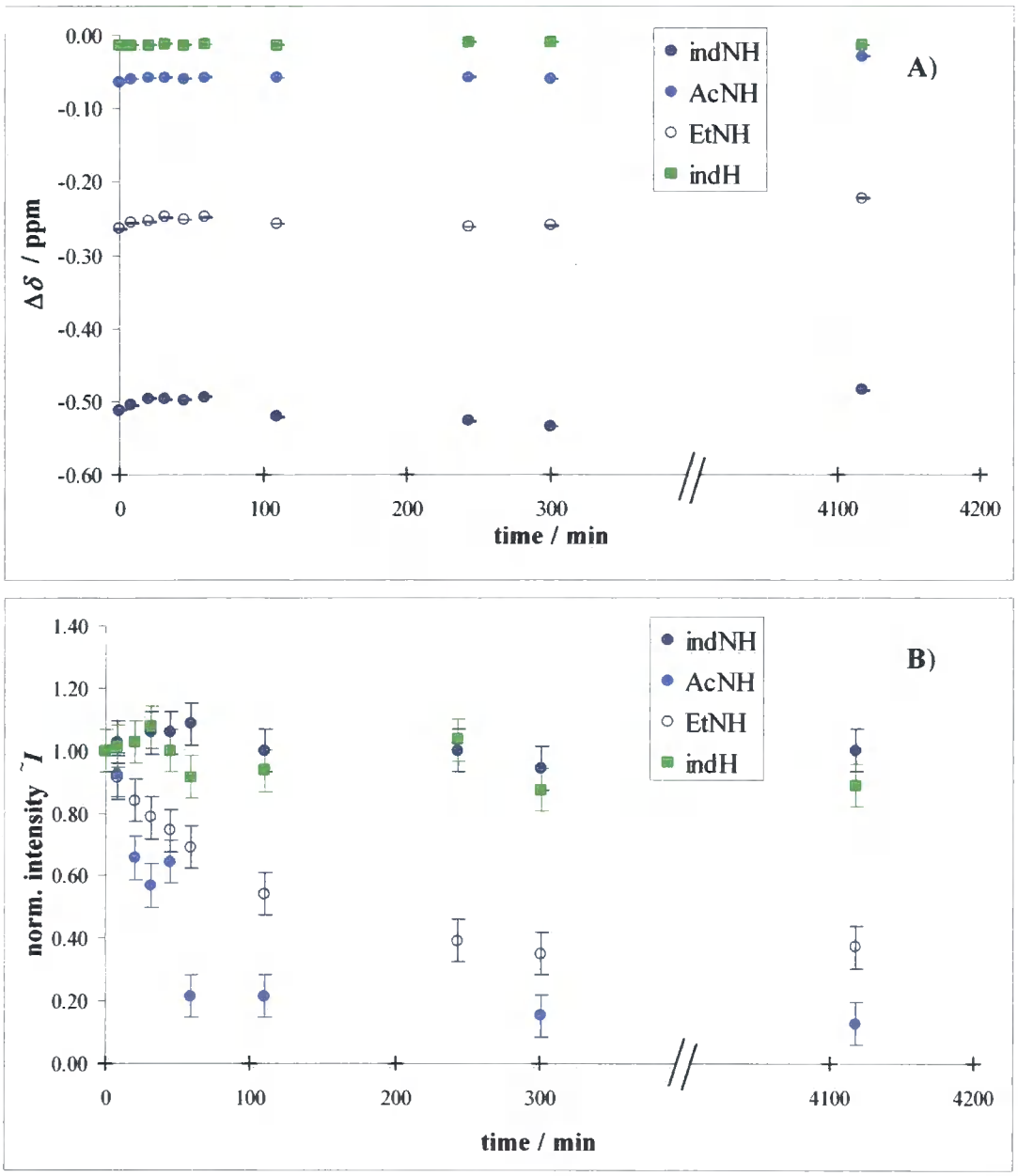


Figure 2.15: Effects of 495 μ*M* D_2O addition on a tryptophan-lipid sample as a function of time. **A** shows the observed chemical shift change $\Delta\delta$ of N^{α}-acetyl-*L*-tryptophan ethyl amide (**6n**) protons in a 1:1 solution of the tryptophan derivative and DMPC in chloroform (55 μ*M* each) as a function of time immediately after addition of 495 μ*M* D_2O. **B** shows the normalised intensities associated to the signals in **A**.

The indole amine proton does not show a further reduction in signal intensity. The drop in intensity of its signal after each successive addition of D_2O indicate a fast proton/deuterium exchange, which is in agreement of previous studies of an exchange of the

indole amine proton.[13] In water the indole amine proton of WALP16 peptides was reported to exchange with a life-time < 3 min.[14]

The data for the chemical shift changes and their associated changes in signal intensity provide an insight into possible kinetic and thermodynamic events that characterise binding. Upon addition of D_2O, protons of free, unbound water are expected to exchange quickly and lower the effective concentration of unbound water. This can be seen directly by a reduction of the signal arising from water protons (-20%). Protons of bound water molecules can potentially undergo exchange as well, especially if the waters are coordinating the ammonium group of the lipid, but unbound water protons are expected to exchange the most rapidly.

This will affect all the association equilibria of H_2O with DMPC and tryptophan-DMPC complexes: bound H_2O will be perturbed in favour of water-free adducts, and D_2O is expected to bind with a higher binding association constant than H_2O or HOD (order D_2O, HOD, H_2O). These fast exchanges were most likely completed before a spectrum could be recorded.

After these fast equilibria are established, the hydrogen bonding protons of the indole N*H* and the two amide functional groups are likely to be exchanged for deuterium. The indole amine (indole NH) exchanged the fastest ($t_{½}$ < 30 s) and had already equilibrated within the time of the mixing of the sample with D_2O and the recording of the first ^{1}H NMR spectrum (~30 s). Although no exchange kinetics (*i.e.* time-dependent shifts in the complexation induced chemical shift of the indole NH proton) were observed, a fast exchange of the proton could be inferred from the recorded drop in signal intensity of the indole NH peak after each successive addition of deuterium.

The acetyl amide proton (AcNH) was the next fastest to exchange. The initial exchange occurs relatively quickly and similarly to the indole amine proton as the AcNH signal intensity dropped markedly between the times D_2O was added to the sample and a spectrum had been run. However, an equilibrium had not been reached with respect to the AcNH signals as its intensity further decreased over time suggesting further exchange reactions with deuterium or unbinding of the deuterated species from adducts that contribute towards the observed complexation induced chemical shift The slow rate of this further decrease in signal intensity could be associated with an estimated half-life of $t_{½}$ ~40 min after addition of 496 m*M* D_2O.

The ethyl amide proton (EtNH) was initially virtually unaffected by the addition of D_2O to the sample. No noticeable drop in intensity was observed with increasing D_2O concentration, suggesting that no immediate exchange had taken place. Only when the sample was studied over a longer period of time ($t > 30$ min) was a gradual decrease in signal intensity with an estimated half-life of $t_{½}$ ~200 min observed. The slow decrease in signal intensity is indicative of a slow proton/deuterium exchange which could be caused by either strong hydrogen bonding interactions of EtNH or by an inaccessibility of the proton for an exchange.

2.3.6 Relationships between experimentally determined ΔG values and physical properties of the tryptophan derivatives (6a-g, 6n) used

Calculated and experimentally determined free energies of the binding interactions were also plotted against parameters of the tryptophan compounds other than the Hammett parameter, such as dipole moment of the indole ring, solvent accessible area, and log P, a measure of relative solubility in octanol and water. All of these parameters were computed for 5-monosubstituted 3-ethyl indoles, which were used as model compounds for the tryptophan analogues, using *ab initio* calculations with the software package GAUSSIAN[15]. The total dipole moment of the compounds was found to point parallel to the indole amine nitrogen-hydrogen bond and assume magnitudes between 1.8 and 7.5 *D* (§2.5.1.b).

No simple pattern was found when any of these parameters were plotted against the free energy of 1:1 or 2:1 complex formation of the derivative with DMPC (for more details see §7.4).

2.4 ITC studies

2.4.1 General Considerations and Theory of the Analysis of Isothermal Titration Calorimetry Data

Isothermal titration calorimetry experiments can potentially give access to thermodynamic data for the association process between N^{α}-acetyl-*L*-tryptophan ethyl amide derivatives interacting and DMPC in chloroform. A rather high level of noise was observed during initial experiments even though the reference cell of the calorimeter was filled with chloroform, which prevented the measurement of the reaction heats of single compounds in dilute

samples. The addition of concentrated lipid solutions to a measuring cell containing tryptophan analogues lead to dissolution and de-micellisation effects that gave intense heat signals with an acceptable signal-to-noise level. The heat associated with these processes could be accounted for by subtractions of the heats of injections from a DMPC blank titration from signals of the DMPC-tryptophan titration. This allowed the heat of reaction due to the interactions between the lipid and tryptophan derivatives to be measured as an added effect on the demicellisation isotherm. Typical values were a heat flow of around 30 μcal s^{-1} for an injection of 8 μl of DMPC (26 m*M*) into neat chloroform and approximately 20 μcal s^{-1} for an injection of 8 μl into a measuring cell containing a tryptophan analogue at about 4 m*M* concentration. This implies that the binding of tryptophan derivatives to DMPC is exothermic, as expected.

Initial data analysis was performed using the MicroCal Systems analysis software (MicroCal Origin, version 5.0). The base line was automatically generated by the program and was used except when it was obviously accepting an errant data point as reference, in which case these base-line points were corrected by hand before further analysis. The software then integrated each peak to yield integrated heats, which were normalized to the moles of lipid in the injectant.

2.4.1.a Model for Fitting

There are three different types of model for self-assembly that allow consideration of dimerisation (with binding constant K_d) and oligomerisation (binding constant K_m). 1) The isodesmic model assumes that all binding constants are equal ($K_d = K_m$); 2) in cooperative binding, the formation of smaller complexes is slower than the formation of larger aggregates ($K_d < K_m$); 3) the anti-cooperative model assumes the opposite of the cooperative model such that the binding constant decreases with growing complex ($K_d > K_m$). In this simplified approach, these models do not account for the aggregation of oligomers or other secondary effects that might influence the binding constants (such as competing processes *e.g.* water binding). However, they do serve as a good model description for the data within the limitations of a reasonable number of variables for fitting. Different self-association models containing adjustable parameters were fitted to the observed heat change *versus* concentration data *via* a Marquardt-Levenberg minimization of χ^2.[16]

2.4.1.b Theory

At the beginning of the experiment, all lipid molecules are contained in the syringe (superscript syr). The total concentration of lipid molecules, L_0^{syr} will be present in micellar and monomer forms and can be accounted for by

$$L_0^{syr} = \sum_{n=1}^{\infty} nL_n^{syr}, \tag{2.12}$$

where L_n^{syr} stands for the concentration of lipid aggregates with size n.

The concentration of all lipid micelles with size n (L_n) can be expressed by applying the mass action law:

$$L_n = K_d K_m^{n-2} (L)^n. \tag{2.13}$$

If micelle-micelle aggregation is neglected, conservation of mass yields a cubic equation

$$K_m(K_m - K_d)L^3 + (2K_d - 2K_m - K_m^2 L_0)L^2 + (1 + 2K_m L_0)L - L_0 = 0 \tag{2.14}$$

from which the free monomer concentration L can be determined, if K_d, K_m, and the overall concentration L_0 are known (see §7.1 for details of how to solve the cubic equation).[17,18] Through equation (2.13), all lipid concentrations can be determined.

During the course of an ITC experiment, the volume of the measuring cell is kept constant at V_0. When i injections of volume V_i have occurred, the total volume injected is

$$V_i = \sum_{j=1}^{i} \Delta V_j \tag{2.15}$$

and the total solute concentration in the cell is $L_0^{cell}(i)$. Let $L_n^{cell}(i)$ be the concentration of micelles with size n in the cell, then the total concentration after the i^{th} injection is:

$$L_0^{cell}(i) = \sum_{n=1}^{\infty} nL_n^{cell}(i). \tag{2.16}$$

Because of the total-fill nature of the measuring cell, each injection drives liquid out of the cell. Consequently, the concentration of the cell is given by equation 2.17,[19] which contains an according correction of the volume:

$$L_0^{cell}(i) = L_0^{syr} \frac{V_i}{V_0}\left(1 - \frac{V_i}{2V_0}\right). \tag{2.17}$$

The heat exchanged [$Q_{obs}(i)$] during the i^{th} injection is the sum of two terms:

$$Q_{obs}(i) = \Delta Q(i) + Q_{dil}. \tag{2.18}$$

$\Delta Q(i)$ results from the dissociation of micelles with size n into smaller sized micelles and monomers; Q_{dil} accounts for both the mechanical heat related to the flow of the solution (which is independent of i) and the heat of dilution of solutes (*i.e.* the micelles). This heat of dilution does not involve dissociation and can be considered to be independent of i, if the concentration in the syringe is so low that nonideality effects are negligible.[20] $\Delta Q(i)$, which corresponds to the injection of volume $\Delta V_i = V_i - V_{i-1}$, can be expressed as:

$$\Delta Q(i) = Q(i) - Q(i-1) + \frac{\Delta V_i}{V_0}\left[\frac{Q(i) + Q(i-1)}{2}\right], \tag{2.19}$$

where $Q(i)$ is the heat exchanged when a micellar solution of volume V_i and concentration L_0^{syr} is injected into the cell containing pure solvent. The last term in equation 2.19 takes the displaced volume into account that overflows the cell but contributes partially to the heat measured.[19] The heat $Q(i)$ corresponds to the dissociation of micelles when the concentration of a solution is reduced from L_0^{syr} to $L_0^{cell}(i)$. It can be evaluated by considering the fictitious process of first diluting the solution from concentration L_0^{syr} to infinite dilution, where only monomers are present, and then concentrating back to $L_0^{cell}(i)$:

$$Q(i) = Q(i)_{L_0^{syr} \to 0} + Q(i)_{0 \to L_0^{cell}(i)}. \tag{2.20}$$

These hypothetical heats are related to the molar enthalpies of association corresponding to dimerisation (ΔH_2) and micellisation (ΔH_m) (Scheme 2.10, with $n = 2$ or $n > 2$, respectively):

$$L + L_{n-1} \underset{}{\overset{K,\, n \geq 2}{\rightleftharpoons}} L_n$$

Scheme 2.10: Self-association of lipid molecules L forming a lipid aggregate L_n with size n and an associated binding constant K (which corresponds to K_d for $n = 2$, and K_m for $n > 2$)

$$Q(i)_{L_0^{syr} \to 0} = -\Delta H_m\left[\sum_{n=3}^{\infty}(n-2)L_n^{syr}V_i\right] - \Delta H_d\left[\sum_{n=2}^{\infty}L_n^{syr}V_i\right], \tag{2.21}$$

$$Q(i)_{0 \to L_0^{cell}(i)} = \Delta H_m\left[\sum_{n=3}^{\infty}(n-2)L_n^{cell}(i)V_i'\right] + \Delta H_d\left[\sum_{n=2}^{\infty}L_n^{cell}(i)V_i'\right]. \tag{2.22}$$

V_i' is a hypothetical volume such that the concentration of solutes fictitiously diluted and concentrated back are identical:

$$L_0^{syr} V_i = L_0^{cell}(i) V_i'. \tag{2.23}$$

Combining Equations 2.13 and 2.20-2.22 yields:

$$\begin{aligned} Q(i) = -\Delta H_m K_d K_m V_i \left\{ \frac{\left(L_1^{syr}\right)^3}{\left(1 - K_m L_1^{syr}\right)^2} - \frac{L_0^{syr}\left[L_1^{cell}(i)\right]^3}{L_0^{cell}(i)\left[1 - K_m L_1^{cell}(i)\right]^2} \right\} \\ - \Delta H_d K_d V_i \left\{ \frac{\left(L_1^{syr}\right)^2}{\left(1 - K_m L_1^{syr}\right)} - \frac{L_0^{syr}\left[L_1^{cell}(i)\right]^2}{L_0^{cell}(i)\left[1 - K_m L_1^{cell}(i)\right]} \right\} \end{aligned} \tag{2.24}$$

which is then introduced in equations 2.19 and 2.18. The observed heat of change, $Q_{obs}(i)$, is then normalised by the amount of solute injected:

$$\tilde{Q}_{obs}(i) = \frac{Q_{obs}(i)}{L_0^{syr} \Delta V_i}. \tag{2.25}$$

Consequently, five unknown constants (K_d, K_m, ΔH_d, ΔH_m, and Q_{dil}) are necessary to describe such a calorimetric experiment completely.

2.4.2 Analysis of Isothermal Titration Calorimetry Data

2.4.2.a Titrations of Hydrated DMPC Solutions into Chloroform

The observed heat in titration experiments of DMPC into chloroform will include contributions from various processes. 1) An endothermic heat exchange arising from the unbinding of water that was associated with the lipid (which is expected to be a major contribution as the lipid used contained approximately 2 water molecules per lipid molecule, see §2.3.2.a). 2) There will be an endothermic process consistent with the disruption of molecular interactions between monomers in the micelle (such as hydrogen bonds, van der Waals and electrostatic interactions) during demicellisation; and 3) heat of dilution. At concentrations far below the critical micelle concentration (CMC) of the lipid, the heat of demicellisation will be negligible.

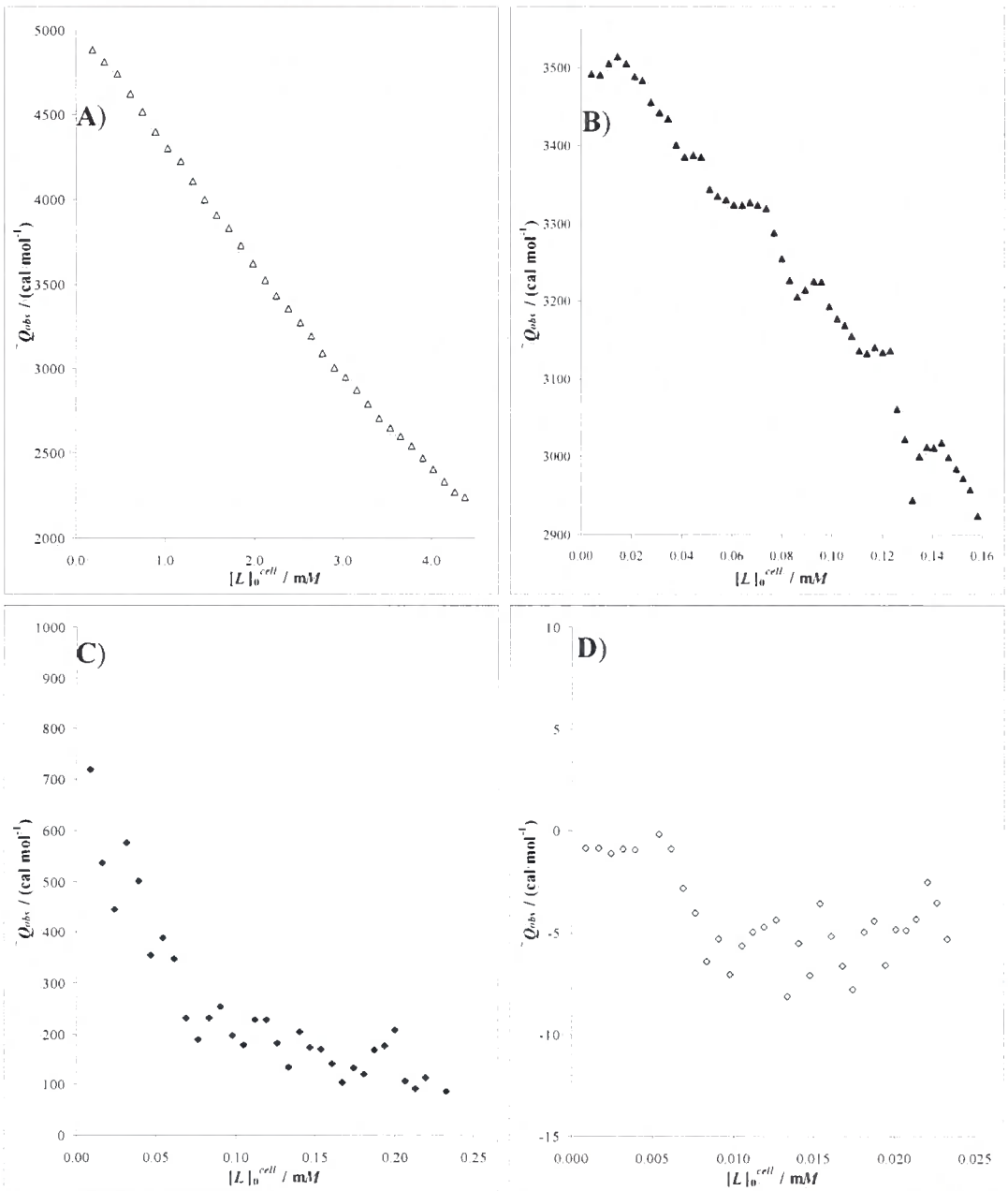


Figure 2.16: Enthalpograms of hydrated DMPC titrated into chloroform, measured at 25 °C as a function of lipid added. The starting lipid concentrations were 26 m*M* for **A**, 14 m*M* for **B**, 1.4 m*M* for **C** and 0.14 m*M* for **D**.

Using dilute solutions of DMPC in chloroform will render the heat of dilution to be relatively low. Therefore, injections of dilute DMPC samples into chloroform allowed an estimation of the contribution of the unbinding of water (Figure 2.16, **D**).

A sigmoidal curve was observed when a very dilute (140 μM) solution of hydrated DMPC was titrated into chloroform. Estimates of the relative concentrations of the species

present could be made using the equations derived for the NMR host-guest titrations (§7.2), suggesting that 2% of the lipid molecules belong to a micellar form, 57% as free, unbound lipid monomers, and ~41% as present in water:lipid adducts. The observed heat of injection dropped by about -6 cal mol^{-1} and may be associated with the heat change of unbinding water from the lipid. The error at this low concentration is likely to be large due to the small heats observed and the critical influence of the accuracy of the starting concentration.

At higher starting concentrations (> 1.4 m*M*) the total observed heat change increased significantly. Now, the contributions from demicellisation and dilution processes are expected to become more and more significant. At a syringe concentration of 1.4 m*M* DMPC, the observed heat was two orders of magnitude larger than the observed heat of the ten times less concentrated starting solution and followed a more exponential trend (Figure 2.16, **C**). In comparison with the syringe concentration of 0.14 m*M* DMPC, the amount of micellar lipid has increased by 11% reducing the amount of water:lipid adducts and free, unbound lipid monomers by 8% and 4% respectively.

At 14 m*M* lipid solution in the syringe, a sigmoidal step-like pattern was observed in the heat flow profile (Figure 2.16, **B**). This pattern is likely to be an artefact from rather than distinct sequential unbinding events. The absolute heat change per injection was approximately one order of magnitude higher than the heat change observed for a ten times less concentrated lipid starting solution (1.4 m*M*).

When 26 m*M* DMPC solution was injected into the sample cell, the maximum response level was achieved (Figure 2.16, **A** and Figure 2.17), *i.e.* the heat flow change upon injection was approaching the upper detection limit of the instrument (40 μcal s^{-1}). The associated observed heat adopted a linear pattern with respect to the moles of injectant (Figure 2.17, **B**) and the sequential sigmoidal shape observed when using 14 m*M* DMPC solution was no longer detected. The concentrations of micellar lipid, unbound, monomeric lipid, and lipid:water adducts were calculated to be 75%, 15% and 11% respectively. The observed endothermic process and associated heats are now likely to be dominated by heats of demicellisation and dilution and add to the heat of water unbinding, which will now be less significant.

With the maximum response level (and the best signal-to-noise ratio) achieved, a concentration of 26 m*M* was chosen as concentration for the DMPC guest solution in tryptophan-lipid host guest titrations. The lipid concentration was approximately 25% of the typical concentration used in the NMR host-guest titrations, which made a comparison of the

obtained results possible. All runs were performed in duplicate to ensure consistency. Five repeat runs were carried out for the DMPC solution at 26 m*M* concentration. A typical run is shown in Figure 2.17.

It is noteworthy that it is not possible to obtain the full sigmoidal curve for DMPC demicellisation in a single experiment.

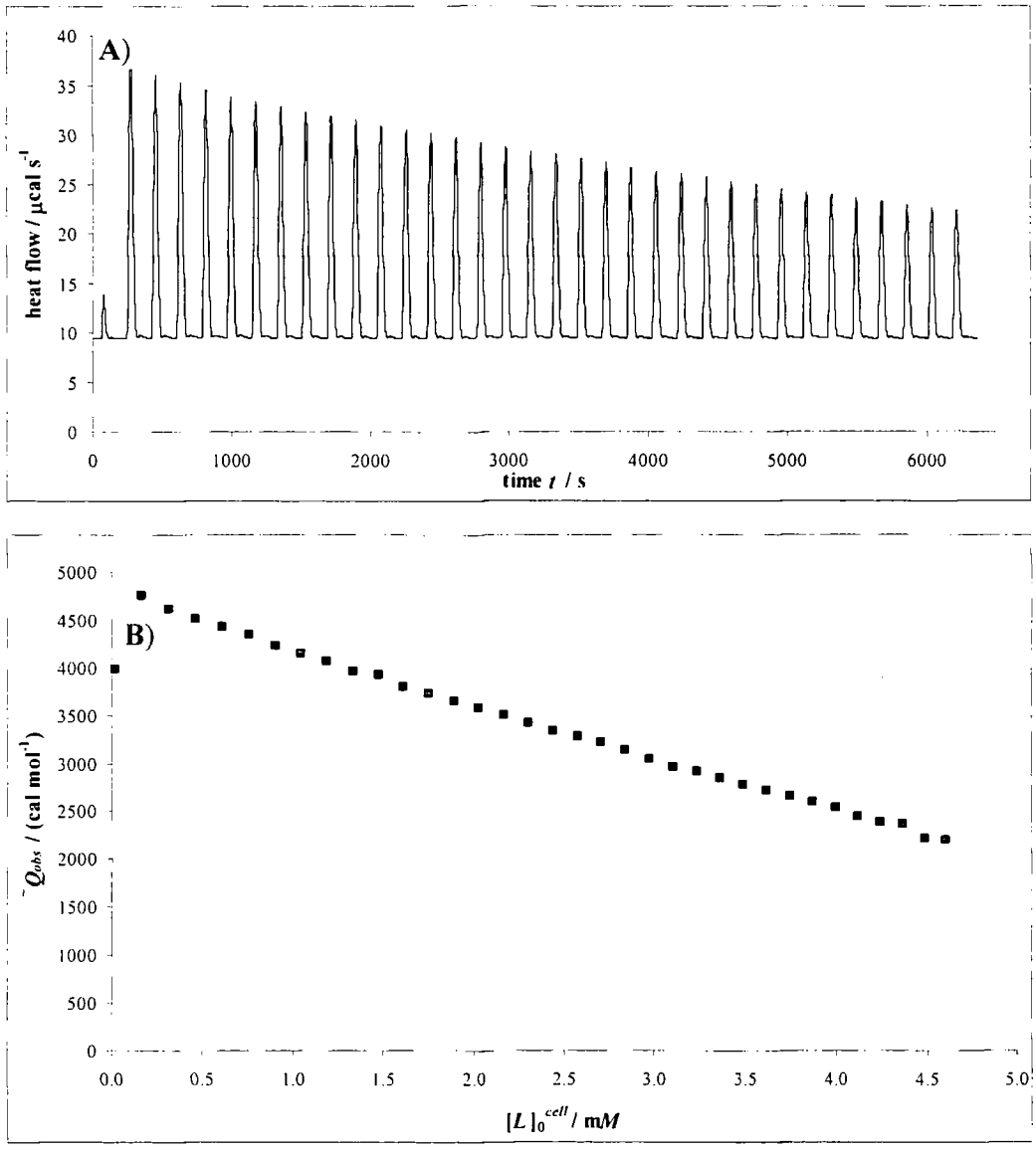


Figure 2.17: Enthalpogram of a titration of hydrated DMPC (26.26 m*M*) into chloroform. **A** depicts the heat flow curve at 25 °C. **B** shows the normalised observed heat associated with the signals in **A**.

2.4.2.b *CMC Determination of DMPC using ITC*

With a complete loading of the syringe at the highest feasible lipid concentration (~26 m*M*), the total concentration of lipid in the measuring cell reached a maximum of about 4.5 m*M*, which is below the critical micelle concentration (CMC) of DMPC in chloroform at 25 °C (16 m*M*, §2.3.2.b). To reach such a concentration range required the injection of highly concentrated (>100 m*M*) DMPC solutions. However, the heat associated with the first injections exceeded the maximum upper detection limit of the instrument at these concentrations when using the original settings (8 µl injections over 16 s). By reducing the injection volume and the duration of the injection, the heat flow peaks could be somewhat lowered through broadening. However, even at 5 µl injections over 25 s, the signals for the first 27 injections exceeded the upper detection limit of the instrument. Any associated heat of these injections is likely to be underestimated, making them less useful for the analysis of the data. This lack of data could be compensated by the injection of less concentrated DMPC solutions. Furthermore, the CMC of DMPC in chloroform is expected to be at lower lipid concentrations at lower temperatures.[9] Therefore, the heat of dilution of concentrated (100 m*M*) lipid solutions was also measured at a lower temperature (21 °C) (Figure 2.18). This also allowed a direct comparison to the CMC obtained by extrapolation of data gained in the NMR titrations of hydrated DMPC into chloroform (§2.3.2.b).

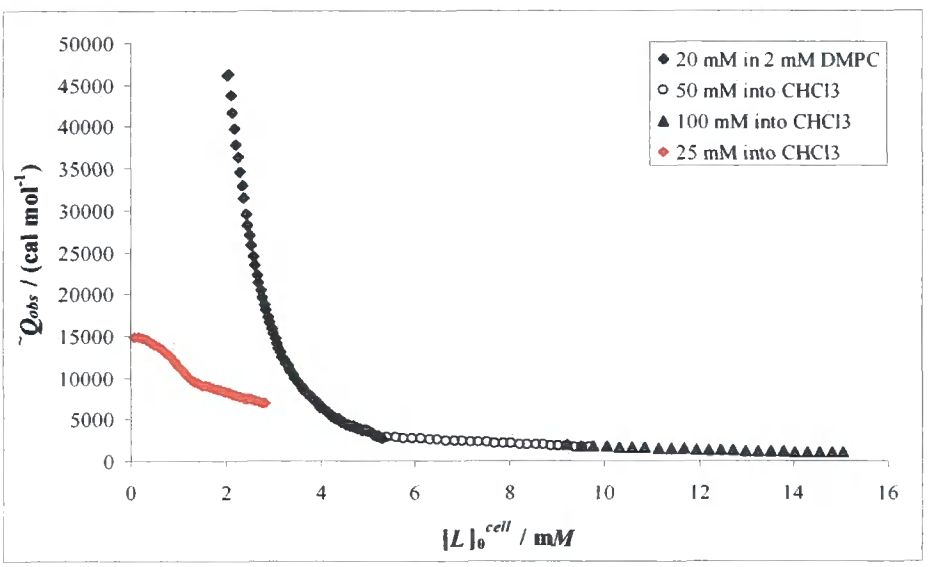


Figure 2.18: Enthalpogram of DMPC solutions being diluted into chloroform, measured at 21 °C as a function of lipid added. A titration of 25 m*M* hydrated DMPC into chloroform was added as reference, however as the heats of injections were too large and exceeded the upper detection limit of the calorimeter they are likely to be underestimated and are therefore highlighted in red.

Due to the intense heats of injections, it proved to be difficult to obtain the full expected sigmoidal curve for titrations of hydrated DMPC into chloroform. The graph was complemented by measurements of the heat of injection for hydrated DMPC solutions into less concentrated DMPC solutions. The curve obtained showed a significant change in the observed heat for injections when the total concentration of lipid in the cell, $[L]_0^{cell}$, was <5 m*M*. This heat change is likely to be associated with the dissociation of water molecules from the lipid, as this concentration is far below the expected CMC (~20 m*M*). The concentration range in which the CMC was anticipated would require injections of highly concentrated lipid solutions into solutions of DMPC with a starting concentration of >14 m*M*. These conditions were well removed from the conditions used for the titration of DMPC to chloroform solutions of tryptophan derivatives.

2.4.2.c Analysis of Tryptophan-Lipid Isothermal Host-Guest Titrations

Tryptophan-lipid isothermal host-guest titrations were performed using 26.26 m*M* DMPC guest solution, containing tryptophan analogue (**6n**, **6a-6g**) at a typical concentration of about 3 m*M* (for full details of the procedure see §4.3.3). The set-up used a concentration of tryptophan analogue similar to the one used for the NMR host-guest titrations (~7 m*M*). The DMPC concentration had to be adjusted to ~26% of the concentration used in the NMR titrations as a higher lipid concentration lead to signals that exceeded the upper detection limit of the ITC instrument. The heat signals arising from demicellisation, dilution and unbinding of water from water-lipid adducts had been determined in five blank titrations of the hydrated lipid into chloroform using the same starting concentration (26.26 m*M*). These heat signals were subsequently subtracted from the heat signals observed during the isothermal titration measurements resulting in heat signals that are likely to be caused by binding of tryptophan compound to lipid and water species (Figure 2.19).

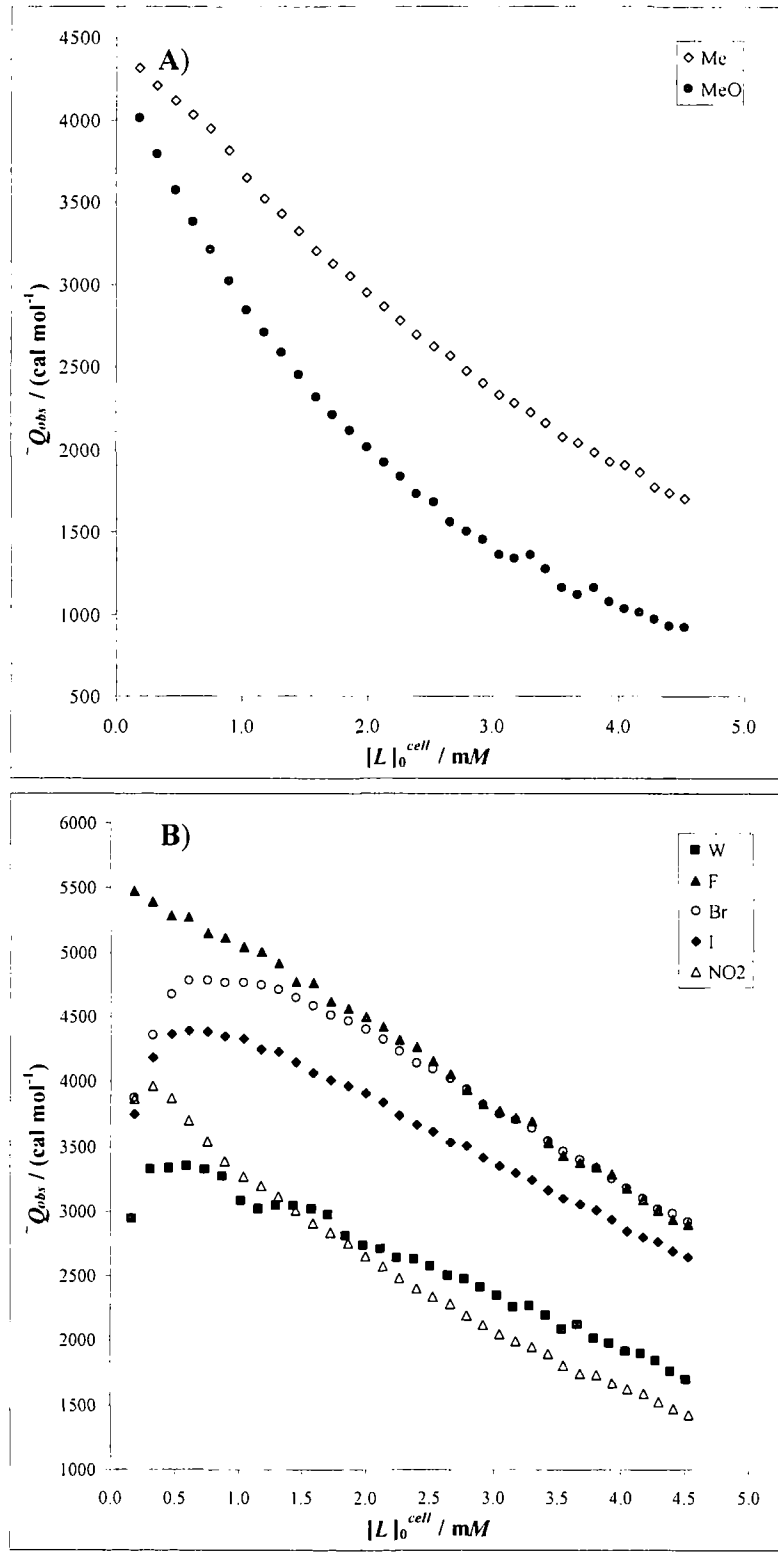


Figure 2.19: Enthalpograms of the normalized observed heats ($\tilde{Q}_{obs}$) of isothermal host-guest titrations tryptophan analogues (**6n**, **6a-6g**) and DMPC lipid at 25 °C as a function of lipid added.

When the normalised heat change $\tilde{}Q_{obs}$ was then plotted against the total lipid concentration in the measurement cell, a near linear pattern of the resulting enthalpograms was obtained. At low lipid concentrations deviations from this linear behaviour were noted: Some of the tryptophan analogues (**6n**, R = H, Figure 2.20 ; **6c**, R = Br Figure 2.20 and **6d**, R = I) exhibited an exothermic incline at lipid concentrations below 0.75 m*M* (Figure 2.19, **A**).

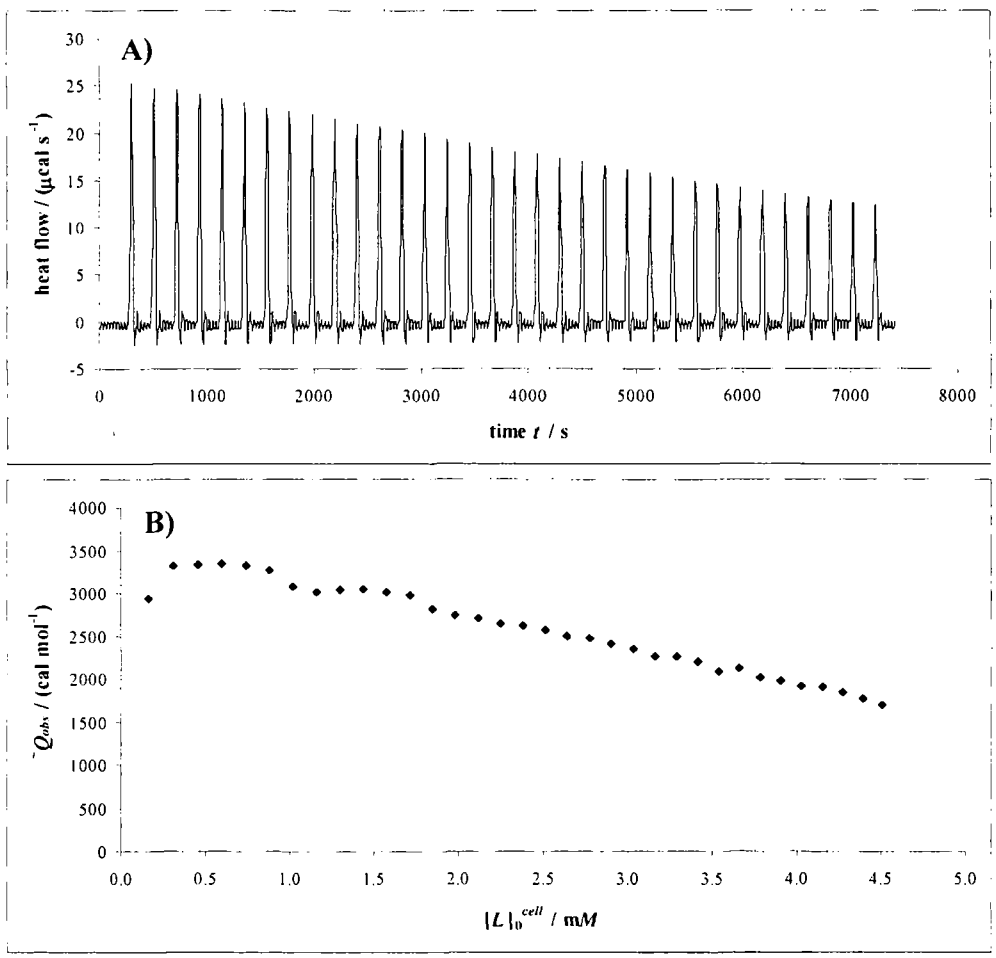


Figure 2.20: Enthalpogram of DMPC (26.26 m*M*) titrated to a solution of N^{α}-acetyl-*L*-tryptophan **6n** (4.9 m*M*) in chloroform at 25 °C. **A** depicts the heat flow curve; **B** shows the normalised observed heat associated with the signals in **A**.

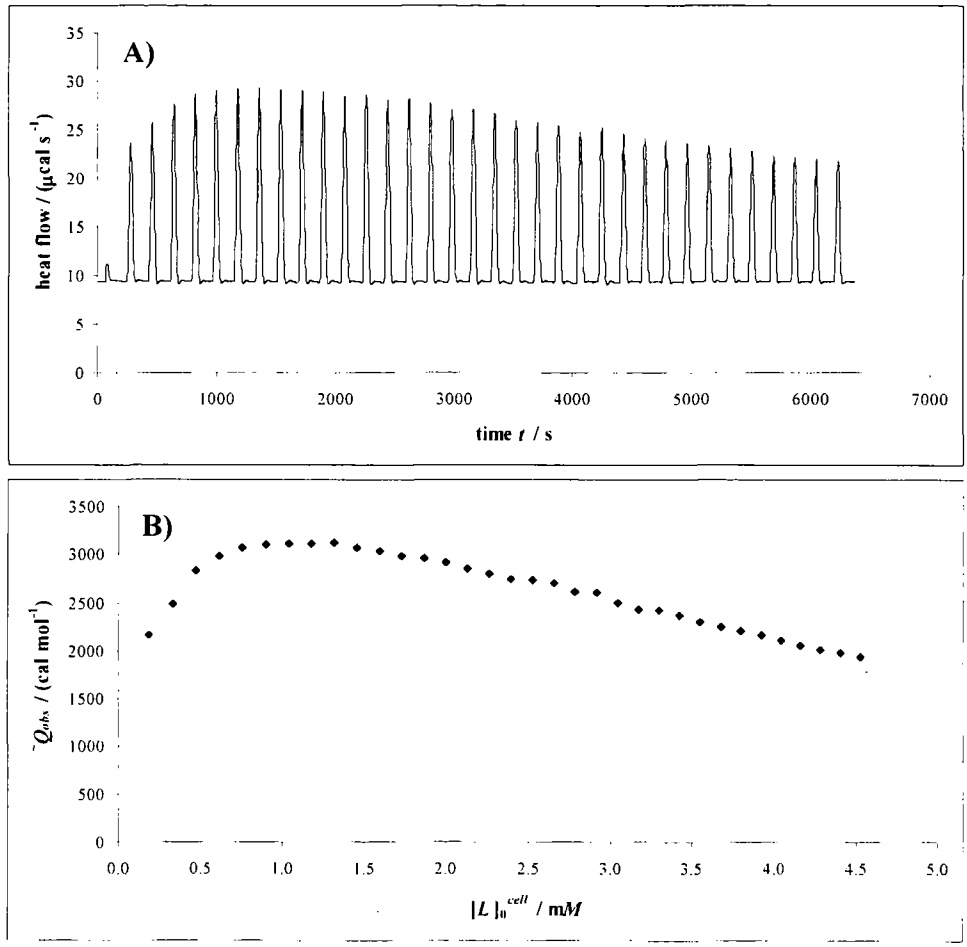


Figure 2.21: Enthalpogram of DMPC (26.26 m*M*) titrated to a solution of N^{α}-acetyl-5-bromo-*L*-tryptophan **6c** (2.2 m*M*) in chloroform at 25 °C. **A** depicts the heat flow curve; **B** shows the normalised observed heat associated with the signals in **A**.

The fluoro-substituted tryptophan analogue (**6a**) showed heat changes that followed a near perfect linear pattern. The heat changes associated arising from titrations with the methyl- (**6e**) and methoxy-substituted (**6f**) amino acid derivatives followed a more exponential pattern (Figure 2.19, **B**; Figure 2.22).

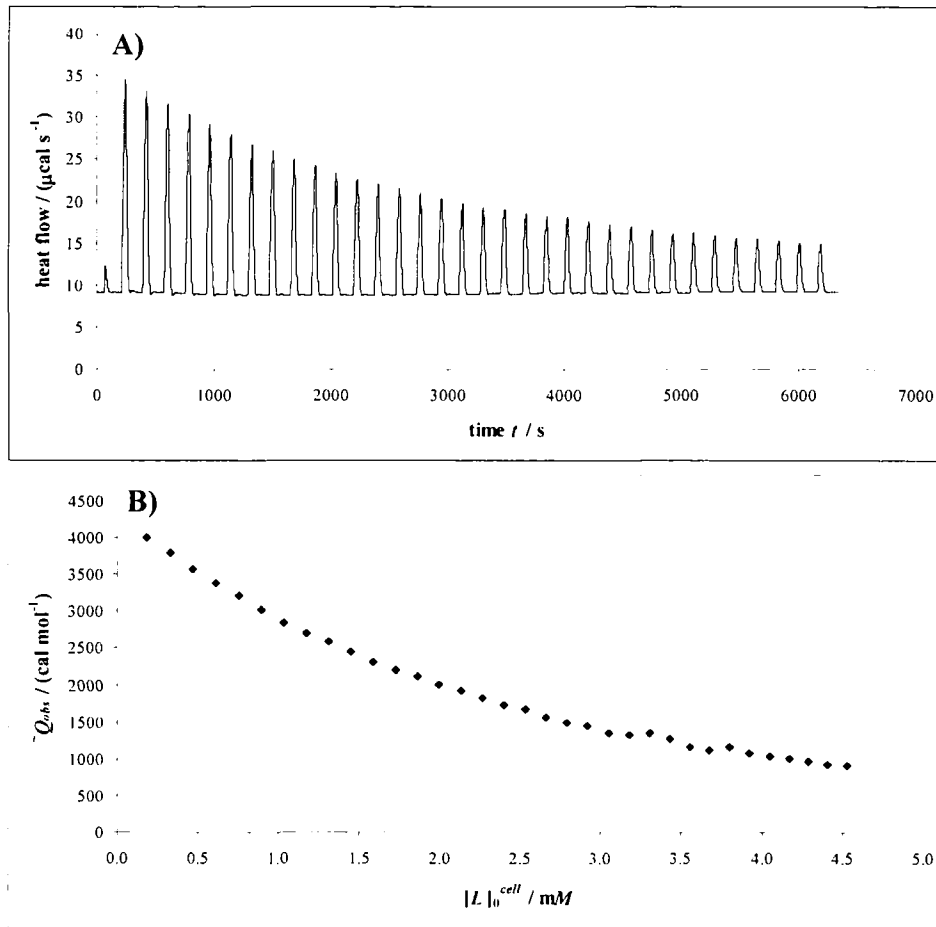


Figure 2.22: Enthalpogram of DMPC (26.26 m*M*) titrated to a solution of N^{α}-acetyl-5-methoxy-*L*-tryptophan **6f** (2.0 m*M*) in chloroform at 25 °C. **A** depicts the heat flow curve; **B** shows the normalised observed heat associated with the signals in **A**.

At higher lipid concentrations (>1 m*M*), the observed heat changes of all compounds analysed could be described by a linear regression with >96% confidence (Table 2.14).

ID	R	Gradient	Intercept	R^2	σ_p
		cal mol^{-1} M^{-1}	cal mol^{-1}		
6n	NO_2	-0.52±0.01	3493±348	0.9907	0.78
6d	I	-0.49±0.02	4828±42	0.9990	0.23
6c	Br	-0.46±0.14	5467±1313	0.9957	0.23
6a	F	-0.63±0.02	5737±582	0.9980	0.06
6n	W	-0.41±0.01	3495±216	0.9615	0
6e	Me	-0.57±0.05	3872±290	0.9885	-0.17
6f	MeO	-0.67±0.03	3142±357	0.9583	-0.27

Table 2.14: Parameters for the linear regression of normalised observed heat signals from tryptophan-lipid isothermal titrations. The regressions were performed for signals at total lipid concentration >1.0 m*M*.

The gradient of the linear regressions might be used as an indicator for the strength of the observed interactions and associated heat signal changes. When plotted against the Hammett parameter, a familiar n-shaped pattern was observed (Figure 2.23).

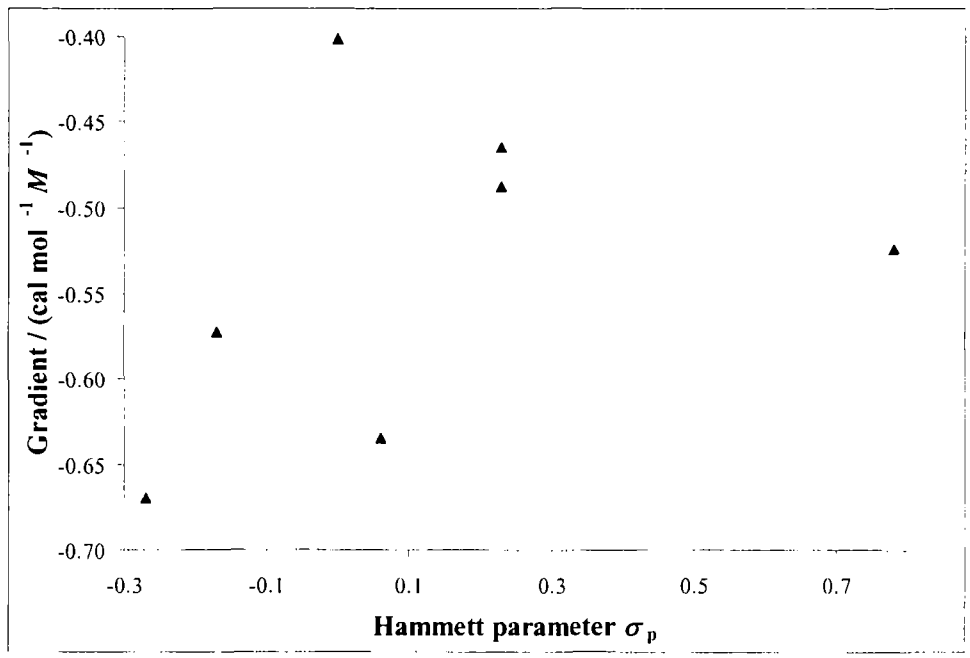


Figure 2.23: Gradient of the linear regression of the heat signals from tryptophan-lipid isothermal titrations plotted against the Hammett parameter.

The fluoro-substituted tryptophan analogue (**6a**) was found not to fit onto an n-shaped pattern. The macroscopic association free energies for tryptophan derivatives binding to DMPC observed in the NMR host-guest titrations (Figure 2.11) compares extraordinarily well with the binding seen in the ITC experiments, suggesting the same approximations made in the data analysis were justified and did not affect the outcome of fitting significantly.

Due to the necessarily low total DMPC concentration during the isothermal titration experiments, a full binding isotherm for tryptophan-lipid interactions could not be obtained. Instead the data was found to lie within the near-linear incline of the binding isotherm prior to the micellisation event and allowed at best a semi-quantitative analysis. A full binding isotherm might be obtained by injections of very small aliquots (<3 μl) of more concentrated lipid solutions (~100 m*M*) using a long injection duration (>90 s) and a large spacing (>240 s) between injections.

Accordingly, the observed heat changes in the ITC experiments are likely to contain contributions due to water equilibria in the system. This was demonstrated with ITC

measurements of hydrated DMPC titrations into chloroform (Figure 2.19). The importance of water in all association events and its competitiveness with tryptophan derivatives for association with the lipid was also shown by the NMR titrations and (as seen later in §2.5.3) in the computational analysis of binding events. The observed heat changes of the ITC experiments are likely to contain an according contribution due to equilibria established with water. It will be difficult to separate the observed heats into the individual contributions.

The unbinding of water from lipid molecules due to competition with the tryptophan compounds, is likely to exhibit an endothermic heat associated with the disruption of molecular interactions. This leads to an intrinsic underestimation of the heat signal arising from tryptophan-lipid interactions. The amount of water and water-associated adducts in the system could be estimated using the equations derived for the NMR titration experiments (§7.2). At the last injection, the highest total water concentration will be reached (19 m*M*). The total concentrations of lipid and the tryptophan analogue at this point will be 4.5 m*M* and ~4 m*M* respectively. Most (79%) of the water molecules were determined be free and unbound in solution, while ~21% are associated with lipid molecules in lipid-water adducts, and only <1% are involved in tryptophan:water:lipid 1:1:1 adduct formation. A perturbance of the water-lipid equilibrium in favour of interactions with tryptophan analogues might therefore expected to be small. With respect to the total tryptophan concentration, 13% of the amino acid analogue was calculated to be present as the 1:1:1 species, compared with 3% and 4% as the 1:1 and 2:1 Trp:lipid adducts respectively.

2.5 *Molecular Dynamics Simulations of Tryptophan - DMPC Interactions*

2.5.1 General Considerations

2.5.1.a Treatment of Charges

Two sets (**A** and **B**, see Table 2.15 and Table 2.16) for the representation of partial charges on the tryptophan analogues (§1.6.6) were evaluated in this study.

rHF - calculated using Jaguar[21] and implicit $CHCl_3$								
	6n	**6a**	**6b**	**6c**	**6d**	**6e**	**6f**	**6g**
C_β	-0.17	-0.29	-0.27	-0.23	-0.21	-0.14	-0.31	-0.24
N1	-0.58	-0.54	-0.57	-0.56	-0.59	-0.49	-0.52	-0.53
C2	-0.02	-0.04	-0.04	-0.04	-0.03	-0.13	-0.08	-0.02
C3	-0.33	-0.20	-0.21	-0.25	-0.24	-0.17	-0.18	-0.29
C3a	0.26	0.23	0.10	0.13	0.00	0.17	0.27	0.31
C4	-0.36	-0.54	-0.19	-0.20	0.02	-0.47	-0.66	-0.45
C5	-0.18	0.42	-0.10	-0.13	-0.40	0.25	0.56	0.10
C6	-0.12	-0.35	0.01	0.03	0.25	-0.28	-0.36	-0.25
C7	-0.39	-0.28	-0.47	-0.47	-0.62	-0.34	-0.32	-0.29
C7a	0.30	0.20	0.35	0.35	0.46	0.22	0.18	0.24
$H_\beta 1$	0.15	0.16	0.16	0.16	0.15	0.11	0.15	0.17
$H_\beta 2$	0.08	0.11	0.11	0.10	0.09	0.07	0.13	0.11
H1	0.43	0.42	0.43	0.43	0.43	0.40	0.42	0.43
H2	0.22	0.23	0.23	0.23	0.23	0.23	0.23	0.23
H4	0.22	0.26	0.19	0.20	0.15	0.24	0.26	0.26
R	0.16	-0.23	-0.12	-0.11	-0.03	-0.42	-0.47	0.80
H6	0.16	0.21	0.14	0.14	0.09	0.19	0.21	0.22
H7	0.20	0.22	0.24	0.24	0.26	0.20	0.22	0.22
X1						0.13	0.02	-0.50
X2						0.12	0.09	-0.50
X3						0.11	0.10	
X4							0.11	

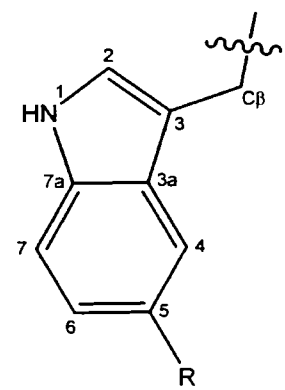


Table 2.15: rHF partial charges used (Set **A**). Charges are given as fractions of the charge on an electron (e); X1-X2 correspond to atoms that are attached onto the substituent atom R. They are X1, X2, X3 = H for **6e**; X1 = O and X2, X3, X4 = H for **6f**; X1, X2 = O for **6g**.

Additionally, two cut-off values (1.2 and 1.9 nm) were used for the treatment of the real-space part of the electrostatic interactions *via* PME. These do not have a direct effect on the treatment of long-range electrostatic interactions (as PME energies and forces are independent of the cutoff),[22] but the truncation for the van der Waals terms was increased accordingly as well which might influence the calculated parameters. The two sets and their influence on binding interactions between tryptophan analogues (**6a-g, 6n**) and a DMPC lipid

molecule were studied in simulations containing both species at equimolar concentrations (28 m*M*) (§5.2.1.a).

MP2 calculated using Gaussian, fitted with the CHelpG scheme								
	6n	**6a**	**6b**	**6c**	**6d**	**6e**	**6f**	**6g**
C_β	0.28	0.26	0.28	0.32	0.30	0.29	0.24	0.27
N1	-0.59	-0.55	-0.56	-0.57	-0.56	-0.59	-0.54	-0.55
C2	0.01	-0.11	-0.08	-0.08	-0.07	-0.08	-0.07	-0.08
C3	-0.41	-0.19	-0.24	-0.25	-0.23	-0.25	-0.20	-0.23
C3a	0.28	0.12	0.11	0.10	0.10	0.15	0.18	0.11
C4	-0.33	-0.39	-0.26	-0.22	-0.22	-0.38	-0.43	-0.21
C5	-0.06	0.40	0.13	0.05	0.02	0.12	0.49	-0.07
C6	-0.18	-0.27	-0.14	-0.11	-0.09	-0.24	-0.30	-0.09
C7	-0.23	-0.23	-0.29	-0.30	-0.29	-0.24	-0.23	-0.30
C7a	0.22	0.21	0.26	0.26	0.26	0.23	0.20	0.29
$H_\beta 1$	0.00	-0.04	-0.04	-0.05	-0.06	-0.05	-0.04	-0.03
$H_\beta 2$	-0.07	-0.02	-0.02	-0.03	-0.05	-0.03	-0.03	-0.01
H1	0.40	0.40	0.39	0.39	0.39	0.40	0.40	0.40
H2	0.17	0.19	0.18	0.18	0.17	0.17	0.18	0.19
H4	0.21	0.24	0.20	0.17	0.17	0.20	0.22	0.21
R	0.10	-0.35	-0.24	-0.15	-0.10	-0.13	-0.58	0.80
H6	0.14	0.18	0.15	0.12	0.11	0.14	0.17	0.16
H7	0.14	0.16	0.16	0.16	0.16	0.14	0.16	0.17
X1						0.04	0.28	-0.50
X2						0.04	0.04	-0.51
X3						0.04	-0.01	
X4							0.00	

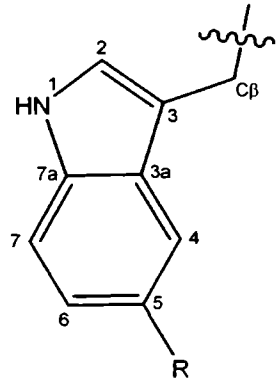


Table 2.16: MP2 partial charges (Set **B**). Charges are given as fraction of the charge on an electron (e); X1-X2 correspond to atoms that are attached onto the substituent atom R. They are X1, X2, X3 = H for **6e**; X1 = O and X2, X3, X4 = H for **6f**; X1, X2 = O for **6g**.

2.5.1.b Dipolar Interactions

Tryptophan analogues carry a considerable dipole moment which was approximated by calculating the permanent dipole moment of 5'-monosubstituted 3-ethyl indole analogues (Figure 2.24) using the program library MOPAC[23] and different Hamiltonians (AM1, PM3, and PM5). These compounds were used as mimics of the tryptophan derivatives, to simplify the computation. Since the backbone structure and polarity will be similar for the analogues this simplification was justified. The permanent dipole moment was calculated to range from 1.6 *D* (R = Me) to 7.5 *D* ($R = NO_2$) (Table 2.17). AM1 dipole moments have an average difference between experimental and observed values of ±0.35 *D*.[24]

3-ethyl 5'-R-indole	Dipole Moment / *D*		
	AM1	PM3	PM5
R = H	1.784	1.532	1.191
R = F	3.517	3.625	3.736
R = Cl	3.250	3.021	2.819
R = Br	3.401	2.252	2.740
R = I	3.385	2.573	2.192
R = Me	1.607	1.350	0.997
R = MeO	2.548	2.066	1.877
R = NO_2	7.548	7.676	8.037

Table 2.17: Permanent dipole moments of 3-ethyl indole derivatives as calculated using MOPAC using AM1, PM3 and PM5 parameters respectively.

The directions of the dipole are roughly parallel to the bond between the nitrogen and the proton in the NH group of indole (Figure 2.24).

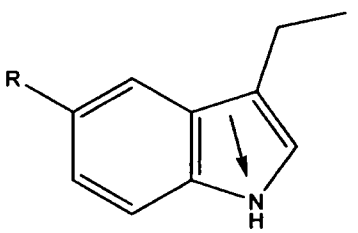


Figure 2.24: Figure showing the permanent dipole of 5'-monosubstituted 3-ethyl indole derivatives.

When plotted against calculated binding constants, no linear trend was found with respect to the dipole moment the molecules carried. (See §7.4, Figure 7.1)

2.5.2 Simulation of Amino Acid – Lipid Interactions at a Ratio of 1:1

Simulations of one tryptophan derivative interacting with one DMPC molecule were performed under the conditions described in §5.2.1. The simulation box contained one molecule of DMPC and one tryptophan analogue (+ 450 chloroform molecules) corresponding to an equimolar concentration of 28 m*M*. This allowed the direct observation of 1:1 adduct formation over the course of a series of 100 ns simulations.

2.5.2.a *Analysis of Binding Events*

Regardless of the initial starting configuration and relative position of the two molecules with respect to each other, all tryptophan analogues (Trp) analysed were found to interact with a DMPC lipid molecule favourably, forming an Trp:DMPC adduct (Figure 2.25). Adducts were defined by the distance of the centre of mass of the tryptophan analogue from the centre of mass of the lipid molecule, r_{AB}:

$$1) \quad r_{AB} \leq 0.5 \text{ nm},$$

$$2) \quad 0.5 \leq r_{AB} \leq 1.0 \text{ nm},$$

$$3) \quad r_{AB} \geq 1.0 \text{ nm}.$$

Here, case 1) corresponds to a bound state, case 3) corresponds to an unbound state, and in case 2) the molecule is considered to be neither bound nor unbound.

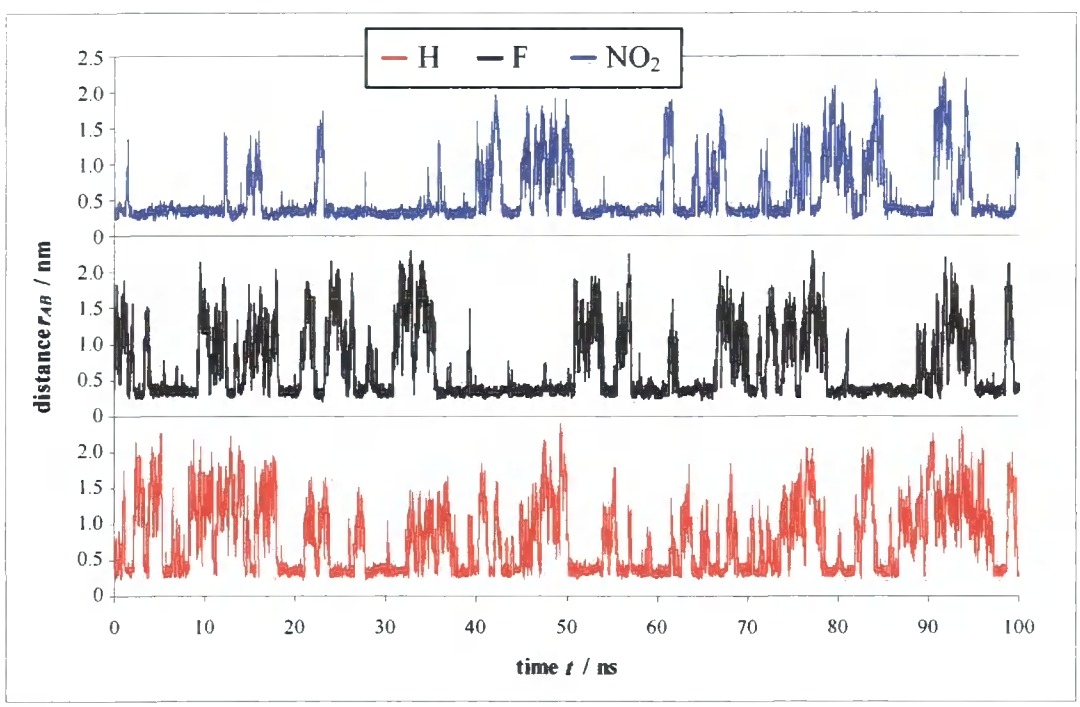


Figure 2.25: Distance profiles showing the distance between the centre of mass of an amino acid derivative (**6n** R = H, **6a** R = F, & **6g** R = NO_2) and the centre of mass of the DMPC molecule as a function of time. The compounds are typical examples for weak (**6n**), intermediate (**6a**) and strong (**6g**) associations.

Figure 2.25 shows the time evolution of r_{AB} for three selected tryptophan derivatives, showing an example of strong binding (R = NO_2, **6g**), intermediate (R = F, **6a**), and weak association behaviour (R = H, **6n**). The graphs show distinct periods of time during which the two molecules remain at approximately the same distance with the durations varying between 0.5 and 12.5 ns for the different amino acids derivatives.

The distance profile allowed the estimation of the minimum distance of r_{AB} by determining the average distance r_{AB} during the times when $r_{AB} \leq 0.5$ nm. It was found, that the interaction distance between the amino acid derivative and the lipid molecule was approximately the same ($r_{AB} = 0.35 \pm 0.05$ nm) for all compounds analysed.

2.5.2.b Radial Distribution Functions for selected atoms in Amino Acid – Lipid 1:1 Adducts

Binding events could be visualised by analyzing the radial distribution functions, g(r), for various atom pairs. The radial distribution function describes the probability of finding atom i within a distance r of atom j relative to an ideal gas. Exemplary, Figure 2.26 shows a 3-dimensional expansion of the radial distribution function for the distance between the centre of mass of N^{α}-acetyl-5-NO_2-*L*-tryptophan ethyl amide and the centre of mass of the DMPC molecules, with time being the third dimension.

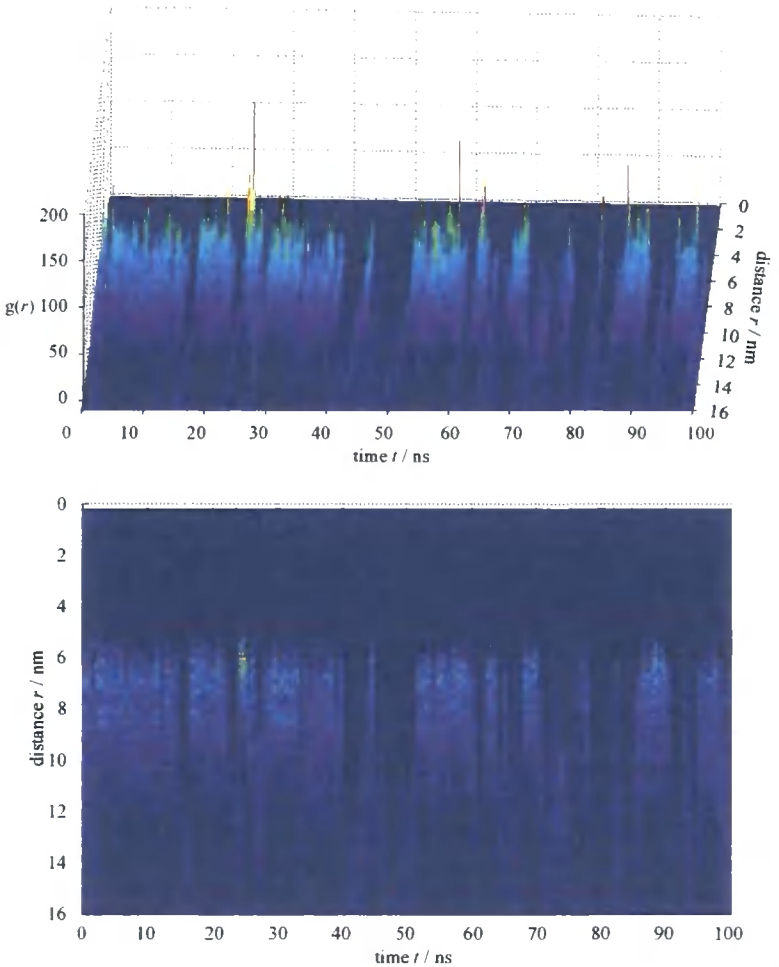


Figure 2.26: Time evolution of g(r). **A** shows the radial distribution function for the N^{α}-acetyl-5-nitro-*L*-tryptophan ethyl amide (**6g**) indole centre of mass with respect to the choline nitrogen of DMPC over a 100 ns molecular dynamics simulation. **B**) shows a contour map of **A**).

The time-evolved radial distribution function helps to identify times when two molecules (or a pair of atoms) are far apart from each other (*i.e.* the probability of finding an atom i within a radius r of atom j is low, corresponding to a valley region in Figure 2.26). Additionally, the analysis of g(r) provides information about the preferred distance between two atoms during times when they are close to each other. A time-averaged radial distribution function helps to identify the distances at which two atoms are most likely found with respect to each other (Figure 2.27).

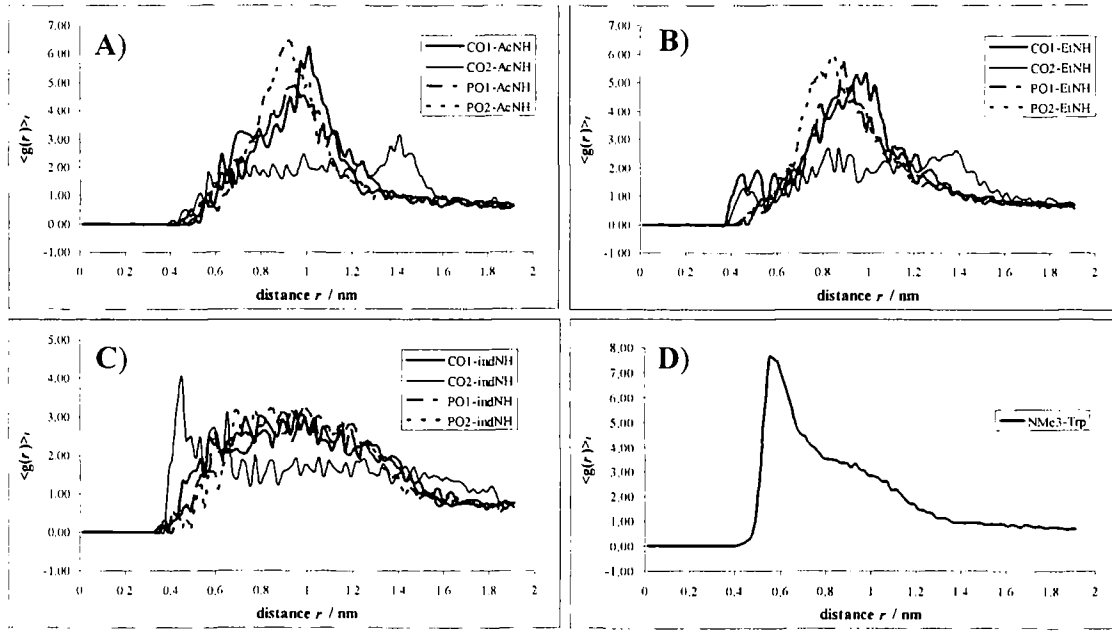


Figure 2.27: Time averaged radial distribution functions ($\langle g(r)\rangle_t$) between hydrogen bond acceptors of the lipid molecule and hydrogen bond donors from N^{α}-acetyl-*L*-tryptophan ethyl amide (**6n**). **A** shows the $\langle g(r)\rangle_t$ of the lipid carbonyl oxygens (denoted as CO1 and CO2 respectively), non-bridging lipid phosphate oxygens (PO1 and PO2 respectively) and the acetyl amide hydrogen (AcNH) of **6n**. **B** and **C** show the $\langle g(r)\rangle_t$ of the lipid atoms as defined in **A** and the ethyl amide proton (EtNH) and the indole amine hydrogen (indNH) respectively. The radial distribution function between the centre of mass of the tryptophan indole ring and the choline nitrogen is shown in **D**.

The time averaged radial distribution function shows, that the minimum distance between all the atom pairs analysed is ≥ 0.35 nm. The $g(r)$ between the tryptophan side chain and the choline head group exhibits a distinct maximum ($r = 0.57$ nm) and a broader shoulder ($r \sim 1.00$ nm) (Figure 2.27, **D**). This is indicative of a specific interaction at 0.57 nm which might include cation-π interactions.

For the remainder of the data depictured, the radial distributions show a broad signal distribution with a roughly Gaussian distribution around $r = 0.9$ nm for interactions of the two amide hydrogens with hydrogen bond acceptors of the DMPC lipid molecule (Figure 2.27, **A-C**). The undefined nature and the lack of a clear peak are suggestive that no hydrogen bonding interactions were established, as hydrogen bonds normally span only a very short distance ($r \leq 0.35$ nm). However, there might be short-lived hydrogen bond formations that are obscured by the averaging of the data. The underlying peak at around 1.0 nm might be an indicator of specific interactions at this distance. Adducts were therefore

analysed *via* cluster analysis and the hydrogen bonding distance evaluated through the software (§5.2.1.e for experimental set-up and §2.5.2.c for results).

A similar picture can be seen when looking at the radial distribution functions of the other tryptophan derivatives with respect to the lipid (Figure 2.28).

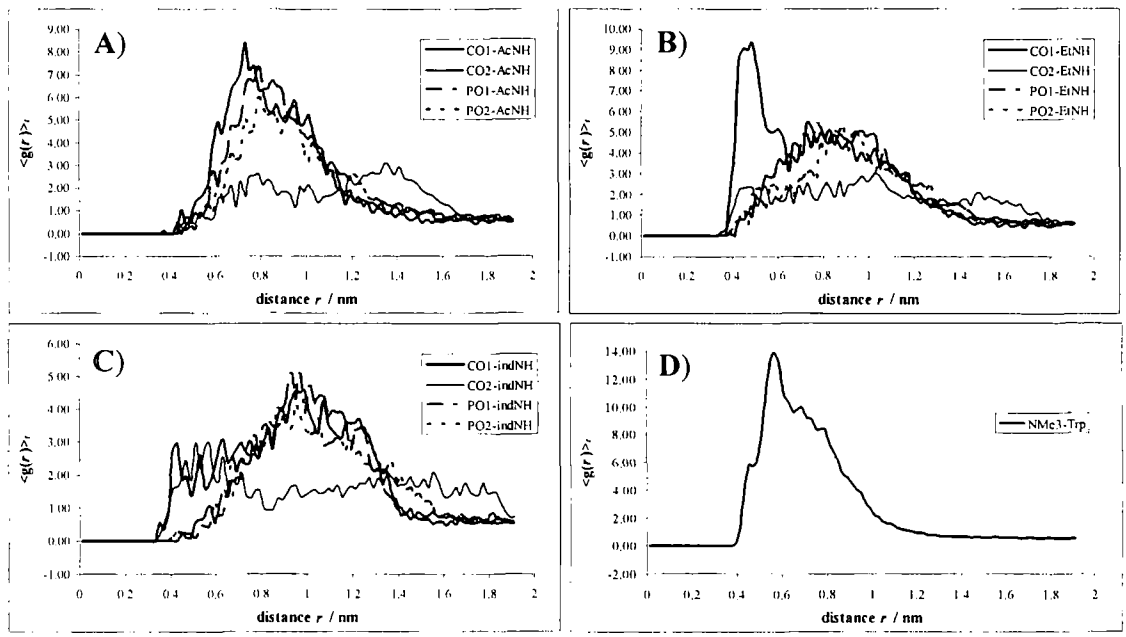


Figure 2.28: $\langle g(r) \rangle_t$ functions between hydrogen bond acceptors of the lipid molecule and hydrogen bond donors of N^{α}-acetyl-5-F-*L*-tryptophan ethyl amide (**6a**). Same descriptions are used as in Figure 2.27.

Here, the same observations about the minimum interaction distance and the interacting character of the indole with the choline head group can be made (Figure 2.28, **D**). A striking difference can be seen in Figure 2.28, **B**. A peak in the g(*r*) at $r = 0.5$ nm (roughly the same distance as for the indole – lipid nitrogen peak) is indicative of a specific interaction. A high probability to find the ethyl amine hydrogen close to one of the lipid carbonyl oxygens of the lipid molecule is suggestive of the establishment of a reasonably strong, long-lasting interaction, which might include hydrogen bonding interactions at short distances. The same results can be seen in more detail for the interactions of N^{α}-acetyl-5-nitro-*L*-tryptophan ethyl amide with DMPC (Figure 2.29).

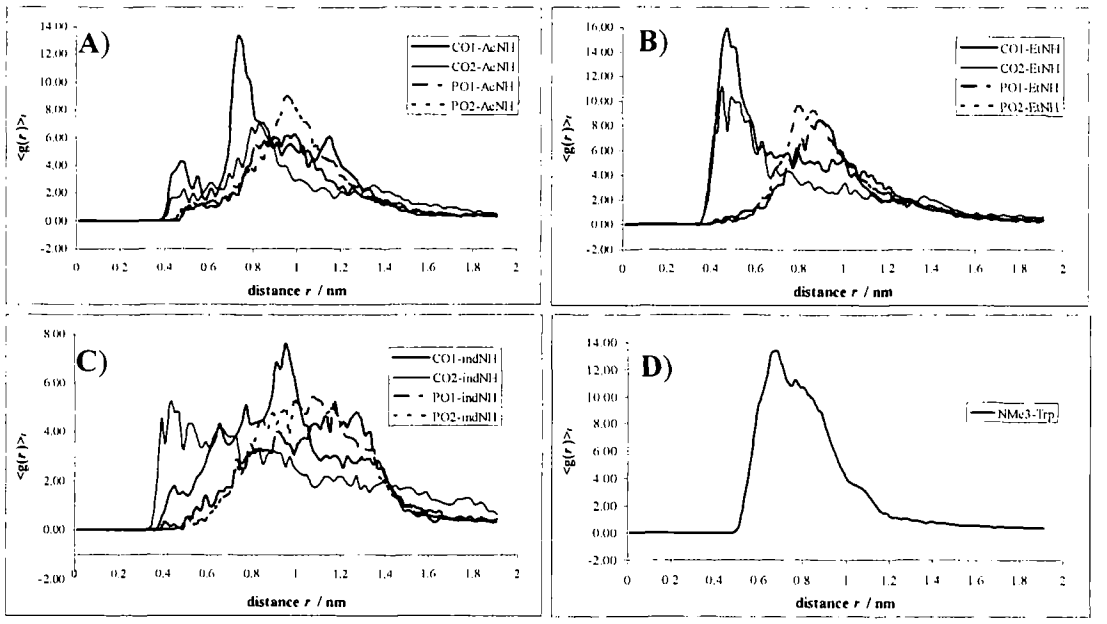


Figure 2.29: $\langle g(r) \rangle_t$ functions between hydrogen bond acceptors of the lipid molecule and hydrogen bond donors from N^{α}-acteyl-5-nitro-*L*-tryptophan ethyl amide (**6g**). Same descriptions are used as in Figure 2.27.

Figure 2.29 exhibits a more refined structure for the radial distribution functions of the amide protons. There are clear maxima in the g(r) (at $r = 0.50 \pm 0.1$ nm) for the distribution of ethyl amide hydrogen and the lipid carbonyl oxygens, and well-defined, broader peaks at r ~0.9 nm (Figure 2.29, **B**). Additionally, the radial distribution function of the acetyl amide proton with respect to a lipid carbonyl oxygen exhibits a peak r ~0.75 nm (Figure 2.29, **A**). No clearly defined structures can be seen for interactions between the indole amine hydrogen and the lipid hydrogen bond acceptors (Figure 2.29, **C**), while the indole-lipid head group (*i.e.* the choline nitrogen) is well-defined.

The time averaged radial distribution function between the hydrogen bond acceptors of the lipid molecule and hydrogen bond donors from the remaining tryptophan derivatives exhibited a similar behaviour to the one seen in Figure 2.28. Preferred distances of various intermolecular distances vary between the different tryptophan analogues and were observed for $0.35 \leq r \leq 1.4$ nm. No clear hydrogen bonding interactions between the tryptophan derivative and the lipid in the 1:1 Trp:lipid complexes were observed, as all time-averaged preferred interaction distances between hydrogen bond donors and acceptors in the tryptophan derivative and lipid respectively were found for distance $r > 0.35$ nm. The centre of mass of the indole side chain of all tryptophan derivatives was found to be at a preferred distance of r ~0.75 nm to the lipid choline nitrogen. This might be indicative of a long-range cation-π interaction.

The specific orientations of the various groups and their interactions could be visualised by using a cluster analysis of the bound complexes as detailed in §2.5.2.c.

2.5.2.c Cluster Analysis of 1:1 Adducts of Tryptophan Analogues with DMPC

The time averaged radial distribution functions of atom-atom distances of the tryptophan derivative and the lipid were indicating preferred interaction distances, which could be associated with preferred orientations of the molecules during the adduct formation. To evaluate this further, a cluster analysis of the bound species was performed, using the procedure described in §5.2.1.e. Figure 2.30 shows representative examples of the most populated (up to 19% for **6a**) preferred binding conformations.

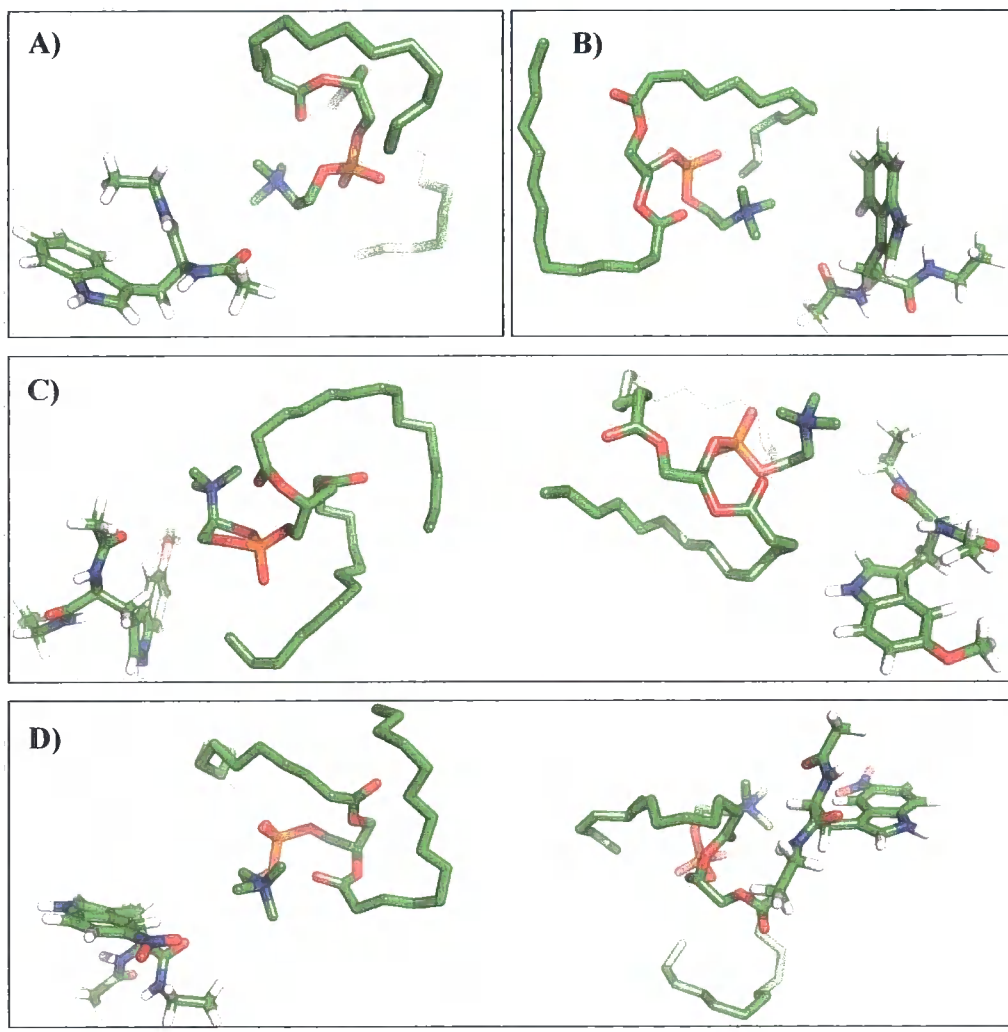


Figure 2.30: Examples of the most populated binding conformations of **6n** (frame **A**), **6a** (frame **B**), **6f** (frame **C**), and **6g** (frame **D**).

In some cases (*e.g.* **6f, 6g**), the shown confirmations are the two most populated states found for the compound (Table 2.18). The remaining binding conformations could not be clustered to a common structure or were populated in less than 2% of the total number of binding conformation.

Cluster	6n	6a	6b	6c	6d	6e	6f	6g
A	4%	19%	2%	11%	5%	3%	2%	3%
B	9%		15%	2%	3%	5%	5%	8%
C	4%			2%	3%	2%	2%	7%
D							5%	2%
E							2%	2%
total	17%	19%	17%	15%	12%	10%	16%	25%

Table 2.18: Most populated Trp:lipid 1:1 clusters as obtained from a 100 ns MD simulation. Cluster A-E do not refer to a common structure of the bound conformations of the 1:1 Trp:lipid complexes, but labels individual clusters according to appearance in time in the individual simulations.

The $\langle g(r) \rangle_t$ distributions of the choline head group nitrogen with respect to the indole ring had already indicated a close relationship between the two, which can now be visualised in the binding conformations, which show the amine nitrogen oriented towards the tryptophan compounds in most cases. No preferred orientations of the ethyl amide proton with respect to the lipid carbonyl oxygen atoms was observed, suggesting that the preferred interaction distance, as seen in the $\langle g(r) \rangle_t$ distributions of these atom pairs, might be an artefact of the preferred interaction distance between the indole ring and the choline head group.

2.5.2.d Calculation of Binding Constants and Associated Free Energies for the 1:1 Adduct Formation of Tryptophan Analogues with DMPC - Theory

The association of molecule *A* and molecule *B* to form the complex *AB* in the solvent (*s*) can be described as

$$A(s) + B(s) \underset{k_{off}}{\overset{k_{on}}{\rightleftharpoons}} AB(s).$$

Scheme 2.11: General scheme for a bimolecular association-dissociation process in a solvent (*s*).

Here, k_{on} and k_{off} refer to the association and dissociation rate constant, respectively. The standard free energy of the association process $\Delta G^0_{AB(s)}$ is given by[25]

$$\Delta G^0_{AB(s)} \equiv \mu^0_{AB(s)} - \mu^0_{A(s)} - \mu^0_{B(s)} = -RT \ln\left(\frac{\gamma_{AB}}{\gamma_A \gamma_B}\frac{C^0 C^{eq}_{AB}}{C^{eq}_A C^{eq}_B}\right) \equiv -RT \ln K^{eq}_{AB}, \quad (2.26)$$

where μ^0_i is the standard chemical potential, C_i and γ_i are, respectively, the concentration and the activity coefficient of species i. A deviation of the activity coefficient γ_i from a value of 1 is in essence due to the interaction with other solutes in solution, choosing γ_i as one is usually a good approximation when the interaction is weak. C^0 refers to the standard concentration, with 1 mol/l standard concentration as 1 molecule/1.66 nm^3. K^{eq}_{AB} refers to the association constant of AB at equilibrium. R is the gas constant and T the absolute temperature. With Equation 2.25, the association and dissociation rate constants can be related to the equilibrium association constant:

$$K^{eq}_{AB} = C^0 \frac{k_{on}}{k_{off}}. \quad (2.27)$$

The unit for C is M, for k_{on} is M^{-1} s^{-1}, and for k_{off} is s^{-1}.

The criteria for the formation of an AB complex is that the distance between the centre of mass of A and B be less than a critical value, *i.e.* $r_{AB} < R_c$. Each plot in the minimum distance profile between tryptophan analogues and DMPC (see Figure 2.25, page 95) shows many distinct association and dissociation events. It appears that the simulations have effectively explored the configurational space open to the two molecules, providing confidence in the statistical significance of the results.

Survival Function

Following a previously reported methodology,[26] the association constant K^{eq}_{AB} can be expressed by the survival function $S_{AB}(t)$ of the AB complex. In the general case, assuming there are initially $n_{AB}(0)$ complexes AB and no separate molecules A and B present in a the solvent s with volume V, the survival function is defined as

$$S_{AB}(t) = \frac{n_{AB}(t)}{n_{AB}(0)}, \quad (2.28)$$

where $n_{AB}(t)$ is the number of complexes left after a time t. The numbers of particles A and B in the system were equal

$$n_A(t) = n_B(t) = n_{AB}(0) - n_A(t). \tag{2.29}$$

With

$$\frac{dC_{AB}}{dt} = k_{on} C_A C_B - k_{off} C_{AB}, \tag{2.30}$$

and equation 2.28 it follows that

$$\frac{dS_{AB}(t)}{dt} = k_{on}(1 - S_{AB}(t))^2 \frac{n_{AB}(0)}{V} - k_{off} S_{AB}(t). \tag{2.31}$$

Here, V is in the units of 1.66 nm^3 and k_{on} is in 1.66 nm^3/ (molecule.s), *i.e.* M^{-1} s^{-1}. At equilibrium $dS(t)/dt = 0$, the association constant becomes

$$K_{AB}^{eq} = C^0 \frac{k_{on}}{k_{off}} = C^0 \frac{S_{AB}^{eq}}{(1 - S_{AB}^{eq})^2} \frac{V}{n_{AB}(0)}. \tag{2.32}$$

S_{AB}^{eq} is the survival function of the complex AB at equilibrium.

Probability of AB Complex Formation

S_{AB}^{eq} can also be interpreted as the average probability for either a molecule A or a molecule B to form the complex AB. A and B can be considered being in a complex when there centres of mass are sufficiently close together (*i.e.* $r_{AB} < R_c$). Assuming both particles are spherical, and one molecule A is positioned in the centre, the probability of finding a molecule B at distance r from A is given by the radial distribution function

$$G(r) = 4\pi r^2 \rho g(r) \tag{2.33}$$

with

$$\int_0^\infty G(r) dr = n_{AB}(0). \tag{2.34}$$

Here, ρ is the bulk density of B and g(r) is the pair distribution function, which can be computed. The probability S_{AB}^{eq} for a molecule A to form the complex AB is the probability for the molecule A to have a nearest neighbour B inside a sphere with R_c:

$$S_{AB}^{eq} = \int_0^{R_c} G(r)dr = 4\pi \frac{n_{AB}(0)}{V} \int_0^{R_c} g(r)r^2 dr. \qquad (2.35)$$

Mean Survival Time – a Measure of the Dissociation Constant k_{off}

Analysing the distance profile, the mean survival time T_{AB}^{mean} of the complex AB can be calculated from the times when the molecules can be considered being associated, *i.e.* $r_{AB} < R_c$. The dissociation rate k_{off} can then be expressed as

$$k_{off} = \frac{1}{T_{AB}^{mean}}. \qquad (2.36)$$

Calculating the Equilibrium Association Constant by the Ratio of Bound and Unbound States

The association constant can also be approximated as the ratio of the number of states in which the tryptophan analogue and the lipid can be considered as being bound (WL) and the number of states when they can be considered unbound (W and L respectively)[27]

$$K_{AB}^{eq} = \frac{[WL]}{[W][L]} = \frac{\gamma}{[L]}, \qquad (2.41)$$

where γ refers to the ratio of the number of states when tryptophan derivatives and DMPC are bound and the number of states when the tryptophan derivative is unbound; the bound and unbound states follow the same criteria as defined above (§2.5.2.a), *i.e.* $r_{AB} <$ 0.5 nm for bound and $r_{AB} >$ 1.0 nm for unbound states. Table 2.19 shows computationally determined binding constants and associated free energies for the 1:1 Trp:lipid adduct formation in chloroform.

ID	K_{AB}^{eq}			ΔG_{AB}^{eq}		
	A	B	C	A	B	C
	M^{-1}	M^{-1}	M^{-1}	kJ mol^{-1}	kJ mol^{-1}	kJ mol^{-1}
6n	48.21 ± 4.29	40.71 ± 12.50	46.43 ± 8.38	-9.54 ± 0.85	-9.13 ± 2.80	-9.45 ± 1.70
6a	52.14 ± 7.86	67.86 ± 17.14	75.66 ± 2.65	-9.74 ± 1.47	-10.38 ± 2.62	-10.65 ± 0.37
6b	64.29 ± 7.50	71.07 ± 20.71	65.73 ± 11.02	-10.25 ± 1.20	-10.5 ± 3.06	-10.31 ± 1.73
6c	105.36 ± 5.36	55.36 ± 13.57	48.79 ± 5.53	-11.47 ± 0.58	-9.88 ± 2.42	-9.57 ± 1.09
6d	206.07 ± 46.79	47.50 ± 15.36	38.15 ± 4.76	-13.12 ± 2.98	-9.51 ± 3.07	-8.97 ± 1.12
6e	78.21 ± 5.36	41.07 ± 8.93	42.89 ± 7.55	-10.73 ± 0.74	-9.15 ± 1.99	-9.25 ± 1.63
6f	66.07 ± 7.14	77.14 ± 16.43	55.18 ± 12.98	-10.32 ± 1.12	-10.7 ± 2.28	-9.87 ± 2.32
6g	57.14 ± 11.79	54.29 ± 8.57	164.58 ± 39.78	-9.96 ± 2.05	-9.83 ± 1.55	-12.57 ± 3.04

Table 2.19: Comparison of calculated Trp:lipid 1:1 binding constants obtained by different treatments of electrostatic interactions. Column **A** describes the binding constants and associated free energies obtained when using Hartree-Fock derived partial charges for the tryptophan compound. Columns **B** and **C** used partial charges derived from MP2 calculations using 1.2 and 1.9 nm as cut-off radius for the truncation of electrostatic interactions respectively.

HF was computed from a set-up where the partial charges of the tryptophan derivative were derived from restricted Hartree-Fock *ab initio* calculations (Table 2.15); MP2a was derived by using partial charge parameters derived from calculations using Møller-Plesset perturbation theory MP2 (Table 2.16); MP2b used the same partial charges as the set-up for the system for MP2a but used a larger cut-off radius (1.9 nm) in the treatment of long-range interactions.

The error estimates reflect the standard deviation of the analysed parameter in its values obtained from the 6 starting configurations. Repeat runs using the same starting configuration typically gave smaller errors (±5%).

Calculating the Equilibrium Association Constant using the Survival Function of the 1:1 Complex

Using Equation 2.32, the association constant can be calculated when the survival function of the complex at equilibrium, S_{AB}^{eq}, the number of initial complexes $n_{AB}(0)$, and the volume V of the solvent are known. In the specific case of this study, $n_{AB}(0) = 1$, as there can only be a 1:1 complex formation resulting from the simulation set-up (see §5.2.1.a); the volume V can be calculated from the dimensions of the simulation box, and the survival function at equilibrium, S_{AB}^{eq}, can be computed using the radial distribution function of the distance

between the centres of mass of the tryptophan analogue with respect to the centre of mass of the DMPC molecule in equation 2.35. The dissociation constant k_{off} was determined using Equation 2.36. With the knowledge of k_{off} and S_{AB}^{eq}, the association constant k_{on} can be calculated as given in equation 2.32. The standard concentration C^0 was expressed in molecules/ 1.66 nm^3. The calculated parameters are summarized in Table 2.20.

ID	V	T_{mean}	k_{off}	S_{AB}^{eq}	k_{on}
	1.66 nm^3	ps	ns^{-1}	x 10^3	1.66 nm^3 / (molecule.ns)
6n	36.04±0.19	411±68	2.58±0.34	10.89±1.53	1.04±0.20
6a	36.12±0.15	386±40	2.65±0.24	16.69±3.69	1.64±0.10
6b	36.13±0.19	347±16	2.92±0.14	15.89±2.44	1.75±0.16
6c	35.96±0.30	417±20	2.42±0.11	15.22±4.08	1.37±0.32
6d	35.86±0.26	449±55	2.34±0.30	12.65±1.59	1.11±0.25
6e	35.96±0.11	460±38	2.23±0.18	11.70±2.55	0.99±0.15
6f	36.14±0.20	442±30	2.38±0.13	24.19±4.24	2.22±0.27
6g	36.14±0.23	494±45	2.63±0.52	24.20±4.29	2.44±0.91

Table 2.20: Calculated dissociation and association constants for the formation of a tryptophan analogue-lipid complex.

The resulting association constants could then be converted into association free energies for the 1:1 adduct formation of a tryptophan derivative and DMPC (ΔG_I^{MP2b2}) using Equation 2.26.

ID	R	σ_P	K_I	K_I^{MP2b2}	ΔG_I	ΔG_I^{MP2b2}
			M^{-1}	M^{-1}	kJ mol^{-1}	kJ mol^{-1}
6g	NO_2	0.78	88.22 ± 1.39	6.28 ± 0.75	-11.03 ± 0.52	-6.98 ± 0.84
6d	I	0.23	34.56 ± 2.21	10.09 ± 0.82	-8.72 ± 0.12	-4.91 ± 0.39
6c	Br	0.23	26.28 ± 2.02	10.54 ± 0.96	-8.05 ± 0.62	-5.14 ± 0.67
6b	Cl	0.18	24.01 ± 0.32	8.08 ± 1.05	-7.83 ± 0.50	-5.80 ± 0.53
6a	F	0.06	17.93 ± 0.85	7.37 ± 0.58	-7.11 ± 0.13	-5.69 ± 0.46
6n	H	0.00	11.50 ± 0.21	8.17 ± 0.86	-6.01 ± 0.35	-4.52 ± 0.54
6e	Me	-0.17	35.59 ± 2.06	17.03 ± 2.06	-8.79 ± 0.18	-5.17 ± 0.54
6f	MeO	-0.27	61.13 ± 1.28	17.03 ± 2.06	-10.13 ± 0.16	-6.98 ± 0.84

Table 2.21: Computed and experimentally determined binding constants for 1:1 adduct formation of a tryptophan analogue with DMPC.

2.5.2.e *Comparison of Computed Binding Constants with the Experimentally Determined Values*

Free Energies of Association Determined by the Ratio of Bound and Unbound States

Experimentally determined free energies of association for the 1:1 adduct formation of tryptophan derivatives with DMPC, ΔG_1, (§2.3.3.c) and computed values were found to be in very good agreement.

The computed values using the partial charge set A (ΔG_1^{rHF}) generally overestimated the experimentally determined free energy values by 1.87 ± 0.85 kJ mol^{-1} (Figure 2.31, **A**). The overall pattern of the free energies observed in the NMR titration experiment was conserved. However, association free energies for the most electron rich and most electron poor compounds (**6f**, R = MeO and **6g**, R = NO_2) were underestimated indicating that the molecular dynamics simulations and the calculation of free energies of association are sensitive to the treatment and size of the partial charges on the indole ring.

This importance was investigated using a partial charge set that had been derived from MP2 calculations (Set **B**). Two sets of runs (Set **Ba** and **Bb**) were performed using this partial charge model. One employed a truncation of long-range interactions at 1.2 nm the other used a truncation at 1.9 nm. The derived free energies of association (ΔG_1^{MP2a} and ΔG_1^{MP2b} respectively) were almost identical except for the electron poor 5-nitro tryptophan derivative (**6g**), where the variation was found to be outside experimental error (Table 2.19, **B** and **C**; Figure 2.31, **B**). Both treatments show an excellent agreement for the free energies of association for electron rich species (**6e**, R = Me and **6f**, R = MeO) and for compounds carrying a heavy halogen atom (**6c**, R = Br and **6d**, R = I) with a deviation of 0.73 ± 0.51 kJ mol^{-1}. The free energies of association of the unsubstituted tryptophan derivative (**6n**), the 5-fluoro (**6a**) and 5-chloro (**6b**) substituted analogues were slightly underestimated by 1.92 ± 0.59 kJ mol^{-1} with respect to the experimental values. The two treatments varied the most in the description of the free energy of association of the 5-nitro amino acid analogue (**6g**). Here, the model employing a shorter truncation radius underestimates, while the model using 1.9 nm as truncation radius overestimates with respect to the values obtained experimentally. However, both lie within experimental error of the latter.

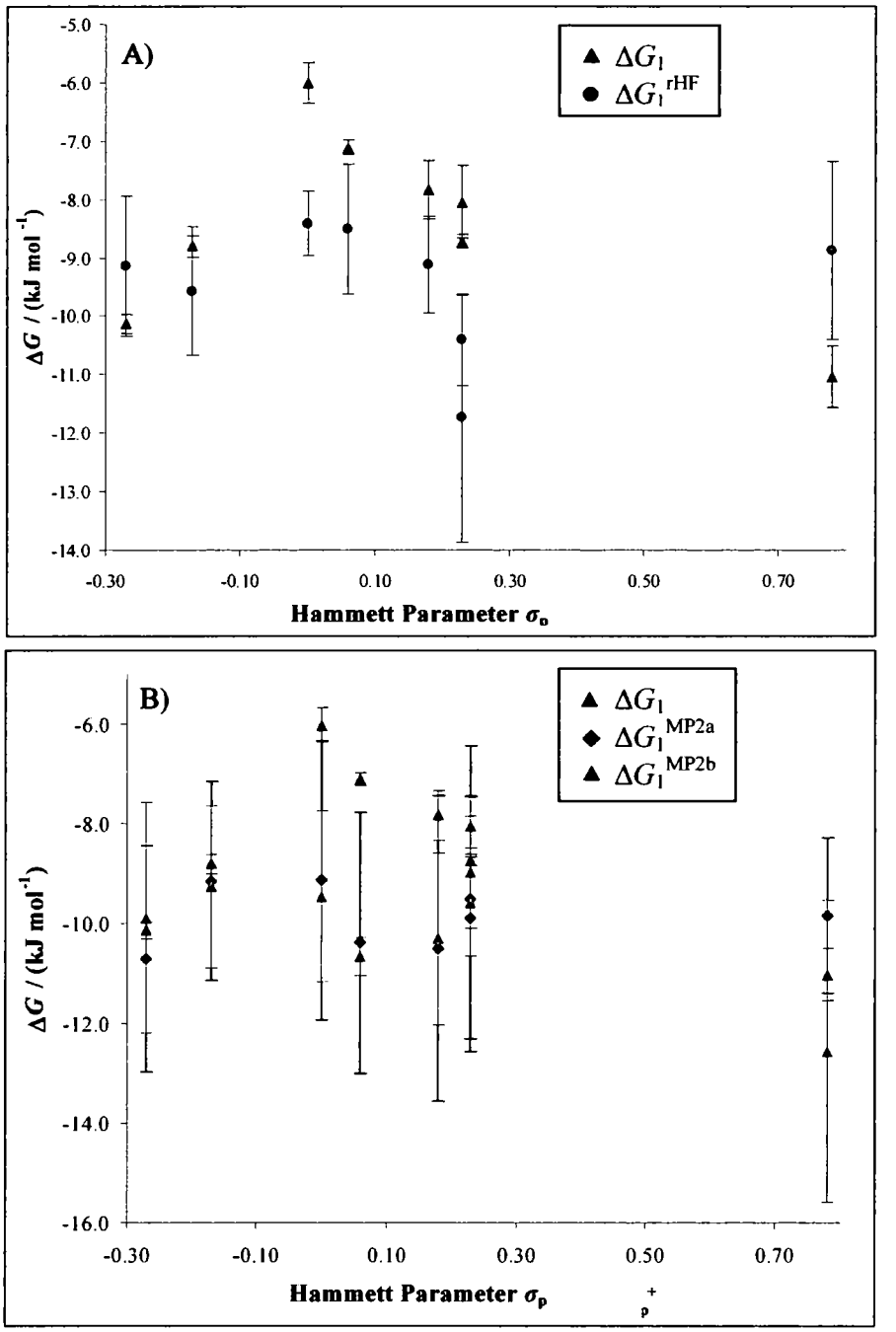


Figure 2.31: Comparison of calculated and experimentally obtained free energies of association for the 1:1 adduct formation of tryptophan derivatives (**6a-g, 6n**) with DMPC in chloroform.

The data for the halogen-substituted tryptophan compounds deviate more from the pattern found in experiment. Within the error of the simulations and the approximations in the force field of the simulations (e.g. lack of polarizability), this might be an exaggerated deviation. Such polarization effects could be incorporated in the description of the system using massless, virtual sites that carry some partial charge of the adjacent polarisable atom to

which they are bound by a harmonic potential. It should be noted however, that the introduction of such virtual sites would increase the computational costs and make the long simulations that were performed in this study prohibitively expensive.

Free Energies of Association Determined by the Survival Function of the 1:1 Complex

As the partial charge Set **B** and simulations using a truncation at 1.9 nm were found to approximate experimentally obtained free energies of association for the 1:1 adduct formation of tryptophan derivatives (**6a-g, 6n**) and DMPC extremely well, the data from these simulations were used to derive association free energies using the survival function of the complex (Equation 2.32, Table 2.21). The computed free energies (ΔG_1^{MP2b2} and ΔG_1 respectively) followed the experimentally determined n-shaped pattern when plotted against the Hammett parameter (Figure 2.32). The calculation generally under-estimated the free energy of association by 2.81 ± 0.98 kJ mol^{-1}. The observed pattern for the calculated data resembles the data for the free energies of associations obtained by the ratio of bound and unbound states (ΔG_1^{MP2a} and ΔG_1^{MP2b} respectively).

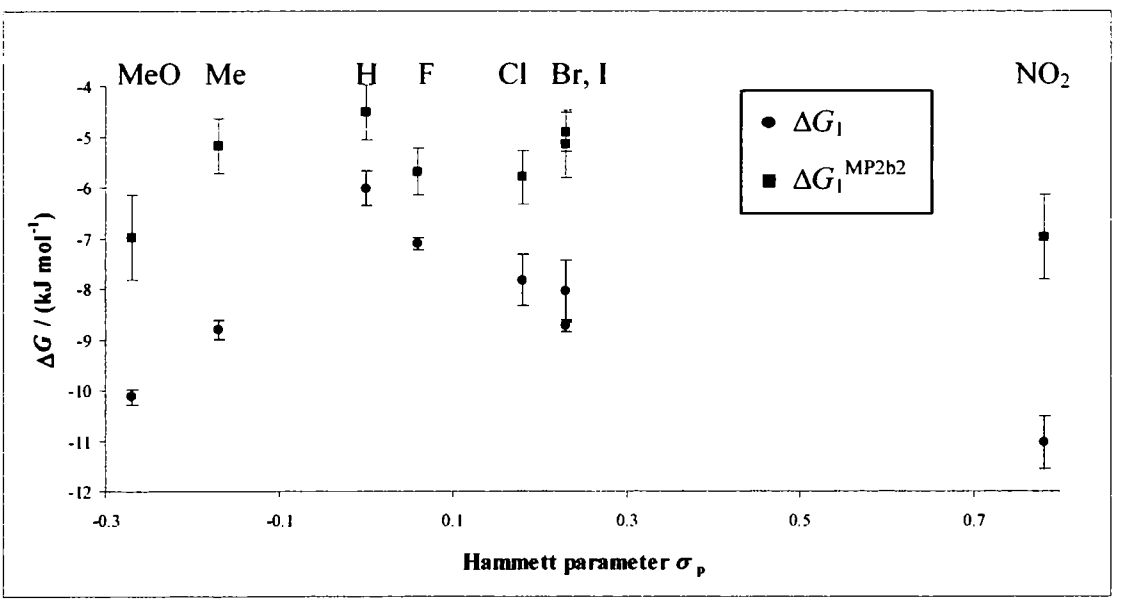


Figure 2.32: Calculated (ΔG_1^{MP2b2}) and experimentally (ΔG_1) determined values of the association free energies for the 1:1 adduct formation of tryptophan derivatives (**6a-g & 6n**) and a DMPC lipid in chloroform at 25 °C.

2.5.3 Simulation of the Influence of Water on the 1:1 Trp:Lipid Adduct Formation

2.5.3.a Evaluation of Minimum Distances between the Molecules

The experimental analysis of the binding of tryptophan derivatives to DMPC in chloroform showed a significant influence of the water present in the system (§2.3.2.a). Therefore, a simulation of an amino acid analogue (**6a**), DMPC and a water molecule was undertaken to investigate the specific molecular interaction of water with the system. A detailed description of the setup can be found in §5.2.1.b. Briefly, a single point charge (SPC) representation of a water molecule was added near the polar head group of the lipid molecule. This resulted in an equimolar concentration of 28 μM for DMPC, tryptophan derivative, and water in chloroform (423 molecules). A molecular dynamics simulation was performed over 100 ns and the relative distances between the centres of mass between the molecules evaluated (Figure 2.33).

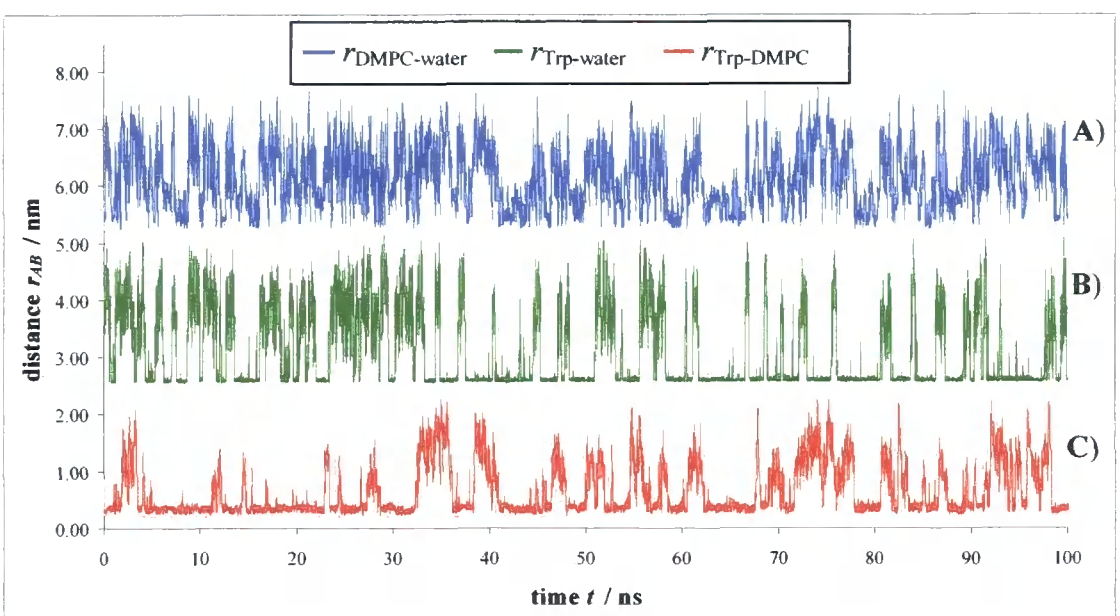


Figure 2.33: Distances between DMPC-water, Trp-water and Trp-lipid. **A** displays the intermolecular distances between N^{α}-acetyl-5-fluoro-*L*-tryptophan ethyl amide (**6a**) and the DMPC molecule; the distances between the water molecule and N^{α}-acetyl-5-fluoro-*L*-tryptophan ethyl amide (**6a**) molecule and the DMPC molecule are shown in **B** and **C** respectively. The graphs in **B** and **A** are shifted by +2.5 and +5 nm respectively to aid visualisation.

As with the distance profiles of the tryptophan analogue in the absence of water (Figure 2.33, **C**), there are distinct time periods (*e.g.* $18.0 \leq t \leq 24.0$ ns) at which the centres of mass of the amino acid derivative and the lipid are sufficiently close together ($r_{Trp\text{-}DMPC} \leq 0.5$ nm) that the two molecules can be considered being associated with each other. The minimal interaction distance between the centres of mass of the tryptophan compound and the lipid respectively was found to be 0.35 ± 0.5 nm, which equals the minimal interaction distance found when no water molecule was present.

Figure 2.33, **B** shows, that there are distinct periods of time, when the water molecule and the amino acid molecule remain at a constant distance ($r_{Trp\text{-}water} = 0.20 \pm 0.5$ nm). This behaviour can be seen at times when the tryptophan analogue is bound to the lipid (*e.g.* $18.0 \leq t \leq 24.0$ ns) and at times when it is not (*e.g.* $8.0 \leq t \leq 10.0$ ns) indicating that an adduct formation between water and amino acid derivative occurs independently from any interactions of the tryptophan compound with the lipid.

Interestingly, the distance between the water and lipid molecules respectively does only fall below 0.5 nm when the amino acid is bound to the lipid molecule at the same time (*e.g.* $18.0 \leq t \leq 24.0$ ns). This suggests that the water and tryptophan binding to DMPC is cooperative, *i.e.* binding of one enhances binding of the other. This is further supported by the observation that there are no hydrogen bonds being formed between the lipid and the water molecule (assuming a maximum hydrogen bond length of 0.35 nm). In contrast, in the water-tryptophan analogue adduct there is at least 1 hydrogen bond present at all times.

Three main clusters, which account for 46.7% of all adduct structures between the water and tryptophan molecule, were identified when analysing the water-tryptophan adduct (Figure 2.34).

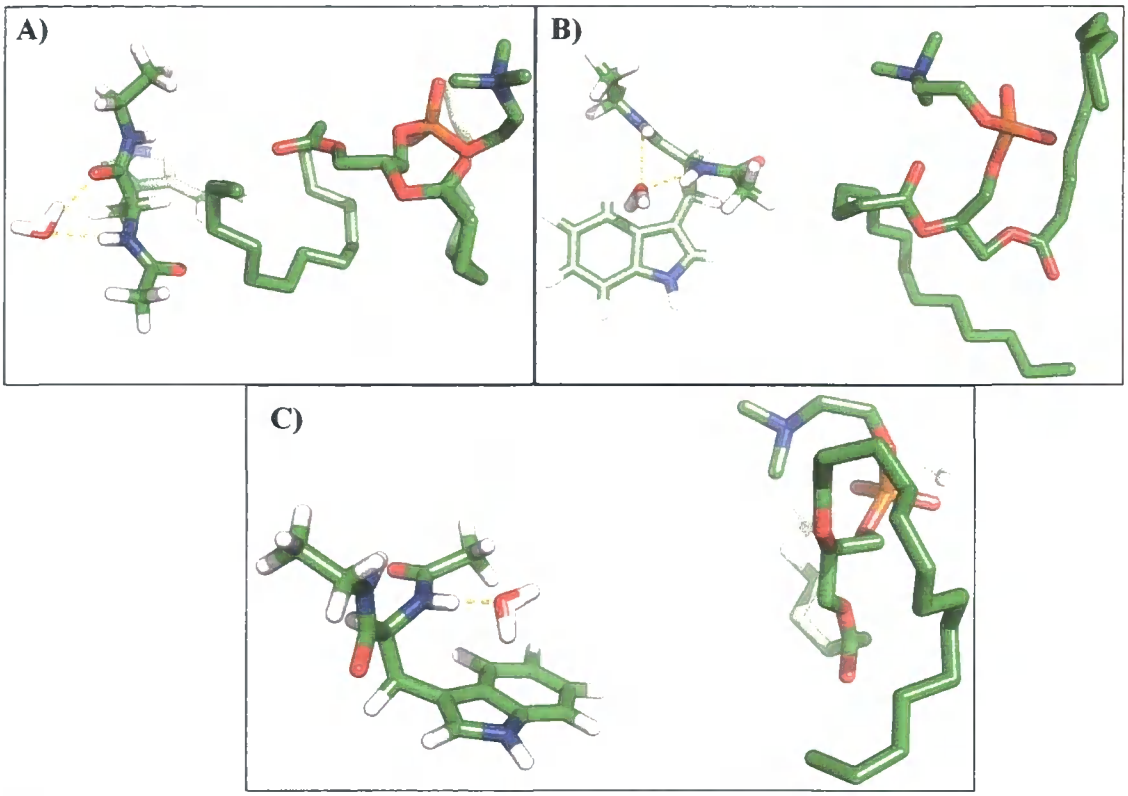


Figure 2.34: Figure showing the three most abundant clusters as found by a cluster analysis (see §5.2.1.e) of the water-amino acid adduct. Cluster **A**) was found in 11.5%, **B**) in 8.7% and **C**) in 26.5% of all analysed adduct structures.

2.5.3.b Hydrogen Bonding Interactions in the Trp:DMPC:water 1:1:1 Complex

The simulations showed a preference of the water molecule towards the tryptophan compound. Hydrogen bonding was seen between the acetyl and ethyl amine hydrogens to the water oxygen atom. The water molecule was also interacting with the π-electron system of the indole ring (as seen in Figure 2.34, **B** and **C**).

Hydrogen bonding interactions could be visualised using the radial distribution function of the water oxygen (OW) to hydrogen bond donors in the amino acid analogue (Figure 2.35).

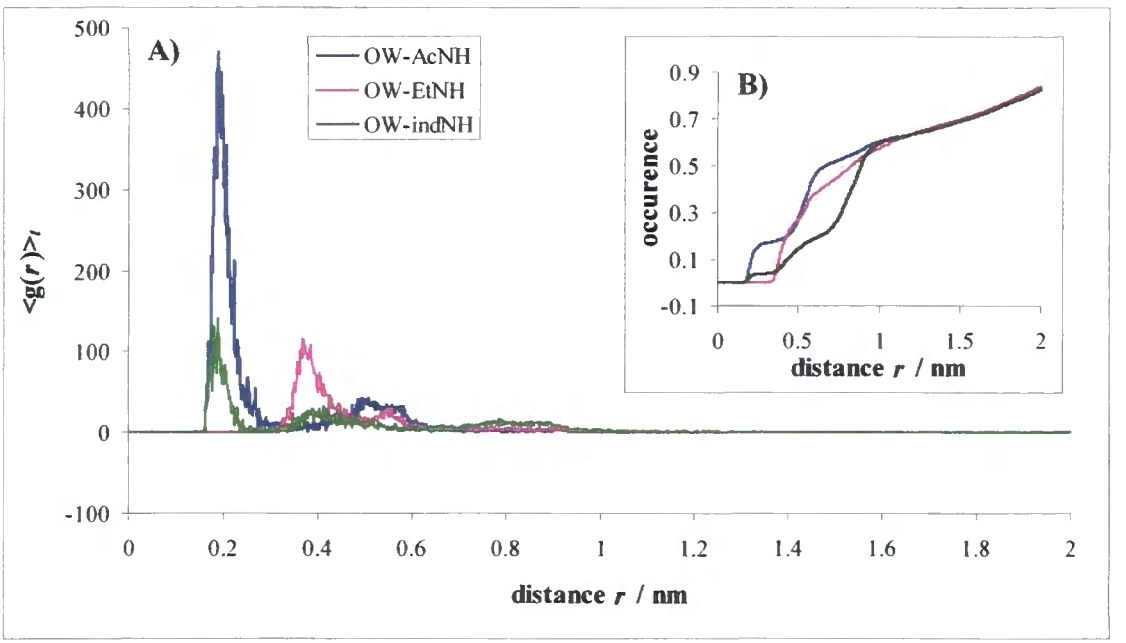


Figure 2.35: Hydrogen bonding interactions between OW and hydrogen bond donors of the amino acid analogue (**6a**). **A** shows the time averaged radial distribution function for the distances between OW, and the acetyl and ethyl amide protons (AcNH and EtNH respectively) as well as to the indole amine hydrogen (indNH). **B** shows the cumulative number of the radial distribution functions displayed in **A**.

Peaks in the time averaged radial distribution $\langle g(r) \rangle_t$ function at short distances ($r = 0.21 \pm 0.03$ nm and $r = 0.20 \pm 0.02$ nm respectively) indicate hydrogen bonding interactions of the water oxygen (OW) to occur preferentially with the acetyl amide and the indole amine hydrogens. Secondary preferred interaction distances were observed at $r = 0.53 \pm 0.05$ nm and $r = 0.44 \pm 0.06$ nm for the distances to the acetyl amide and the indole amine hydrogens respectively. The latter coincides with the preferred interaction distance between the OW and the ethyl amide proton ($r = 0.39 \pm 0.03$ nm). These distances are slightly too large to be considered as classical hydrogen bonds. The second, less pronounced interaction distance between the ethyl amide proton and OW ($r = 0.57 \pm 0.03$ nm) was found at the same distance where a secondary preferred interaction distance between the acetyl amide hydrogen and the OW was observed ($r = 0.53 \pm 0.05$ nm). The $\langle g(r) \rangle_t$ of the distance between the indole amine proton and OW reflects the pronounced favourable interaction distances between OW and the acetyl amide proton with a small elevation at $r = 0.81 \pm 0.08$ nm.

Hydrogen bonding interactions of the water molecule with the lipid were not observed (Figure 2.36, **A** and **C**).

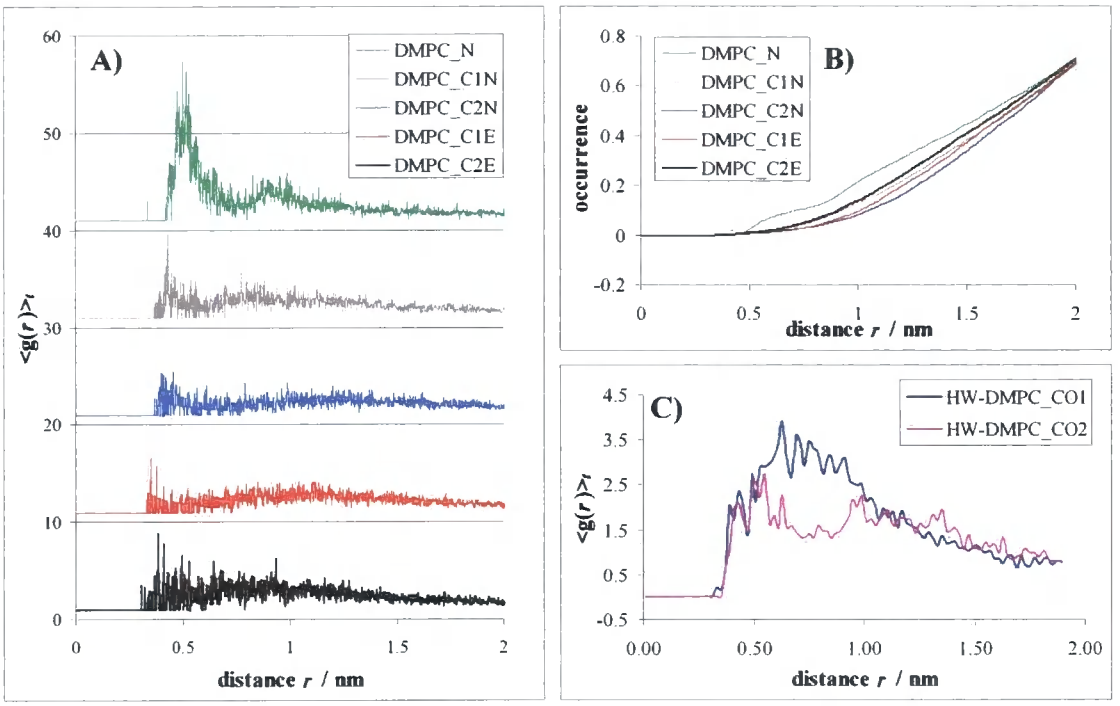


Figure 2.36: Time averaged radial distribution functions between water oxygen (OW) and selected sites in the DMPC lipid molecule. **A** shows the radial distribution function between OW and the lipid choline nitrogen (DMPC_N), the hydrocarbon chain tail atoms (DMPC_C1N and DMPC_C2N) and two lipid hydrocarbon tail midsection atoms (DMPC_C1E and DMPC_C2E). The functions are off-set by +41, 31, 21, 11, and 1 units respectively to aid visibility. **B** shows the cumulative number of the radial distribution functions depicted in **A**. **C** shows the $\langle g(r) \rangle_r$ for the distance between a water hydrogen (HW) and the two lipid carbonyl groups (DMPC_CO1 and DMPC_CO2 respectively).

Only the time averaged radial distribution function between OW and the lipid choline nitrogen shows distinct preferential locations for the two atoms with respect to each other. The more pronounced of these preferred distances was observed at ($r = 0.53 \pm 0.05$ nm), the other, less distinct at $r = 0.93 \pm 0.08$ nm.

These interaction distances are likely to be due to electrostatic interactions of the charged choline head group and the electron negative OW. It is noteworthy that the intensity of these peaks is about an order of magnitude lower than the intensities found in the radial distribution functions of OW and hydrogen bond donors in the tryptophan derivative.

2.5.3.c Determination of Binding Constants

Analogous to the method described above (§2.5.2.d), the binding constant for adduct formation between the lipid and the tryptophan molecules (K_1^{SPC}) and the binding constant for the association of water to DMPC (K_{1wl}^{SPC}) could be determined (Table 2.22).

	exp.	computed		
		with SPC	without*	
K_1	17.93 ± 0.85	45.09 ± 11.41	75.66 ± 2.65	M^{-1}
ΔG_1	-7.11 ± 0.13	-9.31 ± 0.67	-10.65 ± 0.37	kJ mol^{-1}
K_{1wl}	31.86 ± 0.11	5.95 ± 1.53	-	M^{-1}
ΔG_{1wl}	-8.53 ± 0.03	-4.33 ± 0.70	-	kJ mol^{-1}

Table 2.22: Comparison of computed and experimental binding constants and associated free energies for the adduct formation of N^{α}-acetyl-5-fluoro-*L*-tryptophan ethyl amide (**6a**) with DMPC in the absence and presence of water. The first column refers to the binding constants and free energies of association obtained *via* NMR host-guest titrations; column 2 and 3 list these values as obtained by using Equation 2.41 for the simulations of **6a** and DMPC in the presence and absence of water.*Values cited were obtained in the simulations using partial charge Set **B** and 1.9 nm truncation of electrostatic interactions.

The computed values for the binding constant and the association free energy of the adduct formation of N^{α}-acteyl-5-F-*L*-tryptophan ethyl amide (**6a**) with DMPC are improving slightly towards the experimentally obtained value when the presence of water is accounted for in the simulation. The error estimates for K_1^{SPC} and ΔG_1^{SPC} were obtained by three repeat runs of the simulation with identical starting configuration.

Furthermore, the binding constant (K_{1wl}^{SPC}) and association free energy of the 1:1 adduct formation of DMPC and water (ΔG_{1wl}^{SPC}) could be determined. The calculation underestimates the experimentally determined value by ~4 kJ mol^{-1}. It seems likely, that the introduction of more water molecules in the simulation set-up might improve the calculated values further.

2.5.4 Simulation of Amino Acid – Lipid Interactions at a Ratio of 2:1

2.5.4.a Minimum Distance Profiles for Trp1, Trp2 and DMPC

A molecular dynamics simulation of the 2:1 binding of two N^{α}-acetyl-5-fluoro-*L*-tryptophan ethyl amide molecules (**6a**), Trp1 and Trp2, interacting with one molecule of DMPC lipid was performed using the set-up described in (§5.2.1.c). The system with 4024 chloroform molecules in a cubic box with a length of 8.2 nm corresponds to a lipid concentration of

6 m*M* and an amino acid concentration of 3 m*M*. The level of binding at these concentrations could be estimated from the NMR host-guest titrations to be about 13% for the 2:1 and about 14% for the 1:1 species.

The size of the system was necessary to allow full solvation of all compounds, but slowed the computation down to 22 hours per 1 ns simulated time. This limited the run time for the simulations to 50 ns and due to the long-lasting nature of the binding interactions, restricted their assessment to a qualitative analysis.

Adduct formation was analysed by plotting the distance profiles for the mean distance between the centres of mass of the two tryptophan derivatives, $r_{Trp1\text{-}Trp2}$, and their individual distances to the centre of mass of the lipid molecule, $r_{Trp1\text{-}DMPC}$ and $r_{Trp2\text{-}DMPC}$ respectively, as a function of time t (Figure 2.37).

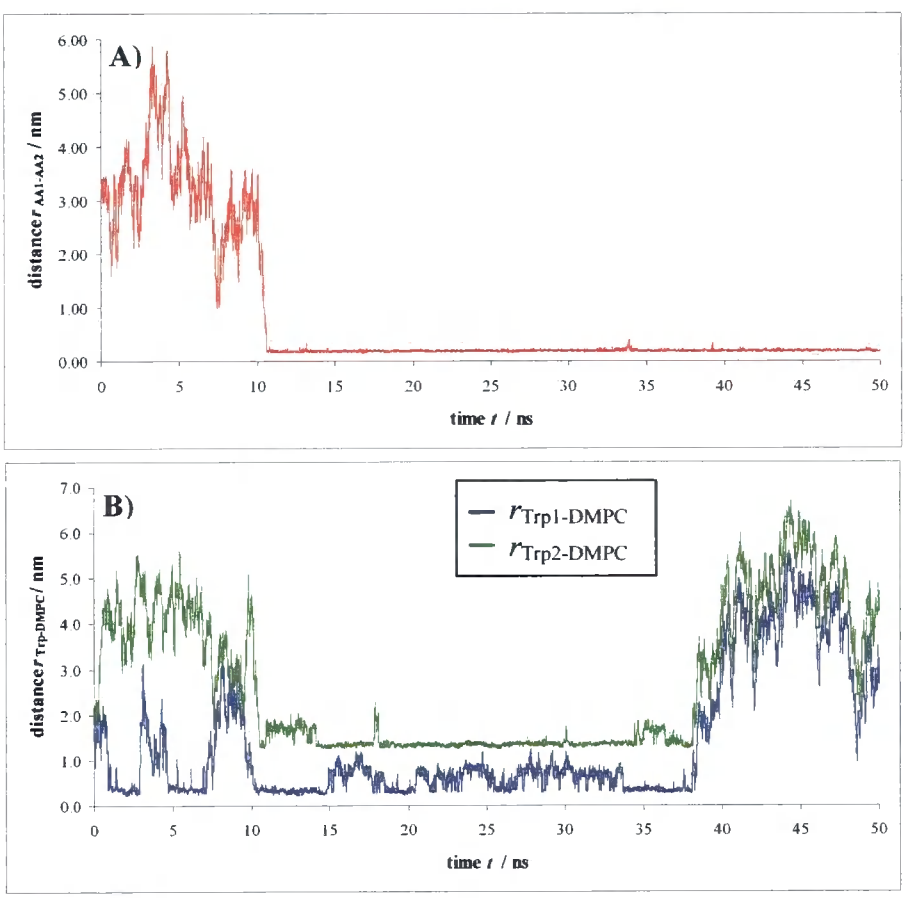


Figure 2.37: Distance profiles for the distances between two tryptophan derivatives (**6a**) and a DMPC molecule respectively. **A** shows the intermolecular distance between the centres of mass of two amino acid analogues, $r_{Trp1\text{-}Trp2}$; **B** shows the distance $r_{Trp1\text{-}DMPC}$ and $r_{Trp2\text{-}DMPC}$ respectively between the centre of mass of Trp1 and Trp2 and the centre of mass of a DMPC molecule over the course of a 50 ns molecular dynamics simulation. The distance $r_{Trp2\text{-}DMPC}$ was shifted by +1.0 nm for clarity.

The $r_{Trp1\text{-}Trp2}$ distance profile (Figure 2.37, **A**) shows that the two amino acid molecules, once they came into close contact ($r_{Trp1\text{-}Trp2} \leq 0.5$ nm), remained at a distinct minimal interaction range of $r_{Trp1\text{-}Trp2} = 0.19 \pm 0.02$ nm. The two molecules maintained this distance for over 39 ns. Figure 2.37, **B** displays the close relationship of the distances between the centres of mass of the two tryptophan derivatives and centre of mass of the lipid molecule ($r_{Trp1\text{-}DMPC}$ and $r_{Trp2\text{-}DMPC}$ respectively).

The minimal interaction distance of the centres of mass of the tryptophan derivatives with respect to the centre of mass of the lipid molecules could be determined as 0.34 ± 0.05 nm. This value equals the minimal interaction distance found in the analysis of the 1:1 binding of the same amino acid derivative with DMPC (§2.5.2). This suggests a similar interaction mechanism for the DMPC-amino acid adduct formation. The occurrence of only a few distinct binding events during the 50 ns simulation meant that the calculation of binding constants for the 2:1 adduct formation was not possible.

Interestingly, the time at which the two amino acid derivatives first move close enough to each other be considered as bound ($t = 10.6$ ns, Figure 2.37, **A**) occurs within 100 ps after the time at which *both* tryptophan compounds *and* the lipid centres of mass are within 0.5 nm of each other. In fact, one of the tryptophan derivatives (Trp1) could already be considered interacting when the second amino acid analogue (Trp2) approaches ($t = 10.5$ ns, Figure 2.38). At this point, the centres of mass of the amino acid analogues are still too far away from each other to be considered being in an associated state ($r_{Trp1\text{-}Trp2} > 0.50$ nm). However, after a short time span (100 ps) the distances between the centres of mass of the amino acids falls below that threshold and stay there for the remainder of the simulation. This suggests that the lipid promotes or even initiates the formation of an amino acid-amino acid complex. To validate this hypothesis further, the two amino acid derivatives were simulated under the same conditions as this simulation but in the absence of DMPC (see §2.5.5).

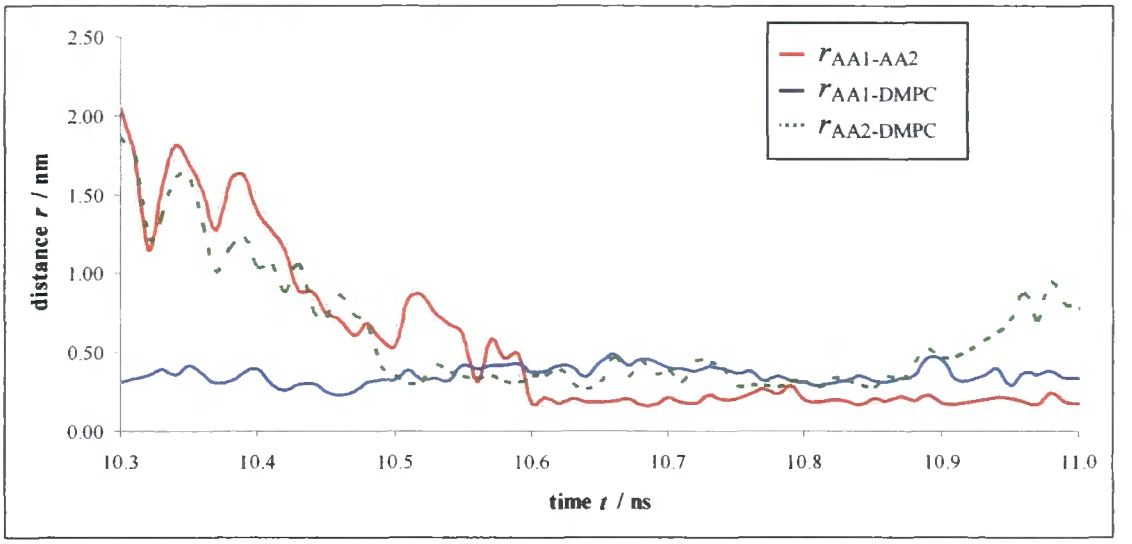


Figure 2.38: Distance profiles for the distances between the centres of mass of the two tryptophan analogues (**6a**) and the DMPC molecules respectively. The region around 10.5 ns is enlarged to allow a more detailed picture of the sequence of association events.

At least two modes of interactions between the amino acid complex and the lipid appear as likely from the radial distribution data: In the first mode (*e.g.* $11.0 \leq t \leq 14.0$ ns), one of the tryptophan derivatives is close to the centre of mass of the lipid molecule ($r_{\text{Trp-DMPC}} \leq 0.34$ nm), and the centre of mass of the other is approximately 0.71 ± 0.14 nm away from the lipid (and the two tryptophan derivatives are 0.37 ± 0.14 nm away from each other). In the second mode (*e.g.* $19.0 \leq t \leq 20.0$ ns), both centres of mass of the amino acid derivatives are approximately equidistant from the centre of mass of the DMPC molecule (both $r_{\text{Trp1-DMPC}}$ and $r_{\text{Trp2-DMPC}} = 0.34 \pm 0.05$ nm respectively) at the same time. The relative occurrence of these modes can be visualised by plotting the root mean squared (rms) difference of the $r_{\text{Trp-DMPC}}$ distances (Figure 2.39).

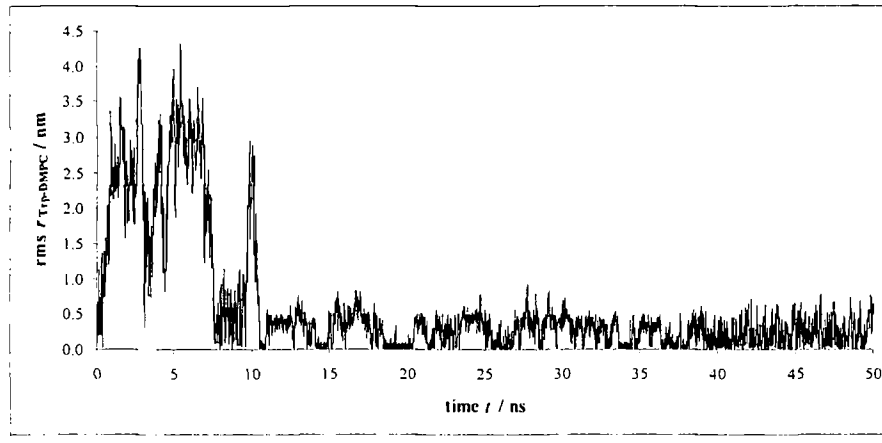


Figure 2.39: Figure displaying the root mean squared distance of the difference in the distances between the centres of mass of the DMPC molecule and the amino acid derivatives Trp1 and Trp2 respectively.

2.5.4.b Analysis of $\langle g(r) \rangle_t$ for distances between Trp1, Trp2 and DMPC atoms

Hydrogen bonding interactions between the two amino acid derivatives were evaluated using radial distribution functions of the distances between hydrogen bond donors (indole amine, acetyl and ethyl amide protons) and hydrogen bond acceptors (acetyl and ethyl carbonyl oxygen) of both (Figure 2.40). Preferred interaction distances between indole amine proton of Trp1 and the ethyl amide proton of Trp2 to the ethyl amide carbonyl of the corresponding other tryptophan derivative were observed at $r_{AB} < 0.35$ nm, which might indicate hydrogen bonding interactions between these atom pairs. Another set of preferred interaction distances for the between these atoms that was seen at larger distances (r_{AB} ~0.65 nm and r_{AB} ~0.95 nm respectively).

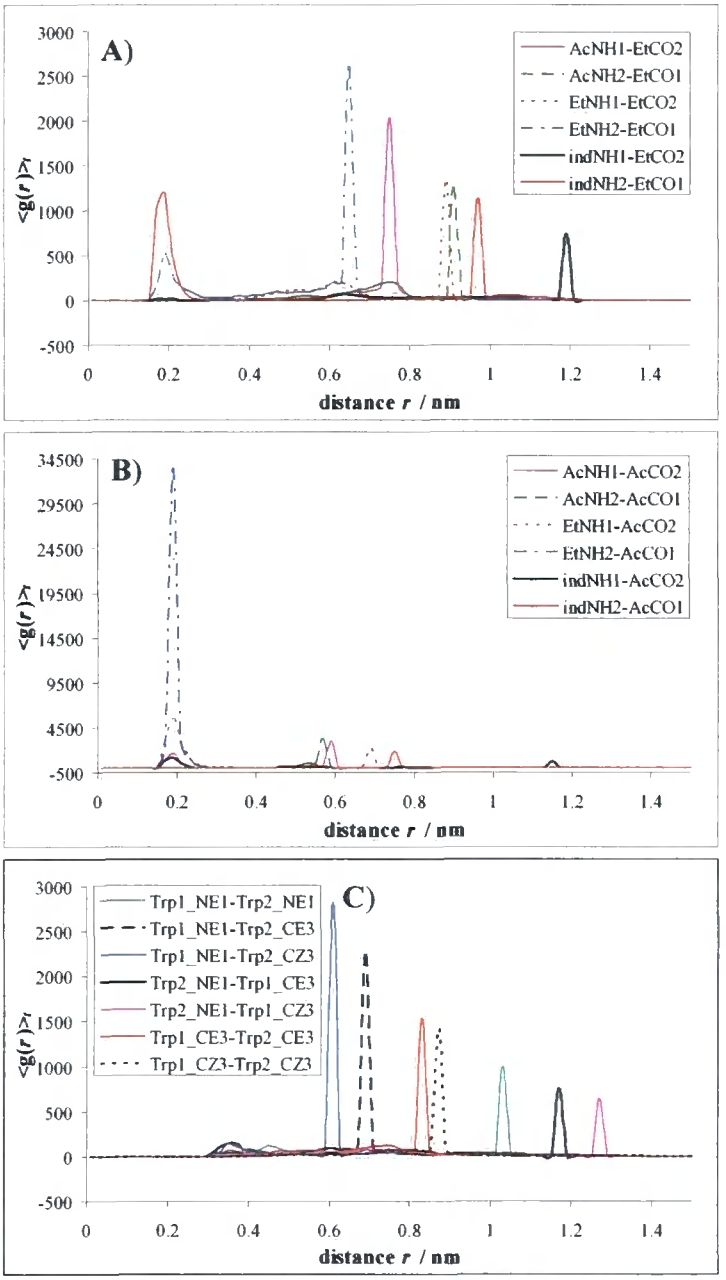


Figure 2.40: $\langle g(r) \rangle_t$ of the distance between hydrogen bond donors and acceptors of the two tryptophan derivatives (**6a**), Trp1 and Trp2. **A** depicts the distances of hydrogen bond donors (the acetyl amide (AcNH), ethyl amide (EtNH) and indole amine protons (indNH) respectively) of Trp1 and Trp2 with respect to ethyl amide carbonyl oxygen (EtCO) of the other Trp molecule. In **B** the distances of the hydrogen bond donors were evaluated with respect to the acetyl amide carbonyl (AcCO) of the other Trp molecule. **C** displays intermolecular distances of selected indole ring atoms of the amino acid analogues with respect to each other. The labeling of atom names follows the one introduced in Table 2.16.

For the ethyl amide proton this second preferred distance can be rationalised by the strongly preferred interaction distance of this proton with the acetyl amide oxygen of Trp1 (Figure 2.40, **B**). These hydrogen bonding interaction are likely to be contribute to interactions that fix the conformation of the two tryptophan derivatives with respect to each other, causing these distinct preferential interaction distances. Various other preferred interaction distances at larger distances (r_{AB} >0.50 nm) were observed in the time averaged radial distribution functions of the intermolecular distances of hydrogen bond donors and acceptors of both tryptophan derivatives, but these interactions are less likely to be involved in hydrogen bonding interactions. However, they indicate specific orientations and interaction distances between the two amino acid analogues.

This occurrence of preferred orientations and interaction distances is further supported by the time averaged radial distribution functions of the distance between indole ring atoms of the two tryptophan derivatives with respect to each other (Figure 2.40, **C**). Preferred interaction distances could be seen at distances r_{AB} ~0.70 ± 0.1 nm, which could indicate π-π interactions between the two aromatic ring systems. The fact that the preferred interaction distances are highly preserved (*i.e.* strong peaks in the $\langle g(r) \rangle_t$) indicates that such interactions are likely.

No strongly preserved preferred interaction distances between hydrogen donors of either of the tryptophan derivatives to hydrogen bond acceptors of the lipid molecule were observed (see §7.5), indicating that no long-lasting hydrogen bond bridges were established between the amino acid analogues and DMPC. However, interactions between the polar choline head group and the indole side chain of at least one of the tryptophan derivatives were observed (Figure 2.41).

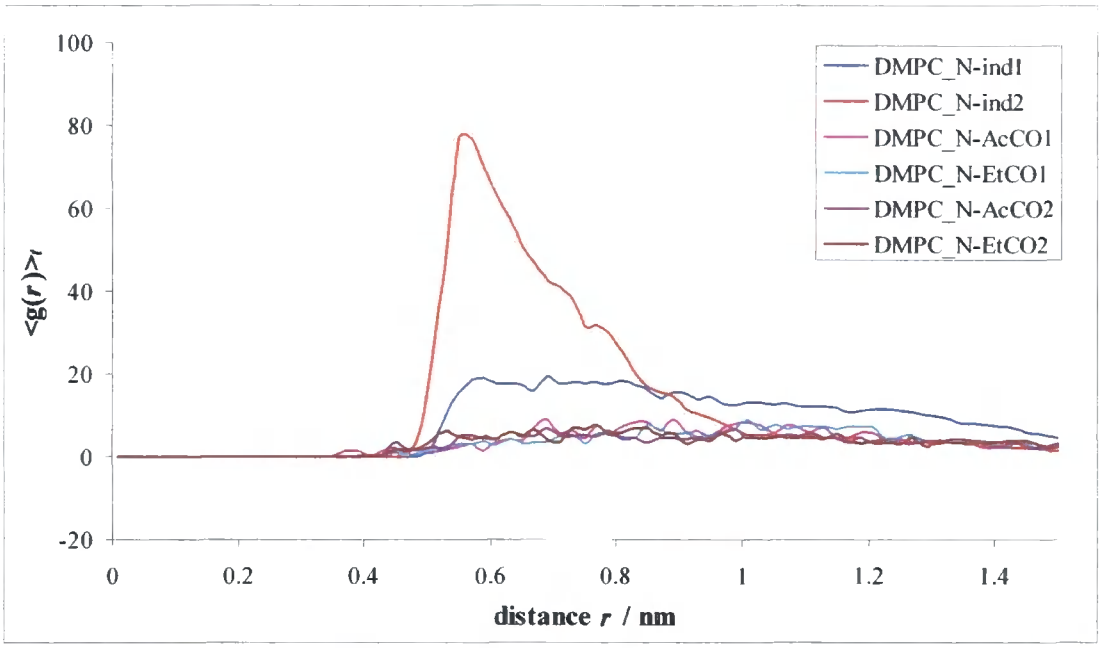


Figure 2.41: Interaction distances between the lipid choline nitrogen and selected sites of two amino acid analogues (**6a**). These are the acetyl amide and ethyl amide carbonyl oxygens (AcCO and EtCO respectively) and the centre of mass of the indole aromatic side chains.

The observed peak for at a distance r_{AB} ~0.60 nm for the distance between the choline head group nitrogen (DMPC_N) and the indole side chain of one of the tryptophan derivatives (Trp1) indicates a preferential interaction distance between these to sites, which is likely to be a cation-π interaction. Interestingly, this preferred distance is only observed for one of the tryptophan compounds while the other shows a broader distribution with an elevation at r_{AB} ~0.60 nm and subsequent gradual decline. This might indicate a less defined interaction distance between the choline nitrogen and the aromatic rind of Trp2.

2.5.4.c Analysis of hydrogen bonds between the Trp1, Trp2 and DMPC Molecules

A hydrogen bond analysis using GROMACS software revealed that up to 4 hydrogen bonds were established between the two amino acid tryptophan derivatives upon complexation (Figure 2.42). No hydrogen bonding interactions between the tryptophan compounds and the lipid were observed at any time

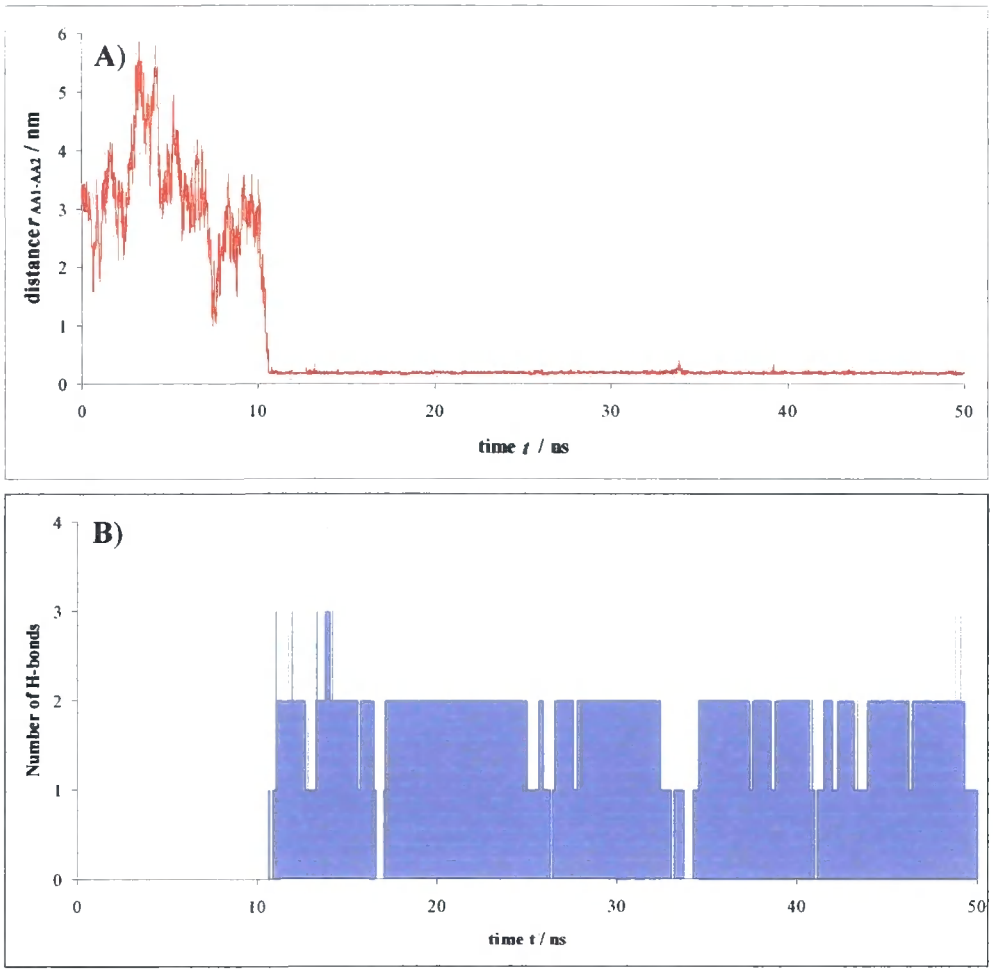


Figure 2.42: Number and occurrence of hydrogen bonds between Trp1 and Trp2. **B** depicts a histogram of the number of hydrogen bonds established between the two amino acid analogues (**6a**). **A** shows the distance between the centres of mass of the two tryptophan derivatives with respect to each other as a reference.

The occurrence of hydrogen bonds being established between the two tryptophan derivatives coincides with the cluster formation of the two molecules as can be seen clearly in Figure 2.42. On average, the cluster is held together by 2 hydrogen bonds. These are likely to involve the hydrogen donor/acceptor atom pairs with preferred interaction distances $r_{AB} < 0.35$ nm (Figure 2.40). Hydrogen bonds between the ethyl amide proton of Trp2 (EtNH2) and the acetyl amide carbonyl oxygen of Trp1 (AcCO1), the indole amine proton of Trp2 (indNH2) and the ethyl amice carbonyl oxygen of Trp1 (EtCO1), and the ethyl amide proton of Trp1 (EtNH1) and the acetyl amide carbonyl oxygen of Trp2 (AcCO2) are likely candidates.

2.5.4.d Analysis of Trp:DMPC 2:1 Clusters

A cluster analysis of the bound conformations observed in the minimum distance profiles of the two tryptophan derivatives Trp1 and Trp1 with respect to each other and with respect to the lipid molecule were analysed for the occurrence of clusters (Figure 2.43). The analysis of clusters revealed structures that exhibited face-to-face (Figure 2.43, **A**) and edge-to face (Figure 2.43, **B**) π-stacking interactions between the two aromatic indole side chains of the tryptophan derivatives (**6a**). The most populated clusters are depicted in Figure 2.43, **C-E** in increasing occurrence (9.5%, 14%, 42% respectively). As expected, distinct hydrogen bonding interactions between the two amino acid analogues can be observed with the ethyl amide proton of one Trp hydrogen bonding with the acetyl amide carbonyl oxygen of the other Trp. To the latter two hydrogen bonds could be established as seen in Figure 2.43, **D**. The visual inspection supports the conclusions drawn in §2.5.4.b.

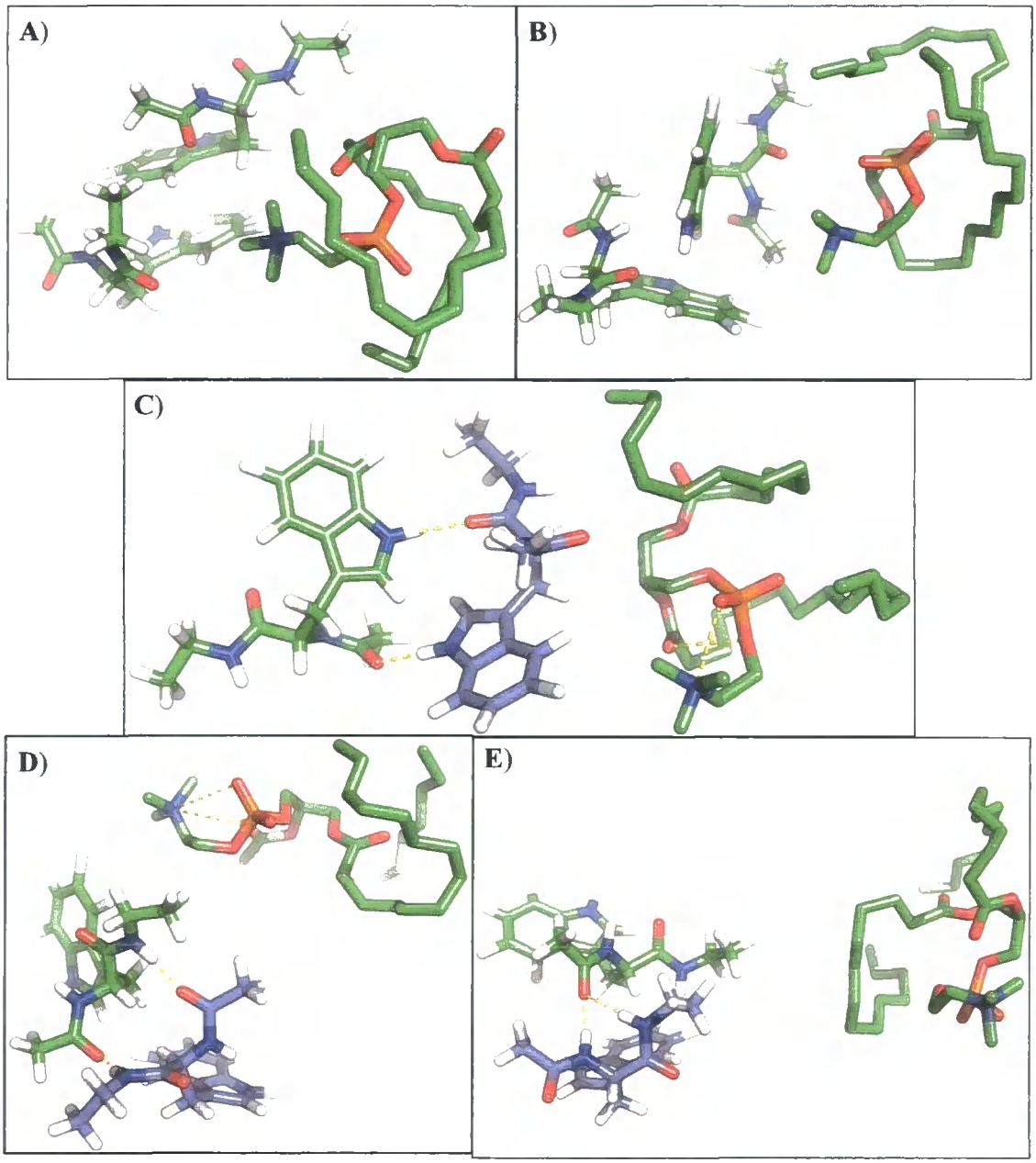


Figure 2.43: Trp:DMPC 2:1 clusters as seen in molecular dynamics simulations. **A** and **B** show clusters that show π-stacking interactions of the two amino acid derivatives (**6a**), face-to-face interactions in **A** and edge-to-face interactions in **B**. **C**, **D**, and **E** display the more common clusters

2.5.5 Simulation of Amino Acid – Amino Acid Interactions at a ratio of 1:1

A 50 ns simulation of two N^{α}-acetyl-5-fluoro-*L*-tryptophan ethyl amide molecules (**6a**) in chloroform was performed using the experimental set-up described in §5.2.1.d. This control experiment was used to determine any background binding between the two analogues and to access the importance of the presence of the lipid. It was found that the two molecules were

not interacting strongly with each other as shown in Figure 2.44, which depicts the average distance between the two centres of mass of the two tryptophan derivatives with respect to each other over time.

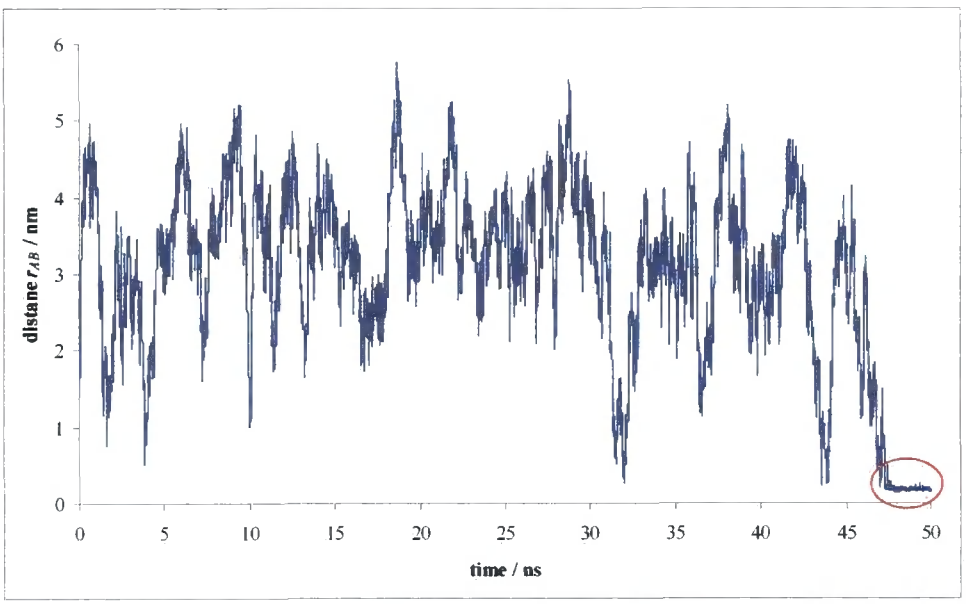
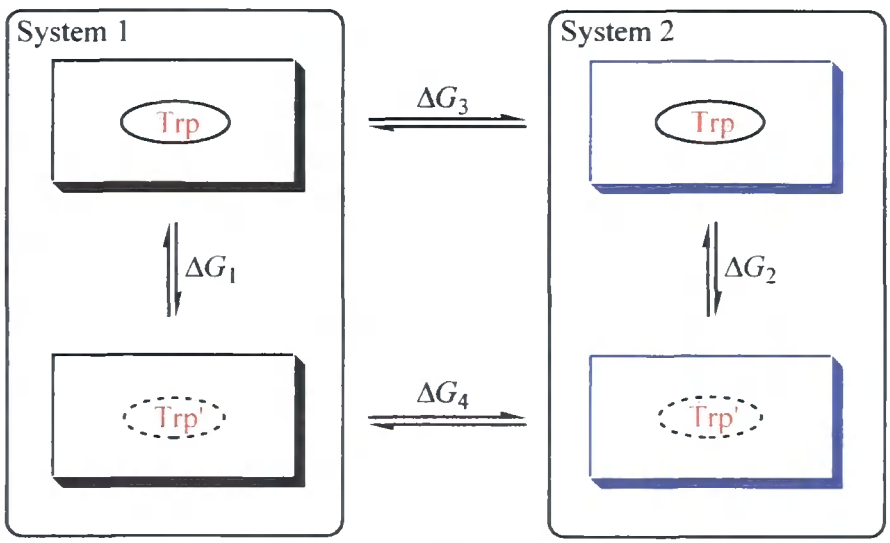


Figure 2.44: Distance profile of the distance between the centres of mass of two N^{α}-acetyl-5-fluoro-*L*-tryptophan ethyl amide molecules (**6a**) over a 50 ns molecular dynamics simulation. Adduct formation towards the end of the analysed time span is highlighted by a red circle.

During 95% of the simulated time span, the two molecules were apart from each other. Only during the last 2.5 ns was an association at an average distance of $r_{AB} = 0.18 \pm 0.02$ nm visible. These data indicate that the two amino acid derivatives were freely solvated in solution and not associated for longer period of times during the time span analyzed. This contrasts with the case when DMPC was present in the simulations, when the two amino acid derivatives were found to be associated during most (78%) of the 50 ns, with the lipid molecule being in close proximity ($r_{AB} = 0.34$ nm). This further supports the finding that the lipid is essential for the fostering and stabilisation of the amino acid – amino acid complex (see also §2.5.3).

2.6 *Free Energy Calculations*

The following section describes the calculation of solvation free energy of tryptophan derivatives in chloroform (§2.6.1) and their free energy of association with a DMPC lipid molecule in chloroform at 25 °C (§2.6.2). As outlined in §5.2.3, free energies were computed by employing a thermodynamic cycle (Scheme 2.12), where the tryptophan analogue is represented as Trp when fully interacting with its environment or as dummy particle Trp' when all its non-bonding interactions with the environment have been turned off:



Scheme 2.12: General scheme for the calculation of a free energy of interest by using the thermodynamic integration method. ΔG_1 and ΔG_2 refer to the free energy associated with the gradual fading of all non-bonded interactions of Trp with its environment in System 1 and 2, respectively; ΔG_4 describes the free energy contribution of the transfer of the dummy Trp' from System 1 to System 2; ΔG_3 describes the free energy contribution of the transfer of the fully interacting Trp from System 1 to System 2.

In the calculation of solvation free energies, the particle were represented once in vacuum and once in solvent (System 1 and System 2 in Scheme 2.12 respectively).

2.6.1 Calculation of Solvation Free Energies of Tryptophan Analogues in $CHCl_3$

The simulations of N^{α}-acetyl-*L*-tryptophan ethyl amide molecules (**6a-g** & **6n**) were set up using the procedure described in (§5.2.3.a). For the representation of System 1 (see Scheme 2.12), the molecules were positioned in a simulation box with infinite box lengths and non-bonding interactions were gradually and linearly turned off over 21 steps. The free energy of

the transition between the fully interacting and the dummy particle was evaluated by integration of the free energies at each step (Equation 4.24, §5.1.1.d). The free energy contribution *in vacuo*, ΔG_1, and the free energy of the transfer from a fully interacting particle in solution into a dummy particle, ΔG_2, could be calculated (Table 2.23) with small estimated errors (< 1% and < 3.5% respectively).The accuracy of the calculations was further assessed by the inversion of the change in λ to produce a simulation for the change of the tryptophan derivative from a dummy to a fully interacting molecule ($\lambda = 1 \rightarrow \lambda = 0$). The obtained values of the reverse-simulation were within experimental error identical to the values for the corresponding simulation where λ was changed from 0 to 1.

It is proposed, [28] that the free energy of transferring a fully non-interacting particle from vacuum into solution, ΔG_4 is zero. Applying Equation 2.42, the free energy of solvation ΔG_{solv} can be expressed as the sum of the free energies (ΔG_1 to ΔG_4):

$$\Delta G_{solv} = \Delta G_1 - \Delta G_2 + \Delta G_4 = \Delta G_3 \qquad (2.42)$$

Assuming $\Delta G_4 = 0$, the free energy of solvation ΔG_{solv} of a compound could be determined (Table 2.23). The calculated free energies of solvation ranged from approximately -11.9 ± 5.46 kJ mol^{-1} (R = Me, **6e**) to -22.2 ± 7.9 kJ mol^{-1} (R = NO_2, **6g**).

ID	ΔG_1	ΔG_2	ΔG_{solv}
	kJ mol^{-1}	kJ mol^{-1}	kJ mol^{-1}
6a	269.48 ± 0.14	283.73 ± 4.80	-14.25 ± 4.80
6b	238.91 ± 0.28	257.50 ± 5.80	-18.59 ± 5.80
6c	220.56 ± 0.46	241.33 ± 5.54	-20.77 ± 5.54
6d	215.51 ± 1.60	236.10 ± 6.55	-20.59 ± 6.55
6e	239.45 ± 0.71	251.36 ± 5.46	-11.91 ± 5.46
6f	296.57 ± 1.44	313.04 ± 5.89	-16.47 ± 5.89
6g	222.97 ± 0.35	245.13 ± 7.89	-22.16 ± 7.89
6n	208.66 ± 0.33	229.60 ± 5.21	-20.94 ± 5.21

Table 2.23: Free energy contributions ΔG_1 and ΔG_2 of N^{α}-acetyl-*L*-tryptophan ethyl amide compounds (**6a-g** & **6n**) being transferred from a fully interacting state to a fully non-interacting state in vacuum and in solution respectively.

When plotted against electron density of the indole ring, the pattern shown in Figure 2.45 was observed. No clear pattern was observed. The negative solvation energies are consistent with a favourable solvation of the compounds in chloroform.

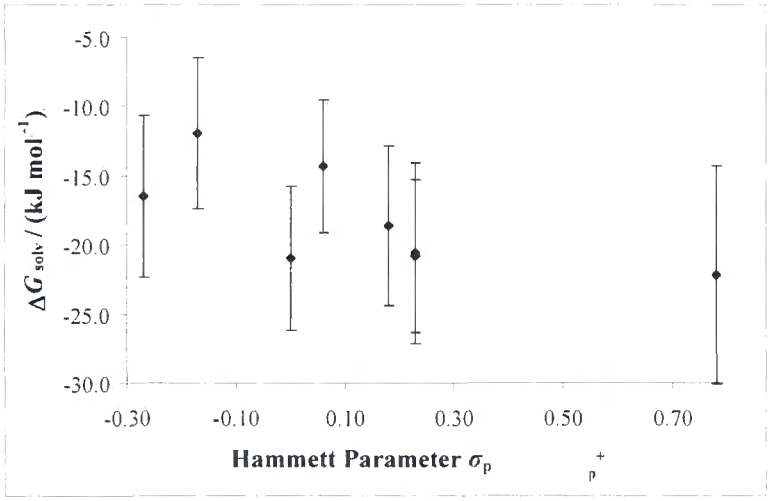


Figure 2.45: Solvation free energies of 5-monosubstituted N^{α}-acetyl-*L*-tryptophan ethyl amides (**6a-g, 6n**) plotted against the Hammett parameter.

Villa *et al.*[29] calculated the free energy of solvation for 3-methylindole as a mimic of the tryptophan side chain in chloroform as being -39.6 ± 1.4 kJ mol^{-1} at 20 °C. The authors used the standard rigid model for chloroform as implemented in the GROMOS96 (43a2) force field.[30] The reported experimental value for the free energy of solvation of 3-methylindole in chloroform at 20 °C (-36.8 kJ mol^{-1}) was somewhat smaller.[31]

In comparison, the calculated free energy of solvation for N^{α}-acetyl-*L*-tryptophan ethyl amide (**6n**) in chloroform (-20.94 ± 5.21 kJ mol^{-1}) is about half the reported value for 3-methylindole. However, the polar backbone of the tryptophan derivative is likely to decrease favourable interactions with the non-polar solvent chloroform, decreasing the free energy of solvation. The free energy contribution of the backbone can be estimated from studies by Daura *et al.*[32] The authors used the united atom GROMOS87 force field to calculate the free energy difference of the alchemistical transformation of *N*-acetyltryptamine into 3-methylindole in chloroform to be 73.3 ± 2.2 kJ mol^{-1} at 27 °C. The implemented partial charges on the *N*-acetyl functional group were smaller than the values for the all-atom description in this study. This difference in partial charges is likely to have a strong effect on calculated free energy differences as a scaling factor of 1.1 for the partial charges on the *N*-

acetyl functional group reportedly decrease the free energy of solvation by -6.02 kJ mol^{-1}.[32] The partial charges used here differed from the values employed in Dauras study by a factor of approximately 2.

Recently, the importance of partial charges for hydration free energies was further investigated by Xu *et al.*[28a] In their study of 3-methylindole, two new sets of partial charges for the aromatic ring atoms were derived from density functional calculations by natural population analysis (NPA) and fitting of the electrostatic surface potential (ESP) (Table 2.24).

The authors suggested reparameterised partial charges that were a combination of the original OPLS-AA force field charges and the partial charges derived by NPA. Partial charges that were derived by ESP were found to be more exaggerated and to vary in a greater range of negative values. This suggested that the NPA charges were more transferable than ESP charges and thus potentially more suitable in nonpolarizable force fields. Furthermore, the calculations in this study were performed at 22 °C accounting for a slight deviation in the ΔG_{solv} value.

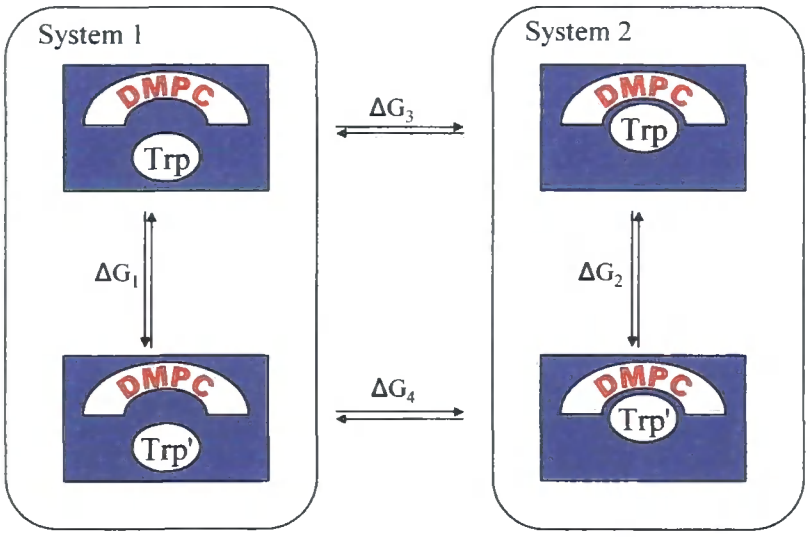
	OPLS-AA		NPA	ESP	OPLS-AA/m		Set A	Set B
CG	0.075		-0.10	0.05	-0.100		-0.331	-0.407
CD1	-0.115		-0.04	-0.16	-0.040		-0.020	0.013
HD1	0.115		0.23	0.16	0.230		0.220	0.166
CD2	-0.055		-0.09	0.04	-0.090		0.258	0.284
NE1	-0.570		-0.57	-0.50	-0.570		-0.582	-0.587
HE1	0.420		0.42	0.39	0.420		0.300	0.402
CE2	0.130		0.15	0.28	0.150		0.298	0.221
CE3	-0.115		-0.24	-0.18	-0.115		-0.356	-0.330
HE3	0.115		0.24	0.13	0.115		0.218	0.208
CZ2	-0.115		-0.24	-0.29	-0.115		-0.388	-0.229
HZ2	0.115		0.24	0.14	0.115		0.204	0.142
CZ3	-0.115		-0.24	-0.18	-0.115		-0.184	-0.061
HZ3	0.115		0.24	0.12	0.115		0.155	0.096
CH2	-0.115		-0.24	-0.08	-0.115		-0.123	-0.179
HH2	0.115		0.24	0.11	0.115		0.163	0.137

Table 2.24: Comparison of indole partial charges. The column OPLS-AA displays the partial charges used in the original force field; columns 2-4 display partial charges used by Xu *et al.*[28a] calculated by natural population analysis (NPA) or fitting of the electronic surface potential (ESP) and their adapted charges obtained by a combination of OPLS-AA and NPA derived partial charges (OPLS-AA/m). The last two columns present the partial charges used in this study as determined by restricted Hartree-Fock *ab initio* calculation (Set **A**) and MP2 calculation with subsequent fitting using CHelpG (Set **B**).

Finally, the free energy simulations will depend to a certain extend on the starting conformation of the molecule. More compact conformations will exhibit a smaller exclusion volume for solvent molecules, decreasing the calculated free energy of solvation. More extended molecules, and particularly exposed polar functional groups, might influence the calculated free energy of solvation. This is also likely to contribute to the somewhat higher error associated with the free energy calculations of the tryptophan derivatives observed in this study in comparison to free energy calculations of the more rigid 3-methylindole, which is very restricted in it conformational space.

2.6.2 Analysis of Binding Free Energies of Amino Acid – Lipid Interactions at a Ratio of 1:1

Similar to the determination of free energies of solvation, binding free energies were calculated using the thermodynamic cycle as shown in Scheme 2.13. Here, all the simulations were performed in chloroform (see §5.2.3.b for more details).



Scheme 2.13: General scheme for the calculation of the association free energy of Trp with DMPC. In System 1, both molecules are freely solvated, whereas in System 2, they are associated with each other. In both systems, the non-bonding interactions of one of the molecules (Trp in this case) are gradually and linearly turned off, resulting in the generation of a dummy molecule (Trp') which no longer interacts with its environment. The free energy associated with both processes (ΔG_1 and ΔG_2) can be computed with relative ease.

ΔG_1 corresponds now to the free energy contribution of the phasing out of non-bonding interactions of a fully solvated, free N^{α}-acetyl-*L*-tryptophan ethyl amide derivative

(**6a**-g, **6n**) in the presence of DMPC and was calculated with an error of <9% (Table 2.25). The free energy contribution of the associated N^{α}-acetyl-*L*-tryptophan ethyl amide derivative losing its non-bonding interactions to the environment ΔG_3 were calculated within an error margin of <5.5%.

ID	ΔG_1	ΔG_2	ΔG_{bind}
	kJ mol^{-1}	kJ mol^{-1}	kJ mol^{-1}
6n	241.52 ± 10.84	236.62 ± 10.68	4.90 ± 10.84
6a	293.44 ± 25.82	312.83 ± 9.79	-19.39 ± 25.82
6b	324.16 ± 12.65	285.95 ± 10.21	38.20 ± 12.65
6c	264.93 ± 10.81	254.11 ± 13.79	10.82 ± 13.79
6d	252.23 ± 11.86	253.91 ± 13.17	-1.68 ± 13.17
6e	273.85 ± 12.02	333.49 ± 10.38	-59.64 ± 12.02
6f	337.85 ± 13.95	320.47 ± 14.31	17.37 ± 14.31
6g	281.88 ± 12.05	258.33 ± 13.22	23.54 ± 13.22

Table 2.25: Free energy contributions ΔG_1 and ΔG_2 of N_a-acetyl-*L*-tryptophan ethyl amide (**6a-g, 6n**) compounds being transferred from a fully interacting state to a fully non-interacting state in vacuum and in solution respectively.

The resulting binding free energy ΔG_{bind} was approximated as the difference between ΔG_1 and $\Delta\Delta G_2$. The obtained free energies of association vary considerably including strongly favourable interactions (ΔG_{bind} = -59.64 ± 12.02 kJ mol^{-1}, **6e**), weak associations (ΔG_{bind} = -1.68 ± 13.17 kJ mol^{-1}, **6d**) and unfavourable interactions (ΔG_{bind} = +38.20 ± 12.65 kJ mol^{-1}, **6b**). The associated error in these free energies was large. This discrepancy might be caused by the limited conformational sampling used in the free energy calculations of the bound complex. Only one average structure of a bound complex was used in each of the calculations, while experimentally observed and calculated free energies of calculations were derived from an ensemble average over a large conformational space. The free energy calculations should be performed with a selection of bound state configurations. Furthermore, biasing potentials that restrain translation, orientation, and conformation of the molecules could help enhance the convergence of the calculations.[33]

2.7 *Molecular Dynamics Simulation of Peptide – Lipid Interactions*

Molecular dynamics simulations of model peptides of the general structure AcWLWLL were performed using the procedure described in §5.2.2. A confomer of the peptide was inserted 1 nm away from a pre-equilibrated bilayer surface, replacing SPC water molecules in its place. Simulations were then run between 20 and 50 ns and the orientation of the peptide with respect to the bilayer normal analysed. A cluster analysis of the peptide revealed confomers relevant for association with the bilayer.

2.7.1 Bilayer Structure

2.7.1.a Area per lipid

In molecular dynamics simulations of lipid bilayers, distances between certain atoms across the membrane, and the orientational order parameter of the lipid hydrocarbon chains are used to benchmark the calculation. Comparison with experimentally obtained values for these parameters can help assess the quality of the simulation at hand. One of the most commonly employed tests is the area per lipid molecule, A, which can be measured accurately *via* both experiment and simulation. In the simulation, the area per lipid can be assessed by following the dimensions of the lipid bilayer plane over time. The division of this area by the number of lipids per leaflet (*i.e.* 64 in this study) gives the area per lipid. Table 2.26 shows the relevant experimental data for the area per lipid A of DMPC bilayers (measured at 27 °C), in comparison to the values obtained in this study:

Entry	Ref	Method used	A / nm^2
I	34	gravimetric x-ray method (GX)	0.652
II	35	corrected gravimetric x-ray method (GXC)	0.617
III	36	corrected value for entry I	0.676
IV	36	corrected value for entry II	0.634
V	37	Start of the equilibration: t_{eq} = 0 ns (Figure 2.46)	0.734
VI		End of the equilibration: t_{eq} = 10 ns	0.586
VII		AcWLWLL production run (Figure 2.48)	0.579 ± 0.03

Table 2.26: Table showing experimentally and computationally determined values of the area per lipid A for a DMPC bilayer at 27 °C. These values are compared to the values obtained at various stages of the molecular simulation, using the procedure described in §5.2.2.

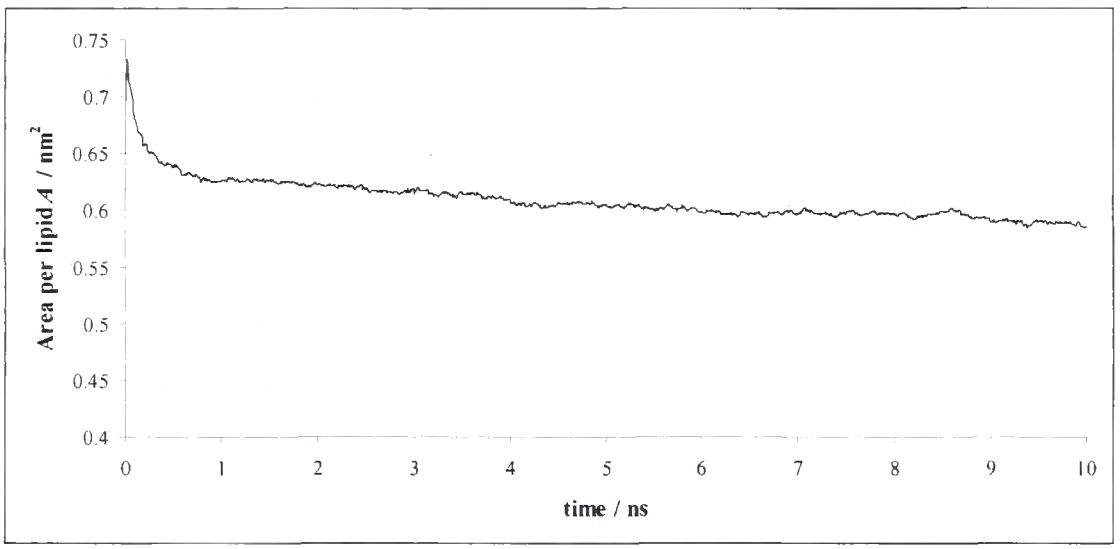


Figure 2.46: The evolution of the area per lipid during the equilibration phase of the calculations.

The starting conformation of the lipid bilayer equilibration was found to represent experimentally obtained values for the area per lipid *A* well (Table 2.26). However, the area per lipid dropped significantly during the first nanosecond of the equilibration, with accompanying expansion of the simulation box in the direction of the bilayer normal and contraction along the bilayer plane.

A variation the experimentally determined area per lipid values of up to 20% is not unknown,[36,38] yet the observed fast and significant reduction in *A* (-16%) is likely to be due to artificial ordering.

The area per lipid is very sensitive to simulation details and large thermal fluctuations in the lipid area around its average value have been observed with correlation times of 1 to 10 ns.[39,40,41] These fluctuations can be attributed to the contribution from a hierarchy of motions, including isomerisation (in saturated lipids), rotation, and diffusion of individual lipids, as well as collective motions of the bilayer itself such as the appearance of undulations. Such thermal fluctuations in *A* could potentially induce a phase transition from the liquid crystalline phase L_β to a more ordered gel-phase (L_α). The reported transition temperature for the phase transition (24 °C, Ref. 42) is not far from the temperature the simulations were performed at (27 °C).

The equilibration of the area per lipid value to a relatively low value, suggests that the force field and long-range treatment of electrostatic interactions used for the description of the bilayer underestimate this parameter. Such artificial ordering of lipid bilayers has been

described previously,[43] and a small decrease in the area per lipid value was to be expected, as small charge groups were used for the representation of the lipid charges. Anézo *et al.*[41] had shown that the area per lipid decreases by 2%, when smaller charge groups were used with a PME treatment of electrostatic interactions. The effect seen in PME can be attributed to the effect of changing the effective cut-off for the Lennard-Jones interactions (as the electrostatic cut-off in PME serves only as a numerical device to separate direct- and reciprocal-space sums).

The lipid parameters used were adapted from Ref. 44 and were originally parameterised for the use of the GROMOS force field gmx.[45] In this study, the lipid parameters were converted into an OPLS-AA notation as the AcWLWLL peptides used were parameterised for this force field. This conversion of the lipid parameters into OPLS-AA notation might have caused a further drop in the area per lipid value due to the slightly different operation of both force fields.

A control run of an identical starting configuration of the hydrated bilayer using the same treatments for long-range interactions using the GROMOS gmx force field lead to an area per lipid value ($A = 0.594 \pm 0.06$ nm^2) that showed considerable fluctuations but that was consistent with experimentally determined values (Figure 2.46).

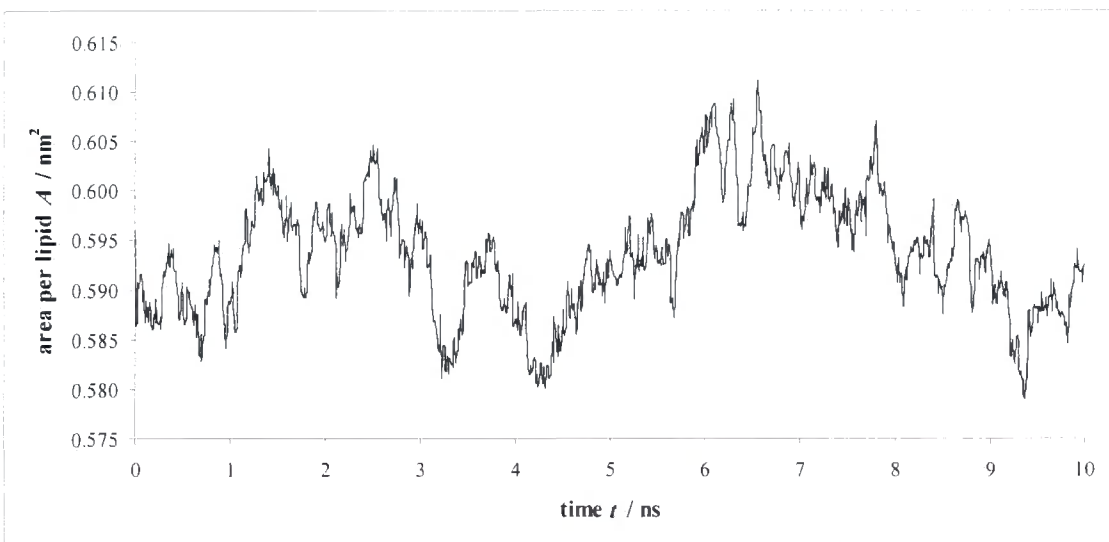


Figure 2.47: Area per lipid A as observed in a control equilibration of a DMPC bilayer using a molecular dynamics simulation employing the GROMOS force field gmx.

Although the area per lipid parameter was not represented with high accuracy, the simulation of AcWLWLL type peptides with the bilayer were pursued using the OPLS-AA description of the lipid bilayer as this force field had been implemented in the set-up of the peptides and the discrepancy between the simulated area per lipid and experimentally obtained values was acceptable given the uncertainty in A of up to 20%.[36,46]

The water to lipid ratio in the simulation was chosen to allow full hydration of the bilayer (3655 molecules of water to 128 molecules of lipid, *i.e.* 57:2), allowing for the replacement of some (~20) water molecules by the peptide. Nagle *et al.*[36] suggested that a ratio of 51:2 water molecules per molecule of lipid would ensure full hydration.

2.7.2 Simulations of Binding Interactions between AcW*LWLL and a DMPC Bilayer

Binding of AcW*LWLL peptides (where W* refers to 5-monosubstituted derivatives of the tryptophan residue) to a DMPC lipid membrane was simulated using the procedure described in §5.2.2. The analysed substituents on the indole ring span the halogen series from F to Cl. Peptides are described by the abbreviation AcXLWLL (where X stands for W, F, Cl, Br, or I) indicating the substituent on the 5-position of the indole ring in Trp^1.

2.7.2.a Area per Lipid during Peptide-Lipid Simulations

During the peptide calculations the area per lipid value was found to be stable and equilibrated ($A = 0.579 \pm 0.03$ nm^2) and not markedly influenced by the presence of the peptides (Figure 2.48). This consistency in A was seen for all the analysed peptide types.

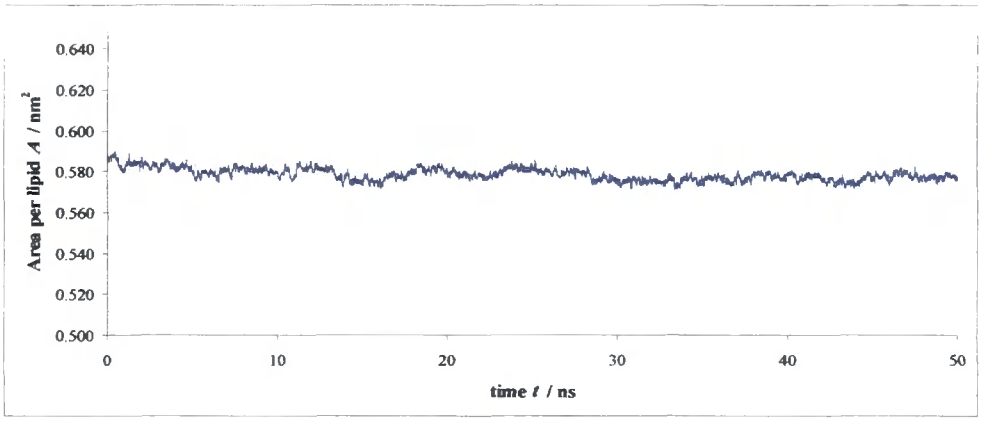


Figure 2.48: Area per lipid development for a 50 ns simulation of the bilayer in the presence of AcWLWLL peptide (see §5.2.2 for experimental conditions).

2.7.2.b Conformations of AcWLWLL

A conformational analysis of the preferred orientations of the peptide was conducted using the procedure as described in §5.1.4.c. Two distinct major clusters (Figure 2.49) were observed in both, the cluster analysis of a molecular dynamics simulation (30 ns) and the cluster analysis using the Maestro software package.

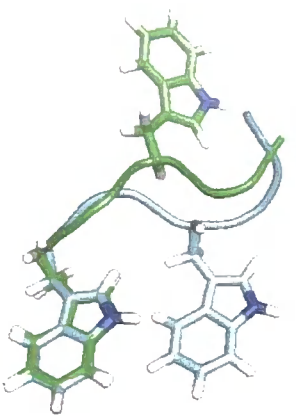


Figure 2.49: Overlay of the two major confomers of AcWLWLL. The confomers were obtained by cluster analysis of the structures of the explicitly solvated peptide structures generated *via* a 30 ns molecular dynamics simulation at 30 °C, 1 bar (§5.2.2). Carbon atoms in confomer 1 are depicted in green, in confomer 2 in blue.

Confomer 1 adopted an N-shaped conformation of the backbone with the two tryptophan sidechains, Trp^1 and Trp^3 being on opposite side to each other. In the other observed conformation, confomer 2, Trp^3 had flipped to the other side (with respect to the orientation of Trp^1) the two were aligned with one another (Figure 2.49).

Influences of the confirmation of the peptide on its binding interactions with the lipid bilayer were evaluated by separated molecular dynamics simulations of both confomers. A snapshot of a simulation containing AcWLWLL can be seen in Figure 2.50.

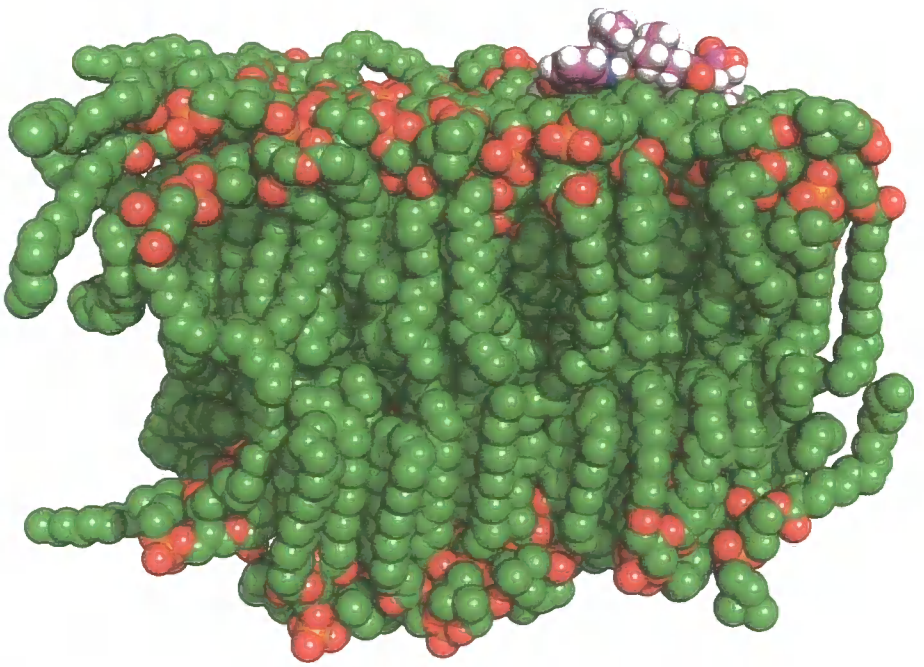


Figure 2.50: Snapshot of a simulation of AcWLWLL interacting with a DMPC bilayer. For clarity, solvent molecules were omitted; lipid molecules are shown as green (united atom CH groups), red (oxygen), orange (phosphor) and blue (nitrogen) spheres that represent the van der Waals radius of the group displayed. AcWLWLL is represented using coloured spheres with van der Waals radii for hydrogen atoms (white), carbon atoms (pink), oxygen (red) and nitrogen (blue). Simulation conditions are described in §5.2.2.

2.7.2.c Peptide Orientation with Respect to the Bilayer

Peptide orientations with respect to the bilayer surface were evaluated by plotting distribution functions, $p(r)$ of backbone atoms to planes through the bilayer which were defined by the average position the phosphate phosphorous atoms P, the choline nitrogens N, and an carbohydrate methyl end group carbons C1N of the leaflet nearest to the peptide respectively (Confomer 2, Figure 2.51, Confomer 1, Figure 2.52).

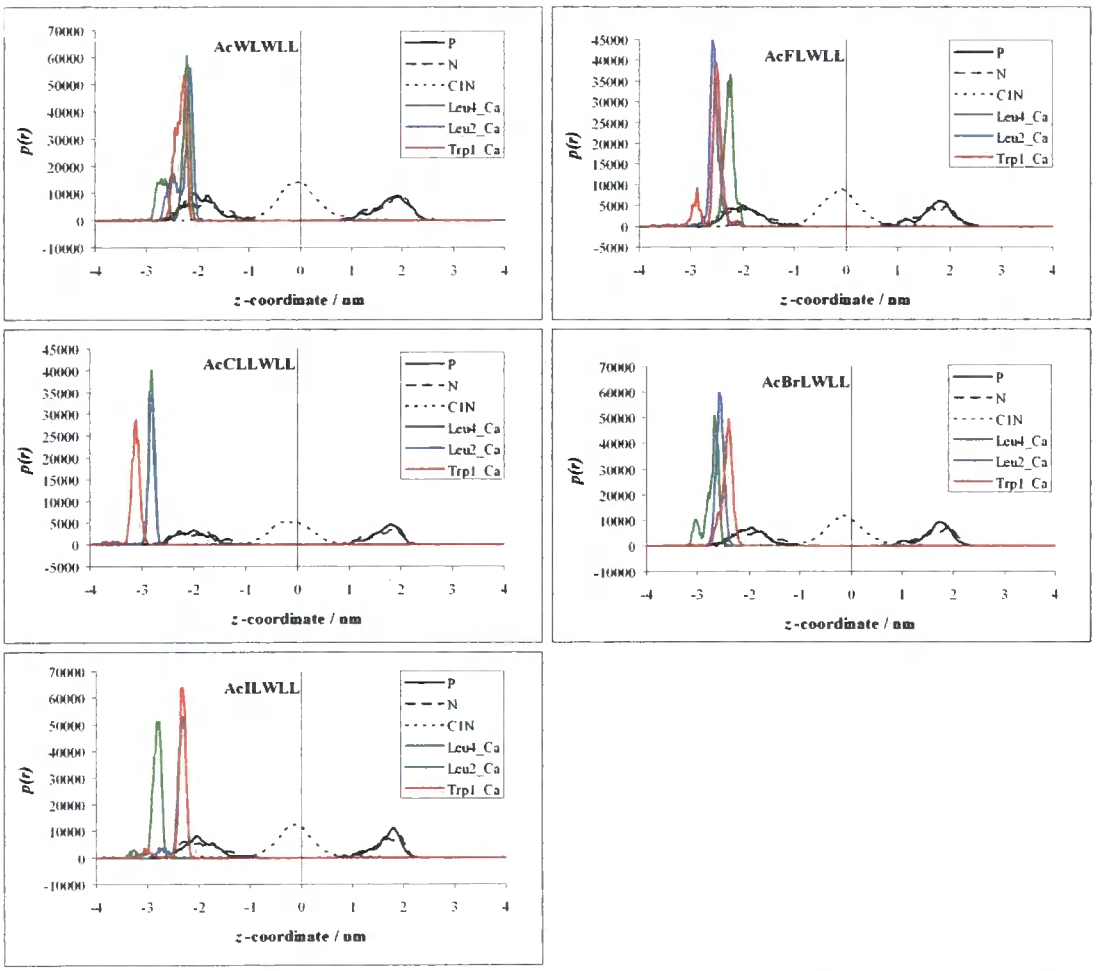


Figure 2.51: Distribution functions for various atoms (C_α of Trp^1, Leu^2 and Leu^4) of the peptide confomer 2 with respect to the lipid bilayer (phosphate P, choline nitrogen N and carbon tail methyl end group C1N). The centre of the bilayer was set to 0 nm. The description AcXLWLL, where X = W, F, Cl, Br, I, refers to substituent on the 5-indole position of Trp^1.

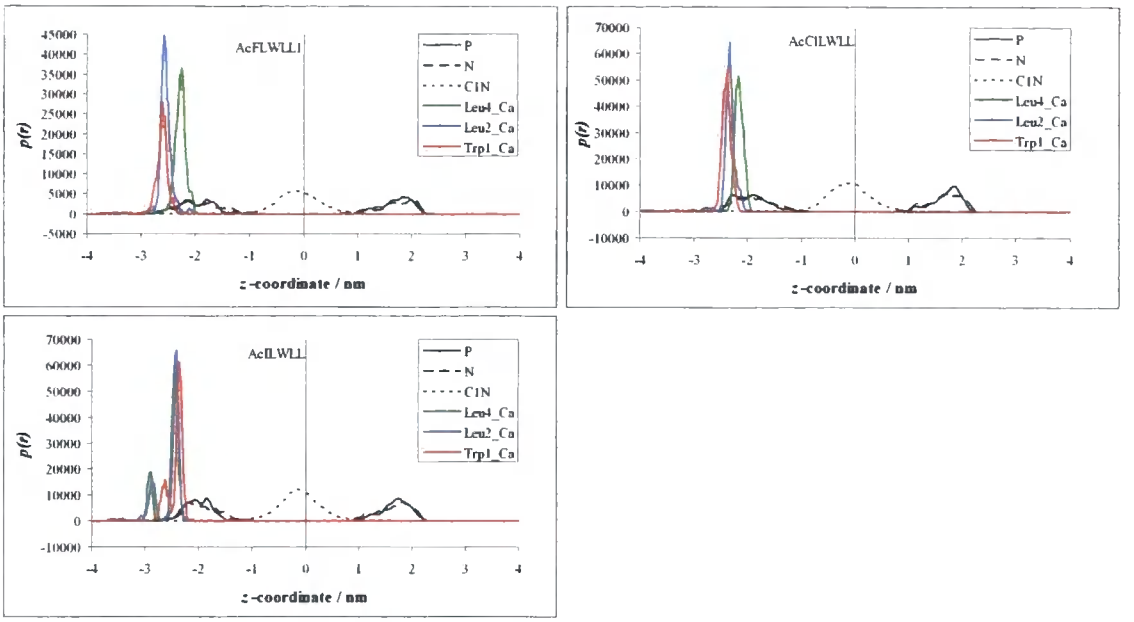


Figure 2.52: Distribution functions for various atoms (C_α of Trp^1, Leu^2 and Leu^4) of the peptide confomer 1 with respect to the lipid bilayer (phosphate P, choline nitrogen N and carbon tail methyl end group C1N). The centre of the bilayer was set to 0 nm. The description AcXLWLL, where X = W, F, Cl, Br, I, refers to substituent on the 5-indole position of Trp^1.

It can be seen, that in most cases (except for the chlorine containing peptide) the peptide is preferably in close inserted in the head group region of the lipid bilayer. AcClLWLL confomer 2, was found preferentially approximately 1 nm away from the peaks in the distribution of some of the lipid head group atoms of the nearest leaflet. This behaviour was exceptional from all the peptides and confomers analyzed and was not seen for the AcClLWLL confomer 1.

The distribution functions of the atoms of some of the peptides (*e.g.* AcWLWLL, AcFLWLL confomers 2, and AcILWLL confomer 1) exhibited two peaks, which is likely to correspond to two preferred locations or orientations of the particular peptide, which might indicate the presence of more than one preferred and stable conformation of the peptide.

The extent of the insertion of the peptides in the bilayer interfacial region was assessed by plotting the minimum distance of the peptide (defined by its minimum *z*-coordinate) to a plane through the nearest leaflet of the bilayer defined the averaged position of the choline nitrogens of the lipid (Figure 2.53).

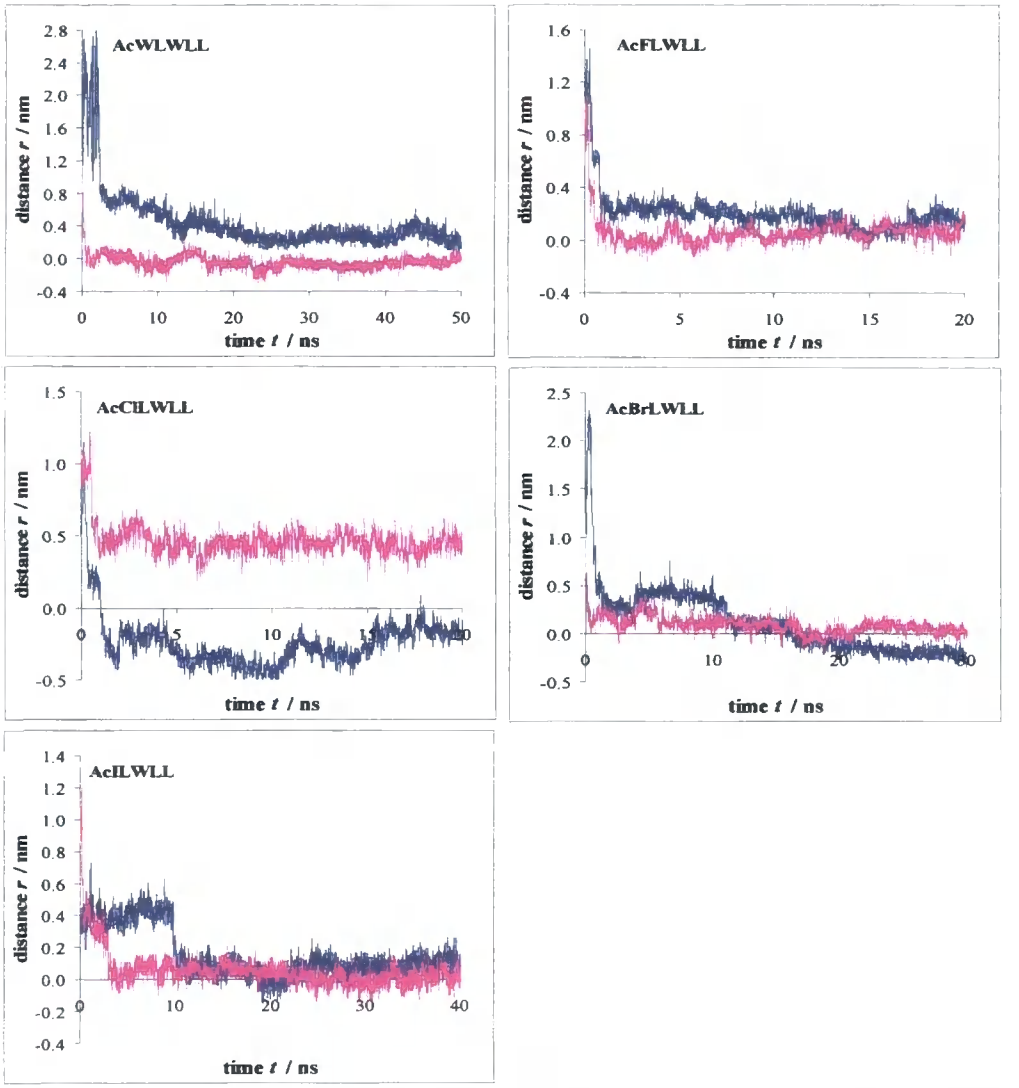


Figure 2.53: Distance from peptide confomers 1 (—) and 2 (—) to lipid bilayer planes defined by the averaged position of the DMPC nitrogen atoms. Distances are displayed as minimum z-coordinate of the peptide to the plane of the nearest leaflet of the bilayer.

The minimum distances between the peptide and the choline head group of the bilayer were found to be very similar between the clusters of a peptide with the exception of AcClLWLL, which had already been noticed in the distribution functions as being an exception to the general trends. The minimum distances of the peptides appear to converge to the same value for both of the confomers of each peptide suggesting that there is a preferred interaction distance between the peptides and the bilayer, which is established over time. This ideal interaction distance was found to approach the average position of the choline head group of the lipid, which could indicate short-range interactions between the peptide and the interfacial lipid atoms. Such favourable interactions could include cation-π

interactions of the tryptophan side chains with the charged choline functionality, hydrophobic effects, and hydrogen bonding interactions between hydrogen bond donors in the peptide (*e.g.* the indole amine) and hydrogen bond acceptors in the lipid molecule (*e.g.* the carbonyl oxygens). A detailed sampling of radial distribution functions of the distances between these atom pairs could elucidate the importance of hydrogen bonding interactions further.

It was further analysed, if the peptide assumes a preferred conformation upon interaction with and inserting in the lipid bilayer. Distances of selected backbone atoms of the peptide (C_α of the residues and the indole nitrogen of the two tryptophans) to a plane through the bilayer which was defined by the average position of the nitrogen atoms respectively were analysed (Figure 2.54).

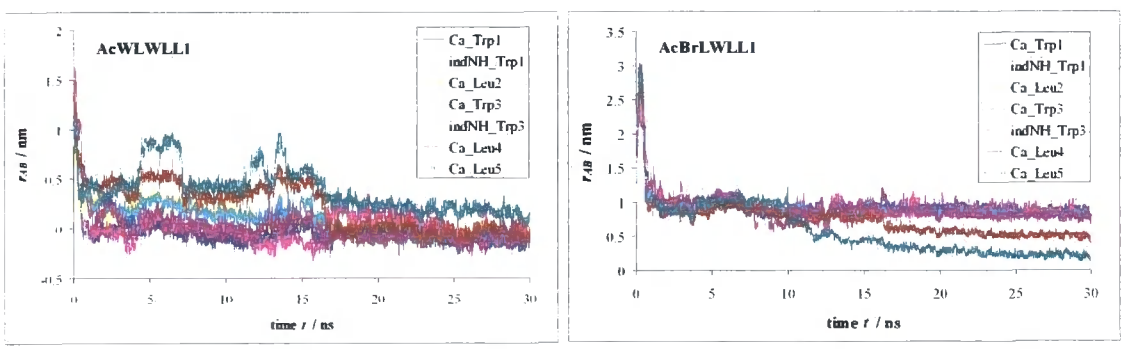


Figure 2.54: Interaction distances between peptide backbone atoms and a plane through the lipid bilayer, which was defined by the average position of the lipid nitrogen atoms. The figure shows the distance profile for the AcWLWLL and AcBrLWLL confomers 1.

The distances between the bilayer and the selected backbone atoms did not vary greatly over time. For the AcWLWLL and AcBrLWLL confomers 1, a slightly increased flexibility was observed for the *C* terminal Leu^4 and Leu^5 residues after about 10 ns was observed, with the residues tending to approach the choline head group region of the bilayer further than the remainder of the peptide which appeared to maintain a constant distance to the plane defined by the average position of the lipid nitrogen atoms (r_{AB} ~0 nm for AcWLWLL1 r_{AB} ~1.0 nm for AcBrLWLL1 respectively).

2.7.2.d Cluster Analysis of Simulated Peptides

The simulated peptides were analysed for the occurrence of clusters in their structures when interacting with the lipid bilayer (Table 2.27, Figure 2.55).

	AcWLWLL		AcFLWLL		AcClLWLL		AcBrLWLL		AcILWLL	
Confomer	1	2	1	2	1	2	1	2	1	2
Cluster1 (%)	13.0	93.7	89.1	15.5	97.4	15.2	26.7	14.6	22.8	3.9
Cluster2 (%)	82.2			81.9		79.2	66.6	84.7	10.7	95.9
Cluster3 (%)									63.8	
sum (%)	95.2	93.7	89.1	97.4	97.4	94.4	93.3	99.3	97.3	99.8

Table 2.27: Peptide clusters as observed in molecular dynamics simulations of AcW*LWLL peptides and DMPC bilayer in water. Parameters for the simulation are given in §5.2.2. Cluster1-3 refer to the occurrence of a particular cluster of the peptide in time and do not indicate that the clusters between the AcW*LWLL peptides are identical.

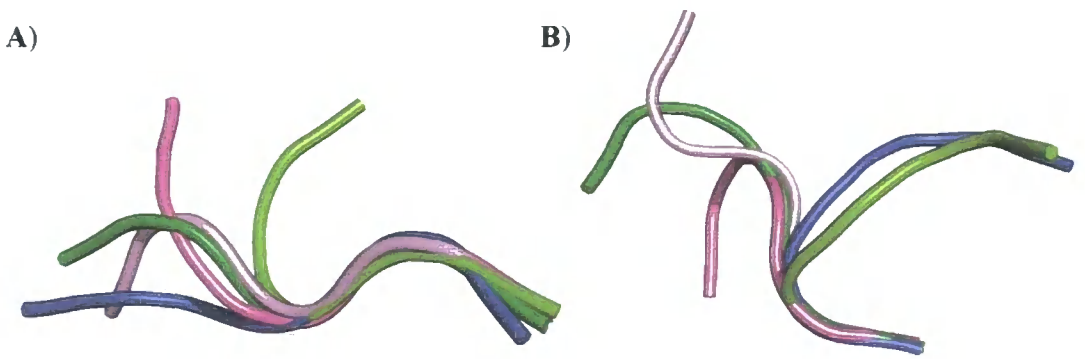


Figure 2.55: Overlay of peptide clusters. **A** and **B** show the overlay of the most populated observed peptide clusters for AcXLWLL (confomer 1 in **A** and confomer 2 in **B**), where X corresponds to a tryptophan residue Trp^1 carrying a monosubstitution R on position 5 of the indole ring. The colour code is the same for both pictures with dark green R = H; blue R = F, pink R = Cl, light green R = Br, and purple R = I. Clusters were superpositioned using the Trp^1 as a reference point.

Confomer 1 of the peptide appears to be well-conserved throughout the homologous series, while confomer 2 displays greater variance. Confomer 2 displayed a greater flexibility particularly in the backbone bonds between residue Leu^2 and Trp^3. A change from confomer 1 into 2 was not seen during the 50 ns of simulation. The peptides appear to mostly preserve their conformation upon interaction with and insertion into the bilayer. The observed clusters accounted for >93% of all the sampled structures indicating that the peptide adopts a stable conformation quickly and efficiently. Minor clusters for some of the peptides analysed were observed during the initial stage of the simulation (t < 5 nm). This is consistent with the observation of minor, secondary peaks in the distribution function of the peptides. (§2.7.2.c).

2.8 References

[1] Du Vigneaud, V., Meyer, C. E., *J. Biol. Chem.*, **1933**, *99*, 143-151
[2] Synder, H. R., MacDonald, J. A., *J. Am. Chem. Soc.*, **1955**, *77*, 1257-1259
[3] Yokoyama, Y.; Hikawa, H., Mitsuhashi, M., Uyama, A., Hiroki, Y., Murakami, Y., *Eur. J. Org. Chem.*, **2004**, *6*, 1244-1253
[4] Reid, G. E., Simpson, R. J., O'Hair, R. A. J., *J. Am. Soc. Mass Spectrom.*, **2000**, *11*, 1047-1060
[5] Yokoyama, Y., Hikawa, H., Mitsuhashi, M., Uyama, A., Hiroki, Y., Murakami, Y., *Eur. J. Org. Chem.*, **2004**, *6*, 1244-1253
[6] Topgi, R., S., Ng., J. S., Landis, B., Wang, P., Behling, J. R., *Bioorg. Med. Chem.*, **1999**, *7*, 2221-2229
[7] Baldaro, E., Fuganti, C., Servi, S., Tagliani, A., Terreni, M. In *Microbial Reagents in Organic Syntheis* Kluwer Academic Publishers, **1992**, pp175
[8] Kondo-Yamada, Y., Okada, C., Yashida, K., Umeda, Y., Arima, S., Sato, N., Kai, T., Takayanagi, H., Harigaya, Y., *Tetrahedron*, **2002**, *58*, 7851-7861
[9] Haque, R., Tinsley, I. J., Schmedding, D., *J. Biol. Chem.*, **1972**, *247*(1), 157-161
[10] Sanderson, J. M., Whelan, E. J., *Phys. Chem. Chem. Phys.*, **2004**, *6*, 1012-1017
[11] Gross, K.C., Seybold, P.G., *J. Org. Chem.*, **2001**, *66*, 6919-6925
[12] a) Schmuck, C., *Chem. Comm.*, **1999**, 843-844
b) Simova, S., Schneider, H. J., *J. Chem. Soc. Perkin Trans. 2*, **2000**, *8*, 1717-1722
[13] Dalton, L. K., Teitei, T., *Aust. J. Chem.*, **1968**, *21*, 2053-2058
[14] Demmers, J. A. A., Haverkamp, J., Heck, A. J. R., Koeppe, R. E., Killian, J. A., *Proc. Nat. Acad. Sci. USA.*, **2000**, *97*(7), 3189-3194
[15] Gaussian 03, Revision 1, Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, Jr., J. A.; Vreven, T.; Kudin, K. N.; Burant, J. C.; Millam, J. M.; Iyengar, S. S.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.; Scalmani, G.; Rega, N.; Petersson, G. A.; Nakatsuji, H.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Klene, M.; Li, X.; Knox, J. E.; Hratchian, H. P.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Ayala, P. Y.; Morokuma, K.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Zakrzewski, V. G.; Dapprich, S.; Daniels, A. D.; Strain, M. C.; Farkas, O.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Ortiz, J. V.; Cui, Q.; Baboul, A. G.; Clifford, S.; Cioslowski, J.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Gonzalez, C.; and Pople, J. A.; Gaussian, Inc., Wallingford CT, 2004
[16] Press, W. H., Flannery, B. P., Teukolsky, S. A., Vetterling, W. T. in *Numerical Recipes in Pascal: The Art of Scientific Computing* , **1989**, Cambridge University Press, Cambridge, MA
[17] Arnaud, A., Bouteiller, L., *Langmuir*, **2004**, *20*, 6858-6863
[18] Jadzyn, J., Stockhausen, M., Zywucki, B., *J. Phys. Chem.*, **1987**, *91*, 754-757
[19] ITC data analysis in Origin: Tutorial Guide, MicroCal, Northhampton, MA., **1998**, p.73
[20] Blandamer, M. J., Butt, M. D., Cullis, P. M., *Thermochim. Acta*, **1992**, *211*, 49-60
[21] Schrodinger, Inc.: Portland, OR, 1991-2000
[22] Hünenberger, P. H., McCammon, J. A., *Biophys. Chem.*, **1999**, *78*(1-2), 68-88
[23] MOPAC2007, James J. P. Stewart, Stewart Computational Chemistry, Colorado Springs, CO, USA, http://OpenMOPAC.net
[24] Stewart, J. J. P., *J. Comp. Chem.*, **1989**, *10*, 221-264
[25] Gilson, M. K., Given, J. A., Bush, B. L., McCammon, J. A., *Biophys.* J., **1997**, *72*, 1047
[26] Zhang, Y., McCammon, J.A., *J. Chem. Phys.*, **2003**, *118*(4), 1821-1827
[27] Ding, L, Zhang, X.X., Chang, W.B., Lin, W., Yang, M., *Anal. Chim. Acta*, **2005**, *543*(1-2), 249-253
[28] For examples see: a) Xu, Z., Luo, H., Tieleman, P., *J. Comput. Chem.*, **2007**, *28*, 689-697
b) MacCallum, J. L., Tieleman, D. P., *J. Comput. Chem.*, **2003**, *24*, 1930-1935
[29] Villa, A., Mark, A. E., *J. Comput. Chem.*, **2002**, *23*, 548-553
[30] Tironi, I. G., van Gunsteren, W. F., *Mol. Phys.*, **1994**, *83*(2), 381-403
[31] Radzicka, A., Wolfenden, R., *Biochemistry*, **1988**, *27*(5), 1664-1670
[32] Daura, X., Hünenberger, P. H., Mark, A. E., Querol, E., Avilés, F. X., van Gunsteren, W. F., *J. Am. Chem. Soc.*, **1996**, *118*, 6285-6294
[33] a) Gilson, M. K., Given, J. A., Bush, B. L., McCammon, J. A., *Biophys. J.*, **1997**, *72*, 1047–1069

b) Boresch, S., Tettinger, F. , Leitgeb, M., Karplus, M., *J. Phys. Chem. B*, **2003**, *107*, 9535–9551
c) Woo, H. J., Roux, B., *Proc. Natl. Acad. Sci. USA.*, **2005**, *102*, 6825–6830
d) Roux, B., Nina, M., Pomes, R., Smith, J. C., *Biophys. J.*, **1996**, *71*, 670–681
e) Hermans, J., Wang, L., *J. Am. Chem. Soc.*, **1997**, *119*, 2707–2714
[34] Lis, L.J., McAlister, M., Fuller, N., Rand, R.P., Parsegian, V.A., *Biophys. J.*, **1982**, *37*, 657-666
[35] Rand, R. P., Parsegian, V. A., *Biochim. Biophys.* Acta, **1989**, *988*, 351-376
[36] Nagle, J.F., Tristram-Nagle, S., *Biochim. Biophys. Acta*, **2000**, *1469*(3) , 159-195
[37] Zubrzycki, I. Z., Xu, Y., Madrid, M., Tang, P., *J. Chem. Phys.*, **2000**, *112*(7), 3437-3441
[38] Gurtovenko, A. A., Patra, M., Karttunen, M., Vattulainen, I., *Biophys. J.*, **2004**, *86*, 3461-3472
[39] Lindahl, E., Edholm, O., *Biophys. J.*, **2000**, *79*, 426-433
[40] Marrink, S.J., Mark, A.E., *J. Chem. Phys. B.*, **2001**, *105*, 6122-6127
[41] Anézo, C., de Vries, A. H., Höltje, H.-D., Tieleman, D. P., Marrink, S.-J., *J. Phys. Chem. B*, **2003**, *107*, 9424-9433
[42] Cevc, G., Marsh, D., in Phospholipid Bilayers: Physical Principles and Models, John Wiley & Sons, New York, **1987**
[43] a) Patra, M., Karttunen, M., Hyvönen, M. T., Falck, E., Vattulainen, I., *J. Phys. Chem. B*, **2004**, *108*(14), 4485-4494
b) Douliez, J.-P., Léonard, A., Dufourc, E. J., *J. Phys. Chem.*, **1996**, *100*, 18450-18457
[44] Berger, O., Edholm, O., Jahnig, F., *Biophys. J.*, **1997**, *72*(5), 2002-2013
[45] a) Berendsen, H. J. C., van der Spoel, D. Van Drunen, R., *Comput. Phys. Commun*, **1995**, *91*, 43-56
b) Lindagl, E., Hess, B., van der Spoel, D., *J. Mol. Modell.*, **2001**, *7*, 306-317
[46] Gurtovenko, A. A., Patra, M., Karttunen, M., Vattulainen, I., *Biophys. J.*, **2004**, *86*, 3461-3472

Chapter 3: Summary and Discussion of Results

3.1 Introduction and Overview

A facile synthesis for tryptophan compounds was employed to produce a wide range of tryptophan derivatives in good yield and optical purity. Indole derivatives were converted to N^{α}-acetyl tryptophan amino acids (**1**) using *L*-serine. The racemic products were then subjected to acylase resolution, generating *L*-tryptophan (**2**) and N^{α}-acetyl-*D*-tryptophan derivatives (**2'**). The *L*-amino acids were then converted to N^{α}-acetyl ethyl amides (**6**) to mimic non-terminal tryptophan residues in proteins and peptides (§2.2.6). The binding of these N^{α}-acetyl ethyl amides (**6a-g**, **6n**) to DMPC molecules was studied by NMR host-guest titrations and isothermal titration calorimetry, and modelled *via* molecular dynamics simulations.

The following sections will summarize and discuss the key findings of this thesis. The use of the binding model employed for the fitting of the experimental data obtained *via* NMR host-guest titrations is evaluated in (§3.2). The importance of hydrogen bonding (§3.3.1) and cation-π interactions (§3.3.2) in Trp:DMPC and DMPC:water complexes is reviewed. Subsequently, the experimentally and computationally obtained binding constants and associated free energies of adduct formation are discussed (§3.4). Data acquired by isothermal titration calorimetry and free energy calculations are then compared with the results from the other techniques for the analysis of peptide-lipid analysis (§3.5 and §3.6 respectively). §3.7 presents a discussion of the results obtained from molecular dynamics simulations of the interactions of AcWLWLL type peptides with a DMPC bilayer in water.

3.2 Discussion of the Binding Model Used for the Analysis of NMR Titration Data

NMR titrations gave detailed information about the micellisation of DMPC in chloroform and could be used to estimate the critical micelle concentration (CMC) of DMPC to be 19.71 ± 3.28 m*M* at 21 °C (§2.3.2.b), which was in good agreement with the previously reported value of 16 m*M* at 24 °C.[9] ITC measurements of the hydrated DMPC into chloroform supported the finding that the CMC is >15 mM at 21 °C (§2.4.2.b). The precise CMC was not determined due to the limitations placed by the experimental set-up. Given the

value determined for the CMC, the injected samples of DMPC in both the NMR host-guest and the isothermal calorimetry titrations were at concentrations above the CMC. This made the inclusion of the self-association (and hence by default also dissociation) of lipid molecules in the binding model desirable.

Job plots of selected tryptophan derivatives (**6f**, R = MeO representing the most electron dense and **6g**, R = NO_2 representing the most electron deficient compounds analysed) determined a preferred association stoichiometry of the interaction between the amino acid analogues and DMPC in chloroform to be approximately 2:1 (§2.3.3.b). This was found to be in good agreement of a previous study that clearly showed N^{α}-acetyl-*L*-tryptophan ethyl amide (**6n**) to form 2:1 adducts with 1,2-diacetyl-*sn*-3-phosphocholine (DAPC) in chloroform.[1]

Water also played an important role for binding interactions as observed by the significant chemical shift of the water proton signal during the course of the NMR guest-host titrations. During the course of a titration, the value shifts towards that of the DMPC:water adduct. This is expected, as by the end of the titration the concentrations of both were significantly higher than that of tryptophan. Given these relative concentrations and the observation that the water chemical shift is close to that for water in $CDCl_3$ at the start of the titrations (*i.e.* in the absence of DMPC, with tryptophan only), water interactions are more significant with DMPC than with tryptophan. This justifies not considering tryptophan:water interactions. It was clearly demonstrated that water in the sample strongly influenced binding interactions and needed to be accounted for in the model. Previously, hydration studies on the closely related 1,2-dipalmitoyl-*sn*-glycero-phosphocholine (DPPC) lipid in chloroform *via* NMR indicated that, contrary to the aggregated DPPC, the monomers of DPPC do not appear to be hydrated.[2] This is supported by the intense observed heat changes in the ITC experiments upon injection of low concentrations of hydrated DMPC into chloroform that are likely to arise from the endothermic unbinding of water from the lipid. Further evidence for the lack of association between lipid monomer and water could be seen in the molecular dynamics simulations of a system containing water, tryptophan derivative (**6a**) and lipid at equimolar concentrations. No binding interactions between the water molecule and the lipid were observed, while association interactions of water to the amino acid analogue were noticed.

The model employed for the fitting of the data to a binding isotherm accounted therefore for lipid self-aggregation and micelle formation, Trp:DMPC 1:1 and 2:1 adducts,

and hydrated species (Water:DMPC 1:1 and 2:1 and Trp:DMPC:water 1:1:1). The corresponding binding constants could be determined using NMR host-guest titrations of hydrated DMPC into chloroform and titrations of the lipid into chloroform solutions of the tryptophan derivatives. Observed chemical shifts of up to seven key protons were plotted against lipid concentration and multiple, simultaneous non-linear fitting of the data allowed the calculation of a binding isotherm. The fitting of a binding isotherm to the experimental data of the NMR titration experiments was found to give best results when an isodesmic model was assumed (*i.e.* the binding constant for lipid-lipid dimerisation $K_d = K_m$, the binding constant for micellisation).

The quality and accuracy of the model could be assessed by the accuracy and reproducibility of the obtained fits and was found to be working extremely well (observed error <8% between repeat runs). Water equilibria were accounted for with high confidence and the observed complexation-induced chemical shift change of the water proton signals could be reproduced within 3% of the experimentally determined values.

3.3 Factors Important for Lipid-Tryptophan Associations

3.3.1 Hydrogen Bonding Interactions

Proton/deuterium exchange experiments pointed out the labile hydrogen atoms in the tryptophan derivatives and provided indications to the relative strength of hydrogen bonding interaction in Trp:DMPC and Trp:water:DMPC adducts. The indole amine hydrogen and the acetyl and ethyl amide protons were determined as being susceptible to deuterium exchange under the experimental conditions used. The data suggest an order of accessibility and availability for the exchanged protons to be indNH > AcNH >> EtNH. This trend could also be related to the involvement of these protons in hydrogen bonding interactions.

In agreement, Bordwell *et al.* determined the pKa for the dissociation of an amide proton for a related compound (*N*-methylacetamide) to be 25.9 in DMSO[3] and the pKa for the dissociation of the indole amine hydrogen as 21.0 in DMSO[4]. This indicates that the amide protons are slightly less likely to exchange for deuterium than the indole amine proton. Additionally, indole amine protons have previously been shown to exhibit particular weak hydrogen bonding interactions to water in measurements of the free energy of transfer of tryptophan analogues from water into cyclohexane.[5] It was also reported that there are no

significant hydrogen bonding interactions between the indole NH and lipid molecules.[6] This was demonstrated by comparing the ordering of indole and *N*-methylindole in both ether and ester lipid bilayers, which was found to be identical with experimental error of the linear dichroism and NMR studies that gave rise to these results.[7]

Further evidence of the lack of hydrogen bonding interactions of the indole amine proton to both water and lipid molecules was shown in the molecular dynamics studies of the system containing water, lipid and tryptophan derivative at equimolar concentrations. No hydrogen bonding interactions between the indole NH and any of the lipid hydrogen bond acceptors nor any significant interactions to the water oxygen were detected. Although, the time averaged radial distribution function, $\langle g(r) \rangle_t$, for the distance between the indole amine proton and the water oxygen (OW) exhibited a peak in the distribution at r_{AB} ~0.20 nm, no such interactions were seen in the most populated clusters of the water-Trp interactions. The peak is likely to be due to weak association of the water molecule to the indole NH as suggested by Ref. 5. In contrast, defined hydrogen bonding interactions were seen for the ethyl amide proton (EtNH) with the water oxygen both in the $\langle g(r) \rangle_t$ for the distances between OW and EtNH and upon visual inspection of the clusters observed.

In the D_2O experiments, the hydrogen of the acetyl amide functionality exchanged with an approximate half-life of $t_{½}$ = 40 min. This is consistent with the strong preference for hydrogen bonding interactions with water as observed in the simulations (§2.5.3.b), which would slow an exchange of the proton down. The exchange of the ethyl amide proton was found to be very slow ($t_{½}$ = 200 min) suggesting that it is involved in strong hydrogen bonding interactions. No such hydrogen bonding interactions were seen in the molecular dynamics simulations with water. However, the D_2O experiments followed the averaged hydrogen signals of bound tryptophan compounds. As the preferred stoichiometry of the Trp:lipid complex was found to be 2:1 (§2.3.3.b), there might be hydrogen bonding interactions between the two tryptophan derivatives in the complex that prevent a fast deuterium exchange. Indeed, the simulation of amino acid-lipid interactions at a ratio of 2:1 (§2.5.4) showed a significant hydrogen bonding interaction between the ethyl amide proton and the acetyl amide carbonyl oxygen of a nearby tryptophan derivative (**6a**) (§2.5.4.c).

Hydrogen bonding between two amino acids derivatives (**6a**) were shown to be a significant driving force of the formation of a Trp:Trp 1:1 complex (§2.5.4.c), appearing at the same time when complex formation was observed (Figure 3.1). A very stable complex was formed, that exhibited 2 hydrogen bonds on average. These could be related to time

averaged radial distribution functions of hydrogen bond acceptors and donors in the respective tryptophan derivatives.

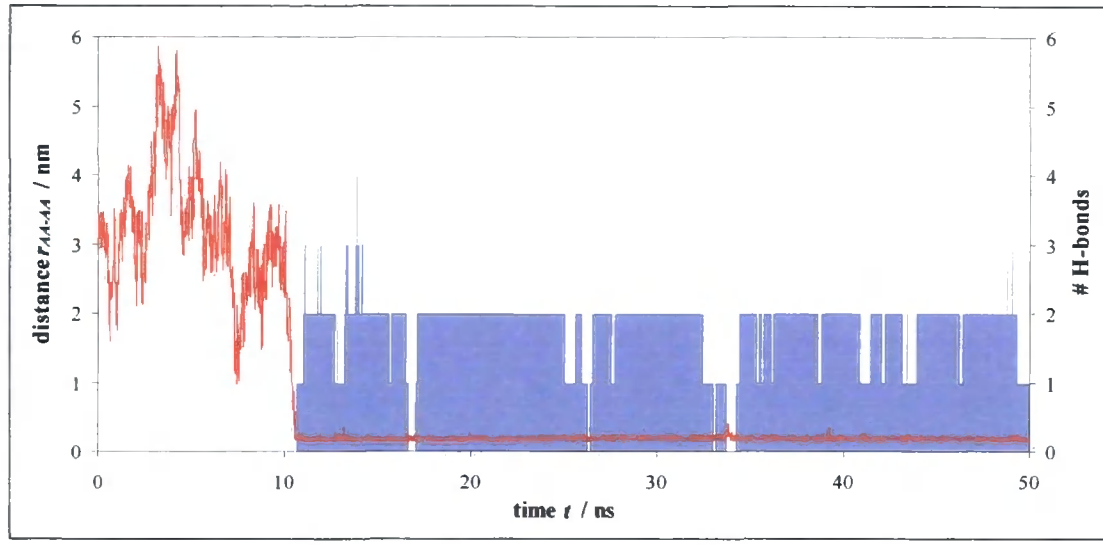


Figure 3.1: Distance between the two centers of mass of two tryptophan derivatives (**6a**) overlaid with a histogram of hydrogen bonds being established between them.

No hydrogen bonding interactions between the lipid and any of the tryptophan derivatives were observed at any time during the simulation. This indicates, that for the 2:1 and the 1:1 Trp:DMPC complexes, hydrogen bonding interactions to the lipid play a minor role.

Interestingly, the self-association of the tryptophan derivatives (**6a**) was only seen in the presence of DMPC. In a control simulation with an identical starting configuration of the two amino acid derivatives but in the absence of the lipid molecules did not show any significant association between the two amino acid analogues. During >95% of the simulations the molecules seemingly diffused freely in the chloroform simulation box, and only at the end of the 50 ns simulation was association between the two molecules observed. This suggests that the lipid molecule enhances the formation of the Trp:Trp 1:1 complex, potentially by providing a local environment that fosters the establishment of hydrogen bonding interactions to each other similar to a nucleation site. However, this result is to be seen with caution due to the restricted analysis time frames. Longer simulations of the control run will be necessary to substantiate the observations made. Furthermore, due to the long-lasting nature of Trp:Trp interactions the analysis was restricted to a qualitative assessment.

In conclusion, hydrogen bonding interactions are likely not significantly important for the interactions in 1:1 Trp:DMPC complexes. Accordingly, hydrogen bonding interactions between single tryptophan residues and lipid molecules in the bilayer might be less significant. This is consistent with observations made in solid-state NMR, where the hydrogen bonding interactions of WALP peptides to DMPC lipid carbonyl groups were found to be of minor importance as the orientation of the peptide were found to be near identical to orientation adopted in the DMPC analogue ditetradecylphosphatidyl-choline (DTPC) that misses the carbonyl linker.[8]

In higher order complexes of the amino acid derivatives, hydrogen bonding interactions were seen to have a significant contribution to binding interactions between the tryptophan analogues. The importance of the presence of DMPC in the establishment of hydrogen bonds in the Trp:Trp 1:1 complex remains to be investigated further.

3.3.2 Cation-π interactions

The importance of cation-π interactions for the preferred localisation of tryptophan residues and analogues of the aromatic amino acid is still debated.

Norman and Nymeyer[9] reported that both hydrogen bonding to lipid carbonyl groups and cation-π and dipolar interactions between the indole and the POPC bilayer played a role in the its preferred localisation at the water/lipid interface. Further support for the involvement of cation-π interactions in the localization of indole in the choline region was given by Gaede *et al.*[10] Cation-π interactions for interfacial tryptophan residues were also reported to depend on both, the tryptophan position relative to the bilayer-water interface, and on the chemical nature of the lipid headgroups seen in molecular dynamics simulations, where ethanolamine headgroups (POPE) were favoured over choline headgroups (POPC).[11]

The strength of the cation-π interaction to a lipid choline head group has been estimated to be ~8.4 kJ mol^{-1} by studying the effects of substitutions on tryptophans indole ring and its implications for the ligand recognition by serotonergic (5-HT3A) and nicotinic acetylcholine receptors in water.[12] Others estimated the strength of cation-π interactions to be ~23.0 kJ mol^{-1} in water using *ab initio* calculations.[13] A typical cut-off for cation-π interactions is a distance of 0.5 nm between the centre of the indole ring and a charged atom.[14] However, due to the relative orientation to each other, this value might vary by

±0.2 nm to include distances between the charged atom and the nearest atom of the aromatic ring. A truncation at 0.7 nm has also been used before by Petersen *et al.*[11]

In contrast in a recent study using linear dichroism spectroscopy, cation-π interactions with lipid head groups (POPC and POPC/POPE) were reported to be less important and that localisation of indole derivatives (indole, 3-methylindole, 1-methylindole and tryptophan octyl ester) instead would be directed by dipole-dipole interactions and steric constraints in the membrane hydrocarbon region.[15]

3.3.2.a Cation-π Interactions in Trp:DMPC complexes

The importance of cation-π interactions between the tryptophan derivatives and the lipid molecules were evaluated using molecular dynamics simulation of the compounds dissolved in chloroform. The strengths of the cation-π interactions are strongly dependent on the dielectric constant of the medium, as would be expected for interactions with large electrostatic components, but cation-π interactions were reported to maintain their strength across a range of solvents.[13] Therefore, conclusions drawn from the data obtained in the molecular dynamics simulations in chloroform, are comparable to information gained about cation-π interactions in aqueous media.

Cation-π interactions between tryptophan derivatives and the lipid were evaluated by analysis of the time averaged radial distribution functions $\langle g(r) \rangle_t$ for the distance between the charged lipid choline nitrogen atom and the centre of mass of the indole side chain of the amino acid analogue.

All of the tryptophan derivatives analysed exhibited a peak in the $\langle g(r) \rangle_t$ for this intermolecular distance with a maximum ~0.6 nm (Figure 3.2). The interaction distance did not change significantly with varying electron density of the indole ring, but slight deviations (±0.2 nm) of the position the maximum peaks were observed. Preferred binding interactions could not be visualized directly *via* cluster analysis of the Trp:DMPC 1:1 complex as the cluster populations for this complex were found to be small (<19%), which complicated a statistically relevant analysis of preferred binding conformations. The peak observed in the $\langle g(r) \rangle_t$, however, is likely to arise from a cation-π interaction between the indole ring and the choline head group.

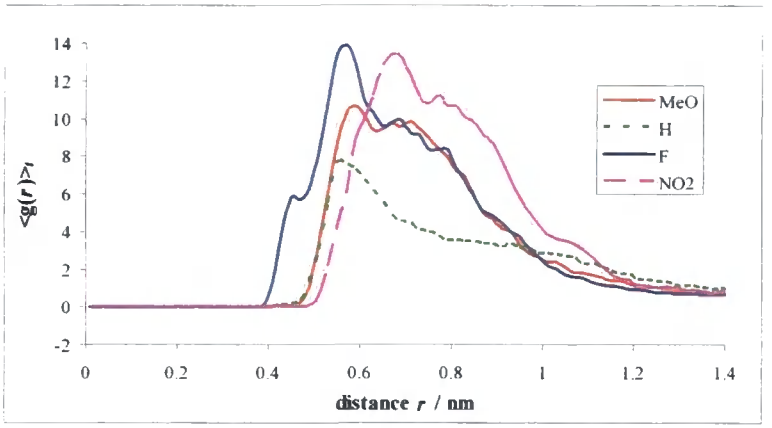


Figure 3.2: Time averaged radial distribution functions of the distance between the DMPC choline nitrogen and the centre of mass of the indole ring of a selection of tryptophan derivatives (**6a**, **6f**, **6g** and **6n**).

The $<g(r)>_t$ for the distance between the centre of mass of the indole ring of one of the two tryptophan derivatives and the lipid choline nitrogen in the 2:1 Trp:DMPC complex displayed a similar pattern to the 1:1 Trp:DMPC complex with a broad peak at r_{AB} ~0.6 nm (§2.5.4.b, Figure 2.41). This is likely to be indicative of a cation-π interaction between the indole ring and the choline head group. The cluster analysis visualised the interaction distance to some extend showing the choline head group in proximity to the aromatic ring of the nearest tryptophan compound (Figure 2.43, p.126).

This adds further evidence to the observation that cation-π interactions might be involved in tryptophan-lipid interactions. Only minor changes in the preferred distance of the peaks in the time averaged radial distribution function and its relative intensity were observed for the varying tryptophan derivatives indicating that the change in electron density on the indole ring did not alter cation-π interactions between the lipid choline head group and the indole side chain significantly.

In conclusion, cation-π interactions between the DMPC and tryptophan derivatives were observed. Interestingly, their strength (as estimated from the intensity of radial distribution functions and minimum interaction distances between the two groups) was found to be only loosely associated with the electron density of the indole ring. The occurrence of lipid-tryptophan cation-π interactions has been suggested to be a contributing factor for the preferred localisation and interaction of tryptophan residues in and with lipid bilayers. This study adds further evidence to this observation adding that the electron density of the indole ring system appears to have only minor influences on strength and average distance of these interactions.

3.4 Discussion of Determined Binding Constants for Adduct Formation

Using titrations of hydrated DMPC into chloroform (§2.3.2) and the binding model detailed in §2.3.2.c, binding constants for the association of water to DMPC molecules could be calculated. The constants of the water:lipid 1:1 and 2:1 adducts could be approximated as $K_{1wl} = 31.86 \pm 0.11\ M^{-1}$ and $K_{2wl} = 121.80 \pm 1.47\ M^{-1}$ respectively. Additionally, the binding constant for the micellisation of DMPC molecules could be determined as $K_m = 144.06 \pm 0.86\ M^{-1}$. These values were used as constants in the calculation of the binding constants of tryptophan derivatives to DMPC molecules and DMPC-water complexes. The binding constants of the sequential addition of water and tryptophan derivatives to the DMPC complexes (K_1, K_2, K_1', K_{1w}') could be established.

3.4.1 Free Energies of Association of the 1:1 Trp:DMPC Adduct

3.4.1.a Comparison between Experimentally Derived and Computed ΔG_1 Values.

Binding constants and associated binding free energies for the Trp:DMPC 1:1 could be calculated using two different approaches (§2.5.2.a). The importance of electrostatic interactions could be evaluated by using two different charge sets (Set **A** and **B**) for the representation of partial charges on the indole side chain (Table 2.15, Table 2.16, p.92). These charge sets had been derived either by fitting the electrostatic potential obtained *via* restricted Hartree-Fock *ab initio* calculations or by Møller-Plesset type MP2 perturbation theory. This allowed a direct evaluation of the influence of partial charges and hence electrostatic interactions for association processes.

The Hammett parameter could be used as an approximation of the electron density of substituted indole rings. When observed and calculated free energies for the Trp:DMCP 1:1 complex formation were plotted against the Hammett parameter, an n-shaped (parabolic) trend was displayed (see §2.5.2.e).

The experimentally determined and computed values of the association free energy for the 1:1 Trp:DMPC were found to be in excellent agreement. Both treatments of the partial charges (using Set **A** and **B** respectively, §Table 2.15 and Table 2.16, p.93) gave comparable results, with the models employing Set **B** were generally representing the experimental data and the n-shaped trend better. The discrepancy of the rHF method to experimentally obtained values was particularly pronounced for the electron dense and deficient tryptophan

compounds (**6f** and **6g** respectively). The treatment of partial charges was seen to directly influence the calculated free energies of binding suggesting that electrostatic interactions are important for their correct representation. The differences in the carbon atoms of the indole were significant (±0.2 e) between the two sets.

Set **B** represented the experimentally observed pattern for the electron rich compounds better. The extension of the treatment of long-range interactions from a truncation value from 1.2 nm to 1.9 nm did not lead to any noticeable improvement. Values derived from both methodologies were identical within experimental error. When binding free energies were derived using the survival function of the complex (MP2b2), the binding was not slightly overestimated but instead underestimated by approximately 1.92 ± 0.59 kJ mol^{-1} with respect to the experimental values. The same pattern as for the free energies determined by both MP2a and MP2b was observed for MP2b2 derived values.

Calculated free energies of association obtained from molecular dynamics simulations using the partial charge set that was derived by Hartree-Fock calculations (Set **A**) was found to imitate the general pattern for the association free energies seen in NMR host-guest titrations very well for tryptophan derivatives with moderate alterations of the electron density of the indole ring (**6a-e**, **6n**). However, electron rich or poor systems (**6f**, **6g**) were approximated less well.

The brute force approach of calculating binding constants from the ratio of bound *versus* unbound states represented experimental data better than the methodology employing the survival function of the complex. The long molecular dynamics simulation allowed a sufficient sampling of the conformational space accessible to the compounds involved in complex formation. The occurrence of many binding and unbinding events with varying durations gave further evidence of a statistically significant sampling, justified the implementation of the calculation of the binding constants *via* the ratio of bound and unbound states. The methodology using the survival function of the complex could be improved by consideration of re-crossing events and better defined cut-off values for the distances that determine when a state can be considered as being bound or unbound.[16] This was not attempted here as the method using the ratio of bound and unbound states was in excellent agreement with the experimental results and no further improvement was seen to be necessary.

3.4.1.b The n-Shaped Pattern of ΔG_1 versus Hammett Parameter Plots

When observed and calculated free energies for the Trp:DMCP 1:1 complex formation were plotted against the Hammett parameter, an n-shaped trend was displayed (**Error! Reference source not found.**). Binding interactions were found to be enhanced regardless of the substituent and electron density of the indole ring when compared to the unsubstituted tryptophan derivative (**6n**). The non-linear shape of the binding free energies ΔG_1 with respect to the electron density on the indole ring of the compound analysed indicates that binding interactions are likely to follow more complex, second-order electrostatic interactions. It is interesting to note that the free energy of binding decreased with increasing electron density from fluorine to iodine, reflecting a trend for more favourable binding for more electron-rich halogen derivatives. For electron-releasing substituents (**6e**, **6f**; R = Me, MeO), increased electron density on the indole ring system also led to more favourable binding in free-energy terms.

Other factors that might influence the free energy of association were investigated in an attempt to elucidate the origin for the observed pattern.

1) The molecular dynamics simulations suggested that binding interactions were not influenced by hydrogen bonding interactions between a single tryptophan molecule and lipid.

2) In the non-polar solvent chloroform, additional forces such as the minimisation of solvent accessible surface area (SAS) to avoid unfavourable exposure of polar functional groups could be envisaged. To assess this, the SAS for 3-methylindole 5-monosubstituted derivatives mimicking the tryptophan compounds analysed was calculated (§2.3.6). When plotted against the observed free energies of Trp:DMPC 1:1 adduct formation, no clear trend was observed (§7.4).

3) Similarly, no clear trends were seen for the relation of the determined free energies to the dipole moments or a measure of hydrophobicity of the compounds, log P (which both could be calculated for the 5-monosubstituted 3-methylindole derivatives as well) (§7.4).

3.4.2 Free Energies of Association for the Higher Order Complexes

The binding free energy for the 2:1 Trp:DMPC complex was found to be 4.91 ± 1.02 kJ mol^{-1} more favourable than the corresponding 1:1 adduct formation (ΔG_1) and comparable in size to the free energies of tryptophan:water:lipid 1:1:1 adduct formation (ΔG_{1w}) and the association of tryptophan to a lipid:water 1:1 complex ($\Delta G_1'$) (Table 2.10, Figure 2.10, p.64).

This is in agreement with the Job plots of selected compounds (**6f**, **6g**) that indicated the predominant stoichiometry for Trp:DMPC to be approximately 2:1 (§2.3.3.b). Further evidence was observed in molecular dynamics simulations of a tryptophan derivatives and DMPC at a 2:1 ratio. Here, strong, favourable binding interactions between the two amino acid analogues and the lipid were observed qualitatively (as the duration of binding events >39 ns prevented a statistically relevant quantitative analysis) (§2.5.4).

The binding free energy associated with the adduct formation of a water:lipid 1:1 complex ($\Delta G_{1w/}$) could also be determined experimentally *via* NMR titrations of hydrated DMPC into chloroform and related to the calculated value obtained from molecular dynamics simulations of tryptophan, water and lipid being dissolved in chloroform with equimolar concentrations (§2.5.3). The simulations underestimated $\Delta G_{1w/}$ by approximately 4 kJ mol^{-1} (Table 2.22 p.116). Within the limitations of the simulation such as the limited amount of water present this is a reasonably good approximation. The same molecular dynamics simulation also allowed evaluating the influence of water on ΔG_1. The calculated value for ΔG_1 in the presence of a water molecule approximated the experimentally observed value slightly better than a calculated value derived from a simulation of Trp-DMPC interactions in the absence of water.

3.5 Isothermal Titration Calorimetry

Isothermal titration calorimetry could be used to estimate the importance of the heat associated with the dissociation of water from hydrated lipid molecules. At very low concentrations (0.14 m*M*) of hydrated DMPC in the injection syringe, heat changes that are likely to be mainly attributed to dehydration of lipid molecules could be determined in a semi-quantitative way (as the errors associated with the small observed heats of injection and the small concentration of the lipid solution are likely to induce a fairly large error of observables). At higher lipid concentrations the observed heats are likely to include contributions arising from the heats of demicellisation, dehydration and dilution. In the titration experiments, the observed heats additionally will contain the heat associated with the interactions of tryptophan derivatives to the species present in the measurement cell and in particular the heat of interest, *i.e.* the heat associated with the binding of the amino acid analogues to DMPC. Blank titrations of hydrated DMPC into chloroform with absolute lipid concentration of 26 m*M* maximised the observed signal-to-noise ratio and established

conditions that were comparable to the experimental conditions used in the NMR host-guest titrations. The enthalpograms obtained by subtraction of the heats of injection of the blank titration from the heats of injection of the titrations involving the tryptophan analogues showed the contributions of heat that could be attributed to the presence of tryptophan. This observed heat will be comprised partially by the heat of the binding of Trp to DMPC, as well as by contributions arising from the competitive nature of this binding and its affect on water:lipid complexes. The heat associated with the disassociation of water from water:lipid complexes can not be accounted for easily and is likely to result in an endothermic contribution to the observed overall change in heat. This meant that the obtained heats of injections are likely to represent an underestimated heat of the binding associated with Trp:DMPC and Trp:DMPC:water complexes.

A fitting of the obtained enthalpograms to a binding isotherm was found to be complicated by the near-linear pattern of it. The software (MicroCal Origin) of the isothermal calorimeter used was unable to account for the number of equilibria and processes likely to be present in the measuring cell and could not be used for data analysis. The observed heats were therefore analysed qualitatively by comparison of the observed difference in heat patterns and of the gradients of first order regressions of the parts of the enthalpograms that were seen to behave near-linearly (typically for heats of injections obtained at a total concentration of lipid in the measurements cell $L_0^{cell} > 1.0$ mM).

The heats of injections for electron dense compounds (**6e**, **6f**) exhibited a steeper decline at $L_0^{cell} < 1.0$ mM (*i.e.* a greater endothermic contribution) than seen for following the near-linear pattern at higher L_0^{cell}. In contrast, heats of injections for more electron deficient compounds (**6a**, **6c-d, 6g,** and **6n**) showed an exothermic contribution at $L_0^{cell} < 1.0$ mM then went through a transition point at concentrations that varied for the compounds analysed and then followed a near-linear pattern at higher $L_0^{cell} > 1.0$ mM.

The difference in the behaviour of the enthalpograms indicates that different interaction processes take place for electron dense and electron deficient tryptophan derivatives respectively. A more detailed analysis of the nature of this difference was complicated due the limited amount of data points at the $L_0^{cell} < 1.0$ mM.

The gradients of the near-linear part of the enthalpograms could be used in the qualitative assessment of the observed heats of injections. When plotted against the Hammett parameter, the gradients exhibited an n-shaped pattern that was comparable to the pattern

observed for the macroscopical binding free energies of the compounds as observed in the NMR host-guest titrations (Figure 2.23, p.90).

3.6 Calculation of Free Energies by a Thermodynamic Cycle

The free energy of solvation and particularly hydration of tryptophan analogues has received some interest in the past. Often, indole derivatives such as methylindole were used as a simple approximation for tryptophan. MacCallum and Tieleman[17] compared experimentally determined hydration free energies of the latter with computed values and found that the GROMOS96 and OPLS-AA force field descriptions they used underestimated the free energies and gave the largest errors for the tryptophan analogue as compared with the other amino acid analogues they had studied. This might be due to the experimentally determined large hydrophobicity and high absolute magnitude of the free energy of solvation of this compound.[18]

Here, free energy calculations of the association free energy of the 1:1 Trp:DMPC adduct formation in chloroform were performed. First, the free energy of solvation of the amino acid analogues was calculated with a reasonable accuracy using a thermodynamic cycle in which the non-bonding interactions of the compounds in vacuum and in solvent (*i.e.* chloroform) were determined. The obtained, slightly negative free energies of solvation reflect the empirically observed poor solubility of the tryptophan compounds in chloroform. The calculation of the free energy of binding had high errors associated with it, arising partially from the difficulty to represent a sufficient number of bound configurations with high accuracy - the clustering of 1:1 Trp:DMPC adducts gave numerous individual binding conformations that, combined into clustered average structures, were only accounting for up to 19% of all observed binding conformations. This made the choice of the appropriate starting conformation essential but difficult. Multiple starting configurations for some of the derivatives were analysed but the overall error could not be reduced significantly. Another reason for the rather poor performance of the free energy calculations is the flexibility of the molecules. For accurate free energy calculations, the full or at least a significant part of the conformational space available to the molecules involved should be undertaken.

3.7 *Interactions of AcWLWLL with a DMPC Bilayer*

Two statistical important conformations for the peptides of the family AcWLWLL were observed in molecular dynamics simulations of the peptide in water. This is in contrast with the lack of secondary structure of this peptide in solution as determined by White and Wimley using circular dichroism (CD).[19] However, the very short time spans used in the simulations with respect to an averaged ensemble sampling in CD might overestimate the importance of short-lived conformations.

Wimley and White used AcWLWLL as a benchmark in their experiments of partitioning of peptides into lipid bilayers.[20] The authors could demonstrate by fluorescence spectroscopy and quenching studies that AcWLWLL inserts and partitions in the bilayer with a near constant partition coefficient over varying peptide concentrations. In another study, White and Wimley could estimate the free energy associated with interaction of AcWLWLL with a POPC bilayer phase ($\Delta G_{water\text{-}POPC}$ = -1.85 ± 0.06 kcal mol^{-1}) by comparison to the free energy of the partitioning of other pentapeptides of the AcWL-X-LL family, where X stands for any of the 20 amino acid residues.[21] The native peptide and its derivatives used here exhibited such favourable partitioning into a lipid bilayer, representing the experimental observations well.

The peptides analysed in this study exhibited preferred binding conformations when interacting with the lipid. These were adopted quickly (<5 ns) and retained for long times (accounting for up to 99% of all sampled configurations.)

3.8 *References*

[1] Sanderson, J. M., Whelan, E. J., *Phys. Chem. Chem. Phys.*, **2004**, *6*, 1012-1017
[2] Kovac, A., Kidric, J., Hadzi, D., *J. Mol. Struct.*, **1992**, *267*, 399-404
[3] Bordwell, F. G., Harrelson, J. A., Lynch, T. Y., *J. Org. Chem.*, **1990**, *55*(10), 3337-3341
[4] Bordwell, F. G., Drucker, G. E., Fried, H. E., *J. Org. Chem.*, **1981**, *46*(3), 632-635
[5] Wimley, W. C., White. S. H., *Biochemistry*, **1992**, *31*(51), 12813–12818
[6] Persson, S., Killian, S. A., Lindblom, G., *Biophys. J.*, **1998**, *75*(3), 1365-1371
[7] Gaede, H. C., Yau, W.-M., Gawrisch, K., *J. Phys. Chem. B.*, **2005**, *109*(26), 13014–13023
[8] Van der Wel, P. C. A., Reed, N. D., Greathouse, D. V., Koeppe, R. E., *Biochemistry*, **2007**, *46*(25), 7514-7524
[9] Norman, K. E., Nymeyer, H., *Biophys. J.*, **2006**, *91*, 2046-2054
[10] Gaede, H. C., Yau, W.-M., Gawrisch, K., *J. Phys. Chem. B.*, **2005**, *109*(26), 13014–13023
[11] Petersen, F. N. R., Jensen, M. Ø., Nielson, C. H., *Biophys. J.*, **2005**, *89*, 3985-3996
[12] Beene, D. L., Brandt, G. S., Zhong, W. G., Zacharias, N. M., Lester, H. A., Dougherty, D. A., *Biochemistry*, **2002**, *41*(32), 10262-1069
[13] Gallivan, J.P., Dougherty, D. A., *J. Am. Chem. Soc.*, **2000**, *122*, 870-874
[14] Gallivan, J. P., Dougherty, D. A., Proc. *Natl. Acad. Sci. USA*, **1999**, *96*, 9459-9464

[15] Esbjörner, E. K., Caesar, C. E. B., Albinsson, B., Lincoln, P., Nordén, B., *Biochim. Biophys. Acta*, **2007**, *361*(3), 645-650
[16] Zhang, Y., McCammon, J.A., *J. Chem. Phys.*, **2003**, *118*(4), 1821-1827
[17] MacCallum, J. L., Tieleman, D.P., *J. Comp. Chem.*, **2003**, *24*(15), 1930-1935
[18] Wolfenden, R., Anderson, L., Cullis, P., Southgate, C., *Biochemistry*, **1981**, *20*, 849-855
[19] Wimley, W. C., Hristova, K., Ladokhin, A. S., Silvestro, L., Axelsen, P. H., White, S. H., *J. Mol. Biol.*, **1998**, *277*(5), 1091-1110
[20] Wimley, W. C., Creamer, T. P., White, S. H., *Biochemistry*, **1996**, *35*(16), 5109-5124
[21] White, S. H., Wimley, W. C., *Biochim. Biophys. Acta*, **1998**, *1376*, 339-352

Chapter 4: Experimental Part

4.1 Overview

The following chapter details the synthesis of tryptophan derivatives and their incorporation into peptides. §4.2 contains the general methods and techniques used in the production and analysis of the compounds. §4.3 describes the experimental procedures for the individual experiments. Experimental data for the compounds prepared are given in §4.4.

4.2 General Methods

4.2.1 Chemicals

Chemicals and reagents were obtained from Sigma Aldrich or Fisher Scientific UK unless stated otherwise. All reagents were of the highest quality and were used without further purification. Deuterated solvents were used as supplied from GOSS. The enzyme "Amano Acylase" was used as purchased from Amano Enzyme Inc., Nagoya, Japan. Indole derivatives, N^{α}-acetyl-5-Me-*D,L*-tryptophan, and general purpose reagents were obtained from Sigma Aldrich. Wang resin (loading 0.75 and 0.30 mmol g^{-1} respectively) was used as purchased from Nova Biochem. For peptide synthesis, solvents were of HPLC grade, with reagents of biochemical purity (> 98 %) and were used as purchased from Fisher Scientific. The lipid DMPC (biochemical grade) was obtained from Avanti Lipids Inc. (Alabama) or Bachem (UK) and used as purchased.

4.2.2 High Pressure Liquid Chromatography (HPLC)

The HPLC set-up consisted of a pump, autosampler, diode array detector, and Peltier column oven (from the Perkin Elmer Series 200), plus an Anachem fraction collector. A Hypersil C_{18} column (250 x 4.6 mm, 4 μ particle size) was used for analytical runs and a Hypersil C_{18} column (250 x 10 mm, 4 μ particle size) for semi-preparative purifications. The column temperature was adjusted to 25 °C and chromatograms were recorded at a wavelength of 254 nm. A linear gradient system between two eluents was used. These were HPLC grade acetonitrile and water respectively, each containing 0.1% biochemical grade (99.6%)

trifluoracetic acid (TFA) as additive, used as purchased from Fisher Scientific UK. Analytical runs used sample injection volumes of 10 µl with a 10 µl air cushion and 2 µl excess-volume. Semi-preparative runs used sample injection volumes of 30 – 150 µl with a 10 µl air cushion and 2 µl excess-volumes.

Optical separation conditions for the purification of N^{α}-acetyl-tryptophan ethyl amides, N^{α}-acetylated tryptophan methyl esters, and Fmoc-protected tryptophan analogues were found by using a linear gradient. Analytical separation was achieved using an elution mixture of 70% water and 30% acetonitrile that was changed linearly to 40% water and 60% acetonitrile over 25 min at a flow rate of 1 ml/min. The initial mixture ratio was restored in a linear fashion over 1 min and eluted for further 5 min. The same gradient system was employed for semi-preparative purification changing the ratio over 45 min at a flow rate of 2 ml/min, restoring initial conditions over 5 min and equilibrating the column for further 10 min.

4.2.3 Melting Point Analysis

Melting point analysis was performed on Griffin melting point apparatus and transition temperatures determined by visual inspection. All measurements were performed in duplicate and are uncorrected.

4.2.4 Polarimetry

The optical rotation of enantiomers was determined using a Jasco P-1020 polarimeter at a wavelength of 589 nm. Aqueous α-glucose solution (10 mg/ml, $c = 1.0$) was used as standard, measuring the optical rotation in a cell of 100 x 1.8 mm dimensions. Sample solutions (typical concentrations $c = 0.15$ to 1.0) were prepared with HPLC grade methanol or other appropriate solvents as specified in the literature. The optical rotation was measured against the pure solvent background. All measurements were performed in duplicate.

4.2.5 Mass Spectrometry (MS)

Liquid chromatography coupled with mass spectroscopy was employed for the mass analysis of compounds. A Micromass LCT mass spectrometer in electrospray (ES+) ionisation mode

and a Waters 2795 LC instrument were used. Samples were analysed at a concentration of approximately 1 mg/ml.

4.2.6 Nuclear Magnetic Resonance (NMR) Spectroscopy

300 MHz and 400 MHz Varian spectrometers were used for compound analysis. The following abbreviations are used to describe multiplicities: s for singlet; d for doublet; t for triplet; q for quartet; m for multiplet; and br for a broad signal. Chemical shift data are reported in parts per million (ppm) with respect to the residual protons of the solvent as internal reference (3.30 ppm for CH_3OH/CD_3OD; 2.50 ppm for $(CD_3)CHD_2SO/ (CD_3)_2SO$; 2.05 ppm for CHD_2COCD_3/CD_3COCD_3; 7.26 ppm for $CHCl_3/CDCl_3$). ^{13}C NMR spectra were referenced to solvent as internal reference (49.0 ppm for CD_3OD; 39.5 ppm for $(CD_3)_2SO$; 29.9 ppm for CD_3COCD_3; 77.2 ppm for $CDCl_3$). COSY and NOESY 2D-NMR spectra were used for the assignments of signals.

4.3 Synthesis and Experimental Descriptions

4.3.1 Synthesis of 5- and 6-monosubstituted Tryptophan Derivatives

4.3.1.a Reaction of Indoles with Serine: General Method[1]

L-serine (21 mg, 0.2 mmol) was dissolved in a solution of the substituted indole (0.1 mmol) in acetic acid (0.24 ml) and acetic anhydride (0.9 mmol). The mixture was stirred under argon at 73 °C for 2 h. Upon cooling, the solution was treated with diethyl ether (2 ml), adjusted to pH 11 using 30% sodium hydroxide solution and further diluted with ether (3 ml). The layers were partitioned and treated separately. The ether layer was extracted with 1 *N* sodium hydroxide solution (2 x 2 ml) and a small amount of sodium thiosulphate $Na_2S_2O_4$ (~5 mg) was added. The aqueous solutions were combined, neutralized to pH 7 using concentrated hydrochloric acid and the volume reduced to 50% by evaporation under vacuum. If precipitation occurred, samples were stored at 4 °C for 24 hours to allow further precipitation. Precipitates were filtered and dried under vacuum. The filtrate was further concentrated to half its original volume and any further precipitates collected. Otherwise the filtrates were acidified to pH 3 with 5% hydrochloric acid, extracted with ethyl acetate (3 x 7 ml) and the combined organic extracts dried over magnesium sulphate. After filtration,

the solvent was removed under vacuum and the residue treated with benzene and re-evaporated to yield the desired mono-substituted *D,L*-tryptophan derivative. The crude product was further purified by silica flash chromatography using a chloroform/methanol mixture (95:5) as eluent.

The ether layer was dried over magnesium sulphate. Removal of the solvent under vacuum recovered unreacted indole starting material.

4.3.1.b Enzymatic resolution of N^{α}-Acetyl-D,L-Tryptophan Analogues

Borate buffer solution A was prepared from sodium tetraborate decahydrate (1.9 g, 0.05 *M*) in deionised water (100 ml). The pH was adjusted to pH 8.00 using concentrated hydrochloric acid. Cobalt(II) chloride hexahydrate (0.5 mg, 3.1 μmol) was added to a final concentration of 0.125 m*M*.

The N^{α}-acetyl-*D,L*-tryptophan analogue (0.40 mmol) was dissolved in borate buffer solution A (5 ml). A solution of amino acylase (100 mg) in borate buffer A (10 ml) at pH 8 was added with stirring. The reaction mixture was stirred at 37 °C for 48 hours and then quenched by adjusting to pH 5 using 10% hydrochloric acid. Subsequently the mixture was filtered through a celite pad and the filtrate extracted into ethyl acetate (3 x 25 ml). The combined organic extracts were dried over magnesium sulphate, filtered and the solvent removed under vacuum to afford the crude unreacted N^{α}-acetylated-*D*-tryptophan derivative. The aqueous phase was purified by ion-exchange chromatography using DOWEX® Resin 50X2-200. The resin was prepared prior to the run through deprotonation and protonation via 0.1 *M* sodium hydroxide solution and 0.1 *M* hydrochloride (HCl) solution respectively. Water and methanol eluents were used to wash the resin before its application. The column was prepared for loading by protonation of the acceptor sites with 0.1 *M* HCl solution. The product, dissolved in the aqueous phase, was then applied to the column and eluted first with water (150 ml) and then with 10% ammonia in methanol (200 ml). Fractions which showed UV activity and responded positively to a ninhydrin test were collected and the solvent removed under vacuum to afford the *L*-enantiomer of the free amino acid.

4.3.1.c Acetylation of L-Tryptophan Analogues

Acetic anhydride (1.4 ml, 1.6 mmol) was added to a solution of the amino acid (0.8 mmol) in acetic acid (5 ml). The reaction mixture was stirred at room temperature for 1 h and subsequently quenched by the addition of ice-cold water (2 ml). The solvents were removed under vacuum and the crude products recrystallized from ethyl acetate/ petroleum ether (60/80). Alternatively, the reaction mixture was adjusted to pH = 2.5 with 0.1 *N* sodium hydroxide solution. The product was then extracted into ethyl acetate (3 x 10 ml) and the organic solvent removed under vacuum to yield the crude product.

4.3.1.d Preparation of Methyl-Esters[2]

The N^{α}-Acetyl-*L*-tryptophan analogue (0.1 mmol) was dissolved in methanol (5 ml). Concentrated sulphuric acid (0.15 ml) was added and the reaction mixture refluxed for 2 hours. After adjustment to pH 7 using 0.5 *M* aqueous sodium carbonate solution, the solvents were removed under vacuum and the residues redissolved in deionised water (2 ml). The product was extracted from the aqueous phase using dichloromethane (3 x 2 ml). The organic layer was separated and washed with 0.5 *M* aqueous sodium carbonate solution (2 ml) and deionised water (2 x 2 ml). The combined dichloromethane extracts were then dried over magnesium sulphate, filtered and the solvent removed under vacuum to yield the crude product, which was used without further purification for the ethyl amidation reaction.

4.3.1.e Ethyl Amidation[3]

The N^{α}-Acetylated-*L*-tryptophan methyl ester (0.1 mmol) was dissolved in a solution of ethylamine in methanol (50% v/v, 5 ml). The solution was stirred at 0 °C for 24 hours. Solvents were removed under vacuum and reaction was repeated. The crude product, following concentration *in vacuo*, was purified by silica flash column chromatography (95% DCM, 5% EtOH).

4.3.1.f Determination of Resin Loading

3.0 mg of leucine-loaded Wang resin were suspended in a solution of 20% (v/v) of piperidine in DMF (3 ml). The solution was transferred into a 3 ml UV/Vis cell and the absorbance was

measured against a solution of 20% (v/v) solution of piperidine in DMF at 290 nm. The measuring cell was agitated and stirred for 2 – 3 minutes and particulate material allowed to settle prior to the measurement. The loading was calculated using Equation 3.1:

$$\text{Loading} = \frac{(\text{Abs}_{\text{sample}} - \text{Abs}_{\text{ref}})}{(1.75 * \text{mass resin [mg]})}, \tag{3.1}$$

where $\text{Abs}_{\text{sample}}$ and Abs_{ref} are the absorbance of sample and reference respectively. The factor 1.75 was derived by using 5253 M^{-1} cm^{-1} as extinction coefficient at 290 nm. The theoretical substitution *A* of a peptidyl resin was calculated from the substitution of the base resin using Equation 3.2

$$A = \frac{B * 1000}{[1000 + (B * (M - 18))]}, \tag{3.2}$$

where *B* is the initial substitution of the starting resin (mmol g^{-1}), and *M* the molecular weight of the target peptide, plus all protecting groups, in mg.

4.3.2 NMR Guest-Host Titrations

4.3.2.a General

NMR titration spectra were recorded on a Varian Unity 200 MHz NMR spectrometer. Tetramethylsilane (TMS) was used as internal standard (0 ppm). The quantities for the stepwise addition of guest solution in the NMR host-guest titrations were 5, 5, 10, 10, 25, 50, 50, 50, 100, 100, 100, 100, 100, and 200 µl. All titrations were performed in duplicate. Data were fitted to a 2:1 binding isotherm that allowed for DMPC micellisation and water adducts at various stages (see §2.3).

4.3.2.b DMPC Blank Titrations

A solution of DMPC (75 mg, 110 m*M*) in $CDCl_3$ (1 ml) was added portionwise to $CDCl_3$ (0.5 ml) in an NMR tube. Following each addition, a ^1H spectrum was recorded, typically providing fifteen data points. Observed complexation-induced NMR signals for the water and lipid protons were integrated when clearly distinct from nearby signals. Titrations were

performed in duplicate using TMS ($\delta = 0.0$) as internal standard. The data were analysed by plotting the intensities of the water protons against calculated lipid concentration enabling the estimation of the free water content of the chloroform and the water per lipid molecule ratio (see §2.3.2.a).

4.3.2.c NMR Job Plots

A series of solutions (0.5 ml each) of constant total molar concentration (10 m*M*) were prepared containing varying molar ratios of lipid (DMPC) and amino acid analogue (**6f** or **6g**). Job plots were then made of the observed chemical shift change (χ x $\Delta\delta_{obs}$) against mole fraction (χ).[4]

4.3.2.d Tryptophan-DMPC Host-Guest Titrations

Host solution was prepared by dissolving N^{α}-acetyl-monosubstituted ethyl *L*-tryptophan derivatives (typical concentration 4 m*M*) in deuterated chloroform (2 ml). Host solution (0.5 ml) was placed in an NMR tube and the remainder used to generate the guest solution by solvation of DMPC in it giving at a typical concentration of 100 m*M*. Guest solution was added stepwise, agitated thoroughly and a ^{1}H NMR spectrum recorded after each addition, typically providing fourteen data points. All NMR titrations were performed in duplicate. Data were analysed by least squares curve fitting to a calculated binding isotherm, as described in §2.3.1.

4.3.2.e Tryptophan-DMPC D_2O Titrations

Host solution was prepared by dissolving N^{α}-acetyl- *L*-tryptophan ethylamide (15 mg, 55 m*M*) and DMPC (37 mg, 55 m*M*) in deuterated chloroform (2 ml). Deuterated water (D_2O) was added in 1 µl steps and ^{1}H NMR spectra recorded immediately after each addition. The titration was repeated as a kinetics experiment, with ^{1}H NMR spectra recorded at 30 second intervals after a single injection of D_2O. Spectra were recorded until changes in the spectrum were undetectable.

4.3.3 Isothermal Titration Calorimetry (ITC)

Isothermal titration calorimetry was performed on a computer-assisted VP 2000 ITC instrument using MicroCal control software (version 7.0). The measuring cell was loaded with a solution of N^{α}-acetyl-monosubstituted *L*-tryptophan ethylamide derivative in chloroform (1.5 ml, typical concentration 4 m*M*). Chloroform was also used in the reference cell. The temperature was adjusted to 25 °C and aliquots of DMPC guest solution (typical concentration 26 m*M*) were injected into the cell with stirring at 307 rpm. A first injection of 2 µl was followed by aliquots of 8 µl DMPC solution over 32 seconds, with 180 seconds between injections, until the syringe loading of approximately 275 µl was exhausted. This resulted typically in 32 injections. The heat of reaction was recorded as a function of time and lipid concentration. The reference power was chosen to be 10 µcal per second. The initial time delay of the first injection was 60 seconds. A blank titration of lipid into neat chloroform was also performed. The subtraction of the signal of this DMPC blank titration from the measured values of DMPC-tryptophan titrations gave the desired heat profiles of the interaction between the two molecules.

4.4 Analytical Data for Compounds Prepared

4.4.1 Overview

The following section lists the physical and spectroscopic properties of all compounds synthesised. Colouration of products change upon hydration and are given here as the colour of the dry compound. Mass spectrometry data is given as mass of identified ion, relative abundance in () brackets and the identified ion in [] brackets. R_f values given were obtained by TLC. Conditions are given in brackets.

4.4.1.a *5-substituted N^{α}-acetyl-D,L-tryptophan derivatives (**1a-f**)*

ID	Substituent	*L*-serine & Ac_2O (eq.)	Yield (%)
1a	R = F	2.0	85
1b	R = Cl	2.0	86
1c	R = Br	2.0	77
1d	R = I	1.5	87
1e	R = MeO	1.0	84
1f	R = NO_2	2.5	44

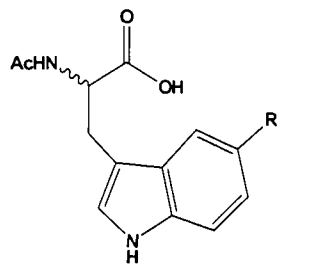


Table 4.1: Table of N^{α}-Ac-5-*D,L*-tryptophan derivatives that were produced *via* the indole synthesis.

N^{α}-Acetyl-5-fluoro-*D,L*-tryptophan (**1a**)

Using the general method above (§4.3.1.a), the title compound was synthesised as an orange-brown powder.

^{1}H NMR (400 MHz, MeOH – d_4): δ 1.90 (s, AcCH_3, 3H), 3.11 (dd, *J* = 7.6, 15.6 Hz, H_β, 1H), 3.26* (d, *J* = 7.6 Hz, H_β, 1H), 4.68 (dd, *J* = 5.2, 7.6 Hz, H_α, 1H), 6.84 (dt, *J* = 2.8, 8.4 Hz, ind*H*6, 1H), 7.14 (s, ind*H*4, 1H), 7.21 (dd, *J* = 2.0, 10.0 Hz, ind*H*7, 1H), 7.26 (dd, *J* = 3.6, 8.4 Hz, ind*H*2, 1H); **^{13}C NMR** (100 MHz, MeOH – d_4): δ 22.38 (AcCH_3), 28.44 (C_β), 54.71 (C_α), 103.77 (ind*C*3), 110.31 (ind*C*4), 110.57 (ind*C*7), 112.95 (ind*C*6), 126.38 (ind*C*2), 129.26 (ind*C*3a), 134.56 (ind*C*7a), 160.11 (ind*C*5), 173.18 (Ac*C*O), 175.12 (*C*OOH); **MS** (ES+): *m/e* 287.1 (100%) $[m+Na]^+$, 288.1 (15%); **Mp**.: 180-183 °C (dec.); **CHN** analysis (calcd %) C, 59.09; H, 4.96; N, 10.60; (found %) C, 58.99; H, 4.99; N, 10.58

*full assignment of coupling constants not possible due to signal overlap with residual protons of solvent.

(Lit.[5] **Mp**.: 189-192 °C)

N^{α}-Acetyl-5-chloro-*D,L*-tryptophan (**1b**)

Using the general method above (§4.3.1.a), the title compound was synthesised as an orange-brown powder.

^{1}H NMR (400 MHz, MeOH – d_4): δ 1.90 (s, AcCH_3, 3H), 3.17 (dd, *J* = 7.0, 15.0 Hz, H_β, 1H), 3.30 (d, *J* = 5.5, 15.0 Hz, H_β, 1H), 4.77 (dd, *J* = 5.5, 13.0 Hz, H_α, 1H), 7.06 (d, *J* = 7.5 Hz, ind*H*6, 1H), 7.24 (s, ind*H*4, 1H), 7.27 (d, *J* = 7.5 Hz, ind*H*7, 1H), 7.60 (s, ind*H*2, 1H); **^{13}C NMR** (100 MHz, MeOH – d_4): δ 22.40 (AcCH_3), 28.32 (C_β), 54.75 (C_α), 108.27

(ind*C*3), 111.51 (ind*C*4), 112.19 (ind*C*7), 120.35 (ind*C*6), 124.35 (ind*C*2), 124.44 (ind*C*3a), 124.75 (ind*C*7a), 136.02 (ind*C*5), 173.18 (Ac*C*O), 175.11 (*C*OOH); **MS** (ES+): *m/e* 281.1 (24%) $[m^{35}Cl+H]^+$, 283.1 (7%) $[m^{37}Cl+H]^+$, 303.1 (100%) $[m^{35}Cl+Na]^+$, 305.2 (35%) $[m^{37}Cl+Na]^+$; **Mp**.: 215-217 °C (dec.); **CHN** analysis (calcd %) C, 55.62; H, 4.67; N, 9.98; (found %) C, 55.60; H, 4.69; N, 9.97

(Lit.[6] **Mp**.: 212 °C)

N^{α}-Acetyl-5-bromo-*D,L*-tryptophan (**1c**)

Using the general method above (§4.3.1.a), the title compound was synthesised as a yellowish white powder.

^{1}H NMR (400 MHz, MeOH – d_4): δ 1.91 (s, AcCH_3, 3H), 3.12 (dd, *J* = 6.4, 12.0 Hz, H_β, 1H), 3.30* (d, *J* = 4.4, 11.6 Hz, H_β, 1H), 4.70 (dd, *J* = 4.4, 6.0 Hz, H_α, 1H), 6.98 (dt, *J* = 1.6, 6.8 Hz, ind*H*6, 1H), 7.10 (s, ind*H*4, 1H), 7.31 (d, *J* = 1.2 Hz, ind*H*7, 1H), 7.49 (d, *J* = 2.8 Hz, ind*H*2, 1H); **^{13}C NMR** (100 MHz, MeOH – d_4): δ 22.41 (AcCH_3), 28.28 (C_β), 54.81 (C_α), 111.02 (ind*C*3), 112.99 (ind*C*4), 113.92 (ind*C*7), 121.94 (ind*C*6), 125.08 (ind*C*2), 125.93 (ind*C*3a), 130.76 (ind*C*7a), 136.62 (ind*C*5), 173.18 (Ac*C*O), 175.23 (*C*OOH); **MS** (ES+): *m/e* 347.1 (100%) $[m^{79}Br+Na]^+$, 349.1 (95%) $[m^{81}Br+Na]^+$, 325.1 (35%) $[m^{79}Br+H]^+$, 327.1 (32%) $[m^{81}Br+H]^+$; **Mp**.: 201-204 °C; **CHN** analysis (calcd %) C, 48.02; H, 4.03; N, 8.62; (found %) C, 47.89; H, 4.11; N, 8.57

(Lit.[7,6] **Mp**.: 212-214 °C, 193 °C)

N^{α}-Acetyl-5-iodo-*D,L*-tryptophan (**1d**)

^{1}H NMR (400 MHz, MeOH – d_4): δ 1.92 (s, AcCH_3, 3H), 3.10 (dd, *J* = 8.0, 14.8 Hz, H_β, 1H), 3.31 (d, *J* = 5.2, 14.8 Hz, H_β, 1H), 4.67 (dd, *J* = 5.2, 8.0 Hz, H_α, 1H), 7.07 (s, ind*H*4, 1H), 7.14 (d, *J* = 8.4 Hz, ind*H*6, 1H), 7.32 (dd, *J* = 1.6, 8.4 Hz, ind*H*7, 1H), 7.88 (d, *J* = 1.6 Hz, ind*H*2, 1H); **^{13}C NMR** (100 MHz, MeOH – d_4): δ 22.46 (AcCH_3), 28.23 (C_β), 54.83 (C_α), 82.75 (ind*C*3), 110.69 (ind*C*4), 114.43 (ind*C*7), 125.58 (ind*C*6), 128.37 (ind*C*2), 130.67 (ind*C*3a), 131.61 (ind*C*7a), 137.01 (ind*C*5), 173.18 (Ac*C*O), 174.95 (*C*OOH); **MS** (ES+): *m/e* 395.1 (100%) $[m+Na]^+$, 396.1 (16%) $[m+Na]^+$; **Mp**.: >280 °C (dec.); **CHN** analysis (calcd %) C, 41.96; H, 3.52; N, 7.53; (found %) C, 42.19; H, 3.60; N, 7.55

N^{α}-Acetyl-5-methoxy-*D,L*-tryptophan (**1e**)

Using the general method above (§4.3.1.a), the title compound was synthesised as a yellow powder.

^{1}H NMR (400 MHz, MeOH – d_4): δ 1.91 (s, AcCH_3, 3H), 3.12 (dd, J = 7.6, 14.8 Hz, H_β, 1H), 3.30 (d, J = 5.2, 14.8 Hz, H_β, 1H), 3.82 (s, CH_3O-Ar, 3H), 4.70 (dd, J = 5.2, 7.2 Hz, H_α, 1H), 6.74 (dd, J = 2.0, 8.8 Hz, ind$H$6, 1H), 7.05 (s, ind$H$4, ind$H$7, 2H), 7.20 (d, J = 8.8 Hz, ind$H$2, 1H); **^{13}C NMR** (100 MHz, MeOH – d_4): δ 22.44 (AcCH$_3$), 30.68 (C_β), 54.74 (C_α), 56.23 (CH$_3$O-Ar), 101.09 (ind$C$3), 110.82 (ind$C$4), 112.72 (ind$C$7), 112.92 (ind$C$6), 125.06 (ind$C$2), 129.06 (ind$C$3a), 133.23 (ind$C$7a), 155.09 (ind$C$5), 173.19 (AcCO), 175.28 (COOH); **MS** (ES+): *m/e* 299.1 (100%) $[m+Na]^+$; **Mp.:** 187-189 °C

(Lit.[8.6] **Mp**.: 182-183 °C, 170 °C); **CHN** analysis (calcd %) C, 60.86; H, 5.84; N, 10.14; (found %) C, 60.79; H, 5.91; N, 10.10

N^{α}-Acetyl-5-nitro-*D,L*-tryptophan (**1f**)

Using the general method above (§4.3.1.a), the title compound was synthesised as a yellow powder.

^{1}H NMR (400 MHz, MeOH – d_4): δ 1.92 (s, AcCH_3, 3H), 3.22 (dd, J = 7.6, 14.8 Hz, H_β, 1H), 3.39 (d, J = 5.2, 14.8 Hz, H_β, 1H), 4.74 (dd, J = 5.2, 7.6 Hz, H_α, 1H), 7.31 (s, ind$H$4, 1H), 7.43 (d, J = 7.2 Hz, ind$H$6, 1H), 8.01 (dd, J = 2.0, 8.8 Hz, ind$H$7, 1H), 8.56 (d, J = 2.4 Hz, ind$H$2, 1H); **^{13}C NMR** (100 MHz, MeOH – d_4): δ 22.39 (AcCH$_3$), 28.14 (C_β), 54.70 (C_α), 112.45 (ind$C$3), 114.26 (ind$C$4), 115.82 (ind$C$7), 116.82 (ind$C$6), 117.84 (ind$C$2), 128.25 (ind$C$3a), 141.01 (ind$C$7a), 142.46 (ind$C$5), 173.21 (AcCO), 174.65 (COOH); **MS** (ES+): *m/e* 314.1 (100%) $[m+Na]^+$, 315.1 (15%) $[m+Na]^+$; **Mp**.: 160-163 °C (dec.); **CHN** analysis (calcd %) C, 64.60; H, 6.20; N, 10.76; (found %) C, 64.19; H, 6.29; N, 10.72

*4.4.1.b 6-substituted N^{α}-acetyl-D,L-tryptophan derivatives (**1g-l**)*

ID	Substituent	*L*-serine & Ac_2O (eq.)	Yield (%)
1g	R = F	2.0	98
1h	R = Cl	2.0	86
1i	R = Br	2.0	73
1j	R = Me	1.5	73
1k	R = MeO	1.0	77
1l	R = NO_2	2.5	58

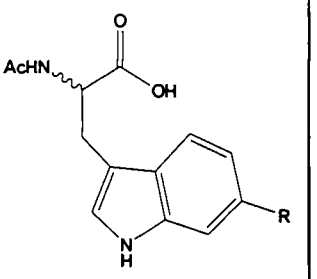


Table 4.2: Table of N^{α}-Ac-6-*D,L*-tryptophan derivatives that were produced *via* the indole synthesis.

N^{α}-Acetyl-6-fluoro-*D,L*-tryptophan (**1g**)

Using the general method above (§4.3.1.a), the title compound was synthesised as a greenish yellow crystalline powder.

^{1}H NMR (500 MHz, MeOH – d_4): δ 1.90 (s, AcCH_3, 3H), 3.12 (dd, *J* = 8.0, 14.4 Hz, H_β, 1H), 3.30* (d, *J* = 8.0 Hz, H_β, 1H), 4.69 (dd, *J* = 5.5, 10.0 Hz, H_α, 1H), 6.79 (dt, *J* = 2.5, 10.0 Hz, ind*H*5, 1H), 7.01 (dd, ind*H*4, *J* = 2.0, 10.0 Hz, 1H), 7.07 (s, ind*H*7, 1H), 7.45 (dd, *J* = 5.5, 9.0 Hz, ind*H*2, 1H); **^{13}C NMR** (125 MHz, MeOH – d_4): δ 22.37 (AcCH_3), 28.43 (C_β), 54.71 (C_α), 108.09 (ind*C*3), 110.14 (ind*C*4), 111.40 (ind*C*7), 120.14 (ind*C*5), 124.79 (ind*C*2), 125.60 (ind*C*3a), 137.87 (ind*C*7a), 160.23 (ind*C*6), 173.18 (Ac*C*O), 175.12 (*C*OOH); **MS** (ES+): *m/e* 287.1 (100%) $[m+Na]^+$, 288.1 (15%) $[m+Na]^+$, 265.1 (26%) $[m+H]^+$, 265.1 (4%) $[m+H]^+$; **Mp**.: 179-182 °C; **CHN** analysis (calcd %) C, 59.09; H, 4.96; N, 10.60; (found %) C, 59.06; H, 4.98; N, 10.59

*full assignment of coupling constants not possible due to signal overlap with residual protons of solvent.

(Lit.[9] **Mp**.: 178-179 °C)

N^{α}-Acetyl-6-chloro-*D,L*-tryptophan (**1h**)

Using the general method above (§4.3.1.a), the title compound was synthesised as a yellow crystalline powder.

^{1}H NMR (500 MHz, MeOH – d_4): δ 1.89 (s, AcCH_3, 3H), 3.12 (dd, *J* = 8.0, 15.0 Hz, H_β, 1H), 3.30* (d, *J* = 8.0 Hz, H_β, 1H), 4.70 (dd, *J* = 5.0, 7.5 Hz, H_α, 1H), 6.98 (dd, *J* = 2.0, 8.5 Hz, ind*H*5, 1H), 7.10 (s, ind*H*7, 1H), 7.31 (d, ind*H*4, *J* = 1.5 Hz, 1H), 7.50 (d, *J* = 6.5 Hz,

ind*H*2, 1H); **^{13}C NMR** (125 MHz, MeOH – d_4): δ 22.38 (Ac*C*H$_3$), 28.33 (C_β), 54.69 (C_α), 111.50 (ind*C*3), 112.03 (ind*C*4), 120.28 (ind*C*7), 120.34 (ind*C*5), 125.30 (ind*C*2), 127.59 (ind*C*3a), 128.29 (ind*C*7a), 138.33 (ind*C*6), 173.17 (Ac*C*O), 175.07 (*C*OOH); **MS** (ES+): *m/e* 303.1 (100%) $[m^{35}Cl+Na]^+$, 305.1 (35%) $[m^{37}Cl+Na]^+$, 281.1 (24%) $[m^{35}Cl+H]^+$, 283.1 (7%) $[m^{37}Cl+H]^+$; **Mp.:** 190-194 °C (dec.); **CHN** analysis (calcd %) C, 55.62; H, 4.67; N, 9.98; (found %) C, 55.58; H, 4.73; N, 9.95

*full assignment of coupling constants not possible due to signal overlap with residual protons of solvent.

N^α-Acetyl-6-bromo-*D,L*-tryptophan (**1i**)

Using the general method above (§4.3.1.a), the title compound was synthesised as a light-brown/ orange crystalline powder.

^{1}H NMR (400 MHz, MeOH – d_4): δ 1.89 (s, Ac*C*H$_3$, 3H), 3.12 (dd, *J* = 8.0, 14.4 Hz, H_β, 1H), 3.30* (d, *J* = 8.0 Hz, H_β, 1H), 4.69 (dd, *J* = 4.2, 8.4 Hz, H_α, 1H), 7.09 (s, ind*H*7, 1H), 7.11 (dd, *J* = 2.0, 8.8 Hz, ind*H*5, 1H), 7.45 (s, ind*H*4, 1H), 7.47 (m, ind*H*2, 1H); **^{13}C NMR** (100 MHz, MeOH – d_4): δ 22.38 (Ac*C*H$_3$), 28.30 (C_β), 54.66 (C_α), 111.80 (ind*C*3), 112.34 (ind*C*4), 120.97 (ind*C*7), 121.26 (ind*C*5), 125.83 (ind*C*2), 127.92 (ind*C*3a), 128.34 (ind*C*7a), 138.55 (ind*C*6), 173.18 (Ac*C*O), 175.00 (*C*OOH); **MS** (ES+): *m/e* 347.2 (100%) $[m^{79}Br+Na]^+$, 349.2 (95%) $[m^{81}Br+Na]^+$; **Mp.:** >320 °C (dec.); **CHN** analysis (calcd %) C, 48.02; H, 4.03; N, 8.62; (found %) C, 47.95; H, 4.10; N, 8.61

*full assignment of coupling constants not possible due to signal overlap with residual protons of solvent.

N^α-Acetyl-6-methyl-*D,L*-tryptophan (**1j**)

Using the general method above (§4.3.1.a) the title compound was synthesised as a dark-brown crystalline powder.

^{1}H NMR (500 MHz, MeOH – d_4): δ 1.89 (s, Ac*C*H$_3$, 3H), 2.40 (s, C*H*$_3$-Ar ,3H), 3.11 (dd, *J* = 8.0, 14.5 Hz, H_β, 1H), 3.30* (d, *J* = 8.0 Hz, H_β, 1H), 4.69 (dd, *J* = 5.0, 8.0 Hz, H_α, 1H), 6.85 (d, *J* = 8.5 Hz, ind*H*5, 1H), 6.99 (s, ind*H*7, 1H), 7.11 (s, ind*H*4, 1H), 7.42 (d, *J* = 9.0 Hz, ind*H*2, 1H); **^{13}C NMR** (125 MHz, MeOH – d_4): δ 21.78 (Ac*C*H$_3$), 22.39 (*C*H$_3$-Ar), 28.55 (C_β), 54.80 (C_α), 110.90 (ind*C*3), 112.12 (ind*C*4), 118.93 (ind*C*7), 121.54 (ind*C*5), 123.61 (ind*C*2), 126.83 (ind*C*3a), 132.02 (ind*C*7a), 138.49 (ind*C*6), 173.18 (Ac*C*O), 175.28

(*C*OOH); **MS** (ES+): *m/e* 283.2 (100%) $[m+Na]^+$, 284.2 (22%) $[m+Na]^+$, 261.2 (23%) $[m+H]^+$; **Mp.:** >320 °C (dec.); **CHN** analysis (calcd %) C, 64.60; H, 6.20; N, 10.76; (found %) C, 64.61; H, 6.41; N, 10.71

*full assignment of coupling constants not possible due to signal overlap with residual protons of solvent.

N^{α}-Acetyl-6-methoxy-*D,L*-tryptophan (**1k**)

Using the general method above (§4.3.1.a) the title compound was synthesised as a yellow crystalline powder. Upon hydration the colour changes to dark blue over green.

^{1}H NMR (500 MHz, MeOH – d_4): δ 1.90 (s, AcCH_3, 3H), 3.09 (dd, *J* = 8.0, 14.5 Hz, H_β, 1H), 3.29 (dd, *J* = 5.5, 14.5 Hz, H_β, 1H), 3.78 (CH_3O-Ar), 4.67 (dd, *J* = 5.5, 7.5 Hz, H_α, 1H), 6.67 (d, *J* = 8.0 Hz, ind*H*5, 1H), 6.80 (d, *J* = 8.5 Hz, ind*H*2, 1H), 7.36 (s, ind*H*4, 1H), 7.63 (d, ind*H*7, 1H); **^{13}C NMR** (125 MHz, MeOH – d_4): δ 22.44 (Ac*C*H$_3$), 30.68 (C_β), 55.18 (C_α), 57.12 (*C*H$_3$O-Ar), 94.50 (ind*C*3), 108.62 (ind*C*4), 110.01 (ind*C*7), 118.78 (ind*C*5), 121.64 (ind*C*2), 125.06 (ind*C*3a), 136.30 (ind*C*7a), 155.21 (ind*C*6), 169.27 (Ac*C*O), 173.63 (*C*OOH); **MS** (ES+): *m/e* 301.2 (100%) $[m+Na]^+$, 302.2 (16%) $[m+Na]^+$; **Mp.:** 178-183 °C (dec.); **CHN** analysis (calcd %) C, 60.86; H, 5.84; N, 10.14; (found %) C, 60.79; H, 5.91; N, 10.10

(Lit.[9] **Mp**.: 178-179 °C)

N^{α}-Acetyl-6-nitro-*D,L*-tryptophan (**1l**)

Using the general method above (§4.3.1.a), the title compound was synthesised as a yellow powder.

^{1}H NMR (500 MHz, MeOH – d_4): δ 1.91 (s, AcCH_3, 3H), 3.23 (dd, *J* = 8.0, 14.5 Hz, H_β, 1H), 3.39 (dd, *J* = 6.0, 14.5 Hz, H_β, 1H), 4.70 (dd, *J* = 6.0, 7.5 Hz, H_α, 1H), 7.34 (d, *J* = 8.0 Hz, ind*H*5, 1H), 7.40 (d, *J* = 8.5 Hz, ind*H*2, 1H), 7.96 (s, ind*H*4, 1H), 8.33 (d, ind*H*7, 1H); **^{13}C NMR** (125 MHz, MeOH – d_4): δ 22.40 (Ac*C*H$_3$), 28.18 (C_β), 55.08 (C_α), 112.50 (ind*C*3), 114.62 (ind*C*4), 117.03 (ind*C*7), 118.58 (ind*C*5), 120.13 (ind*C*2), 128.37 (ind*C*3a), 142.20 (ind*C*7a), 143.21 (ind*C*6), 173.25 (Ac*C*O), 174.55 (*C*OOH); **MS** (ES+): *m/e* 314.1 (100%) $[m+Na]^+$, 315.1 (15%) $[m+Na]^+$; **Mp.:** 290-293 °C (dec.); **CHN** analysis (calcd %) C, 64.60; H, 6.20; N, 10.76; (found %) C, 64.52; H, 6.25; N, 10.74

4.4.1.c *5-substituted N^{α}-acetyl-D-tryptophan derivatives (**2'a-g**)*

ID	Substituent	Yield (%)
2'a	R = F	58
2'b	R = Cl	71
2'c	R = Br	59
2'd	R = I	58
2'e	R = Me	59
2'f	R = MeO	65
2'g	R = NO_2	60

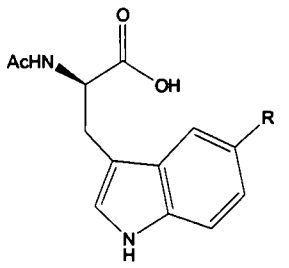


Table 4.3: Table of N^{α}-Ac-5-*D*-tryptophan derivatives that were produced *via* enzymatic resolution of a racemic mixture of the parent tryptophan compound.

N^{α}-Acetyl-5-fluoro-*D*-tryptophan (**2'a**)

Using the general method above (§4.3.1.b), the title compound was synthesised as a cream-white powder.

^{1}H NMR (400 MHz, MeOH – d_4): δ 1.90 (s, AcCH_3, 3H), 3.11 (dd, *J* = 7.6, 14.8 Hz, H_{β}, 1H), 3.26* (d, *J* = 6.0 Hz, H_{β}, 1H), 4.68 (dd, *J* = 5.2, 7.6 Hz, H_{α}, 1H), 6.84 (dt, *J* = 2.4, 9.2 Hz, ind*H*6, 1H), 7.14 (s, ind*H*4, 1H), 7.21 (dd, *J* = 2.4, 10.4 Hz, ind*H*7, 1H), 7.27 (dd, *J* = 4.8, 8.8 Hz, ind*H*2, 1H); **^{13}C NMR** (100 MHz, MeOH – d_4): δ 22.38 (Ac*C*H$_3$), 28.45 (C_{β}), 54.71 (C_{α}), 103.77 (ind*C*3), 110.31 (ind*C*4), 110.34 (ind*C*7), 112.96 (ind*C*6), 126.38 (ind*C*2), 129.27 (ind*C*3a), 134.57 (ind*C*7a), 157.81 (ind*C*5), 173.18 (Ac*C*O), 175.13 (*C*OOH); **MS** (ES+): *m/e* 265.2 (89%) $[m+H]^+$, 287.2 (16%) $[m+Na]^+$; **Mp.:** 187-190 °C; $[\alpha]_d^{25}$: -50.9° (*c* = 1.0, EtOAc)

*full assignment of coupling constants not possible due to signal overlap with residual protons of solvent.

(Lit.[7] **Mp.:** 192 °C; $[\alpha]_d^{20}$: +54.7° (*c* = 1.0, EtOAc) (data for *L*-isomer))

N^{α}-Acetyl-5-chloro-*D*-tryptophan (**2'b**)

Using the general method above (§4.3.1.b), the title compound was synthesised as a cream-white crystalline powder.

^{1}H NMR (500 MHz, MeOH – d_4): δ 1.91 (s, AcCH_3, 3H), 3.11 (dd, *J* = 10.0, 18.5 Hz, H_{β}, 1H), 3.29 (dd, *J* = 5.5, 18.5 Hz, H_{β}, 1H), 4.69 (dd, *J* = 6.5, 10.5 Hz, H_{α}, 1H), 7.11 (dt, *J* = 3.0, 11.5 Hz, ind*H*6, 1H), 7.15 (s, ind*H*4, 1H), 7.23 (dd, *J* = 3.0, 12.0 Hz, ind*H*7, 1H), 7.68

(dd, J = 6.0, 11.0 Hz, ind$H$2, 1H); **^{13}C NMR** (125 MHz, MeOH – d_4): δ 22.42 (AcCH$_3$), 28.23 (C_β), 54.76 (C_α), 110.91 (ind$C$3), 112.98 (ind$C$4), 113.90 (ind$C$7), 121.89 (ind$C$6), 125.06 (ind$C$2), 125.91 (ind$C$3a), 130.68 (ind$C$7a), 136.53 (ind$C$5), 173.18 (AcCO), 175.27 (COOH); **MS** (ES+): *m/e* 282.2 (100%) $[m+H]^+$, 284.2 (38%) $[m+H]^+$; **Mp.:** 210-213 °C; **$[\alpha]_d^{25}$**: -40.4° (c = 1.0, EtOAc)

(Lit.[7] **Mp.:** 212 °C; **$[\alpha]_d^{20}$**: +48.6° (c = 1.0, EtOAc) (data for *L*-isomer))

N^α-Acetyl-5-bromo-*D*-tryptophan (**2'c**)

Using the general method above (§4.3.1.b), the title compound was synthesised as a cream-white powder.

^{1}H NMR (400 MHz, MeOH – d_4): δ 1.91 (s, AcCH_3, 3H), 3.11 (dd, J = 8.0, 14.4 Hz, H_β, 1H), 3.30 (dd, J = 5.2, 14.8 Hz, H_β, 1H), 4.69 (dd, J = 5.2, 7.6 Hz, H_α, 1H), 7.12 (s, ind$H$4, 1H) 7.16 (dd, J = 2.0, 8.8 Hz, ind$H$6, 1H), 7.24 (d, J = 8.8 Hz, ind$H$7, 1H), 7.68 (d, J = 1.6 Hz, ind$H$2, 1H); **^{13}C NMR** (100 MHz, MeOH – d_4): δ 22.41 (AcCH$_3$), 28.27 (C_β), 54.78 (C_α), 111.00 (ind$C$3), 112.99 (ind$C$4), 113.93 (ind$C$7), 121.93 (ind$C$6), 121.09 (ind$C$2), 125.09 (ind$C$3a), 130.75 (ind$C$7a), 136.61 (ind$C$5), 173.19 (AcCO), 174.98 (COOH); **MS** (ES+): *m/e* 347.2 (100%) $[m^{79}Br+Na]^+$, 349.2 (95%) $[m^{81}Br+Na]^+$; **Mp.:** 124-127 °C; **$[\alpha]_D^{25}$**: -14.5° (c = 2.0, MeOH)

(Lit.[10] **Mp.**: 120-122 °C dec.; **$[\alpha]_D^{20}$**: -15.1° (c = 2.0, MeOH))

N^α-Acetyl-5-iodo-*D*-tryptophan (**2'd**)

Using the general method above (§4.3.1.b), the title compound was synthesised as a white crystalline powder.

^{1}H NMR (400 MHz, MeOH – d_4): δ 1.92 (s, AcCH_3, 3H), 3.11 (dd, J = 8.0, 14.8 Hz, H_β, 1H), 3.32 (dd, J = 5.2, 14.8 Hz, H_β, 1H), 4.69 (dd, J = 5.2, 8.0 Hz, H_α, 1H), 7.07 (s, ind$H$4, 1H) 7.14 (d, J = 8.4 Hz, ind$H$6, 1H), 7.32 (dd, J = 1.6, 8.4 Hz, ind$H$7, 1H), 7.88 (d, J = 1.6 Hz, ind$H$2, 1H); **^{13}C NMR** (100 MHz, MeOH – d_4): δ 22.48 (AcCH$_3$), 28.25 (C_β), 54.85 (C_α), 82.73 (ind$C$3), 110.71 (ind$C$4), 114.40 (ind$C$7), 125.61 (ind$C$6), 128.39 (ind$C$2), 130.69 (ind$C$3a), 131.65 (ind$C$7a), 137.05 (ind$C$5), 173.20 (AcCO), 174.91 (COOH); **MS** (ES+): *m/e* 395.1 (100%) $[m+Na]^+$, 396.1 (16%) $[m+Na]^+$; **Mp.:** 139-144 °C; **$[\alpha]_D^{25}$**: -5.2° (c = 0.5, MeOH)

N^{α}-Acetyl-5-methyl-*D*-tryptophan (**2'e**)

Using the general method above (§4.3.1.b), the title compound was synthesised as a cream-white powder.

^{1}H NMR (400 MHz, MeOH – d_4): δ 1.90 (s, AcCH_3, 3H), 2.40 (s, CH_3-Ar, 3H), 3.10 (dd, J = 8.0, 14.8 Hz, H_β, 1H), 3.31 (dd, J = 4.8, 14.8 Hz, H_β, 1H), 4.69 (dd, J = 5.2, 8.0 Hz, H_α, 1H), 6.91 (dd, J = 2.0 8.0 Hz, ind*H*6, 1H), 7.03 (s, ind*H*4, 1H), 7.19 (d, J = 8.0 Hz, ind*H*7, 1H), 7.33 (t, J = 0.8 Hz, ind*H*2, 1H); **^{13}C NMR** (100 MHz, MeOH – d_4): δ 21.69 (*C*H_3-Ar), 22.39 (Ac*C*H_3), 28.49 (C_β), 54.84 (C_α), 110.53 (ind*C*3), 111.95 (ind*C*4), 118.86 (ind*C*7), 124.00 (ind*C*6), 124.38 (ind*C*2), 128.80 (ind*C*3a), 129.10 (ind*C*7a), 136.40 (ind*C*5), 173.18 (Ac*C*O), 175.31 (*C*OOH); **MS** (ES+): *m/e* 261.2 (94%) $[m+H]^+$; **Mp.:** 174-176 °C; **$[\alpha]_d^{25}$**: -40.8° (c = 1.0, EtOAc)

(Lit.[7] **Mp.:** 171 °C; **$[\alpha]_d^{20}$**: +43.8° (c = 1.0, EtOAc) (data for *L*-isomer))

N^{α}-Acetyl-5-methoxy-*D*-tryptophan (**2'f**)

Using the general method above (§4.3.1.b), the title compound was synthesised as a cream-white powder.

^{1}H NMR (400 MHz, MeOH – d_4): δ 1.91 (s, AcCH_3, 3H), 3.12 (dd, J = 8.0, 14.8 Hz, H_β, 1H), 3.30 (dd, J = 4.8, 14.8 Hz, H_β, 1H), 3.81 (s, CH_3O-Ar, 3H), 4.69 (dd, J = 5.2, 8.0 Hz, H_α, 1H), 6.74 (dd, J = 2.4 8.8 Hz, ind*H*6, 1H), 7.05 (s, ind*H*4 and ind*H*2, 2H), 7.20 (d, J = 8.8 Hz, ind*H*7, 1H); **^{13}C NMR** (100 MHz, MeOH – d_4): δ 20.75 (Ac*C*H_3), 28.52 (C_β), 54.73 (C_α), 56.24 (*C*H_3O-Ar), 101.10 (ind*C*3), 110.81 (ind*C*4), 112.72 (ind*C*7), 125.07 (ind*C*6), 126.17 (ind*C*2), 129.17 (ind*C*3a), 133.22 (ind*C*7a), 155.08 (ind*C*5), 173.19 (Ac*C*O), 175.28 (*C*OOH); **MS** (ES+): *m/e* 299.1 (100%) $[m+Na]^+$; **Mp.:** 178-180 °C (dec.); **$[\alpha]_d^{25}$**: -10.5° (c = 1.0; MeOH)

(Lit.[7] **Mp.:** 176 °C; **$[\alpha]_d^{20}$**: +14.3° (c = 1.0; MeOH) (data for *L*-isomer))

N^{α}-Acetyl-5-nitro-*D*-tryptophan (**2'g**)

Using the general method above (§4.3.1.b), the title compound was synthesised as a yellow crystalline powder.

^{1}H NMR (400 MHz, MeOH – d_4): δ 1.91 (s, AcCH_3, 3H), 3.21 (dd, J = 7.6, 14.8 Hz, H_β, 1H), 3.37 (dd, J = 5.6, 14.8 Hz, H_β, 1H), 4.76 (dd, J = 5.6, 7.6 Hz, H_α, 1H), 7.32 (s, ind*H*4,

1H), 7.43 (d, J = 7.2 Hz, ind*H*6, 1H), 8.01 (dd, J = 2.4, 8.4 Hz, 1H), 8.57 (d, J = 2.4 Hz, ind*H*7, 1H); **^{13}C NMR** (100 MHz, MeOH – d_4): δ 22.41 (Ac*C*H$_3$), 28.16 (C_β), 54.74 (C_α), 112.52 (ind*C*3), 114.28 (ind*C*4), 115.85 (ind*C*7), 116.84 (ind*C*6), 117.86 (ind*C*2), 128.28 (ind*C*3a), 141.05 (ind*C*7a), 142.44 (ind*C*5), 173.24 (Ac*C*O), 174.63 (*C*OOH); **MS** (ES+): *m/e* 314.1 (100%) [m+Na]$^+$; **Mp.:** 172 °C; **$[\alpha]_d^{25}$**: -14.5° (c = 1.0; MeOH)

4.4.1.d *6-substituted N^α-acetyl-D-tryptophan derivatives (**2'h-m**)*

ID	Substituent	Yield (%)
2'h	R = F	74
2'i	R = Cl	59
2'j	R = Br	71
2'k	R = Me	64
2'l	R = MeO	59
2'm	R = NO_2	63

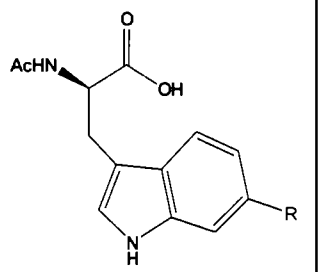


Table 4.4: Table of N^α-Ac-6-*D*-tryptophan derivatives that were produced *via* enzymatic resolution of the parent N^α-Ac-5-R-*D,L*-tryptophan compound.

N^α-Acetyl-6-fluoro-*D*-tryptophan (**2'h**)

Using the general method above (§4.3.1.b), the title compound was synthesised as a cream-white powder.

^{1}H NMR (500 MHz, DMSO – d_6): δ 1.78 (s, Ac*H*$_3$, 3H), 2.96 (dd, J = 9.5, 15.5 Hz, H_β, 1H), 3.11 (dd, J = 5.0, 15.5 Hz, H_β, 1H), 4.43 (m, H_α, 1H), 6.83 (dd, J = 5.0, 9.5 Hz, ind*H*5, 1H), 7.09 (dd, ind*H*4, J = 2.5, 10.0 Hz, 1H), 7.12 (d, J = 2.5 Hz, ind*H*7, 1H), 7.48 (dd, J = 5.0, 7.0 Hz, ind*H*2, 1H), 8.15 (d, J = 10.0 Hz, indN*H*, 1H), 10.90 (s, br, COO*H*, 1H); **^{13}C NMR** (125 MHz, DMSO – d_6): δ 20.39 (Ac*C*H$_3$), 27.03 (C_β), 52.89 (C_α), 97.19 (ind*C*3), 97.39 (ind*C*4), 106.71 (ind*C*7), 106.90 (ind*C*5), 119.19 (ind*C*2), 124.08 (ind*C*3a), 135.90 (ind*C*7a), 157.86 (ind*C*6), 169.20 (Ac*C*O), 173.47 (*C*OOH); **MS** (ES+): *m/e* 287.2 (100%) [m+Na]$^+$; **Mp.:** 173 °C

N^α-Acetyl-6-chloro-*D*-tryptophan (**2'i**)

Using the general method above (§4.3.1.b), the title compound was synthesised as a cream-white powder.

^{1}H NMR (500 MHz, DMSO – d_6): δ 1.78 (s, AcCH_3, 3H), 2.96 (dd, J = 9.0, 15.0 Hz, H_β, 1H), 3.12 (dd, J = 5.0, 15.0 Hz, H_β, 1H), 4.42 (dd, J = 5.0, 8.0, H_α, 1H), 6.98 (dd, J = 2.0, 8.5 Hz, ind*H*5, 1H), 7.17 (d, ind*H*4, J = 2.0 Hz, 1H), 7.36 (d, J = 2.5 Hz, ind*H*7, 1H), 7.51 (d, J = 8.5 Hz, ind*H*2, 1H), 8.14 (d, J = 2.0 Hz, indN*H*, 1H), 10.98 (s, br, COO*H*, 1H); **^{13}C NMR** (125 MHz, DMSO – d_6): δ 20.78 (Ac*C*H$_3$), 26.95 (C_β), 52.91 (C_α), 110.39 (ind*C*3), 110.98 (ind*C*4), 118.69 (ind*C*7), 119.60 (ind*C*5), 124.74 (ind*C*2), 125.67 (ind*C*3a), 126.06 (ind*C*7a), 136.42 (ind*C*6), 170.37 (Ac*C*O), 173.45 (*C*OOH); **MS** (ES+): *m/e* 281.1 (34%) [m^{35}Cl+H]$^+$, 283.1 (10%) [m^{37}Cl+H]$^+$, 303.1 (100%) [m^{35}Cl+Na]$^+$, 305.2 (36%) [m^{37}Cl+Na]$^+$; **Mp.:** 142 °C

N^α-Acetyl-6-bromo-*D*-tryptophan (**2'j**)

Using the general method above (§4.3.1.b), the title compound was synthesised as a cream-white powder.

^{1}H NMR (500 MHz, DMSO – d_6): δ 1.78 (s, AcCH_3, 3H), 2.96 (dd, J = 9.0, 15.0 Hz, H_β, 1H), 3.11 (dd, J = 5.0, 14.0 Hz, H_β, 1H), 4.42 (dd, J = 7.0, 11.0, H_α, 1H), 7.10 (dd, J = 2.0, 8.5 Hz, ind*H*5, 1H), 7.16 (d, ind*H*4, J = 2.5 Hz, 1H), 7.47 (d, J = 9.0 Hz, ind*H2*, 1H), 7.50 (d, J = 1.5 Hz, ind*H*7, 1H), 8.11 (d, J = 7.5 Hz, indN*H*, 1H), 10.65 (s, br, COO*H*, 1H); **^{13}C NMR** (125 MHz, DMSO – d_6): δ 20.76 (Ac*C*H$_3$), 26.92 (C_β), 52.89 (C_α), 110.40 (ind*C*3), 113.68 (ind*C*4), 113.90 (ind*C*7), 120.00 (ind*C*5), 121.21 (ind*C*2), 124.66 (ind*C*3a), 126.29 (ind*C*7a), 136.92 (ind*C*6), 170.34 (Ac*C*O), 173.40 (*C*OOH); **MS** (ES+): *m/e* 347.1 (100%) [m^{79}Br+Na]$^+$, 349.1 (97%) [m^{81}Br+Na]$^+$; **Mp.:** 176 °C

N^α-Acetyl-6-methyl-*D*-tryptophan (**2'k**)

Using the general method above (§4.3.1.b), the title compound was synthesised as a cream-white powder.

^{1}H NMR (500 MHz, DMSO – d_6): δ 1.78 (s, AcCH_3, 3H), 2.36, (s, CH_3-Ar, 3H), 2.93 (dd, J = 8.5, 14.5 Hz, H_β, 1H), 3.10 (dd, J = 5.5, 15.0 Hz, H_β, 1H), 4.42 (dd, J = 7.0, 9.5, H$_\alpha$, 1H), 6.80 (dd, J = 4.0, 8.0 Hz, ind*H*5, 1H), 7.02 (d, ind*H*4, J = 2.0 Hz, 1H), 7.10 (s, ind*H7*, 1H), 7.38 (d, J = 8.0 Hz, ind*H*2, 1H), 8.11 (d, J = 9.5 Hz, indN*H*, 1H), 10.65 (s, br, COO*H*, 1H); **^{13}C NMR** (125 MHz, DMSO – d_6): δ 20.77 (Ac*C*H$_3$), 22.41 (*C*H$_3$-Ar), 27.22 (C_β), 52.98 (C_α), 109.80 (ind*C*3), 111.18 (ind*C*4), 117.88 (ind*C*7), 120.12 (ind*C*5), 122.77 (ind*C*2),

125.16 (ind*C*3a), 129.87 (ind*C*7a), 136.53 (ind*C*6), 170.36 (Ac*C*O), 173.59 (*C*OOH); **MS** (ES+): *m/e* 283.2 (100%) $[m+H]^+$, 284.2 (16%) $[m+H]^+$; **Mp.:** 210 °C

N^{α}-Acetyl-6-methoxy-*D*-tryptophan (**2'l**)

Using the general method above (§4.3.1.b), the title compound was synthesised as a dark-green crystalline powder.

^{1}H NMR (500 MHz, DMSO – d_6): δ 1.78 (s, AcCH_3, 3H), 2.94 (dd, *J* = 8.5, 14.5 Hz, H_β, 1H), 3.06 (dd, *J* = 5.5, 14.5 Hz, H_β, 1H), 3.73, (s, CH_3O-Ar, 3H), 4.40 (dd, *J* = 7.0, 9.5, H_α, 1H), 6.63 (dd, *J* = 3.0, 11.0 Hz, ind*H*5, 1H), 6.81 (d, ind*H*4, *J* = 3.0 Hz, 1H), 7.30-743 (m, ind*H7* and ind*H*2, 2H), 8.11 (d, *J* = 9.5 Hz, indN*H*, 1H), 10.62 (s, br, COO*H*, 1H); **^{13}C NMR** (125 MHz, DMSO – d_6): δ 20.43 (Ac*C*H$_3$), 27.24 (C_β), 55.15 (C_α), 59.79 (*C*H$_3$O-Ar), 94.45 (ind*C*3), 108.60 (ind*C*4), 109.98 (ind*C*7), 118.77 (ind*C*5), 121.60 (ind*C*2), 122.09 (ind*C*3a), 136.79 (ind*C*7a), 155.48 (ind*C*6), 170.39 (Ac*C*O), 173.62 (*C*OOH); **MS** (ES+): *m/e* 299.2 (100%) $[m+Na]^+$, 300.2 (16%) $[m+Na]^+$; **Mp.:** 138 °C

N^{α}-acetyl-6-nitro-*D*-tryptophan (**2'm**)

Using the general method above (§4.3.1.b), the title compound was synthesised as a dark-yellow crystalline powder.

^{1}H NMR (500 MHz, DMSO – d_6): δ 1.78 (s, AcCH_3, 3H), 2.95 (dd, *J* = 8.0, 15.0 Hz, H_β, 1H), 3.09 (dd, *J* = 5.0, 15.0 Hz, H_β, 1H), 4.41 (dd, *J* = 7.0, 9.0, H_α, 1H), 6.63 (dd, *J* = 3.0, 11.0 Hz, ind*H*5, 1H), 6.81 (d, ind*H*4, *J* = 3.0 Hz, 1H), 8.34 (m, ind*H7* and ind*H*2, 1H), 8.89 (d, *J* = 9.0 Hz, indN*H*, 1H), 10.64 (s, br, COO*H*, 1H); **^{13}C NMR** (125 MHz, DMSO – d_6): δ 20.43 (Ac*C*H$_3$), 27.24 (C_β), 55.15 (C_α), 94.45 (ind*C*3), 108.60 (ind*C*4), 109.98 (ind*C*7), 118.77 (ind*C*5), 121.60 (ind*C*2), 122.09 (ind*C*3a), 136.79 (ind*C*7a), 155.48 (ind*C*6), 170.39 (Ac*C*O), 173.62 (*C*OOH); **MS** (ES+): *m/e* 299.2 (100%) $[m+Na]^+$, 300.2 (16%) $[m+Na]^+$; **Mp.:** 158 °C

4.4.1.e *5-substituted L-tryptophan derivatives (**2a-g**)*

ID	Substituent	Yield (%)
2a	R = F	99
2b	R = Cl	82
2c	R = Br	96
2d	R = I	81
2e	R = Me	99
2f	R = MeO	99
2g	R = NO_2	70

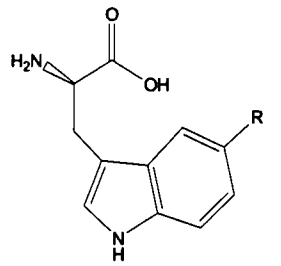


Table 4.5: Table of 5-*L*-tryptophan derivatives that were produced *via* enzymatic resolution of a racemic mixture of the parent N^{α}-Ac-5-R-*D,L*-tryptophan compound (**1a-d**, **1f** & **1g**). The parent compound of **2e**, N^{α}-Ac-5-Me-*D,L*-tryptophan, was purchased from Sigma Aldrich.

5-Fluoro-*L*-tryptophan (**2a**)

Using the general method above (§4.3.1.b), the title compound was synthesised as a cream-white crystalline powder.

^{1}H NMR (400 MHz, MeOH – d_4): δ 3.14 (dd, *J* = 8.8, 15.2 Hz, H_β, 2H), 3.82 (dd, *J* = 4.0, 9.2 Hz, H_α, 1H), 6.88 (dt, *J* = 1.6, 9.2 Hz, ind*H*6, 1H), 7.24 (s, ind*H*2, 1H), 7.31 (dd, *J* = 4.8, 8.4 Hz, ind*H*7, 1H), 7.40 (dd, *J* = 2.4, 10.0 Hz, ind*H*4, 1H); **^{13}C NMR** (100 MHz, MeOH – d_4): δ 28.34 (C_β), 56.54 (C_α), 104.05 (ind*C*4), 109.73 (ind*C*3), 110.71 (ind*C*6), 113.24 (ind*C*7), 127.18 (ind*C*2), 128.83 (ind*C*3a), 134.91 (ind*C*7a), 160.26 (ind*C*5), 174.33 (*C*OOH); **MS** (ES+): *m/e* 223.1 (100%) $[m+H]^+$; **Mp.:** 277-279 °C, $[\alpha]_d^{25}$: +7.2° (*c* = 1; aq. HCl); **CHN** analysis (calcd %) C, 59.45; H, 4.99; N, 12.61; (found %) C, 58.85; H, 5.19; N, 12.50

(Lit.: **^{1}H NMR**[11] (300 MHz, MeOH– d_4) δ 3.14 (H_β), 3.84 (H_α), 6.87 (ind*H*6), 7.21 (ind*H*2), 7.27 (ind*H*7), 7.33 (ind*H*4), **^{13}C NMR**[11] (60 MHz, MeOH– d_4) δ 104.2 (ind*C*4), 109.8 (ind*C*3), 110.9 (ind*C*6), 113.2 (ind*C*7), 127.2 (ind*C*2), 128.9 (ind*C*3a), 134.9 (ind*C*7a), 159.1 (ind*C*5); **Mp.:** 275-276 °C[12]; $[\alpha]_d^{20}$: +5.6° (0.98 g/100 ml; aq. HCl)[12])

5-Chloro-*L*-tryptophan (**2b**)

Using the general method above (§4.3.1.b), the title compound was synthesised as a white crystalline powder.

^{1}H NMR (500 MHz, MeOH – d_4): δ 3.14 (dd, J = 9.0, 15.0 Hz, H_β, 1H), 3.43 (dd, J = 4.0, 15.0 Hz, H_β, 1H), 3.83 (dd, J = 4.5, 9.0 Hz, H_α, 1H), 7.07 (d, J = 2.0, 9.0 Hz, ind*H*6, 1H), 7.25 (s, ind*H*2, 1H), 7.32 (d, J = 9.0 Hz, ind*H*7, 1H), 7.73 (d, J = 2.0 Hz, ind*H*4, 1H); **^{13}C NMR** (100 MHz, MeOH – d_4): δ 28.22 (C_β), 57.9 (C_α), 109.47 (ind*C*3), 113.64 (ind*C*4), 118.95 (ind*C*7), 122.87 (ind*C*6), 126.00 (ind*C*2), 126.94 (ind*C*3a), 129.68 (ind*C*7a), 136.71 (ind*C*5), 174.36 (*C*OOH); **MS** (ES+): 239.1 (100%) $[m^{35}Cl + H]^+$, 241.1 (36%) $[m^{37}Cl + H]^+$; **Mp.:** 272-275 °C, $[\alpha]_d^{25}$: +23.2° (c = 1; aq. HCl); **CHN** analysis (calcd %) C, 55.36; H, 4.65; N, 14.85; (found %) C, 54.82; H, 4.79; N, 14.71

(Lit.: **^{1}H NMR**[13] (300 MHz, MeOD) δ 2.90 (H_α), 3.26 (H_β), 3.54 (H_β), 7.05 (ind*H*6), 7.20 (ind*H*2), 7.31 (ind*H*7), 7.73 (ind*H*4); **MS** (EI) *m/e* 238 $[m]^+$, 240 $[m]^+$; **Mp.**: 268-270 °C[14]; $[\alpha]_d^{20}$: +21.8° (1.00 g/100 ml; aq. HCl)[14])

5-Bromo-*L*-tryptophan (**2c**)

Using the general method above (§4.3.1.b), the title compound was synthesised as a sand-coloured powder.

^{1}H NMR (400 MHz, MeOH – d_4): δ 3.08 (dd, J = 8.8, 15.4 Hz, H_β, 1H), 3.43 (dd, J = 4.0, 15.4 Hz, H_β, 1H), 3.80 (dd, J = 4.4, 9.2 Hz, H_α, 1H), 7.19 (d, J = 2.0 Hz, ind*H*6, 1H), 7.21 (m, ind*H*2, 1H), 7.28 (d, J = 8.8 Hz, ind*H*7, 1H), 7.90 (d, J = 1.6 Hz, ind*H*4, 1H); **MS** (ES+): *m/e* 283.0 (46%) $[m^{79}Br+H]^+$, 285.1 (44%) $[m^{81}Br+H]^+$, 305.1 (100%) $[m^{79}Br+Na]^+$, 307.1 (95%) $[m^{81}Br+Na]^+$; **Mp.:** 281-283 °C, $[\alpha]_d^{25}$: +34.3° (c = 1; aq. HCl); **CHN** analysis (calcd %) C, 46.66; H, 3.92; N, 9.89; (found %) C, 46.27; H, 4.23; N, 9.80

(Lit.[7] **Mp.:** 288-291 °C; $[\alpha]_d^{20}$: +31.8° (1.0 g/100 ml; aq. HCl))

5-Iodo-*L*-tryptophan (**2d**)

Using the general method above (§4.3.1.b), the title compound was synthesised as a yellowish-white crystalline powder.

^{1}H NMR (500 MHz, MeOH – d_4): δ 3.11 (dd, J = 9.0, 15.0 Hz, H_β, 1H), 3.43 (dd, J = 4.0, 15.0 Hz, H_β, 1H), 3.83 (dd, J = 4.0, 9.0 Hz, H_α, 1H), 7.18 (d, J = 2.0 Hz, ind*H*6, 1H), 7.19 (s, ind*H*2, 1H), 7.36 (dd, J = 2.0, 9.0 Hz, ind*H*7, 1H), 8.09 (d, J = 1.5 Hz, ind*H*4, 1H); **^{13}C NMR** (100 MHz, MeOH – d_4): δ 28.20 (C_β), 56.56 (C_α), 83.13 (ind*C*3), 109.10 (ind*C*4), 114.59 (ind*C*7), 126.35 (ind*C*6), 128.43 (ind*C*2), 130.00 (ind*C*3a), 131.06 (ind*C*7a), 137.36 (ind*C*5), 174.40 (*C*OOH); **MS** (ES+): *m/e* 331.1 (100%) $[m+H]^+$; **Mp.:** 170-172 °C, $[\alpha]_d^{25}$: +40.4°

(c = 1; aq. HCl); **CHN** analysis (calcd %) C, 40.02; H, 3.36; N, 8.49; (found %) C, 39.72; H, 3.51; N, 8.40

(Lit.[15] **Mp.:** 165-166 °C; $[\alpha]_d^{20}$: -12.3° (1.08 g/100 ml; CF_3CO_2H))

5-Methyl-*L*-tryptophan (**2e**)

Using the general method above (§4.3.1.b), the title compound was synthesised as fine white needle-like crystals.

^{1}H NMR (300 MHz, MeOH – d_4): δ 2.42 (s, CH_3-Ar, 3H), 3.09 (dd, J = 9.0, 15.2 Hz, H_β, 1H), 3.49 (dd, J = 4.5, 15.2 Hz, H_β, 1H), 3.84 (dd, J = 4.5, 9.3 Hz, H_α, 1H), 6.94 (d, J = 4.8 Hz, ind$H$6, 1H), 7.13 (s, ind$H$2, 1H), 7.23 (d, J = 4.8 Hz, ind$H$7, 1H), 7.50 (s, ind$H$4, 1H); **^{13}C NMR** (125 MHz, MeOH – d_4): δ 21.64 (CH$_3$-Ar), 28.54 (C_β), 56.74 (C_α), 109.08 (ind$C$3), 112.15 (ind$C$4), 118.95 (ind$C$7), 124.39 (ind$C$6), 125.18 (ind$C$2), 128.68 (ind$C$5), 129.26 (ind$C$3a), 136.77 (ind$C$7a), 174.45 (COOH); **MS** (ES+): 219.1 (100%) $[m+H]^+$, **Mp.**: 280-283 °C; $[\alpha]_d^{25}$: +9.4° (c = 1.0; aq. HCl); **CHN** analysis (calcd %) C, 66.04; H, 6.47; N, 12.84; (found %) C, 65.81; H, 6.53; N, 12.77

(Lit.: **^{1}H NMR**[16] (300 MHz, MeOD) δ 2.43 (CH_3-Ar), 2.85 (H_β), 3.32 (H_β), 3.57 (H_α), 6.94 (ind$H$6), 7.09 (ind$H$2), 7.25 (ind$H$7), 7.51 (ind$H$4); **MS** (EI) *m/e* 218 $[M^+]$[12]; **Mp.**: 275-277 °C[14]; $[\alpha]_d^{25}$: +10.6° (1.00 g/100 ml; aq. HCl)[14])

5-Methoxy-*L*-tryptophan (**2f**)

Using the general method above (§4.3.1.b), the title compound was synthesised as a brown powder.

^{1}H NMR (400 MHz, MeOH – d_4): δ 3.10 (dd, J = 5.6, 15.2 Hz, H_β, 1H), 3.47 (dd, J = 3.6, 15.2 Hz, H_β, 1H), 3.83 (dd, J = 4.5, 9.3 Hz, H_α, 1H), 3.83 (s, CH_3O-Ar, 1H), 6.76 (dd, J = 2.4, 8.8 Hz, ind$H$6, 1H), 7.15 (s, ind$H$2, 1H), 7.23 (d, J = 0.8 Hz, ind$H$7, 1H), 7.24 (d, J = 4.8 Hz, ind$H$4, 1H); **^{13}C NMR** (100 MHz, MeOH – d_4): δ 28.62 (C_β), 56.25 (CH$_3$O-Ar), 56.51 (C_α), 101.09 (ind$C$3), 109.34 (ind$C$4), 113.13 (ind$C$7), 113.24 (ind$C$6), 125.85 (ind$C$2), 128.73 (ind$C$3a), 133.53 (ind$C$7a), 155.36 (ind$C$5), 174.63 (COOH); **MS** (ES+): *m/e* 234.1 (100%) $[m+H]^+$; **Mp.:** 253-256 °C, $[\alpha]_d^{25}$: -33.7° (c = 1.0; H_2O); **CHN** analysis (calcd %) C, 61.53; H, 6.02; N, 11.96; (found %) C, 61.34; H, 6.11; N, 11.93

(Lit.[14] **Mp.:** 250-252 °C; $[\alpha]_d^{20}$: -28.9° (c = 1.0; H_2O))

5-Nitro-*L*-tryptophan (**2g**)

Using the general method above (§4.3.1.b), the title compound was synthesised as a bright yellow crystalline powder.

^{1}H NMR (500 MHz, MeOH – d_4): δ 3.27 (dd, 1H, J = 8.5, 15.5 Hz), 3.50 (dd, 1H, J = 4.0, 15.5 Hz), 3.88 (dd, 3H, J = 4.5, 8.5 Hz), 7.42 (s, 1H), 7.45 (s, 1H), 8.04 (d, 1H, J = 9.0 Hz), 8.78 (d, 1H, J = 2.0 Hz); **^{13}C NMR** (125 MHz, MeOH – d_4): δ 27.91 (C_β), 56.52 (C_α), 110.56 (ind*C*3), 112.56 (ind*C*4), 117.09 (ind*C*7), 118.14 (ind*C*6), 126.70 (ind*C*2), 128.07 (ind*C*3a), 129.02 (ind*C*7a), 142.71 (ind*C*5), 174.08 (*C*OOH); **MS** (ES+): *m/e* 250.2 (100%) $[m+H]^+$; **Mp.:** 267-269 °C, $[\alpha]_d^{25}$: +45.4° (*c* = 1.0; aq. HCl); **CHN** analysis (calcd %) C, 53.01; H, 4.45; N, 16.86; (found %) C, 52.95; H, 4.49; N, 16.84

(Lit.[14] **Mp.:** 266-268 °C; $[\alpha]_d^{20}$: 49.1° (1.0 g/100 ml; aq. HCl))

4.4.1.f *6-substituted L-tryptophan derivatives (**2h-m**)*

No.	Substituent	Yield (%)
2h	R = F	67
2i	R = Cl	54
2j	R = Br	99
2k	R = Me	94
2l	R = MeO	65
2m	R = NO_2	72

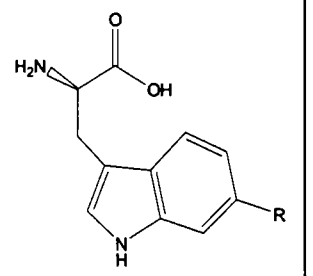


Table 4.6: Table of 6-*L*-tryptophan derivatives that were produced *via* enzymatic resolution of the parent N^α-Ac-6-R-*D,L*-tryptophan compound (**1h-m**).

6-Fluoro-*L*-tryptophan (**2h**)

Using the general method above (§4.3.1.b), the title compound was synthesised as a pinkish-white crystalline powder.

^{1}H NMR (400 MHz, MeOH – d_4): δ 3.11 (dd, *J* = 9.2, 15.6 Hz, H_β, 1H), 3.45 (dd, *J* = 4.4, 15.6 Hz, H_β, 1H), 3.80 (dd, *J* = 4.0, 9.2 Hz, H_α, 1H), 6.82 (dt, *J* = 2.0, 6.1 Hz, ind*H*5, 1H), 7.04 (dd, *J* = 2.4, 10.0 Hz, ind*H*7, 1H), 7.17 (s, ind*H*2, 1H), 7.65 (dd, *J* = 5.2, 8.8 Hz, ind*H*4, 1H); **^{13}C NMR** (100 MHz, MeOH – d_4): δ 28.69 (C_β), 56.67 (C_α), 98.34 (ind*C*7), 108.36 (ind*C*5), 108.61 (ind*C*3), 120.35 (ind*C*4), 125.26 (ind*C*3a), 125.62 (ind*C*2), 135.26 (ind*C*7a), 161.45 (ind*C*6), 174.91 (*C*OOH); **MS** (ES+): *m/e* 235.1 (100%) $[m+H]^+$; **Mp.:** 262-265 °C,

$[\alpha]_d^{25}$: -17.7° (*c* = 1.0; H_2O); **CHN** analysis (calcd %) C, 59.45; H, 4.99; N, 12.61; (found %) C, 59.27; H, 5.11; N, 12.57

(Lit.: **^{1}H NMR**[16] (300 MHz, MeOH – d_4): δ 3.13 (H_β), 3.47 (H_β), 3.84 (H_α), 6.82 (ind*H*5), 7.05 (ind*H*7), 7.17 (ind*H*2), 7.64 (ind*H*4); **^{13}C NMR** (60 MHz, MeOH – d_4): δ 98.3 (ind*C*7), 108.5 (ind*C*5), 109.8 (ind*C*3), 120.4 (ind*C*4), 125.2 (ind*C*3a), 125.6 (ind*C*2), 138.3 (ind*C* 7a), 161.4(ind*C*6); **Mp**.: 255-258 °C[17]; **$[\alpha]_D^{20}$**: -19.5° (*c* = 1.0, H_2O)[17])

6-Chloro-*L*-tryptophan (**2i**)

Using the general method above (§4.3.1.b), the title compound was synthesised as a white crystalline powder.

^{1}H NMR (400 MHz, MeOH – d_4): δ 3.15 (dd, *J* = 9.2, 15.6 Hz, H_β, 1H), 3.46 (dd, *J* = 4.0, 15.6 Hz, H_β, 1H), 3.82 (dd, 1H, *J* = 4.0, 9.2 Hz), 7.00 (dd, 1H, *J* = 2.0, 8.8 Hz), 7.21 (s, 1H), 7.35 (d, 1H, *J* = 1.6 Hz), 7.65 (d, 1H, *J* = 8.4 Hz); **^{13}C NMR** (100 MHz, MeOH – d_4): δ 28.26 (C_β), 56.58 (C_α), 110.02 (ind*C*7), 115.26 (ind*C*3), 116.18 (ind*C*4), 120.96 (ind*C*6), 123.21 (ind*C*2), 126.09 (ind*C*3a), 127.53 (ind*C*5), 139.19 (ind*C*7a), 174.28 (*C*OOH); **MS** (ES+): *m/e* 239 (100%) $[m^{35}Cl + H]^+$, 241.0 (36%) $[m^{37}Cl + H]^+$, **Mp**.: 270-273 °C; **$[\alpha]_D^{25}$**: -25.4° (*c* = 1.0, MeOH); **CHN** analysis (calcd %) C, 55.36; H, 4.65; N, 14.85; (found %) C, 55.22; H, 4.68; N, 14.83

(Lit.: **^{1}H NMR**[13] (300 MHz, MeOH – d_4): δ 2.93, 3.27, 3.55, 6.99, 7.18, 7.35, 7.67; **^{13}C NMR**[13] (60 MHz, MeOH – d_4): δ 32.5, 57.9, 111.9, 112.9, 120.9, 125.5, 127.8, 128.2, 138.5, 182.2; **MS:** *m/e* (EI) 238.0 $[m^+]$[18]; **Mp**.: 264-265 °C (*D*-isomer)[19]; **$[\alpha]_D^{23}$**: 28° (*c* = 1.0, MeOH) (*D*-isomer)[19])

6-Bromo-*L*-tryptophan (**2j**)

Using the general method above (§4.3.1.b), the title compound was synthesised as a pinkish-white crystalline powder.

^{1}H NMR (400 MHz, MeOH – d_4): δ 3.15 (dd, *J* = 7.2, 12.4 Hz, H_β, 1H), 3.46 (dd, *J* = 3.2, 12.4 Hz, H_β, 1H), 3.82 (dd, *J* = 3.6, 7.2 Hz, H_α, 1H), 7.14 (dd, *J* = 1.2, 6.8 Hz, ind*H*5, 1H), 7.19 (s, ind*H*7, 1H), 7.52 (d, *J* = 1.2 Hz, ind*H*2, 1H), 7.61 (d, *J* = 6.8 Hz, ind*H*4, 1H); **^{13}C NMR** (100 MHz, MeOH – d_4): δ 28.21 (C_β), 56.49 (C_α), 109.89 (ind*C*3), 112.22 (ind*C*4), 120.54 (ind*C*3), 120.60 (ind*C*6), 126.31 (ind*C*2), 127.23 (ind*C*3a), 128.59 (ind*C*5), 138.68 (ind*C*7a), 174.57 (*C*OOH); **MS** (ES+): *m/e* 283.2 (100%) $[m^{79}Br + H]^+$, 285.2 (95%)

$[m^{81}Br+H]^+$; **Mp**.: 120-122 °C; $[\alpha]_D^{25}$: -11.3° (*c* = 0.38, AcOH); **CHN** analysis (calcd %) C, 46.66; H, 3.92; N, 9.89; (found %) C, 46.56; H, 3.96; N, 9.87

(Lit. **Mp**.: 125-127 °C[1]; $[\alpha]_D^{20}$: -12.1° (*c* = 0.38, AcOH)[20])

6-Methyl-*L*-tryptophan (**2k**)

Using the general method above (§4.3.1.b), the title compound was synthesised as an orange-white crystalline powder.

^{1}H NMR (400 MHz, MeOH – d_4): δ 3.08 (dd, *J* = 8.0, 12.0 Hz, H_β, 1H), 3.34 (s, Ar-CH_3, 3H), 3.48 (dd, *J* = 2.8, 12.0 Hz, H_β, 1H), 3.83 (dd, *J* = 3.2, 7.6 Hz, H_α, 1H), 6.88 (d, *J* = 6.4 Hz, ind*H*5, 1H), 7.10 (s, ind*H*7, 1H), 7.15 (s, ind*H*2, 1H), 7.57 (d, *J* = 6.4 Hz, ind*H*4, 1H); **^{13}C NMR** (125 MHz, MeOH – d_4): δ 21.79 (Ar-*C*H_3), 28.57 (C_β), 56.68 (C_α), 109.41 (ind*C*7), 112.29 (ind*C*3), 119.07 (ind*C*4), 121.87 (ind*C*5), 124.48 (ind*C*2), 126.39 (ind*C*3a), 132.47 (ind*C*7a), 138.87 (ind*C*6), 174.48 (*C*OOH); **MS** (ES+): *m/e* 219.1 (13%) $[m+H]^+$, 241.1 (100%) $[m+Na]^+$; **Mp**.: 290-293 °C; $[\alpha]_D^{25}$: -18.4° (*c* = 0.4, MeOH); **CHN** analysis (calcd %) C, 66.04; H, 6.47; N, 12.84; (found %) C, 65.96; H, 6.50; N, 12.81

(Lit.: **^{1}H NMR**[21] (300 MHz, MeOH – d_4): δ 2.43, 2.88, 3.31, 3.56, 6.86, 7.06, 7.14, 7.58; **^{13}C NMR** (60 MHz, MeOH – d_4): δ 112.0 (ind*C*7), 112.3 (ind*C*3), 119.4 (ind*C*4), 121.4 (ind*C*5), 124.8 (ind*C*2), 127.0 (ind*C*3a), 131.7 (ind*C*7a); **Mp**.: 286 °C (*D*-isomer)[9]; $[\alpha]_D^{25}$: 19.8° (*c* = 0.4, MeOH) (*D*-isomer)[9])

6-Methoxy-*L*-tryptophan (**2l**)

Using the general method above (§4.3.1.b), the title compound was synthesised as white planar crystals.

^{1}H NMR (400 MHz, MeOH – d_4): δ 3.11 (dd, *J* = 6.0, 15.2 Hz, H_β, 1H), 3.44 (dd, *J* = 4.0, 15.2 Hz, H_β, 1H), 3.79 (s, Ar-OCH_3, 3H), 3.80-3.84 (m, H_α, 1H), 6.76 (dd, *J* = 2.4, 8.4 Hz, ind*H*5, 1H), 6.86 (s, ind*H*7, 1H), 7.24 (d, *J* = 0.8 Hz, ind*H*2, 1H), 7.26 (d, *J* = 4.8 Hz, ind*H*4, 1H); **^{13}C NMR** (100 MHz, MeOH – d_4): δ 28.85 (C_β), 55.21 (C_α), 58.24 (Ar-O*C*H_3), 101.15 (ind*C*7), 109.40 (ind*C*3), 113.18 (ind*C*4), 113.27 (ind*C*5), 125.83 (ind*C*2), 128.76 (ind*C*3a), 133.52 (ind*C*7a), 155.46 (ind*C*6), 174.65 (*C*OOH); **MS** (ES+): *m/e* 235.2 (100%) $[m+H]^+$, 259.2 $[m+Na]^+$; **Mp**.: 264-268 °C; $[\alpha]_D^{25}$: -23.7° (*c* = 0.1, MeOH); **CHN** analysis (calcd %) C, 61.53; H, 6.02; N, 11.96; (found %) C, 62.04; H, 6.25; N, 12.03

(Lit.[9] **Mp**.: 263 - 268 °C)

6-Nitro-*L*-tryptophan (**2m**)

Using the general method above (§4.3.1.b), the title compound was synthesised as a dark-yellow crystalline powder.

^{1}H NMR (400 MHz, MeOH – d_4): δ 3.09 (dd, *J* = 5.6, 15.2 Hz, H_β, 1H), 3.42 (dd, *J* = 5.2, 15.2 Hz, H_β, 1H), 3.86 (dd, *J* = 5.6, 8.4 Hz, H_α, 1H), 7.15 (d, *J* = 0.8 Hz, ind*H*2, 1H), 7.44 (d, *J* = 4.8 Hz, ind*H*4, 1H), 8.01(dd, *J* = 2.0, 8.4 Hz, ind*H*5, 1H), 8.11 (s, ind*H*7, 1H); **^{13}C NMR** (100 MHz, MeOH – d_4): δ 30.10 (C_β), 56.83 (C_α), 105.46 (ind*C*7), 111.26 (ind*C*3), 116.17 (ind*C*5), 119.48 (ind*C*4), 123.83 (ind*C*2), 130.17 (ind*C*6), 130.92 (ind*C*3a), 135.79 (ind*C*7a), 174.75 (*C*OOH); **MS** (ES+): *m/e* 235.2 (12%) $[m+H]^+$, 259.2 (100%) $[m+Na]^+$, **Mp**.: 270 °C; **$[\alpha]_D^{25}$**: -21.6° (*c* = 0.1, MeOH); **CHN** analysis (calcd %) C, 53.01; H, 4.45; N, 16.86; (found %) C, 52.96; H, 4.48; N, 16.84

(Lit.[22] **Mp**.: 265 °C (dec.))

*4.4.1.g 5-substituted N^α-acetyl-L-tryptophan derivatives (**4n**, **4f**, **4g**)*

N^α-Acetyl-*L*-tryptophan (**4n**)

Using the general method above (§4.3.1.c), the title compound was synthesised as a yellowish-white crystalline powder.

^{1}H NMR (400 MHz, MeOH – d_4): δ 1.91 (s, AcCH_3, 3H), 3.13 (dd, *J* = 8.0, 15.2 Hz, H_β, 1H), 3.30 (dd, *J* = 4.8, 15.2 Hz, H_β, 1H), 3.81 (dd, *J* = 5.2, 8.0 Hz, H_α, 1H), 6.97 (t, *J* = 2.4 Hz, ind*H*5, 1H), 7.13 (t, *J* = 2.4 Hz, ind*H*6, 1H), 7.18 (d, *J* = 1.2 Hz, ind*H*2, 1H), 7.34 (d, *J* = 2.4 Hz, ind*H*7, 1H), 7.58 (dd, *J* = 2.4, 4.8 Hz, ind*H*4, 1H); **^{13}C NMR** (100 MHz, MeOH – d_4): δ 21.89 (AcCH_3), 29.14 (C_β), 56.13 (C_α), 109.82 (ind*C*3), 111.35 (ind*C*7), 117.22 (ind*C*4), 119.46 (ind*C*6), 122.75 (ind*C*5), 123.40 (ind*C*2), 127.67 (ind*C*3a), 135.20 (ind*C*7a), 174.80 (*C*OOH); **MS** (ES+): *m/e* 247.2 (14%) $[m+H]^+$, 270.2 (100%) $[m+Na]^+$, **Mp**.: 178-181 °C; **$[\alpha]_D^{25}$**: +24.7° (*c* = 1.00, MeOH)

(Lit. **Mp**.:[23] 174-179 °C , **$[\alpha]_D^{20}$**: +26.0° (*c* = 1.00, MeOH)[24])

N^α-Acetyl-5-methoxy-*L*-tryptophan (**4f**)

Using the general method above (§4.3.1.c), the title compound was synthesised as a light-brown crystalline powder.

^{1}H NMR (400 MHz, MeOH – d_4): δ 1.91 (s, AcCH_3, 3H), 3.12 (dd, J = 8.0, 14.8 Hz, H_β, 1H), 3.30 (dd, J = 4.8, 14.8 Hz, H_β, 1H), 3.81 (s, Ar-OCH_3, 3H), 4.69 (dd, J = 5.2, 8.0 Hz, H_α, 1H), 6.74 (dd, J = 2.4, 8.8 Hz, ind$H$6, 1H), 7.05 (s, ind$H$2 & ind$H$7, 2H), 7.20 (d, J = 8.8 Hz, ind$H$4, 1H); **^{13}C NMR** (100 MHz, MeOH – d_4): δ 20.75 (AcCH$_3$), 28.52 (C_β), 54.73 (C_α), 56.24 (Ar-OCH$_3$), 101.10 (ind$C$7), 110.81 (ind$C$3), 112.72 (ind$C$4), 122.34 (ind$C$5), 125.07 (ind$C$2), 129.17 (ind$C$3a), 133.22 (ind$C$7a), 155.08 (ind$C$6), 173.19 (AcCO), 175.28 (COOH); **MS** (ES+): *m/e* 276.2 (15%) $[m+H]^+$, 298.2 (100%) $[m+Na]^+$; **$[\alpha]_D^{25}$:** 14.2° (c = 1.22, MeOH)
(Lit.[14] **$[\alpha]_D^{16}$:** 16° (c = 1.22, MeOH))

N^α-Acetyl-5-nitro-*L*-tryptophan (**4g**)
Using the general method above (§4.3.1.c), the title compound was synthesised as a yellow crystalline powder.
^{1}H NMR (400 MHz, MeOH – d_4): δ 1.92 (s, AcCH_3, 3H), 3.15 (dd, J = 8.4, 14.8 Hz, H_β, 1H), 3.31 (dd, J = 5.2, 14.8 Hz, H_β, 1H). 4.65 (dd, J = 5.6, 8.4 Hz, H_α, 1H), 7.19 (dd, J = 2.4, 8.8 Hz, ind$H$6, 1H), 7.45 (d, J = 2.4 Hz, ind$H$2, 1H), 8.08 (s, ind$H$4, 1H), 8.13 (d, J = 8.8 Hz, ind$H$7, 1H); **^{13}C NMR** (100 MHz, $CDCl_3$): δ 23.74 (AcCH$_3$), 29.97 (C_β), 55.18 (C_α), 109.79 (ind$C$7), 112.15 (ind$C$3), 114.60 (ind$C$4), 122.52 (ind$C$5), 125.78 (ind$C$2), 129.12 (ind$C$3a), 132.72 (ind$C$7a), 144.16 (ind$C$6), 171.67 (AcCO), 174.44 (COOH); **MS** (ES+): *m/e* 314.1 (100%) $[m+Na]^+$

The remaining N^α-acetylated tryptophan derivatives were not further analyzed. This was justified as the most electron-rich and the most electron deficient tryptophan derivatives both reacted successfully in this standard acetylation. The electron density is not thought to have a strong impact on the reaction as the reaction takes place at the rather remote N^α position.

*4.4.1.h 5-subsituted N^α-acetyl-L-tryptophan methyl ester derivatives (**5a**, **5f**+**g**, **5n**)*

N^α-Acetyl-*L*-tryptophan methyl ester (**5n**)
Using the general method above (§4.3.1.d), the title compound was synthesised as a light-brown crystalline powder.

^{1}H NMR (500 MHz, MeOH – d_4): δ 1.91 (s, AcCH_3, 3H), 3.14 (dd, J = 8.0, 14.5 Hz, H_β, 1H), 3.27 (dd, J = 6.0, 14.5 Hz, H_β, 1H), 3.64 (s, Ar-CH_3, 3H), 4.71 (dd, J = 6.0, 7.5 Hz, H_α, 1H), 7.00 (dt, J = 1.0, 7.5 Hz, ind$H$6, 1H), 7.08 (dt, J = 1.0, 7.5 Hz, ind$H$2, 1H), 7.31 (d, J = 8.0 Hz, ind$H$7, 1H), 7.50 (d, J = 8.0 Hz, ind$H$4, 1H); **^{13}C NMR** (125 MHz, MeOH – d_4): δ 22.31 (*C*OO*C*H$_3$), 28.50 (C_β), 52.64 (Ac*C*H$_3$), 55.02 (C_α), 110.73 (ind*C*3), 112.30 (ind*C*4), 119.09 (ind*C*7), 119.80 (ind*C*6), 122.43 (ind*C*2), 124.32 (ind*C*5), 128.72 (ind*C*3a), 138.04 (ind*C*7a), 173.21 (Ac*C*O), 174.05 (*C*OOH); **Mp**.: 144 °C; **CHN**: 63.40% C, 5.73% H, 11.38% N (theoretical 62.04% C, 5.87% H, 11.03% N)
(Lit.[25] **Mp**.: 155-156 °C), **R$_f$**: 0.52 (using EtOH 5% v/v in $CHCl_3$)

N^α-Acetyl-5-fluoro-*L*-tryptophan methyl ester (**5a**)
Using the general method above (§4.3.1.d), the title compound was synthesised as a light-brown crystalline powder.
^{1}H NMR (500 MHz, $CDCl_3$): δ 1.97 (s, AcCH_3, 3H), 3.23 (dd, J = 5.0, 15.0 Hz, H_β, 1H), 3.28 (dd, J = 5.5, 15.0 Hz, H_β, 1H), 3.69 (s, COOCH_3, 3H), 4.92 (quart, J = 5.5, 15.0 Hz, H_α, 1H), 6.12 (d, J = 7.0 Hz, AcNH, 1H), 6.92 (dt, J = 2.5, 9.0 Hz, ind$H$6, 1H), 7.00 (d, J = 2.5 Hz, ind$H$2, 1H), 7.13 (dd, J = 2.5, 10.0 Hz, ind$H$4, 1H), 7.24 (t, J = 4.5 Hz, ind$H$7, 1H), 8.46 (s, br, indNH, 1H); **^{13}C NMR** (125 MHz, $CDCl_3$): δ 23.16 (Ac*C*H$_3$), 27.60 (C_β), 52.41 (COO*C*H$_3$), 52.95 (C_α), 103.44 (ind*C*3), 110.13 (ind*C*4), 110.49 (ind*C*7), 112.017 (ind*C*6), 124.55 (ind*C*2), 128.07 (ind*C*3a), 132.57 (ind*C*7a), 158.84 (ind*C*5), 169.88 (Ac*C*O), 172.33 (*C*OOH); **MS** (ES+): *m/e* 279.1 (14%) $[m+H]^+$, 301.1 (100%) $[m+Na]^+$; **R$_f$**: 0.68 (using EtOH 5% v/v in DCM)

N^α-Acetyl-5-methoxy-*L*-tryptophan methyl ester (**5f**)
Using the general method above (§4.3.1.d), the title compound was synthesised as a brown crystalline powder.
^{1}H NMR (400 MHz, $CDCl_3$): δ 1.98 (s, AcCH_3, 3H), 3.24-3.34 (m, H_β, 2H), 3.71 (s, AcCH_3, 3H), 3.82 (s, Ar-OCH_3, 3H), 4.95 (dd, J = 2.4, 6.8 Hz, H_α, 1H), 6.02 (d, J = 8.8 Hz, AcNH, 1H), 6.86 (dd, J = 2.4, 8.8 Hz, ind$H$6, 1H), 6.95 (d, J = 2.4 Hz, ind$H$2, 1H), 6.98 (d, J = 2.4 Hz, ind$H$4, 1H), 7.24 (d, J = 8.8 Hz, ind$H$7,1H), 8.01 (s, br, indNH, 1H); **^{13}C NMR** (100 MHz, $CDCl_3$): δ 25.30 (Ac*C*H$_3$), 29.61 (C_β), 54.40 (COO*C*H$_3$), 55.00 (C_α), 56.57 (Ar-O*C*H$_3$), 102.13 (ind*C*3), 111.85 (ind*C*4), 114.01 (ind*C*7), 114.68 (ind*C*6), 125.33 (ind*C*2),

130.14(ind*C*3a), 133.18 (ind*C*7a), 156.29 (ind*C*5), 171.75 (Ac*C*O), 174.42 (*C*OOH); **MS** (ES+): *m/e* 291.1 (9%) $[m+H]^+$, 313.1 (100%) $[m+Na]^+$; $\mathbf{R_f}$: 0.52 (using EtOH 5% v/v in DCM)

N^{α}-Acetyl-5-nitro-*L*-tryptophan methyl ester (**5g**)

Using the general method above (§4.3.1.d), the title compound was synthesised as a yellow crystalline powder.

^{1}H NMR (400 MHz, $CDCl_3$): δ 1.97 (s, AcCH_3, 3H), 3.22-3.32 (m, H_β, 2H), 3.68 (s, AcCH_3, 3H), 4.96 (dd, J = 2.4, 6.8 Hz, H_α, 1H), 7.18 (d, J = 8.8 Hz, AcN*H*, 1H), 7.38 (dd, J = 2.4, 8.8 Hz, ind*H*6, 1H), 7.44 (d, J = 2.4 Hz, ind*H*2, 1H), 8.05 (s, ind*H*4, 1H), 8.11 (d, J = 8.8 Hz, ind*H*7, 1H), 8.33 (s, br, indN*H*, 1H); **^{13}C NMR** (100 MHz, $CDCl_3$): δ 23.98 (AcCH_3), 30.01 (C_β), 52.43 (COOCH_3), 54.81 (C_α), 109.74 (ind*C*3), 112.11 (ind*C*4), 114.57 (ind*C*7), 122.43 (ind*C*6), 125.87 (ind*C*2), 129.15 (ind*C*3a), 132.68 (ind*C*7a), 144.13 (ind*C*5), 171.64 (Ac*C*O), 174.43 (*C*OOH); **MS** (ES+): *m/e* 306.1 (11%) $[m+H]^+$, 328.1 (100%) $[m+Na]^+$; $\mathbf{R_f}$: 0.42 (using EtOH 5% v/v in DCM)

The remaining N^{α}-acetylated tryptophan methyl esters were not further analyzed but used directly for the ethyl amidation.

*4.4.1.i 5-substituted N^{α}-acetyl-L-tryptophan ethyl amide derivatives (**6n**, **6a-g**)*

N^{α}-Acetyl-*L*-tryptophan ethyl amide (**6n**)

Using the general method above (§4.3.1.e), the title compound was synthesised as a cream-coloured crystalline powder.

^{1}H NMR (200 MHz, $CDCl_3$): δ 0.85 (t, J = 7.4 Hz, CH_3CH_2N, 3H), 1.96 (s, AcCH_3, 3H), 3.02 – 3.13 (m, CH_3CH_2N & 1H_β, 3H), 3.27 (dd, J = 5.2, 14.2 Hz, H_β, 1H), 4.65 (quart., J = 5.2 Hz, H_α, 1H), 5.55 (s, br, EtN*H*, 1H), 6.46 (d, J = 5.2 Hz, AcN*H*, 1H), 7.02 (d, J = 2.4 Hz, ind*H*2, 1H), 7.10 (dt, J = 1.2, 7.0 Hz, ind*H*5, 1H), 7.18 (dt, J = 1.2, 7.0 Hz, ind*H*6, 1H), 7.34 (d, J = 7.2 Hz, ind*H*4, 1H), 7.68 (d, J = 7.2 Hz, ind*H*7, 1H), 8.26 (s, br, indN*H*, 1H); **^{13}C NMR** (125 MHz, MeOH – d_4): δ 14.81 (NCH_2CH_3), 22.56 (AcCH_3), 30.12 (C_β), 54.25 (NCH_2CH_3), 54.71 (C_α), 110.73 (ind*C*3), 111.38 (ind*C*4), 119.09 (ind*C*5), 119.82 (ind*C*6), 122.45, 124.30, 128.75, 138.07, 173.25 (Ac*C*O), 174.07 (*C*ONEt); **MS** (ES+): *m/e* 296.1

(100%) $[m+Na]^+$; **HRMS** (FT-MS) calcd for $C_{15}H_{19}N_3O_2Na_1$: 296.1370, found: 296.1372; **Mp.**: 160-162 °C; $[\alpha]_d^{25}$: (MeOH, *c* = 0.7) +4.1°; R_f: 0.41(using MeOH 10% v/v in DCM)

N^{α}-Acetyl-5-fluoro-*L*-tryptophan ethyl amide (**6a**)

^{1}H NMR (500 MHz, $CDCl_3$): δ 0.87 (t, *J* = 7.5 Hz, CH_3CH_2N, 3H), 2.00 (s, $AcCH_3$, 3H), 3.01 (dd, *J* = 9.5, 14.5 Hz, H_β, 1H), 3.04-3.17 (m, CH_3CH_2N, 2H), 3.28 (dd, *J* = 5.0, 14.5 Hz, H_β, 1H), 3.87 (3H, s), 4.60-4.65 (m, H_α, 1H), 5.39 (s, br, EtN*H*, 1H), 6.39 (d, *J* = 7.5 Hz, AcN*H*, 1H), 6.85 (dd, *J* = 2.5, 9.0 Hz, ind*H*6, 1H), 7.01 (d, *J* = 2.5 Hz, ind*H*2, 1H), 7.22 (d, *J* = 2.5 Hz, ind*H*4, 1H), 7.24 (d, *J* = 9.0 Hz, ind*H*7, 1H), 8.00 (s, br, indN*H*, 1H); **^{13}C NMR** (100 MHz, $CDCl_3$): δ 16.43 (NCH_2CH_3), 25.36 ($AcCH_3$), 30.96 (C_β), 54.34 (C_α), 55.81 (NCH_2CH_3), 102.28 (ind*C*3), 112.88 (ind*C*4), 114.01 (ind*C*7), 115.01 (ind*C*6), 125.58 (ind*C*2), 130.14 (ind*C*3a), 133.17 (ind*C*7a), 156.35 (ind*C*5), 170.56 (Ac*C*O), 172.87 (*C*ONEt); **MS** (ES+): *m/e* 314.1 (100) $[m+Na]^+$, 314.1 (17) $[m+Na]^+$; **HRMS** (FT-MS) calcd for $C_{15}H_{19}N_3O_2F_1Na_1$: 314.1277, found: 312.1275; **Mp.**: 190-195 °C (dec.); $[\alpha]_d^{25}$: (MeOH, *c* = 0.1) +29.8°; R_f: 0.39 (using EtOH 5% v/v in DCM)

N^{α}-Acetyl-5-chloro-*L*-tryptophan ethyl amide (**6b**)

^{1}H NMR (500 MHz, $CDCl_3$): δ 0.90 (t, *J* = 7.5 Hz, CH_3CH_2N, 3H), 2.01 (s, $AcCH_3$, 3H), 3.00-3.14 (m, CH_3CH_2N & 1H_β, 3H), 3.32 (dd, *J* = 9.5, 14.5 Hz, H_β, 1H), 4.95 (quart, *J* = 5.5, 13.0 Hz, H_α, 1H), 6.12 (d, *J* = 7.5 Hz, AcN*H*, 1H), 7.01 (d, *J* = 2.0 Hz, ind*H*2, 1H), 7.14 (dd, *J* = 2.5, 9.0 Hz, ind*H*6 , 1H), 7.28 (d, *J* = 3.5, ind*H*4, 1H), 7.49 (d, *J* = 2.0 Hz, ind*H*7, 1H), 8.45 (s, br, indN*H*, 1H); **^{13}C NMR** (125 MHz, $CDCl_3$): δ 15.37 (NCH_2CH_3), 23.19 ($AcCH_3$), 27.53 (C_β), 52.44 (C_α), 52.94 (NCH_2CH_3), 109.83 (ind*C*3), 112.93 (ind*C*4), 118.12 (ind*C*7), 122.48 (ind*C*6), 124.16 (ind*C*2), 125.42 (ind*C*3a), 128.76 (ind*C*7a), 134.40 (ind*C*5), 169.84 (Ac*C*O), 172.25 (*C*ONEt); **MS** (ES+): *m/e* 330.1 (100%) $[m^{35}Cl+Na]^+$, 331.1 (17%) $[m^{35}Cl+Na]^+$, 332.1 (33%) $[m^{37}Cl+Na]^+$, 333.1 (6%) $[m^{37}Cl+Na]^+$; **HRMS** (FT-MS) calcd for $C_{15}H_{19}N_3O_2Cl_1Na_1$: 330.0980, found: 330.0982; $[\alpha]_d^{25}$: (MeOH, *c* = 0.1) +24.3°; R_f: 0.34 (using EtOH 5% v/v in DCM)

N^{α}-Acetyl-5-bromo-*L*-tryptophan ethyl amide (**6c**)

^{1}H NMR (500 MHz, $CDCl_3$): δ 0.93 (t, *J* = 7.5 Hz, CH_3CH_2N, 3H), 1.98 (s, $AcCH_3$, 3H), 3.07-3.20 (m, CH_3CH_2N & 1H_β, 4H), 4.65 (dd, *J* = 7.5, 14.0 Hz, H_α, 1H), 5.98 (t, *J* = 6.0

Hz, EtN*H*, 1H), 6.56 (d, *J* = 8.0 Hz, AcN*H*, 1H), 7.03 (d, *J* = 2.0 Hz, ind*H*6, 1H), 7.19-7.25 (m, ind*H*4 & ind*H*7, 2H), 7.76 (s, ind*H*2, 1H), 8.47 (s, br, indN*H*, 1H); **^{13}C NMR** (125 MHz, $CDCl_3$): δ 14.38 (NCH_2CH_3), 23.23 (AcCH_3), 26.47 (C_β), 34.43 (NCH_2CH_3), 54.00 (C_α), 108.74 (ind*C*3), 110.59 (ind*C*4), 112.71 (ind*C*7), 112.94 (ind*C*6), 121.42 (ind*C*2), 124.32 (ind*C*6), 129.20 (ind*C*5), 131.85 (ind*C*3a), 134.71 (ind*C*7a), 170.21 (Ac*C*O), 170.98 (*C*ONEt); **HRMS** (FT-MS) calcd for $C_{15}H_{19}N_3O_2Br_1Na_1$: 374.0475, found: 374.0473; **$[\alpha]_d^{25}$**: (MeOH, *c* = 0.1) +19.2°; **R_f**: 0.36 (using EtOH 5% v/v in DCM)

N^α-Acetyl-5-iodo-*L*-tryptophan ethyl amide (**6d**)

^{1}H NMR (200 MHz, MeOH – d_4): δ 0.94 (t, *J* = 7.2 Hz, CH_3CH_2N, 3H), 1.94 (s, AcCH_3, 3H), 3.00-3.20 (m, CH_3CH_2N & 1H_β, 4H), 4.52 (t, *J* = 7.0 Hz, H_α, 1H), 7.08 (s, ind*H*2, 1H), 7.14 (d, *J* = 8.6 Hz, ind*H*7, 1H), 7.34 (dd, *J* = 1.6, 8.6 Hz, ind*H*6, 1H), 7.93 (d, *J* = 1.6 Hz, ind*H*4); **^{13}C NMR** (100 MHz, $CDCl_3$): δ 15.45 (NCH_2CH_3), 24.23 (AcCH_3), 31.56 (C_β), 35.65 (NCH_2CH_3), 54.17 (C_α), 105.81 (ind*C*3), 114.84 (ind*C*4), 115.58 (ind*C*7), 117.31 (ind*C*6), 123.64 (ind*C*2), 126.44 (ind*C*3a), 133.37 (ind*C*7a), 136.35 (ind*C*5), 170.39 (Ac*C*O), 171.11 (*C*ONEt); **MS** (ES+): *m/e* 422.1 (100%) $[m+Na]^+$; **HRMS** (FT-MS) calcd for $C_{15}H_{18}N_3O_2I_1Na_1$: 422.0336, found: 422.0338; **$[\alpha]_d^{25}$**: (MeOH, *c* = 0.5) +28.6°; **R_f**: 0.40 (using EtOH 5% v/v in DCM)

N^α-Acetyl-5-methyl-*L*-tryptophan ethyl amide (**6e**)

^{1}H NMR (400 MHz, $CDCl_3$): δ 0.91 (t, *J* = 7.2 Hz, CH_3CH_2N, 3H), 2.01 (s, AcCH_3, 3H), 2.47 (s, Ar-CH_3), 3.00-3.12 (m, CH_3CH_2N & 1H_β, 3H), 3.30 (dd, *J* = 5.2, 13.6 Hz, H_β, 1H), 4.62-4.67 (m, H_α, 1H), 5.44 (s, br, EtN*H*, 1H), 6.34 (d, *J* = 6.8 Hz, AcN*H*, 1H), 7.03-7.05 (m, ind*H*6 & ind*H*7, 2H), 7.24-7.26 (m, ind*H*4, 1H), 7.51 (s, ind*H*2, 1H), 8.00 (s, br, indN*H*, 1H); **^{13}C NMR** (100 MHz, $CDCl_3$): δ 15.31 (NCH_2CH_3), 23.23 (Ar-CH_3), 27.60 (AcCH_3), 32.42 (C_β), 53.06 (NCH_2CH_3), 55.85 (C_α), 100.13 (ind*C*3), 109.80 (ind*C*4), 112.02 (ind*C*7), 112.69 (ind*C*6), 123.34 (ind*C*2), 128.14 (ind*C*3a), 131.16 (ind*C*7a), 154.32 (ind*C*5), 170.01 (Ac*C*O), 172.35 (*C*ONEt); **MS** (ES+): *m/e* 310.2 (100%) $[m+Na]^+$; **HRMS** (FT-MS) calcd for $C_{16}H_{21}N_3O_2Na_1$: 310.1526, found: 310.1528; **Mp**.: 228 °C; **$[\alpha]_d^{25}$**: (MeOH, c = 0.4) +25.6 °; **R_f**: 0.36 (using EtOH 5% v/v in DCM)

N^{α}-Acetyl-5-methoxy-*L*-tryptophan ethyl amide (**6f**)

^{1}H NMR (400 MHz, $CDCl_3$): δ 0.89 (t, J = 7.4 Hz, CH_3CH_2N, 3H), 2.02 (s, $AcCH_3$, 3H), 3.03 (dd, J = 9.2, 14.4 Hz, H_β, 1H), 3.06-3.19 (m, CH_3CH_2N, 2H), 3.30 (dd, J = 4.8, 14.4 Hz, H_β, 1H), 3.89 (s, Ar-OCH_3, 3H), 4.62-4.67 (m, H_α, 1H), 5.41 (s, br, EtN*H*, 1H), 6.41 (d, J = 7.2 Hz, AcN*H*, 1H), 6.87 (dd, J = 2.4, 8.8 Hz, ind*H*6, 1H), 7.03 (d, J = 2.4 Hz, ind*H*2, 1H), 7.24 (d, J = 2.4 Hz, ind*H*4, 1H), 7.26 (d, J = 8.8 Hz, ind*H*7, 1H), 8.02 (s, br, indN*H*, 1H); **^{13}C NMR** (100 MHz, $CDCl_3$): δ 16.43 (NCH_2CH_3), 25.36 ($AcCH_3$), 30.96 (C_β), 36.34 (NCH_2CH_3), 54.34 (C_α), 55.81 (Ar-OCH_3), 102.28 (ind*C*3), 112.88 (ind*C*4), 114.01 (ind*C*7), 115.01 (ind*C*6), 125.58 (ind*C*2), 130.14 (ind*C*3a), 133.17 (ind*C*7a), 156.35 (ind*C*5), 170.57 ($AcCO$), 172.87 (*C*ONEt); **MS** (ES+): *m/e* 326.1 (100%) $[m+Na]^+$; **HRMS** (FT-MS) calcd for $C_{16}H_{21}N_3O_3Na_1$: 326.1475, found: 326.1477; **$[\alpha]_d^{25}$**: (MeOH, c = 0.1) +32.8°; **R_f**: 0.40 (using EtOH 5% v/v in DCM)

N^{α}-Acetyl-5-nitro-*L*-tryptophan ethyl amide (**6g**)

^{1}H NMR (500 MHz, MeOH – d_4): δ 0.94 (t, J = 7.5 Hz, CH_3CH_2N, 3H), 1.93 (s, $AcCH_3$, 3H), 3.01-3.15 (m, CH_3CH_2N & $1H_\beta$, 3H), 3.27 (dd, J = 7.0, 7.5 Hz, H_β, 1H), 4.59 (t, J = 7.0 Hz, H_α, 1H), 7.30 (s, ind*H*2, 1H), 7.43 (d, J = 9.0 Hz, ind*H*7, 1H), 8.02 (dd, J = 2.0, 9.0 Hz, ind*H*6, 1H), 8.62 (d, J = 2.0 Hz, ind*H*4); **^{13}C NMR** (125 MHz, MeOH – d_4): δ 14.53 (NCH_2CH_3), 22.51 ($AcCH_3$), 28.89 (C_β), 35.25 (NCH_2CH_3), 55.69 (C_α), 112.42 (ind*C*3), 114.24 (ind*C*4), 117.03 (ind*C*7), 117.85 (ind*C*6), 128.35 (ind*C*2), 131.03 (ind*C*3a), 132.48 (ind*C*7a), 142.52 (ind*C*5), 173.08 ($AcCO$), 173.30 (*C*ONEt); **MS** (ES+): *m/e* 341.3 (65%) $[m+Na]^+$, 342.3 (12%) $[m+Na]^+$; **HRMS** (FT-MS) calcd for $C_{15}H_{18}N_4O_4Na_1$: 341.1223, found: 341.1225; **Mp**.: 220-224 °C; **$[\alpha]_d^{25}$**: (MeOH, c = 0.15) +8.7°; **R_f**: 0.27 (using EtOH 5% v/v in DCM)

*4.4.1.j 9-Fluorenylmethyloxy-carbonyl (Fmoc) Protection of **3a** and **3f***[26,27]

Free amino acid (0.5 mmol) was dissolved in water (1.9 ml). A solution of sodium bicarbonate (50 mg, 0.5 mmol) containing 9-fluorenylmethyloxycarbonyl succinimide (170 mg, 0.5 mmol) in acetone (1.9 ml) was added. The suspension was stirred at room temperature overnight and the clear solution subsequently quenched by addition of concentrated hydrochloric acid to pH 2, forming a white precipitate. The acetone was removed under vacuum and the product was extracted into chloroform (2 x 4 ml). The

combined organic extracts were washed with 1 *N* hydrochloric acid (2 x 4 ml) and deionised water (2 x 4 ml), dried over magnesium sulphate and filtered. The solvent was evaporated under vacuum affording the crude Fmoc-protected amino acid. Samples were purified using semi-preparative HPLC (§4.2.2).

Fmoc-5-fluoro-*L*-tryptophan (**3a**)

Using the general method above (§4.4.1.j), the title compound was synthesised as a white crystalline powder.

^{1}H NMR (300 MHz, MeOH – d_4): δ 3.11 (dd, *J* = 2.7, 14.7 Hz, H_β, 1H), 4.12-4.33 (m, Fmoc*H*, FmocCH_2, 3H), 4.47 (dd, *J* = 6.9, 14.1 Hz, H_α, 1H), 6.85 (dt, *J* = 2.4, 9.3 Hz, ind*H*, 1H), 7.14 (s, ind*H*, 1H), 7.21-7.30 (m, 2 x ind*H*, Fmoc*H*, 4H), 7.36 (t, *J* = 7.5, Fmoc*H*, 2H), 7.56 (d, *J* = 7.5, Fmoc*H*, 2H), 7.76 (d, *J* = 7.5, Fmoc*H*, 2H); **^{13}C NMR** (75MHz, MeOH – d_4): δ 28.84 (C_β), 56.12 (C_α), 56.41 (Fmoc*C*), 67.81 (Fmoc*C*), 108.6 (ind*C*2), 111.86 (ind*C*), 112.24 (ind*C*), 113.55 (ind*C*), 121.74 (ind*C*2), 125.34 (Fmoc*C*), 128.37 (ind*C*3a), 128.78 (Fmoc*C*), 128.91 (Fmoc*C*), 131.73 (ind*C*7a), 141.16 (Fmoc*C*), 143.61 (Fmoc*C*), 156.72 (ind*C*5), 161.14 (Fmoc*C*O), 174.56 (*C*OOH); **MS** (ES+): *m/e* 445.4 (100%) $[m+Na]^+$; **Mp**.: 139-143 °C; $[\alpha]_d^{25}$: (MeOH, *c* = 1.0) -16.8°; **CHN** analysis (calcd %) C, 70.26; H, 4.76; N, 6.30; (found %) C, 70.25; H, 4.78; N, 6.29

Fmoc-5-methoxy-*L*-tryptophan (**3f**)

Using the general method above (§4.4.1.j), the title compound was synthesised as a white crystalline powder.

^{1}H NMR (400 MHz, MeOH – d_4): δ 3.08-3.13 (m, H_β, 1H), 4.15 (m, Fmoc*H*, FmocCH_2, 3H), 4.68 (m, H_α, 1H), 6.72 (dd, *J* = 2.2 Hz, 4.8 Hz, ind*H*, 1H), 7.05 (s, ind*H*, 1H), 7.20-7.30 (m, Fmoc*H*, 2x ind*H*, 3H), 7.34 (m, Fmoc*H*, 2H), 7.55 (d, *J* = 1.6 Hz, Fmoc*H*, 2H), 7.57 (d, *J* = 1.9Hz, Fmoc*H*, 2H); **^{13}C NMR** (100MHz, MeOH – d_4): δ 28.71 (C_β), 54.73, 56.22, 56.36, 68.04 (Fmoc*C*), 101.18 (ind*C*), 112.66 (ind*C*), 112.94 (ind*C*), 120.87 (ind*C*), 125.21 (ind*C*), 126.27 (ind*C*), 126.36 (Fmoc*C*), 128.17 (Fmoc*C*), 128.74 (Fmoc*C*), 142.52 (Fmoc*C*), 145.25 (Fmoc*C*), 156.21 (ind*C*5), 160.10 (Fmoc*C*O), 174.35 (*C*OOH); **MS** (ES+): *m/e* 465.2 (100%) $[m+Na]^+$; **Mp**.: 82-85 °C; $[\alpha]_d^{25}$: (MeOH, *c* = 1.0) +-24.5°; **CHN** analysis (calcd %) C, 71.04; H, 5.30; N, 6.14; (found %) C, 71.01; H, 5.38; N, 6.13

4.4.2 Peptide Synthesis

4.4.2.a Manual Coupling of Leucine to a Wang-resin

In the manual peptide synthesis, the Fmoc-protected amino acids were activated with diisopropylcarbodiimide (DIC). Apart from this, the deprotection and coupling protocols were identical to the one used in the automated peptide synthesis (§4.4.2.b): Wang resin (400 mg, 0.3 mmol/g free sites = 0.12 mmol) was suspended in a mixture of dichloromethane (DCM) (5.4 ml) and dimethylformamide (DMF) (0.6 ml) under a nitrogen atmosphere. A solution of Fmoc-*L*-leucine (85 mg, 0.24 mmol, 2 eq) and 1-hydroxybenzotriazole hydrate (HOBt) (37 mg, 0.24 mmol, 2 eq.) was added. Di-isopropylcarbodiimide (DIC) (38 µl, 0.24 mmol) and a solution of dimethyl-aminopyridine (DMAP) (1.6 mg, 1.2 µmol) in DMF (0.3 ml) were added and the reaction mixture stirred under a nitrogen atmosphere at room temperature for 2 to 3 hours. Excess reagents were filtered off under reduced pressure. The resin was then treated with acetic anhydride (22.4 µl, 0.24 mmol) and pyridine (20 µl, 0.24 mmol) and mixed for another 30 minutes. Filtration and washing with DCM (3 x 5 ml), and DMF (3 x 5 ml) and subsequent drying through suction afforded the 432 mg of Fmoc-Leu-Wang resin (75% yield). Resin loading was determined using the procedure described in §4.3.1.f.

4.4.2.b Synthesis of AcWLWLL using a Solid Phase Peptide Synthesiser

An Advanced Chemtech peptide synthesiser 348Ω was used for the automated peptide synthesis with a 40-well Teflon reaction block. Superfluous reagents and waste fluids were removed by filtration under positive nitrogen pressure at 5 bar, applied for 2 min. The syringe used for injections was cleaned with methanol (0.3 ml) between individual additions. Peptides were prepared in parallel using two or more reaction wells (total capacity 5 ml) and no more than 100 mg Wang resin in each well.

Wang resin (50 mg), pre-loaded with *L*-leucine (0.75 mmol g^{-1}, 37.5 µmol) was swollen and agitated in DMF (2 ml). Standard Fmoc-coupling procedures were employed using commercially available Fmoc-Leu-OH (0.188 mmol, 5 eq) and Fmoc-Trp(Boc)-OH (0.188 mmol, 5 eq) in solutions of anhydrous *N*-hydroxybenzotriazole (HOBt) (0.5 *M*) in DMF (375 µl each). The substituted Fmoc-protected tryptophan derivatives were used

without Boc protection on the indole nitrogen. The coupling procedure consisted of cycles of deprotection, washing, coupling, washing, second coupling, and final wash.

Deprotection was achieved by addition of piperidine in DMF (20% v/v) (1 ml) and agitation for 5 min. The well was then emptied by positive nitrogen pressure for 2 min. The deprotection procedure was then repeated, agitating the suspension of resin in piperidine/DMF solution (1 ml) for 10 min this time. The removal of the solvents was followed by a wash step. First, the resin was re-suspended in DMF (3 ml) and agitated for 1 min. Then, waste solvents were removed by applying positive nitrogen pressure for 2 min. The washing protocol was repeated 4 more times.

Coupling was achieved by addition of the next Fmoc-protected residue in HOBt solution (375 µl). The protected amino acids were coupled using *N*-*N*-diisopropylethylamine (DIPEA) (0.375 mmol, 5 eq) and (benzotriazol-1-yloxy)tripyrrolidinophosphonium hexafluoro-phosphate (PyBOP) (0.375 mmol, 5 eq) both dissolved in DMF (375 µl each). The resin was agitated for 60 min and solvents removed under positive pressure for 2 min. The coupling procedure was repeated after another washing cycle. After the attachment of the last residue and its deprotection, the *N*-terminus of peptide was capped with an acetyl group by reaction with acetylimidazol (0.188 mmol, 5 eq) in DMF (375 µl) for 30 min. The resin was then washed thoroughly with DMF (2.5 ml) and then washed with DMF (3 x 3 ml), methanol (4 x 3 ml) and dichloromethane (5 x 3 ml).

4.4.2.c Cleavage of AcWLWLL from the Wang-Resin

Wang-resin (59 mg with a loading of 0.3 mmol/g) was mixed with trifluoracetic acid (TFA) (1.4 ml), HPLC grade water (0.08 ml), phenol (0.08 ml) and triisopropylsilane (0.04 ml) and shaken for 2 hours. The solvents were removed and collected by suction filtration. The resin was washed with TFA (2 x 1.5 ml) and the washings combined with the filtrate. The solution was concentrated to a third of its volume *in vacuo* and a 10-fold excess of ice-cold diethyl ether (15 ml) was added drop wise. The solution was stored at 4 °C for 12 hours. Following centrifugation at 3000 rpm for 15 min, the supernatant was removed, concentrated under vacuum and any further precipitate collected by centrifugation. The collected peptide was washed with ether (2 x 5 ml) and centrifuged. The white pellet was dried under vacuum to afford 12 mg of peptide (88% yield).

4.4.2.d Ninhydrin Test

Ninhydrin (1,2,3-triketo-hydrindene hydrate) was .dissolved in ethanol (0.2% w/v, 100ml) and acidified slightly with conc. sulphuric acid (2 drops) to produce a staining dip. TLC plates were stained, dried with a heat gun for 30 seconds and analysed by visual inspection.

4.5 References

[1] Kondo-Yamada, Y., Okada, C., Yashida, K., Umeda, Y., Arima, S., Sato, N., Kai, T., Takayanagi, H., Harigaya, Y., *Tetrahedron*, **2002**, *58*, 7851-7861
[2] Baer, E., Mahadevan, V., *J. Am. Chem. Soc.*, **1959**, *81*, 2494-2498
[3] Zeeberger, B., Caplow, M., *J. Biol. Chem.*, *248*, **1973**, 5887-5891
[4] Connors, K. A.in *Binding Constants: The Measurement of Molecular Complex Stability*, John Wuiley & Sons, **1987**
[5] Porter, J., Dyckert, J., Rivier,J., *Int. J. Peptide Prot. Res.*, **1987**, *30*, 13-21
[6] Rinderknecht; Niemann, *J. Am. Chem. Soc.*, **1950**, *72*, 2296
[7] Allen, M. C., Brundish, D. E., Fullerton, J. D., Wade, R., *J. Chem. Soc. Perkin Trans. 1*, **1980**, 1928-1932
[8] Cook *et al.*, *J. Chem. Soc.*, **1951**, 1203-1206
[9] Bergmann, E. D., Hoffmann, E., *J. Chem. Soc.*, **1962**, *7*, 2827-2829
[10] Allen, M. C., Brundish, D. E., Fullerton, J. D., Wade, R., *J. Chem. Soc. Perkin Trans. 1*, **1986**, 989-1003
[11] Lee, M., Phillips, R.S., *Magn. Res. Chem.*, **1992**, *30*(11), 1035-1040
[12] Ma, C.; Liu, X.; Li, X.; Flippen-Anderson, J.; Yu, S.; Cook, J. M., *J. Org. Chem.*, **2001**, *66*(13) 4525 - 4542
[13] Phillips, R. S., Lee, M., *J. Heterocyclic Chem.*, **1994**, *31*, 711-716
[14] Irie, K.; Ishida, A.; Nakamura, T.; Oh-ishi, T., *Chem. Pharm. Bull.*, **1984**, *32*(6), 2126-2139
[15] Zembower, D. E., Ames, M. M., *Synthesis*, **1994**, 1433-1436
[16] Lee, M., Phillips, R.S., *J. Heterocyclic Chem.*, **1992**, *29*(5), 1181-1187
[17] Gerig, J. T., Klingenborg, J. C., *JACS*, **1980**, *102*(12), 4267-4268
[18] Li, X., Yin, W., Sarma, P.V.V.S., Zhou, H., Ma, J., *Tetrahedron Lett.*, **2004**, *45*(46), 8569-8573
[19] Perry, C. W., Brossi, A.,; Deitcher, K. H., Tautz, T., Teitel, S., *Synthesis*, **1977**, *7*, 482-494
[20] Schmidt, U., Lieberknecht, A., Griesser, H., Boekens, H., *Tet. Lett.*, **1982**, *23*(47), 4911-4914
[21] Lee, M., Phillips, R.S., *J. Heterocyclic Chem.*, **1992**, *29*(5), 1181-1187
[22] De Fazi, R., Berti, G., Da Settimo, A., *Gazz. Chim. Ital.*, **1959**, *89*, 2238-2242
[23] Kurihara, T., Nasu, M. I., Ishida, T., *Chem. Pharm. Bull.*, **1982**, *279*, 203
[24] Baldwin, H. R., Berg, C. P., *J. Nutr.*, **1949**, *39*(2), 203-218
[25] Ishida, A., Nakamura,T., Irie, K., Oh-Ishi, T., *Chem. Pharm. Bull.*, **1985**, *33*(8), 3237-3249
[26] Paquet, A., *Can. J. Chem.*, **1982**, *60*, 976-980
[27] Djedaïni-Pilard, F., Vernet, D., Perty, B., *J. Chem. Soc.* (Perkin Trans 2), **1995**, *4*, 723-730

Chapter 5: Computational Chapter

This chapter describes the parameters, forces and equations that were used in the molecular dynamics simulations and free energy calculations detailed in §2.5 - §2.7.

5.1 Interaction Functions and Force Fields

5.1.1 Molecular Dynamics Simulations

Molecular dynamics (MD) simulations solve Newton's equations of motions for a molecular system using force fields, such as the one detailed in this chapter to describe intra- and intermolecular interactions. These are split into non-bonded interactions (§5.1.1.a) between an atom pair within a certain pre-defined cut-off radius and bonded interactions (§5.1.1.b) between pairs of atoms sharing a chemical bond. Terms needed for free energy calculations are given in §5.1.1.d. It follows a description of the solvent (§5.1.3) and structural representations of the analytes (§5.1.4). The computational procedures are detailed in §5.2.

5.1.1.a Non-bonded Interactions

Non-bonded terms include van der Waals interactions (modelled by a Lennard-Jones type potential) and electrostatic, Coulomb interactions. These interactions were computed based on a neighbour list, which was updated during the course of the simulation at pre-defined intervals. The list contains all non-bonded atoms within a certain defined radius.

Lennard-Jones Interaction

The Lennard-Jones interaction, V_{LJ}, between two atoms i and j can be described as

$$V_{LJ}\left(r_{ij}\right)=\frac{C_{ij}^{(12)}}{r_{ij}^{12}}-\frac{C_{ij}^{(6)}}{r_{ij}^{6}}, \qquad (4.1)$$

where i and j are separated by a distance r_{ij} and $C_{ij}^{(12)}$ and $C_{ij}^{(6)}$ are the so-called Lennard-Jones parameters. The force of this potential is obtained by differentiation of Equation 4.1 and can be written as

$$\mathbf{F}_i(\mathbf{r}_{ij}) = -\frac{\partial V_{LJ}(r_{ij})}{\partial r_{ij}} = \left(12\frac{C_{ij}^{(12)}}{r_{ij}^{12}} - 6\frac{C_{ij}^{(6)}}{r_{ij}^{6}}\right)\frac{\mathbf{r}_{ij}}{r_{ij}}, \qquad (4.2)$$

where $\mathbf{F}_i$ is the force on atom i arising from the interactions of atoms i and j, $\mathbf{r}_{ij}$ is the separation vector between i and j. The Lennard-Jones parameters for different atoms were defined in the force field and are given in §7.6.

Lennard-Jones type interactions between water oxygen (OW) and the lipid acyl chain (*i.e.* the OW-CH_2 and OW-CH_3 interactions) influence the interface region strongly. A variety of Lennard-Jones parameters for these interactions have been used so far.[1] The original GROMOS parameter resulted in a solubility of decane in water that was far too high. Therefore, the values were adjusted (according to the adjustments introduced made by Berger *et al.*[2]) to $\varepsilon = 0.529$ and $\varepsilon = 0.637$ (OW-CH_2 and OW-CH_3 respectively) with $\sigma = 0.310$ for both and implemented as such in the simulations containing lipid-water interactions (*i.e.* in the bilayer simulations §5.2.2 and in the simulation of water-lipid-tryptophan adduct formation §5.2.1.b).

Coulomb Interactions

The Coulomb interaction between two charged particles i and j is given by

$$V_C(r_{ij}) = f\frac{q_i q_j}{\varepsilon_r r_{ij}}, \qquad (4.3)$$

where q_i and q_j are the partial charges on atom i and j respectively, ε_r is the relative dielectric constant of the medium in which the atoms are dispersed, and f is the Faraday constant. The Faraday constant corresponds to

$$f = \frac{1}{4\pi\varepsilon_0} = 138.935485 \text{ kJ mol}^{-1} \text{ nm e}^{-2}, \qquad (4.4)$$

where ε_0 is the dielectric constant in a vacuum (permittivity of free space); and e is the charge on an electron.

Coulombic Interactions Treated by Ewald Summation

The total electrostatic energy of N particles and the periodic images are given by

$$V = \frac{f}{2}\sum_{i,j}^{N}\sum_{n_x}\sum_{n_y}\sum_{n_z}^{*}\frac{q_i q_j}{\mathbf{r}_{ij,\mathbf{n}}}, \tag{4.5}$$

where $(n_x,n_y,n_z) = \mathbf{n}$ is the box index vector, the star indicates that terms with $i = j$ should be omitted when $(n_x,n_y,n_z) = (0, 0, 0)$, and $\mathbf{r}_{ij,\mathbf{n}}$ is the real distance between the charges q_i and q_j and not the minimum image. The force can be evaluated faster when it is split into a sum of its direct V_{dir} and reciprocal sums V_{rec} as well as a constant term V_0:

$$V = V_{dir} + V_{rec} + V_0. \tag{4.6}$$

The individual parts can be expressed by:

$$V_{dir} = \frac{f}{2}\sum_{i,j}^{N}\sum_{n_x}\sum_{n_y}\sum_{n_z}^{*}\frac{\operatorname{erfc}(\beta r_{ij,\mathbf{n}})q_i q_j}{r_{ij,\mathbf{n}}},$$

$$V_{rec} = \frac{f}{2\pi V}\sum_{i,j}^{N} q_i q_j \sum_{m_x}\sum_{m_y}\sum_{m_z}^{*}\frac{\exp\left(-(\pi \mathrm{m}/\beta)^2 + 2\pi i \mathrm{m}\cdot(\mathrm{r}_i - \mathrm{r}_j)\right)}{\mathrm{m}^2}, \tag{4.7}$$

$$V_0 = -\frac{f\beta}{\sqrt{\pi}}\sum_{i}^{N} q_i^{\,2}.$$

Here, V_{dir} is the direct sum, V_{rec} is the reciprocal sum, V_0 is a constant term for the evaluation of the total electrostatic energy of N particles being a distance r_{ij} apart from each other, where f is the Faraday constant, β is a parameter that determines the relative weight of the direct and reciprocal sums and $(m_x,m_y,m_z) = \mathbf{m}$ is the highest magnitude of wave vectors in each direction.

In the particle-mesh Ewald (PME) summation method, charges are assigned to a grid using cardinal B-spline interpolation. The grid is subsequently Fourier transformed with a 3D FFT algorithm and the reciprocal energy term obtained by a single sum over the grid in k-space. Equation 4.3 needs to be adjusted for the evaluation of the short-range coulomb potential in the Ewald summation to

$$V_E(r) = f\frac{\mathrm{erfc}(\beta r_{ij})q_i q_j}{r_{ij}}, \tag{4.8}$$

which gives the short-range electrostatic potential, V_E, between two particles i and j a distance r_{ij} apart from each other, where q_i and q_j are the partial charges on atom i and j respectively, β is the relative weight between space sum and the reciprocal space sum for PME summation, erfc(x) is the complementary error function, and f is the Faraday constant.

PME has the disadvantage of enhancing artefacts caused by periodic boundary conditions. Alper *et al.*[3] suggested that a simple cut-off treatment for systems of lipids with neutral total charge could be used. This also produces artefacts especially when ions are present, but works well for PC lipids. Anézo *et al.*[4] reviewed the effects of long-range electrostatic treatments of the molecular dynamics simulations of DPPC bilayers and found PME to outperform both a reaction field treatment and a simple cut-off approach.

5.1.1.b Bonded Interactions

Bonded interactions are defined as bond-stretching, angle bending as well as improper and proper dihedral forces. These interactions are generally approximated by harmonic potentials as described in detail in the following. Bonded interactions are defined during the set-up of a simulation and are fixed thereafter.

Bond-Stretching Potential

The bond-stretching potential, V_b, can be described by a harmonic oscillator with a spring constant, k_{ij}^b, the distance r_{ij} between two particles i and j and their ideal bonding distance b_{ij}:

$$V_b(r_{ij}) = \frac{1}{2}k_{ij}^b(r_{ij} - b_{ij})^2 \tag{4.9}$$

The corresponding force acting on particle i can then be written as the derivative of the potential

$$\mathbf{F}_i(\mathbf{r}_{ij}) = k_{ij}^b(r_{ij} - b_{ij})^2\frac{\mathbf{r}_{ij}}{r_{ij}}. \tag{4.10}$$

Angle Potential

The angle potential V_a of three sequentially, covalently bound atoms i, j, and k (with j being in the middle) can also be described by a harmonic approximation:

$$V_a(\theta_{ijk}) = \frac{1}{2} k_{ijk}^{\theta} (\theta_{ijk} - \theta_{ijk}^0)^2, \tag{4.11}$$

where k_{ijk}^{θ} is a spring constant, θ_{ijk} the angle between the three atoms i, j, and k, and θ_{ijk}^0 is their ideal bond angle. The angle-bending forces acting on the three atoms are given by:

$$\begin{aligned} F_i &= -\frac{dV_a(\theta_{ijk})}{dr_i}, \\ F_k &= -\frac{dV_a(\theta_{ijk})}{dr_k}, \\ F_j &= -F_i - F_k. \end{aligned} \tag{4.12}$$

The angle can be written in terms of the distances between the atoms as:

$$\theta_{ijk} = \arccos \frac{(r_{ij} \bullet r_{kj})}{r_{ij} r_{kj}} \tag{4.13}$$

For the three atoms i, j, and k, where r_{ij} and r_{kj} are the distances between the atoms i and j and the distance between atoms k and j respectively; $\mathbf{r}_{ij}$ and $\mathbf{r}_{kj}$ are the vectors between the atoms.

Improper Dihedrals

Improper dihedrals impose a force on a group of atoms i, j, k, and l, which keeps them planar, or prevents the 4 atoms from flipping to their mirror images. Improper dihedrals were treated by an harmonic potential:

$$V_{id}(\xi_{ijkl}) = \frac{1}{2} k_{\xi} (\xi_{ijkl} - \xi_0)^2, \tag{4.14}$$

where k_{ξ} is a spring constant, ξ_{ijkl} is the improper dihedral and ξ_0 is the ideal improper dihedral.

Proper Dihedrals

Proper dihedral angles are defined according to the IUPAC/IUB convention, where ϕ_{ijkl} is the angle between the *ijk* and *jkl* planes of the atoms *i*, *j*, *k*, and *l*, with zero corresponding to the *cis* configuration (*i* and *l* on the same side). The proper dihedral potential was described by the Ryckaert-Bellemans function

$$V_{pd}(\phi_{ijkl}) = \sum_{n=0}^{5} C_n(\cos(\psi))^n, \qquad (4.15)$$

$$\psi = \phi - 180^\circ,$$

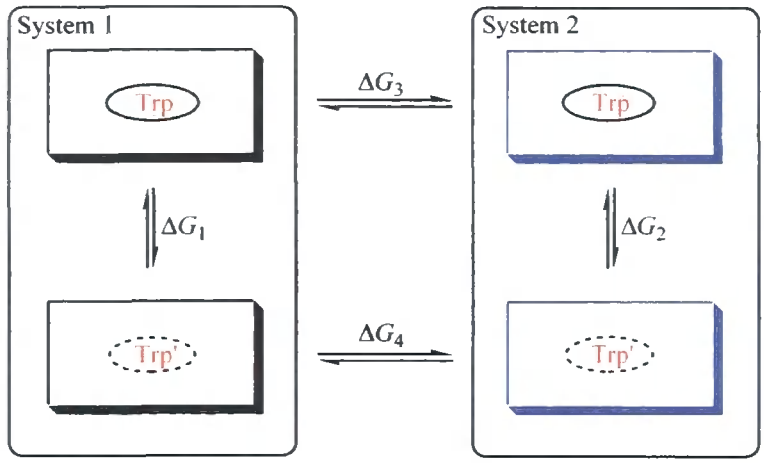
with coefficients C_n.

Restraints

Several parameters in a simulation can be restrained to fixed values based on experimental data (*e.g.* NMR or x-ray crystallographic distance restraints) or to simplify a simulation (*e.g.* restraining of solvent bond-stretching and angle-bending forces). Typically, restraints can be imposed on positions, angles, distances, orientation, and dihedrals. Restraints are defined during the set-up of a run and are not altered during the simulation.

5.1.1.c Free Energy Simulations – Theory

Free energy calculations were performed using the so-called slow growth method in a thermodynamic cycle, which was employed as illustrated in Scheme 2.12.



Scheme 5.1: General scheme for the calculation of a free energy of interest by using the thermodynamic integration method. A detailed description can be found in Scheme 2.12.

It is difficult to calculate solvation and binding free energies directly using molecular dynamics simulations. However, with the use of a thermodynamic cycle, a molecule can be transferred into a non-interacting particle, a so-called dummy molecule (indicated in Scheme 2.12 by a superscript '). Such a transition can be achieved by a gradual, linear decrease of non-bonding interaction potentials of the studied particle during the course of a series of molecular dynamics simulations. The generated free energy differences can be computed more easily and using a thermodynamic cycle one can determine the binding free energy of the system of interest indirectly by applying Equation 4.16:

$$\Delta G_1 - \Delta G_2 = \Delta G_3 - \Delta G_4 . \qquad (4.16)$$

Here, the partial free energies (ΔG_1 to ΔG_4) are associated with an equilibrium between the two states. By gradually changing from one state to the other, the free energies associated with this change can be calculated. This is achieved by perturbing the Hamiltonian H^A of the starting model state A gradually into the Hamiltonian of the target system B, H^B. The change has to be small and reversible so that the system can be assumed to remain in equilibrium during the process. This means that the reverse calculation of changing the system from state B as a starting point into state A will yield the same results. A coupling parameter λ is introduced, so that the Hamiltonian H is now a function depending on the coordinates, p, the charged interactions q, and λ

$$H = H(p,q;\lambda) . \qquad (4.17)$$

System A is now described by $\lambda = 0$ while system B is described when $\lambda = 1$:

$$H(p,q;0) = H^{A}(p,q) \;\; ; \;\; H(p,q;1) = H^{B}(p,q). \tag{4.18}$$

In GROMACS, the functional form of the λ-dependence is different for the various force-field contributions (see also §5.1.1.d).

The knowledge of the partition function Δ of an *NpT* ensemble allows the calculation of the Gibbs free energy of the process:

$$G(\lambda) = -k_B T \ln \Delta \tag{4.19}$$

$$\Delta = c \int\int\int \exp[-\beta H(p,q;\lambda)] dq dp dV \tag{4.20}$$

where $\beta = (k_B T)^{-1}$, and $c = (N! h^{3N})^{-1}$. These integrals over phase space cannot be evaluated from a simulation directly, but it is possible to evaluate the derivative with respect to λ as an ensemble average

$$\frac{dG}{d\lambda} = \frac{\int\int\int \left(\frac{\delta H}{\delta \lambda}\right) \exp[-\beta H(p,q;\lambda) - \beta pV] dq dp}{\int\int\int \exp[-\beta H(p,q;\lambda) - \beta pV] dq dp} = \left\langle \frac{\delta H}{\delta \lambda} \right\rangle_{NpT;\lambda}. \tag{4.21}$$

The difference in free energy between state A and B can be found by integrating the derivative of *H* with respect to λ

$$G^{B}(p,T) - G^{A}(p,T) = \int_0^1 \left\langle \frac{\delta H}{\delta \lambda} \right\rangle_{NpT;\lambda} d\lambda. \tag{4.22}$$

Therefore several calculations at different λ values can be used to determine the free energy change between the two states A and B using the $\left\langle \frac{dG}{d\lambda} \right\rangle$ at a number of well-chosen intermediate values of λ. It has been found that approximately 16 to 20 steps are necessary for an accurate and reversible free energy calculation. A proper error estimate can be made for each value of $\frac{dG}{d\lambda}$ from the fluctuation of $\frac{\delta H}{\delta \lambda}$.

5.1.1.d Free Energy Simulations – Description of Forces

Harmonic bond potentials were used to model bond, angles and improper dihedrals. These were interpolated linearly from state A to state B using the generic potential function

$$V_x = \frac{1}{2}\left((1-\lambda)k_x^A + \lambda k_x^B\right)\left(x - (1-\lambda)x_0^A - \lambda x_0^B\right)^2. \quad (4.23)$$

V_x describes a harmonic potential that changes with a coupling parameter, λ, from a fully interacting state A to a fully non-interacting state B, where k_x^A and k_x^B are the spring constants at state A and B, x the harmonic parameter, and x_0^A and x_0^B the ideal values for the harmonic parameter in state A and B respectively. The force is then given by the derivative of the potential function and can be exemplified by the bond-stretching force

$$\frac{\partial V_b}{\partial \lambda} = \frac{1}{2}\left(k_b^B - k_b^A\right)\left[\left(b - (1-\lambda)b_0^A + b_0^B\right)^2 + \left(b_0^A - b_0^B\right)\left(b - (1-\lambda)b_0^A - \lambda b_0^B\right)\right], \quad (4.24)$$

where V_b is given with respect to the coupling parameter λ, b the bond distance, and b_0^A and b_0^B the ideal values bond lengths in state A and B respectively.

Proper Dihedrals

Proper dihedrals were interpolated using the potential function and its derivative:

$$V_d = \left((1-\lambda)k_d^A + \lambda k_d^B\right)\left(1 + \cos\left(n_\phi\phi - (1-\lambda)\phi_s^A + \lambda\phi_s^B\right)\right)$$

$$\frac{\partial V_d}{\partial \lambda} = \left(k_d^B - k_d^A\right)\left[1 + \cos\left(n_\phi\phi - \left[(1-\lambda)\phi_s^A + \phi_s^B\right]\right) - \left((1-\lambda)k_d^A + k_d^B\right)\left(\phi_s^A - \phi_s^B\right)\sin\left(n_\phi\phi - \left[(1-\lambda)\phi_s^A - \lambda\phi_s^B\right]\right)\right], \quad (4.25)$$

where k_d^A and k_d^B are the spring constants at state A and B, ϕ the dihedral angle, and ϕ_s^A and ϕ_s^B the ideal values dihedral angles in state A and B respectively, and n_ϕ the multiplicity of the dihedral.

Coulomb Interaction

The Coulomb interaction between two particles i and j whose charge varies with λ was represented by

$$V_C = \frac{f}{\varepsilon_{rf} r_{ij}} \left[\left((1-\lambda) q_i^A + \lambda q_i^B \right) \cdot \left((1-\lambda) q_j^A + \lambda q_j^B \right) \right],$$

with

$$\frac{\partial V_C}{\partial \lambda} = \frac{f}{\varepsilon_{rf} r_{ij}} \left[\left(q_j^B - q_j^A \right) \left((1-\lambda) q_i^A + q_i^B \right) + \left(q_i^B - q_i^A \right) \left((1-\lambda) q_j^A + q_j^B \right) \right], \qquad (4.26)$$

where q_i^A and q_j^B and q_j^A and q_j^B are the partial charges of particles i and j in state A and B respectively.

Lennard-Jones Interaction

The Lennard-Jones potential between two particles of which the atom type varies with λ can be written as:

$$V_{LJ} = \frac{\left((1-\lambda) C_{12}^A + \lambda C_{12}^B \right)}{r_{ij}^{12}} - \frac{\left((1-\lambda) C_6^A + \lambda C_6^B \right)}{r_{ij}^6},$$

with

$$\frac{\partial V_{LJ}}{\partial \lambda} = \frac{C_{12}^A - C_{12}^B}{r_{ij}^{12}} - \frac{C_6^A - C_6^B}{r_{ij}^6}, \qquad (4.27)$$

where C_{12}^A and C_{12}^B and C_6^A and C_6^B are the Lennard-Jones parameter of the interaction in state A and B respectively.

Kinetic Energy

When the mass of a particle is perturbed during the course of a free energy calculation, there is also a change in the contribution of the kinetic energy to the free energy that needs to be modelled:

$$E_k = \frac{1}{2} \frac{p^2}{(1-\lambda) m^A + \lambda m^B}. \qquad (4.28)$$

Here, E_k is the change in the kinetic energy with respect to the coupling parameter λ, p is the momentum of the particle, and m^A and m^B its mass in state A and B respectively. The derivative gives then the overall contribution:

$$\frac{\partial E_k}{\partial \lambda} = -\frac{1}{2} \frac{p^2 \left(m^B - m^A \right)}{\left((1-\lambda) m^A + \lambda m^B \right)^2} = -\frac{1}{2} v^2 \left(m^B - m^A \right), \qquad (4.29)$$

where p is the momentum of the particle, v its velocity, and m^A and m^B its mass in state A and B respectively. However, the free energy calculations in this study did not involve any changes in the mass of perturbed atoms. Consequently the kinetic energy in eqn. (4.28) collapses to the usual expression for kinetic energy and $\frac{\partial E_k}{\partial \lambda} = 0$ in eqn. (4.29).

Constraints

As constraints are formally part of the Hamiltonian, they contribute to the free energy of a system. Therefore, they need to be adjusted during the course of a free energy simulation.

Soft-Core Interactions

Due to the linear interpolation of the Lennard-Jones and Coulomb potentials, the free energy calculations for values of the coupling parameter λ close to 0 or 1 are problematic as the simulations can approach singularities. Soft-core potentials V_{SC} were introduced to circumvent these problems and smooth the Lennard-Jones and Coulomb potentials when they are close to state A ($\lambda = 0$) and state B ($\lambda = 1$),

$$V_{SC}(r) = (1-\lambda)V^A(r_A) + \lambda V^B(r_B). \tag{4.30}$$

This gives the soft-core potential, V_{SC}, as a function of distance r, where λ is the coupling parameter, V^A and V^B the normal "hard core" Lennard-Jones or Coulomb potentials in state A and B respectively, and r_A and r_B are the distances in state A and B respectively. The distances r_A and r_B in state A and B are given by:

$$r_A = \left(\alpha\sigma_A^6\lambda^p + r^6\right)^{\frac{1}{6}},$$

$$r_B = \left(\alpha\sigma_B^6(1-\lambda)^p + r^6\right)^{\frac{1}{6}}, \tag{4.31}$$

where α is the soft-core parameter, p is the soft-core λ power, and σ is the radius of the interaction, which is $(C_{12}/C_6)^{1/6}$ or a predefined value if C_6 or C_{12} is zero.

5.1.2 Simulation of Lipid Bilayers

A triclinic box containing a DMPC lipid bilayer in the central plane was filled with 3655 SPC water molecules. This resulted in 16853 atomistic sites in total. The system was energy-minimized and equilibrated for 10 ns. One computed nanosecond took four 3.73 GHz Xeon Intel Processors 11.43 hours to compute using the GROMACS (Version 3.3.1) code.[5, 6]

A system size of 128 lipid molecules (64 per leaflet) was chosen as a compromise between feasible computational cost and reduction of artefacts arising through periodicity. Additionally, the conformational space of a single lipid molecule, with two flexible acyl chains, is very large. A bigger system will sample these different conformations more accurately.

5.1.2.a Degree of Solvation

The amount of water present in the system has a strong influence on the bilayer properties. It has been noted that the ratio of about 26 water molecules per lipid is the minimum required to allow for a fully solvated bilayer structure.[7] In this study the bilayer structures were initially equilibrated with a solvation ratio of 57 to 2 as the insertion of peptides reduced this value somewhat.

5.1.2.b Pressure Coupling

The type of the pressure-coupling appears to have no significant effect on the equilibrium properties of bilayers.[8] Although the Parinello-Rahman method is to be preferred on theoretical grounds as it generates a well defined ensemble in comparison to the Berendsen method, the latter scheme is more practical because it damps large oscillations in box dimensions that might occur. This is especially important during the equilibration time. The treatment of the pressure-coupling is not very computationally demanding in either of the methods. To improve the accuracy of the simulations the Parinello-Rahman method was chosen for the free energy calculations of the amino acid derivatives in chloroform. The Berendsen model was chosen for the bigger lipid bilayer systems. To account for the anisotropic nature of the bilayer, our simulations were performed in a way that an isotropic constant pressure (1 atm) was applied both in the bilayer plane and in the direction of the bilayer normal. The pressure was coupled independently with a 1.0 ps time constant using a Berendsen barostat.

5.1.2.c Temperature

The temperature must be chosen carefully as not to coincide, or to be close to, the lipid transition temperature of DMPC. At the transition temperature the lipid changes its phase, structure and over all ordering. AFM studies of DMPC bilayer microstructure suggested a transition temperature of 301 ± 0.1 K.[9] However, other researchers report the transition from the liquid-crystalline L_α phase which the lipid adopts at a temperature above 302.6 K.[10] The gel-phase $P_{\beta'}$ is adopted at a temperature below 295.2 K. Enders *et al.*[11] determined the main transition temperature of DMPC using temperature-controlled AFM to be at 296.5 ± 0.1 K.

The temperature in the bilayer simulations was adjusted to be 303 K. Weak temperature coupling using the Berendsen algorithm with a coupling constant of 0.1 ps was used independently on all differing components of the simulated system. *I.e.* the temperature was coupled independently to the lipid bilayer, the solvent, and the solvated molecules.

5.1.3 Solvent Models

5.1.3.a Chloroform Simulation Boxes

Chloroform molecules were represented by the model of Jorgensen *et al.*[12] (Figure 5.1), which includes a united CH unit.

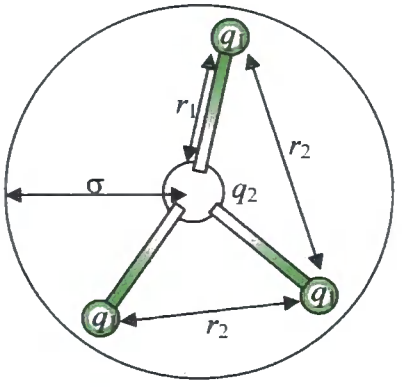


Figure 5.1: Representation of a united atom chloroform molecule of the model used.

Since the electron density for the hydrogen atom is shifted well along the bond toward the carbon atom, this provides a reasonable approximation. The potential successfully reproduces the experimentally determined heat of vaporisation[12] as well as the density and diffusion constant[13] for chloroform. Configuration and partial charge distribution were adapted from literature values (Table 5.1).[14]

Chloroform Model	Lennard-Jones Parameter				C-Cl Dist.	Cl-Cl Dist.	Charges	
Parameter	σ_C	σ_{Cl}	ε_C	ε_{Cl}	r_1	r_2	q_1	q_2
Units	nm		kcal mol^{-1}		nm		e	
Value	0.380	0.340	0.3347	0.1252	0.1758	0.2903	-0.14	+0.42

Table 5.1: Table presenting the values used in the chloroform representation.[14]

Two cubic simulation boxes were generated: one with 512 chloroform molecules (3.9 nm per side) the other with 4096 (8.2 nm per side). Both were equilibrated for 10 ns before used in the set-up of further simulations.

5.1.3.b Choice of Water Model for Lipid-Water Simulations

The single-point charge (SPC)[15] model (Figure 5.2) was used in this study as it provides a computationally more economical representation of water model than the related extended single-point charge (SPC/E)[16] or transferrable intermolecular potential functions TIP[17] representations. Although the differences in the charges and geometry for a single molecule are small, they accumulate to give noticeable effects on the total system. SPC/E was originally developed because previous models did not take into account the self-energy due to polarization.[16] It was found that it gives a better density, radial distribution function, self-diffusion constant and dielectric constant than SPC.[18] However, this comes at a computational expense that is prevalent specifically in larger systems such as common bilayer simulations. Additionally, SPC tends to produce larger area per lipid values and a broader lipid-water interface.[19] This is favourable as experimentally determined area per lipid values and width of the interface are larger than the ones found in simulations.

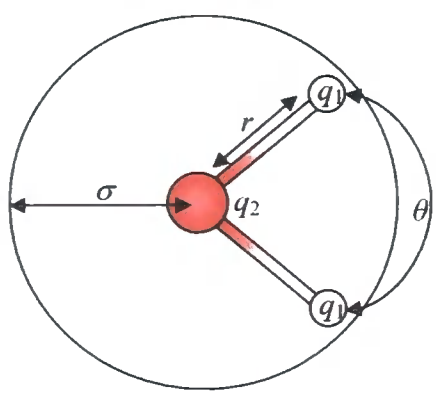


Figure 5.2: Representation of a water molecule using the SPC model.

The SPC water model employs the parameters listed in Table 5.2:

SPC Model	Lennard-Jones Parameter		O-H distance	Angle	Charges	
Parameter	σ	ε	r	θ	q_1	q_2
Units	nm	kJ mol^{-1}	nm	°	e	e
Value	0.317	0.650	0.1	109.47	+0.41	-0.82

Table 5.2: Table presenting the values used in the SPC representation of water molecules.[15]

5.1.4 Representation of Analytes

5.1.4.a Generation of Tryptophan Derivatives

Structures for N^{α}-acetyl-*L*-tryptophan ethyl amide derivatives were generated using the Maestro[20] software package from Schrödinger Inc. The geometry was optimized by *ab-initio* Jaguar[21] calculations. Partial charges were computed with a 6-31G** basis set and assigned initially by using restricted Hartree-Fock (rHF) *ab-initio* Jaguar calculations (Set A). The electrostatic potential was fitted to the atom centres of the molecules using a quadrupole moment constraint and a spherical grip representation of the electrostatic potential. In a second set (Set B), partial charges were evaluated for 3-ethyl-indole derivatives using second-order perturbation theory (MP2) calculations in the GAUSSIAN[22] software package. Partial charges were fitted to atoms using the CHelpG (CHarges from ELectrostatic Potentials using a Grid based method)[23] derivate of the electron density so that they gave the best representation of the latter. The derived partial charges of the substituted indole ring were used to represent the heterocycle of the equivalent tryptophan derivative. The partial charges for the backbone of the N^{α}-acetyl-*L*-tryptophan ethyl amide derivatives were adopted from the optimized parameters for liquid systems (OPLS_AA)[24] force field. Applied on the molecule, this lead to deviations from a net charge of zero for the entire compound ($\pm$0.011 e^-). To regain an uncharged molecule, the charge difference was split and redistributed over the heavy atoms of the indole ring by adding the obtained fraction to the MP2 derived partial charges. Structures were subsequently transferred into the format of the Groningen machine for chemical simulation package GROMACS, 3.3.1.[25,6]

The charge groups used for the representation of the tryptophan derivatives are shown in Figure 5.3.

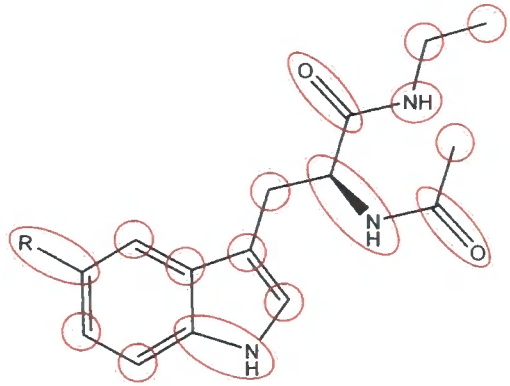


Figure 5.3: Representation of charge groups of the tryptophan derivatives (**6a-g, 6n**).

5.1.4.b 1,2-dimyristoyl-sn-glycerol-3-phosphocholine (DMPC) Representation

Following an adapted model of Anézo *et al.*[4], the DMPC molecule was represented in a united-atom fashion combining the hydrogen atoms with the heavy atom they are bound to (Figure 5.4).

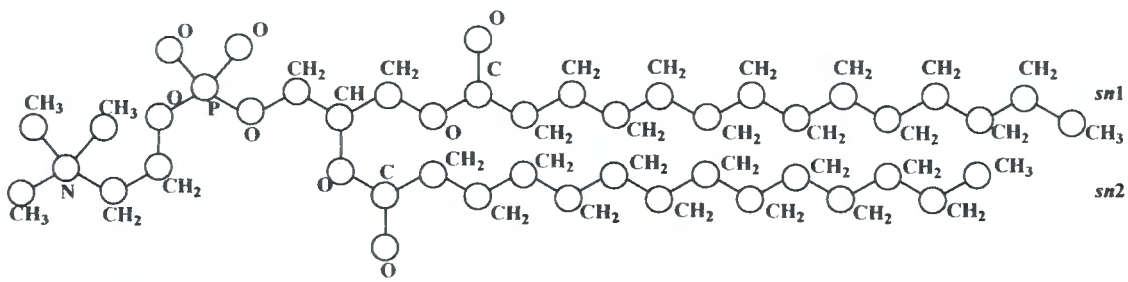


Figure 5.4: United atom representation of a DMPC molecule.

This representation is commonly employed and has been used in several other studies.[26,27] Partial charges, bonded and non-bonded parameters of the lipid were adapted from those that have been established by Berger *et al.*[28] The authors showed that the volume per lipid agreed within 1% of the experimentally obtained value of 1.232 nm^3. Similarly, the lipid order parameter, total volume of the lipid and certain distances between lipid molecules were represented within a 3% error of the experimentally determined values.[28]

The charge groups used for the DMPC representation in all simulations are displayed in Figure 5.5.

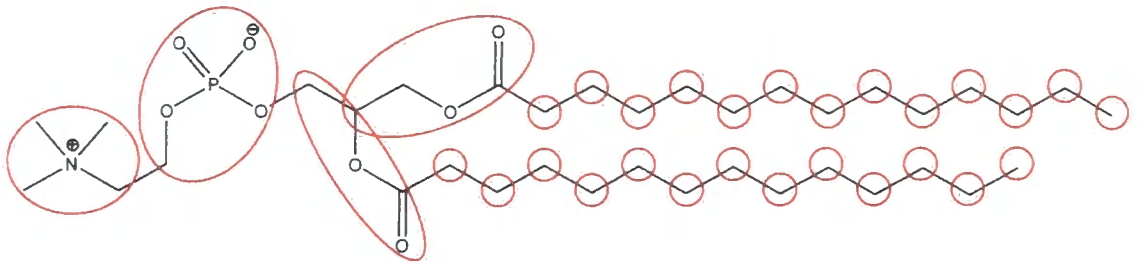


Figure 5.5: Charge groups used in for the representation of DMPC molecules.

A single DMPC molecule was inserted into a solvent box containing 512 chloroform molecules. Solvent molecules that lay within the van der Waals radius of the lipid molecule were removed using a self-written FORTRAN 95 program. Typically, this left around 496 solvent molecules in a cubic box with a length of 3.9 nm per side. The energy of the system was minimized and subsequently the set-up equilibrated for 10 ns. This resulted in the starting structure for all the molecular dynamics simulations of the lipid in chloroform.

5.1.4.c AcWLWLL Representation

The structure of the pentapeptide AcWLWLL was generated using the Maestro software package.[20] The geometry was minimized and partial charges computed and assigned using Jaguar *ab-initio* calculations (6-31G** basis set). The peptide was transferred into GROMACS format and inserted into a pre-equilibrated solvent box (5.6 nm per side) containing 5832 SPC water molecules, of which 119 were removed due to their overlapping with the peptide. Representative conformations of AcWLWLL at 300 K were sampled by two methods: A conformational search was run in Maestro using explicit solvent. 10000 conformations of the peptide were sampled using MCMM tensional sampling. The energy window for the saving of structures was set to 50 kJ mol^{-1} and the maximum distance between atoms in equal structures to 0.025 nm. Secondly, a 10 ns GROMACS molecular dynamics calculation was run, with subsequent cluster analysis to sample conformers. As detailed in §2.7.2.b, both revealed two statistically significant conformations: one in which the pentapeptide adopts a more helical form, and the other in which it is more stretched and W-shaped.

5.1.4.d Representations of AcWLWLL Derivatives

Derivatives of the parent AcWLWLL peptide were generated by substituting the hydrogen on the 5-position on the W_1 indole ring by one of the halogen atoms, F, Cl, Br, or I, where W_1 stands for the first tryptophan residue. The structures were the energy-minimised using the GROMACS software.

The charge groups used for the representation of the pentapeptides are shown in Figure 5.6.

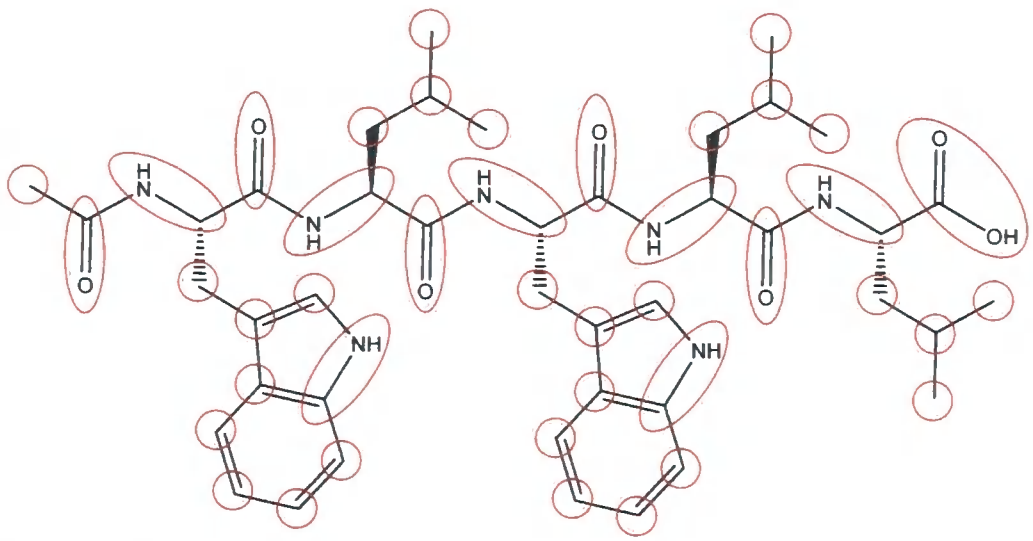


Figure 5.6: Charge groups used in the representation of AcWLWLL peptides.

5.2 Run Parameters

5.2.1 Molecular Dynamics Simulations – Amino Acid

The force field used in the simulations was based on OPLS_AA[29]. Simulations were performed in the constant *NpT* ensemble with the temperature controlled at 295 K using the weak coupling algorithm of a Berendsen thermostat[30] with a coupling constant of 0.1 ps. The pressure was set to 1 bar and coupled every 1.0 ps using a Berendsen barostat[31]. DMPC, tryptophan derivative and the solvent were treated and coupled independently. The smooth particle mesh Ewald (PME) summation (Equations 4.6-4.7)[32,33] was used to evaluate the electrostatic interactions with a real space cut-off of 1.2 nm, fourth-order spline interpolation, and a grid spacing of 0.12 nm. Van der Waals interactions were calculated using a Lennard-Jones potential (equation 4.1) with a truncation at 1.9 nm. The electrostatic interactions were

calculated using PME. All cut-offs are given as atom based distances. Bond lengths of all molecules were constrained using the LINCS algorithm[34]. Using such constraints allowed a time step of 2 fs to be used. Coordinates and velocities of all atoms were recorded every 10 ps.

5.2.1.a Simulation of Amino Acid – Lipid interactions at a ratio of 1:1

Various starting configurations were chosen to reduce the importance of the initial arrangement: one N^{α}-acetyl-*L*-tryptophan ethyl amide derivative was placed on one of 6 different positions on a sphere with 2 nm radius around the centre of a DMPC molecule. Overlapping solvent molecules were removed (typically leaving 450 chloroform molecules) and the starting configurations equilibrated for 10 ns before data was recorded. Both, the partial charge distributions of Set A and Set B were used in independent simulations. Production runs with the parameters were performed over 100 ns. A typical run would require about 2.14 hours per ns of simulation on four 3.73 GHz Xeon Intel Processors.

5.2.1.b Simulation of Amino Acid – Lipid – Water interactions at a ratio of 1:1:1

A set-up for the simulation of amino acid – lipid interactions at a ratio of 1:1 was used as the starting configuration (§5.2.1.a). An additional water molecule, represented with the SPC model, was inserted within 0.5 nm of the lipid head group. Overlapping solvent molecules were removed, leaving 434 chloroform molecules and the system pre-equilibrated for 10 ns. A production run with the parameters as described above was then performed over 100 ns requiring about 2.29 hours per ns of simulation on four 3.73 GHz Xeon Intel Processors.

5.2.1.c Simulation of Amino Acid – Lipid interaction at a ratio of 2:1

Two molecules of N^{α}-acetyl-5-fluoro-*L*-tryptophan ethyl amide (**6n**) were placed opposite to each other 2 nm away from the centre of a central DMPC molecule. Set B was used for the treatment of partial charges. The three molecules of interest were placed in the centre of a simulation box containing 4024 chloroform molecules. Overlapping solvent molecules were removed and the system pre-equilibrated for 10 ns. Several production runs with the parameters as described above were then performed over 10 ns each for a total of 50 ns. A

typical run would require about 25.1 hours per ns of simulation on four 3.73 GHz Xeon Intel Processors.

5.2.1.d Simulation of Amino Acid – Amino Acid interactions at a ratio of 1:1

Analogous to the simulation of amino acid - lipid interactions at a ratio of 2:1 (5.2.1.c), two N^{α}-acetyl-5-fluoro-*L*-tryptophan ethyl amide molecules (**6n**) were placed 2 nm away from each other into a simulation box containing 4024 chloroform molecules. Overlapping solvent molecules were removed and the system pre-equilibrated for 10 ns. Set B was used for the treatment of partial charges. Several production runs, with the parameters as described above, were then performed over 10 ns each for a total of 50 ns. A typical run would require about 21.93 hours per ns of simulation on four 3.73 GHz Xeon Intel Processors.

5.2.1.e Cluster Analysis Adducts

The trajectory from the molecular dynamics runs obtained through the protocols above (§5.2.1.a) were analysed for frames, when the distance between the centres of mass of the two molecules to be analysed (*e.g.* tryptophan derivative and the DMPC lipid) were ≤ 0.5 nm. These frames were combined into a new trajectory using GROMACS internal software (trjcat). The extracted structures were then subjected to a cluster analysis tool (g_cluster), which assigned them to a cluster when the root mean square difference in their coordinates was ≤ 0.1 nm (measured by comparing $atom_{i1}$-$atom_{i2}$ distances, where 1 and 2 stand for structures 1 and 2, and i is an atom descriptor.)

5.2.2 Molecular Dynamics Simulations – Peptides

One of the two statistically significant conformations of the AcWLWLL pentapeptide was placed 1.5 nm away from a DMPC lipid bilayer face. Overlapping solvent molecules were removed, typically leaving 3640 SPC molecules, and the system pre-equilibrated for 1 ns, followed by a production run over 50 ns. An OPLS_AA force field with incorporated Ryckaert-Bellemans parameters for the treatment of the lipid was used. The simulations were performed in the constant *NpT* ensemble. The pressure was coupled semi-isotropically at 1 bar with a coupling constant of 1.0 ps both, in the plane of the lipid bilayer and to its

normal using a Berendsen barostat, which applied two pressure couplings with independent coupling constants to the bilayer normal and the bilayer plane respectively. The temperature was set to 300 K and coupled to the lipid molecules, the solvent and any inserted peptides independently every 0.1 ps through a Berendsen thermostat. The electrostatic interactions were modelled by PME summation with a real-space cut-off at 1.2 nm, fourth-order spline interpolation, and a grid spacing of 0.12 nm. Every 10^{th} step, the truncation for the PME summation was set to 1.5 nm. A Lennard-Jones potential with a cut-off at 1.5 nm was used to calculate van der Waals interactions. Water representations were restricted using the SETTLE algorithm[35]; bond lengths of the other molecules were constrained using the LINCS algorithm[34]. The simulations were run with a time step of 2 fs. Coordinates and velocities of all atoms were recorded every 10 ps. A typical run would require about 11.77 hours per ns of simulation on four 3.73 GHz Xeon Intel Processors. The other conformer and peptides analysed were treated in an analogous fashion.

5.2.3 Molecular Dynamics Simulations – Free Energy

5.2.3.a Free Energies of Solvation – N^{α}-Acetyl-L-Tryptophan Ethyl Amide Derivatives in Chloroform

Free energy calculations of N^{α}-acetyl-*L*-tryptophan ethyl amide derivatives were performed both *in vacuo* and when the molecules were dissolved in chloroform. The 1,4-interactions with neighbouring molecules of the tryptophan analogue were gradually turned off over 21 linear steps. After a short pre-equilibration period at each step (100 ps), the free energy of the system was then calculated in a 400 ps production run. The OPLS_AA force field was used. Temperature and pressure were set constant at 295 K and 1 bar, both being coupled independently to the different molecule types *via* Berendsen algorithms with coupling constants of 0.1 ps and 2.0 ps respectively. All bonds were restrained using the LINCS algorithm. Electrostatic interactions were computed by the PME summation[33] using a cut-off at 1.9 nm, fourth-order spline interpolation, and a grid spacing of 0.12 nm. Van der Waals interactions were modelled *via* a Lennard-Jones potential with a truncation at 1.9 nm, updating the neighbour list every 5 steps. Energies were recorded every 0.2 ps. Non-bonded interactions were interpolated linearly with the coupling parameter λ. When going from the fully interacting state A ($\lambda = 0$) to the fully non-interacting state B ($\lambda = 1$), all bonded

interactions were interpolated by linear interpolation of the interaction parameters. The soft-core parameters were: $\alpha = 1.51$, soft core power $p = 1$, and a radius of interaction $\sigma = 0.3$ nm.

5.2.3.b *Free Energies of Binding between N^{α}-Acetyl-L-Tryptophan Ethyl Amide Derivatives and DMPC in Chloroform*

Simulations for the calculation of binding free energies used the same parameters as described in §5.2.3.a. The starting configuration for System 1 were obtained from the initial configuration used in the 1:1 association simulations for the tryptophan derivatives and DMPC (as detailed in §5.2.1.a). The starting configuration of the Trp:lipid 1:1 adducts for System 2 were chosen by evaluating the trajectory of the unperturbed molecular dynamics simulations of the 1:1 set-up (§2.5.2). States in which the two molecules could be considered as being bound ($r_{AB} \leq 0.5$ nm) were extracted and a cluster analysis performed according to the procedure described in §5.2.1.e. The average structures of the most populated obtained cluster were used as starting configuration to evaluate free energies of association.

5.3 *References*

[1] a) Lagüe, P., Pastor, R. W., Brooks, B. R., *J. Phys. Chem. B.*, **2004**, *108*(1), 363-368
b) van Buuren, A. R., Marrink, S. J., Berendsen, H. J. C., *J. Phys. Chem.*, **1993**, *97*, 9206-9212
[2] Berger, O., Edholm, O., Jahnig, F., *Biophys. J.*, **1997**, *72*(5), 2002-2013
[3] Alper, H., Bassolino, D., Stouch, T., *J. Chem. Phys.*, **1993**, *98*, 9788
[4] Anézo, C., de Vries, A. H., Höltje, H.-D., Tieleman, D. P., Marrink, S.-J., *J. Phys. Chem. B*, **2003**, *107*, 9424-9433
[5] Berendsen, H. J. C., van der Spoel, D., van Drunen, R., *Comput. Phys.Commun.*, **1995**, *91*, 43-56.
[6] Lindahl, E., Hess, B., van der Spoel, D., *J. Mol. Modell.*, **2001**, *7*, 306-317.
[7] Nagle, J.F., Tristram-Nagle, S., *Biochim. Biophys. Acta*, **2000**, *1469*(3), 159-195
[8] Anézo, C., de Vries, A. H., Höltje, H.-D., Tieleman, D. P., Marrink, S.-J., *J. Phys. Chem. B*, **2003**, *107*, 9424-9433
[9] Schuy, S., Janshoff, A., *Chem. Phys. Chem.*, **2006**, *7*(6), 1211-1214
[10] Koynova, r., Caffrey, M., *Biochim. Biophys. Acta*, **1998**, *1376*, 91-145
[11] Enders, O., Ngezahayo, A., Wiechmann, M., Leisten, F., Kolb, H.A., *Biophys. J.*, **2004**, *87*, 2522-2531
[12] Jorgensen, W. L., Briggs, J. M., Contreras, M. L., *J. Phys. Chem.*, **1990**, *94*(4), 1683-1686
[13] Tironi, I. L., van Gunsteren, W. F., *Mol. Physics*, **1994**, *83*(2), 381-403
[14] Jorgensen, W. L., Madura, J. D., Swenson, C. J., *J. Am. Chem. Soc.*, **1984**, *106*, 6638-6646
[15] Berendsen, H. J. C., Postma, J. P. M., van Gunsteren, W. F., Hermans, J., in *Intermolecular Forces*; Pullman, B., Ed., Reidel: Dordrecht, **1981**, pp. 331-342
[16] Berendsen, H. J. R., Grigera, J., Straatsma, T., *J. Phys. Chem.*, **1987**, *91*, 6299
[17] Jorgensen, W. L., *J. Am. Chem. Soc.*, **1981**, *103*, 355
[18] Guissani, Y., Gut, B., *J. Chem. Phys.*, **1993**, *98*, 8221
[19] Tieleman, D. P., Berendsen, H. J., R., *J. Chem. Phys.*, **1996**, *105*(11), 4871-4880
[20] Maestro, Schroedinger Inc., www.schrodinger.com
[21] Schrodinger, Inc.: Portland, OR, 1991-2000

[22] Gaussian 03, Revision 1, Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, Jr., J. A.; Vreven, T.; Kudin, K. N.; Burant, J. C.; Millam, J. M.; Iyengar, S. S.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.; Scalmani, G.; Rega, N.; Petersson, G. A.; Nakatsuji, H.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Klene, M.; Li, X.; Knox, J. E.; Hratchian, H. P.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Ayala, P. Y.; Morokuma, K.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Raghavachari, K.; Foresman, J. B.; Ortiz, J. V.; Cui, Q.; Baboul, A. G.; Clifford, S.; Cioslowski, J.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Gonzalez, C.; and Pople, J. A.; Gaussian, Inc., Wallingford CT, 2004

[23] Breneman, C. M., Wiberg, K. B., *J. Comp. Chem.*, **1990**, *11*(3), 361-373

[24] Jorgensen, W., Tirado-Rives, J., *J. Am. Chem. Soc.*, **1988**, *110*, 1657-1666

[25] Berendsen, H. J. C., van der Spoel, D., van Drunen, R., *Comput. Phys.Commun.*, **1995**, *91*, 43-56.

[26] Tieleman, D.P., MacCallum, J.L., Ash, W.L., Kandt, C., Xu, Z., Monticelli, L., *J. Phys.: Condens. Matter.*, **2006**, *18*, S1221–S1234

[27] Tieleman, D.P., MacCallum, J.L., *J. Comput. Chem.*, **2003** , *24*, 1930–1935

[28] Berger, O., Edholm, O., Jahnig, F., *Biophys. J.*, **1997**, *72*(5), 2002-2013

[29] Jorgensen, W. L., Maxwell, D. S., TiradoRives, J., *J. Am. Chem. Soc.*, **1996**, *118*, 11225-11236

[30] Berendsen, H. J. C., Postma, J. P. M., DiNola, A., Haak, J. R., *J. Chem.Phys.* **1984**, *52*, 1695-1697

[31] Berendsen, H. J. C., Postma, J. P. M., van Gunsteren, W. F., DiNola, A., Haak, J. R., *J. Chem. Phys.*, **1984**, *81*(8), 3684-3690

[32] Darden, T.; York, D.; Pedersen, L. *J. Chem. Phys.* **1993**, *98*, 10089-10092.

[33] Essmann, U., Perera, L., Berkowitz, M. L., Darden, T., Lee, H., Pedersen, L. G., *J. Chem. Phys.*, **1995**, *103*, 8577-8592

[34] Hess, B., Bekker, H., Berendsen, H. J. C., Fraaje, J. G. E. M., *J. Comp. Chem.*, **1997**, *18*, 1463-1472

[35] Miyamoto, S., Kollman, P.A., *J. Comp. Chem.*, **1992**, *13*, 952–962

Chapter 6: Conclusions

It was shown that single-molecule binding interactions between lipid and tryptophan derivatives in the non-competitive solvent chloroform can be studied successfully using a variety of techniques such as NMR host-guest and isothermal calorimetry titrations (ITC) and molecular dynamics (MD) simulations. This enabled the importance of electrostatic interactions to be assessed in a model system where the number of secondary or multiple interaction pathways could be either avoided or accounted for. The specific interactions between DMPC lipid and derivatised tryptophan compounds were analysed, the latter having been produced in good to excellent yields using an improved synthesis of Snyder *et al.*[1] The prepared free amino acid derivatives could be converted to 5-monosubstituted N^{α}-acetyl-*L*-tryptophan ethyl amides, which mimic the non-terminal tryptophan residue in proteins and peptides. The implemented binding model accounted for lipid micellisation, water-lipid and water-lipid-tryptophan equilibria extremely well. With it, the critical micelle concentration of DMPC in chloroform (19.71 ± 3.28 m*M* at 21 °C) could be determined *via* NMR dilution titrations of hydrated DMPC into chloroform, which also revealed the water content of the lipid as purchased (2 molecules of water per lipid) and highlighted water as an association partner of DMPC in chloroform. These observations were supported through ITC experiments, where an observed heat of 6 cal mol^{-1} could be attributed to the release of water from the lipid and the CMC for DMPC at 21 °C estimated to lie above 15 m*M* in agreement with available literature data.[2]

The fitting of observed chemical shift changes of the water proton signals to a binding isotherm could be achieved with excellent accuracy (error between calculated and observed values <3%). The importance of water associations was also seen in the NMR host-guest titrations with the tryptophan derivatives (**6a-g, 6n)** as host and DMPC as guest. Here, water was found to associate with the lipid in a competitive manner with respect to the amino acid analogue. Binding interactions between water and lipid and Trp:DMPC complexes could be determined with a high accuracy as indicated by the small observed errors (<5%) in the calculated binding constants of repeat runs.

The binding of the unsubstituted tryptophan analogue (**6n**) to DMPC was found to be the least favoured (ΔG_1 = -6.05 ± 0.35 kJ mol^{-1}) while both, the association free energies for the formation of a 1:1 Trp:DMPC adduct with 5-MeO (**6f**) and 5-NO_2-substituted (**6g**)

compounds were determined to be approximately twice that energy (ΔG_1 = -10.20 ± 0.16 and ΔG_1 = -11.10 ± 0.53 kJ mol^{-1} respectively). ΔG_1 values derived from MD simulations gave highly consistent results for all models analysed. The two sets of partial charges used to represent charges on the indole ring resulted in slight deviations from the n-shaped pattern indicating that it is sensitive to the differences in the partial charges (±0.2 e). However, the absolute values obtained for the association free energies were still within 1.9 ± 0.9 kJ mol^{-1} to the experimentally obtained values. The truncation of long-range interactions in the system under investigation from 1.9 to 1.2 nm only had a minor effect on the obtained free energies of association. The treatment using the larger truncation was found to approximate experimentally obtained free energies of association the best with an average deviation of 1.14 ± 0.79 kJ mol^{-1}. Therefore, this model and partial charge set was used for all subsequent simulations of the tryptophan derivatives.

The calculation of binding constants and associated free energies for complex formation using the ratios of bound and unbound tryptophan derivatives approximated experimentally obtained values extremely well, slightly overestimating the free energies of compounds with moderate electron density on the indole ring (**6a-b**, **6n**). The free energies of association were underestimated to some extend (2.81 ± 0.98 kJ mol^{-1}) when calculated by the survival function of the Trp:lipid complex. The calculated free energy can be represented schematically using:

$$\Delta G_1^{rHF} < \Delta G_1^{NMR} < \Delta G_1^{MP2a} \sim \Delta G_1^{MP2b} < \Delta G_1^{MP2b2},$$

where the superscripts indicate the methods used to determine the value.

An n-shaped pattern was observed when the free energies of association of tryptophan derivatives and DMPC were plotted against the Hammett parameter. The pattern obtained on a microscopic level was widely conserved on the macroscopic level, which could be approximated by summation of the individual contributions of the free energies of 1:1 and 2:1 Trp:DMPC adduct formation. In the analysis of the observed heats associated with the binding of the tryptophan derivatives to DMPC, ITC experiments also exhibited a similar pattern, which suggests that there is no first-order relationship between the electron density on the indole moiety of tryptophan compounds and their binding strength to DMPC molecules, but instead a more complex interplay of forces and second-order electrostatic interactions. The observed break in the trends either side of tryptophan (R = H) is consistent

with as least two factors influencing the observed binding. For higher order complexes these factors are likely to include hydrogen bonding and cation-π electron interactions between amino acid analogues. However, no significant direct influence of the electronic density on the strength and preferred interaction range of cation-π interactions between the lipid choline head group and the indole side chain of the tryptophan derivatives was observed in both the 1:1 and the 2:1 Trp:DMPC binding interactions. This was also reflected in the poor clustering of binding conformations between the lipid and the amino acid analogue in the 1:1 Trp:DMPC complex. The latter was shown not to be supported by hydrogen bonding interactions between the two molecules in agreement with observations made in the literature.[3,4] However, as seen in both deuterium exchange NMR titrations and molecular dynamics simulations of a Trp:DMPC 2:1 complex, hydrogen bonding becomes more importance in higher order complexes. The D_2O titrations illustrated a relationship for the strength and accessibility of hydrogen bonds that followed indNH < AcNH << EtNH with the indole amine proton (indNH) being the most easily removed and the most accessible one in relation to the acetyl amide and ethyl amide protons (AcNH and EtNH respectively). The MD simulations suggested that hydrogen bonding interactions are a major driving force in the formation and stabilisation of a 1:1 Trp:Trp complex, whose formation appeared to be enhanced and possibly fostered by DMPC molecules present in solution, as simulations of two tryptophan derivatives in the absence of lipid did not show any significant association between the two amino acid analogues in the time span analysed.

Isothermal titration calorimetry showed that the binding of tryptophan analogues to DMPC molecules can be studied in the solvent chloroform, which is a non-standard solvent for ITC experiments. The data analysis of the binding isotherms was complicated through the underlying signal of the heat of dilution of micellar DMPC solutions. However, the concentration of lipid solutions employed was necessary not only to increase the signal to noise level (which is high for measurements in chloroform), but also to approximate the conditions used in the NMR host-guest titrations. The non-sigmoidal response of the observed heat of reaction between the analogues (**6a-g** & **6n**) and the level of detail required, rendered the data analysis tools that the provided analysis software (MicroCal) offered unfeasible. Instead the data was analysed in more qualitative terms by observation of the heat flow curves and associated enthalpograms. However, the qualitative evaluation of the ITC experiments provided data that was in good qualitative agreement with the data obtained

from NMR titrations and simulations. *I.e.* the same trend as seen in the binding free energies of the compounds studied interacting with DMPC in chloroform could be obtained for the heat signals of corresponding ITC experiments.

Free energy calculations for the solvation of the tryptophan derivatives in chloroform and for 1:1 Trp:DMPC adduct formation were performed. The obtained solvation free energies reflected the empirically observed poor solubility of the derivatives in chloroform. The free energy calculations of adduct formations had too high an error associated with them to be sensibly compared to the data obtained experimentally.

Finally, peptide-lipid interactions between peptides of the AcWLWLL family to a fully hydrated DMPC bilayer were investigated *via* molecular dynamics simulations. The conformational sampling of the peptide revealed two statistically significant conformations that were analysed independently in subsequent simulations. Associative interactions of model peptides of the AcW*LWLL family with a DMPC lipid bilayer were observed. Here, W* stands for a tryptophan residue that carries a monosubstitution of F, Cl, Br, or I on the 5-position of its indole ring. The peptides quickly (<5 ns) dock at the bilayer interface, remain a preferred mean distance from it (averaged over all: ~1.76±0.19 nm from the centre of the bilayer), and generally adopt a preferred, extremely stable conformation (occurrence >90%). These conformations show some degree of freedom around the central backbone segments of the peptide and relatively flexible Leu^4 and Leu^5 residues. An altered electron density of the indole moiety of the tryptophan residue Trp^1 was not seen to affect the association profile of the peptide with the lipid significantly and no trends can be seen within experimental error. The subtle change was not found to be sufficient to influence either the formation of binding conformations or the strength of the peptide-lipid interactions significantly.

In summary, this research shows that the association of tryptophan analogues with DMPC lipids in chloroform appears to be an interplay of many factors, including cation-π interactions, hydrogen bonding (for higher order complexes), and secondary factors (such as water-binding). Binding free energies might be predicted for tryptophan compounds whose electron density is similar to the compounds analysed in this study. The simulation of binding events and the computation of their associated free energies appear to be very promising, as the general trend of experimental values was reproducible and free energies of binding could be determined with a high accuracy given a reasonably long time of sampling (>50 ns). Binding of compounds could be analysed on a nanosecond timescale with atom resolution

allowing the evaluation of preferred binding conformations, hydrogen bonding as well as self-association of the tryptophan compounds and interactions with water.

The obtained results reflect the observations made for tryptophan residues and derivatives in bilayer and membrane research suggesting their transferability. They gave further insight and support to the observation that hydrogen bonding interactions between tryptophan analogues and lipid molecules are negligible. The individual interactions between the tryptophan derivatives and lipid molecules seemed also not to be governed or significantly influenced by the permanent dipole of the amino acid analogues. The cation-π interaction of the indole ring to the charged lipid head group was also seen in this research. Although it may not be linearly influenced by the electron density of the aromatic ring, binding affinities of tryptophan compounds can still be altered and adjusted using substituents on the indole ring. Such alterations might be of predictable strength for compounds with Hammett parameters with the range analysed in this study.

6.1 Suggestions for Further Work

The scope of the improved synthesis for tryptophan derivatives, as demonstrated in this thesis, should allow the production of a variety of derivatives carrying multiple substitutions and the synthesis of 4- and 7-monosubstituted tryptophan derivatives. These could then be used to study additive effects of such alterations on the indole ring and their effect on binding to DMPC (and other lipids), The significance of inductive and mesomeric effects (if any) in tryptophan derivatives for lipid-peptide binding could then be investigated. The synthesis of a variety of such tryptophan derivatives is currently in progress in the Sanderson lab.

The effect of multiple substitutions and the importance of the location of an altered tryptophan residue in model peptides of the type AcWLWLL could also be assessed with minor alterations to the simulation set-ups that were used in this study. The validity and quality of such computations could be assessed further by experimental analysis of the designed peptides (*e.g. via* ITC). The starting materials for the synthesis of these peptides and the proof of concept that such a synthesis is feasible were provided in the preparation for this thesis.

The use of monosubstituted tryptophan analogues as fluorescent probes and the intrinsic change in their emission spectrum upon binding to DMPC could provide an

interesting tool for the multiple, simultaneous structure determination of proteins and peptides and their adopted conformations upon binding to lipid bilayers. (Substituted tryptophan compounds have been shown to exhibit different absorption/emission profiles, which could be used to distinguish between individual tryptophan residues in a protein sequence).

6.2 References

[1] Snyder, H. R., MacDonald, J. A., *J. Am. Chem. Soc.*, **1955**, *77*, 1257-1259
[2] Haque, R., Tinsley, I. J., Schmedding, D., *J. Biol. Chem.*, **1972**, *247*(1), 157-161
[3] Persson, S., Killian, S. A., Lindblom, G., *Biophys. J.*, **1998**, *75*(3), 1365-1371
[4] Gaede, H. C., Yau, W.-M., Gawrisch, K., *J. Phys. Chem. B.*, **2005**, *109*(26), 13014–13023

Chapter 7: Appendix

7.1 Solving a Cubic Equation

A cubic equation $ax^3 + bx^2 + cx + d = 0$ can be solved by substituting x with $x = y - \frac{b}{3a}$. This results in the depression of the cubic equation to the quadratic expression:

$$ay^3 + \left(c - \frac{b^2}{3a}\right)y + \left(d + \frac{2b^3}{27a^2} - \frac{bc}{3a}\right) = 0. \qquad (7.1)$$

Let $q = \frac{9abc - 27a^2d - 2b^3}{54a^3}$

$r = \sqrt{\left(\frac{3ac - b^2}{9a^2}\right) + q^2}$.

$s = \sqrt[3]{q + r}$, and

$t = \sqrt[3]{q - r}$.

The solutions for x are then:

$$x_1 = s + t - \frac{b}{3a},$$

$$x_2 = -\frac{1}{2}(s + t) - \frac{b}{3a} + \frac{\sqrt{3}}{2}(s - t)i,$$

$$x_3 = -\frac{1}{2}(s + t) - \frac{b}{3a} - \frac{\sqrt{3}}{2}(s - t)i.$$

Applied to equation 2.1.4, the factors a, b, c, and d can be expressed as:

$$a = K_m(K_m - K_d),$$

$$b = \left(2K_d - 2K_m - K_m{}^2 L_o^{syr}\right),$$

$$c = \left(1 + 2K_m L_o^{syr}\right),$$

$$d = -L_o^{syr}.$$

7.2 *Calculations of the Concentrations of Species*

7.2.1 NMR Titrations of hydrated DMPC into Chloroform

7.2.1.a Concentration of Water:Lipid Adducts

The concentrations of water:lipid adducts and free, unbound species were calculated iteratively by solving the following equations:

$$[W] = [W]_0 - [WL] - 2[W_2L]$$

$$[WL] = \frac{x - \sqrt{x^2 - 4K_{1wl}{}^2 Z^2 ([W]_0 - 2[W_2L])([L]_0 - [W_2L])}}{2K_{1wl}Z}$$

where $x = 1 + 2K_{1wl}Z([L]_0 - 3[W_2L] + [W]_0)$ and

$$Z = 1 + [L]^2 K_m{}^2 - 2[L]K_m$$

$$[W_2L] = \frac{x - \sqrt{x^2 - 16K_{2wl}{}^2 ([W]_0 - [W])([W]_0 - [WL])}}{8K_{2wl}}$$

where $x = 1 + 2K_{2wl}(2[W]_0 - [W] - [WL])$

$$[L] = \frac{1 + x - \sqrt{(1+x)^2 - (2K_m x)^2}}{xK_m{}^2}$$

where $x = 2K_m([L]_0 - [WL] - [W_2L])$

$$[L]_2 = K_m[L]^2$$

$$[L]_n = [L]_0 - [L] - [WL] - [W_2L] - 2[L]_2$$

7.2.1.b Fitting of the Observed Chemical Shift for DMPC blank titrations

The observed complexation-induced chemical shift for the water protons $\delta_{obs,W}$ could then be fitted to

$$\delta_{obs,W} = \left[\frac{K_{1wl}[L](\delta_{bound\,1w} - \delta_f) + 2K_{2wl}[WL](\delta_{bound\,2w} - \delta_f)}{1 + K_{1wl}[L] + 2K_{2wl}[WL]}\right] + \delta_f,$$

where K_{1wl} and K_{2wl} are the binding constants for the formation of WL and W_2L complexes respectively with their associated limiting chemical shifts $\delta_{bound1w}$ and $\delta_{bound2w}$; δ_f refers to the chemical shift of the free, unbound water species W.

The observed complexation-induced chemical shift for the lipid protons $\delta_{obs,L}$ could then be fitted to

$$\delta_{obs,L} = \left[\frac{K_{1wl}[W](\delta_{bound\,1w} - \delta_f) + K_{1wl}K_{2wl}[W](\delta_{bound\,2w} - \delta_f) + K_{1wl}[W](\delta_{bound\,2w} - \delta_f)}{1 + K_{1wl}[W] + K_{1wl}K_{2wl}[W]}\right] + \delta_f,$$

where K_{1wl} and K_{2wl} are the binding constants for the formation of WL and W_2L complexes respectively with their associated limiting chemical shifts $\delta_{bound1w}$ and $\delta_{bound2w}$; δ_f refers to the chemical shift of the free, unbound water species L.

7.2.2 NMR Host-Guest Titrations of Tryptophan Derivatives with DMPC

7.2.2.a Concentration of Trp:Lipid and Water:Lipid Adducts

The concentrations of 1:1 and 2:1 Trp:lipid adducts and water:lipid adducts were calculated iteratively by solving the following equations:

$$[LT] = \frac{x - \sqrt{x^2 - 4K_1^2([L]_0 - [WLT] - [WL] - [W_2L] - [LT_2])([T]_0 - 2[LT_2] - 2[LTW])}}{2K_1}$$

where $x = 1 + K_1([L]_0 - 2[WLT] - [WL] - [W_2L] - 3[LT_2] + [T]_0)$

$$[LT_2] = \frac{x - \sqrt{x^2 - 8K_2^2([L]_0 - [L] - [WLT] - [WL] - [W_2L])([T]_0 - [LT] - [LTW])}}{4K_2}$$

where $x = 1 + K_2([T]_0 - [LT] - 3[WLT] + 2([L]_0 - [L] - [WL] - [W_2L]))$

$$[WLT] = \frac{x - \sqrt{x^2 - 4K'^2_1([L]_0 - [L] - [LT] - [LT_2] - [W_2L])([T]_0 - [LT] - 2[LT_2])}}{2K'_1}$$

where $x = 1 + K'_1([T]_0 - 2[LT] - 3[LT_2] + [L]_0 - [L] - [W_2L])$

$$[WL] = \frac{x - \sqrt{x^2 - 4K^2_{1wl}([W]_0 - 2[W_2L] - [WLT])([T]_0 - [LT] - [WLT] - [W_2L][LT_2])}}{2K_{1wl}}$$

where $x = 1 + K_{1wl}([W]_0 - 3[W_2L] - 2[WLT] - [LT_2] + [L]_0 - [LT])$

$$[W_2L] = \frac{x - \sqrt{x^2 - 8K^2_{2wl}([L]_0 - [L] - [WL] - [WLT] - [LT_2])([W]_0 - [WL] - [WLT])}}{4K_{2wl}}$$

where $x = 1 + K_{2wl}([W]_0 - [WL] - 3[WLT] + 2([L]_0 - [L] - [LT] - [LT_2]))$

$$[L]_n = [L]_0 - [LT] - [LT_2] - [WLT] - [WL] - [W_2L]$$

7.2.2.b Concentration of Free, Unbound Species

The concentrations of free tryptophan, lipid and water were calculated according to the concentration differences:

$$[T] = [T]_0 - [LT] - 2[LT_2] - [WLT]$$
$$[L] = [L]_0 - [LT] - [LT_2] - [WLT] - [WL] - [W_2L] - [L_n]$$
$$[W] = [W]_0 - [WLT] - [WL] - 2[W_2L]$$

7.2.2.c Fitting of the Observed Chemical Shifts

The observed chemical shift δ_{obs} could then be fitted to

$$\delta_{obs} = \frac{K_1[L]\left(\delta_{bound1} - \delta_{free}\right) + K_1'[WL]\left(\delta_{bound1'} - \delta_{free}\right) + K_2[LT]\left(\delta_{bound2} - \delta_{free}\right)}{\left(1 + K_1[L] + K_1'[WL] + K_2[LT]\right)} + \delta_{free}$$

where δ_{free}, δ_{bound1}, δ_{bound2}, and $\delta_{bound1'}$ are the chemical shifts of the tryptophan analogue as free species, in the 1:1 and 2:1 Trp:lipid adduct and the 1:1:1 Trp:water:lipid adducts respectively with K_1, K_2, and K_1' describing the corresponding binding constants at these states.

7.3 *NMR Chemical Shift Data*

7.3.1 NMR Titration of Hydrated DMPC into Chloroform

	H_2O	$-N(CH_3)_3$	*sn* $-CH_2$	*sn* $-CH_2$	*sn* -CH	
δ_f	1.468	3.402	4.294	4.501	5.256	ppm
δ_{1wl}	3.326	3.430	2.450	3.619	5.151	ppm
δ_{2wl}	5.759	3.270	3.934	4.272	5.202	ppm
δ_m	-	3.399	4.504	4.616	5.180	ppm
$\Delta\delta_{1wl}$	1.858	0.028	-1.844	-0.882	-0.105	ppm
$\Delta\delta_{2wl}$	4.291	-0.132	-0.359	-0.229	-0.054	ppm
$\Delta\delta_m$	-	-0.003	0.210	0.115	-0.076	ppm
K_{1wl}	31.73	31.73	32.05	31.73	31.73	M^{-1}
K_{2wl}	122.09	122.09	117.99	122.09	122.10	M^{-1}
K_m	143.78	143.78	146.16	143.79	143.73	M^{-1}

Table 7.1: Chemical shift data for the titration of hydrated DMPC into chloroform. δ_f refers to the chemical shift of the signal of the free unbound species; δ_{1wl}, δ_{2wl} and δm are the limiting chemical shifts for the 1:1 and 2:1 water:lipid adducts and micellar lipid respectively with their associated binding constants K_1, K_2, and K_m. $\Delta\delta_{1wl}$, $\Delta\delta_{2wl}$ and $\Delta\delta_m$ stand for the chemical shift change of the adducts with respect to δ_{free}.

7.3.2 NMR Host-Guest Titrations with N^{α}-acetyl-*L*-tryptophan derivatives (6a-g, 6n)

The chemical shift data of the free, uncomplexed tryptophan derivative, δ_{free}, the limiting chemical shifts δ_{bound1} and δ_{bound2} for the 1:1 and 2:1 Trp:lipid adduct formation with the respective binding constants K_1 and K_2 are given in this section. The calculated values for the limiting chemical shift for the association of a tryptophan derivative to a 1:1 DMPC:water

adduct, $\delta_{bound1'}$, and the associated binding constant for this process, K_1' are presented along with K_{1w}', the association constant for the binding of water to a 1:1 Trp:lipid complex. Relative chemical shift changes of the limiting chemical shifts of adducts with respect to δ_{free} are shown as $\Delta\delta_1$, $\Delta\delta_2$, and $\Delta\delta_1'$ respectively.

N^{α}-Acetyl-*L*-tryptophan ethyl amide (6n)											
δ_{free}	8.166	7.761	7.394	5.460	2.004	6.361	0.885				
$\delta_{bound\ 1}$	6.880	8.546	8.289	6.946	1.846	2.706	0.873				
$\delta_{bound\ 2}$	9.360	7.532	6.806	4.967	2.012	3.291	0.858				
$\delta_{bound\ 1'}$	13.821	7.353	7.460	9.392	1.889	8.959	1.121				
$\Delta\delta_1$	-1.286	0.786	0.895	1.487	-0.158	-3.655	-0.011				
$\Delta\delta_2$	1.194	-0.228	-0.588	-0.493	0.008	-3.071	-0.026				
$\Delta\delta_1'$	5.655	-0.407	0.066	3.933	-0.115	2.598	0.237				
K_1	11.5	11.4	11.2	11.7	11.3	11.7	11.7	11.57	±	0.20	M^{-1}
K_2	114.3	114.5	115.1	113.5	115.1	112.6	112.9	113.40	±	1.11	M^{-1}
K_1'	163.4	165.0	158.7	167.8	165.2	167.5	167.1	165.68	±	3.33	M^{-1}
K_{1w}'	452.7	459.1	450.2	456.5	464.1	453.8	452.8	454.36	±	4.95	M^{-1}

N^{α}-Acetyl-5-fluoro-*L*-tryptophan ethyl amide (6a)											
δ_{free}	8.097	7.112	6.279	5.483	2.005	6.952	0.909				
$\delta_{bound\ 1}$	7.384	8.461	5.168	5.472	2.011	8.009	0.985				
$\delta_{bound\ 2}$	11.209	6.970	6.673	7.789	1.882	6.628	1.032				
$\delta_{bound\ 1'}$	15.607	7.514	9.285	10.781	1.834	6.595	1.165				
$\Delta\delta_1$	-0.713	1.350	-1.111	-0.011	0.006	1.057	0.076				
$\Delta\delta_2$	3.112	-0.142	0.394	2.306	-0.123	-0.324	0.123				
$\Delta\delta_1'$	7.510	0.402	3.005	5.298	-0.171	-0.358	0.255				
K_1	17.7	19.4	16.7	18.6	17.4	17.6	18.3	18.04	±	0.84	M^{-1}
K_2	274.9	269.7	293.9	278.8	277.8	273.6	271.0	277.24	±	9.01	M^{-1}
K_1'	158.5	158.4	158.4	158.4	158.4	158.4	163.1	158.50	±	1.91	M^{-1}
K_{1w}'	283.5	258.6	301.2	270.4	289.2	285.2	282.9	282.17	±	14.51	M^{-1}

N^{α}-Acetyl-5-chloro-*L* -tryptophan ethyl amide (6b)								
δ_{free}	8.142	7.681	2.009	0.923				
$\delta_{bound\ 1}$	8.229	7.327	1.899	6.858				
$\delta_{bound\ 2}$	13.255	7.504	1.958	0.064				
$\delta_{bound\ 1'}$	13.782	7.428	1.873	0.625				
$\Delta\delta_1$	0.087	-0.354	-0.109	5.935				
$\Delta\delta_2$	5.114	-0.177	-0.051	-0.859				
$\Delta\delta_1'$	5.640	-0.253	-0.136	-0.297				
K_1	24.3	23.9	24.2	23.8	24.00	±	0.33	M^{-1}
K_2	237.3	232.4	236.7	215.8	237.13	±	4.78	M^{-1}
K_1'	255.6	261.4	254.4	255.6	255.89	±	5.51	M^{-1}
K_{1w}'	333.6	346.8	333.9	340.6	334.16	±	12.63	M^{-1}

N^{α}-Acetyl-5-bromo-*L* -tryptophan ethyl amide (6c)											
δ_{free}	8.161	7.072	6.258	5.517	2.006	7.814	0.923				
$\delta_{bound\ 1}$	8.947	7.722	4.704	4.865	1.953	8.676	1.066				
$\delta_{bound\ 2}$	12.298	6.959	7.724	8.579	1.967	7.523	1.120				
$\delta_{bound\ 1'}$	14.144	7.452	8.901	9.214	1.885	7.527	1.108				
$\Delta\delta_1$	0.786	0.650	-1.554	-0.652	-0.052	0.862	0.143				
$\Delta\delta_2$	4.137	-0.113	1.466	3.062	-0.039	-0.291	0.198				
$\Delta\delta_1'$	5.983	0.379	2.642	3.698	-0.120	-0.287	0.185				
K_1	24.4	26.7	26.3	30.2	29.2	28.4	28.1	27.10	±	1.68	M^{-1}
K_2	231.7	193.2	175.3	199.5	204.9	196.3	212.0	211.02	±	18.92	M^{-1}
K_1'	249.9	258.0	218.7	298.1	304.7	304.0	286.4	261.09	±	35.46	M^{-1}
K_{1w}'	325.5	306.1	264.1	313.4	331.4	339.4	323.9	311.99	±	26.89	M^{-1}

N^{α}-Acetyl-5-methyl-*L* -tryptophan ethyl amide (6e)											
δ_{free}	7.992	2.014	7.062	7.025	7.520	2.469	5.331	7.020			
$\delta_{bound\ 1}$	8.283	2.091	8.567	5.738	7.905	3.194	7.531	7.569			
$\delta_{bound\ 2}$	10.849	1.901	6.065	7.672	7.246	2.018	1.718	6.689			
$\delta_{bound\ 1'}$	13.797	1.886	6.591	7.564	7.176	2.215	4.663	6.710			
$\Delta\delta_1$	0.291	0.077	1.505	-1.287	0.385	0.724	2.200	0.549			
$\Delta\delta_2$	2.857	-0.113	-0.997	0.646	-0.274	-0.452	-3.613	-0.331			
$\Delta\delta_1'$	5.805	-0.128	-0.470	0.539	-0.344	-0.254	-0.668	-0.310			
K_1	31.8	35.1	31.8	36.1	35.1	39.9	36.1	35.1	35.36	±	2.16	M^{-1}
K_2	159.4	151.9	157.5	156.7	151.8	159.7	166.7	151.8	161.36	±	5.05	M^{-1}
K_1'	191.8	249.9	238.4	255.1	249.9	249.9	258.9	249.9	212.31	±	26.95	M^{-1}
K_{1w}'	191.6	225.6	238.1	224.2	225.7	198.7	227.7	225.7	203.31	±	19.88	M^{-1}

N^{α}-Acetyl-5-methoxy-*L* -tryptophan ethyl amide (6f)												
δ_{free}	7.967	2.024	5.365	6.388	6.855	7.036	3.893	0.896				
$\delta_{bound\ 1}$	8.640	1.751	6.125	6.372	6.851	7.398	3.673	0.797				
$\delta_{bound\ 2}$	8.799	2.031	5.952	6.446	6.816	7.001	3.895	0.954				
$\delta_{bound\ 1'}$	14.995	1.928	10.656	9.274	6.556	7.398	3.814	1.233				
$\Delta\delta_1$	0.673	-0.273	0.760	-0.016	-0.004	0.362	-0.221	-0.100				
$\Delta\delta_2$	0.832	0.006	0.587	0.058	-0.039	-0.035	0.002	0.058				
$\Delta\delta_1'$	7.028	-0.096	5.291	2.886	-0.299	0.362	-0.079	0.336				
K_1	60.7	59.4	61.0	60.9	60.8	62.3	61.3	61.0	60.94	±	0.68	M^{-1}
K_2	296.7	298.6	294.6	304.6	296.7	297.4	295.9	296.5	296.22	±	3.61	M^{-1}
K_1'	168.4	162.9	164.7	166.1	168.5	170.6	167.8	168.9	166.85	±	2.65	M^{-1}
K_{1w}'	88.1	87.0	85.7	86.6	87.9	87.0	86.9	87.9	87.01	±	0.84	M^{-1}

N^{α}-Acetyl-5-nitro-*L* -tryptophan ethyl amide (6g)												
δ_{free}	8.417	8.630	8.143	2.013	0.963	7.375	6.224					
$\delta_{bound\ 1}$	12.596	8.261	7.429	1.514	1.834	8.375	9.676					
$\delta_{bound\ 2}$	12.095	8.539	8.019	1.885	0.963	7.382	6.779					
$\delta_{bound\ 1'}$	14.921	8.540	8.051	1.874	0.881	7.427	7.879					
$\Delta\delta_1$	4.180	-0.369	-0.714	-0.499	0.871	1.000	3.452					
$\Delta\delta_2$	3.678	-0.091	-0.124	-0.129	0.000	0.007	0.555					
$\Delta\delta_1'$	6.505	-0.090	-0.092	-0.139	-0.082	0.053	1.655					
K_1	89.7	89.1	89.2	83.6	91.4	89.0	86.2		88.40	±	2.44	M^{-1}
K_2	488.0	489.0	496.7	492.7	477.1	493.1	482.9		487.82	±	10.10	M^{-1}
K_1'	379.9	380.2	384.9	369.1	397.3	378.7	390.0		381.56	±	13.31	M^{-1}
K_{1w}'	134.4	135.4	137.0	140.0	137.9	135.1	143.5		136.06	±	4.81	M^{-1}

7.4 *Physical Constants of Tryptophan Analogues Plotted against Observed Free Energies of Association*

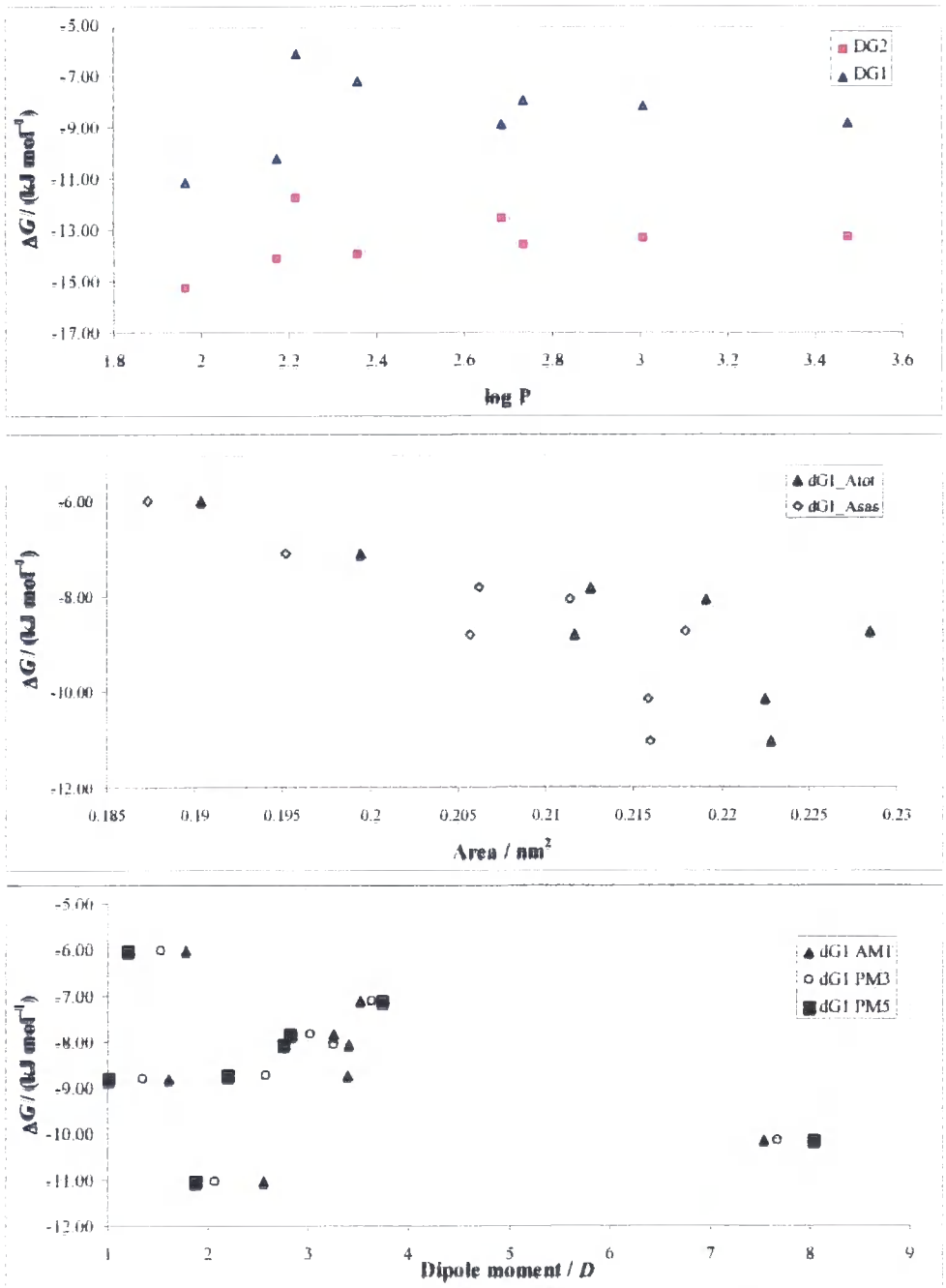


Figure 7.1: Physical constants of tryptophan analogues plotted against observed free energies of association.

7.5 $\langle g(r) \rangle_t$ of Trp1 and Trp2 to DMPC

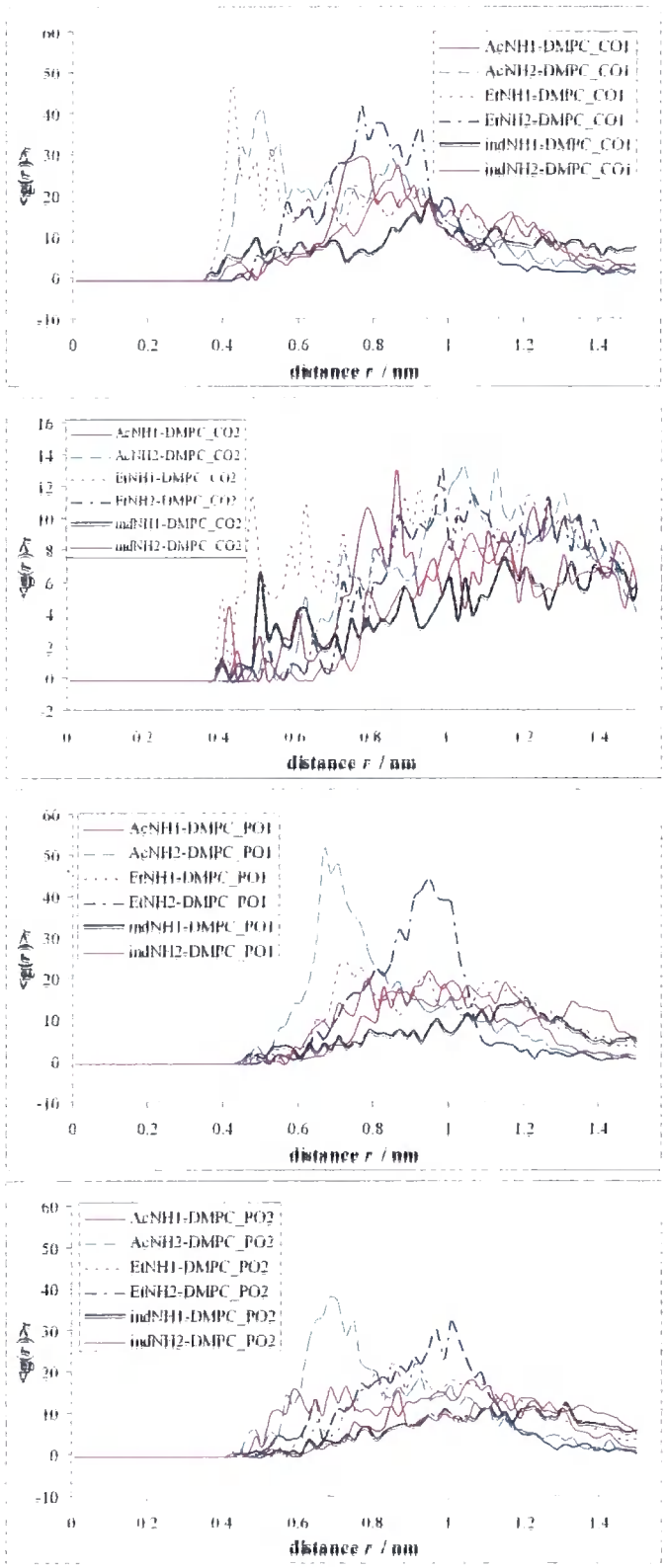


Figure 7.2: $\langle g(r) \rangle_t$ of intermolecular distances between Trp1 and Trp2 hydrogen bond donors (acetyl amide (AcNH), ethyl amide (EtNH) and indole amide protons (indNH) respectively) to hydrogen bond acceptors in the DMPC molecule (the lipid carbonyl oxygens (DMPC_CO1 and DMPC_CO2) and non-bridging phosphate oxygen atoms (DMPC_PO1 and DMPC_PO2) respectively.

7.6 *Parameter Sets:*

a)	Set A	Set Ba	Set Bb	Bilayer	
dt	2	2	2	2	fs
nstlist	10	10	10	10	
r_{list}	1.0	1.0	1.9	1.2	nm
Coulomb-type	PME				
ε_r	1	1	1	1	
$r_{coulomb}$	1.2	1.2	1.9	1.2	nm
vdW-type	Cut-Off				
r_{vdW}	1.9	1.9	1.9	1.5	nm
fourierspacing	0.12	0.12	0.12	0.12	nm
PME order	4	4	4	4	
Ewald r_{tol}	10^{-5}	10^{-5}	10^{-5}	10^{-5}	
Ewald geometry	3D				
T coupling	Berendsen				
T	295	295	295	300	K
τ_T	0.1	0.1	0.1	0.1	ps
p coupling	Berendsen				
p coupling type	isotropic	isotropic	isotropic	semi-isotropic	
p	1.0	1.0	1.0	1.0	bar
τ_p	5.0	5.0	2.0	1.0	ps
compressibility	4.5×10^{-5}	4.5×10^{-5}	4.5×10^{-5}	4.5×10^{-5}	
constraints	all-bonds				
constraints algorithm	LINCS				
LINCS order	4	4	4	4	
LINCS iteration	6	6	6	6	
LINCS warn angle	30	30	30	30	°

Table 7.2: Table containing the parameter used in the MD simulations. *Parameters are defined in the text below.*

dt refers to the time step used in the calculation; nstlist gives the number of timesteps after which the neighbour list will be updated; r_{list}, $r_{coulomb}$ and r_{vdW} give the cut-off radii for the truncation of the neighbouring list, calculation of coulombic interactions and van der Wals interactions respectively; Coulomb type refers to the treatment of long-range electrostatic interactions used; T and p coupling refer to the coupling of the temperature T and the pressure p with respective coupling constants τ_T and τ_p and the pressure coupling type, p coupling; ε_r refers to the relative dielectric constant; fourierspacing refers to the maximum grid spacing for the Fourier function transformation grid for PME, which uses PME order with a tolerance defined by Ewald r_{tol}.