DEPARTMENT OF PURE & APPLIED CHEMISTRY

University of Strathclyde Science

Modelling Fmoc-Dipeptide Nanostructures: The Synergistic Effect of Combining Computational and Experimental Methods

By

Iván Ramos Sasselli

Thesis presented to the Department of Pure and Applied Chemistry, University of Strathclyde, in fulfilment of the requirements for the degree of Doctor of Philosophy.

2015

Declaration

This thesis is the result of the author's original research. It has been composed by the author and has not been previously submitted for examination which has led to the award of a degree. The copyright of this thesis belongs to the author under the terms of the United Kingdom Copyright Acts as qualified by University of Strathclyde Regulation 3.50. Due acknowledgement must always be made of the use of any material contained in, or derived from, this thesis.

Signed: Date:

They did not know it was impossible so they did it

Mark Twain

Acknowledgements

First of all I would like to thank my supervisors, Rein and Tell, for giving me the chance to do this PhD. They always gave me their support when I needed it, and the flexibility and trust to be creative in my projects.

My thanks also to Peter J. Halling, for the wonderful idea which propitiated the work presented in Chapter 4, and to Tong Wang, for his awesome EM images (Chapter 6).

I also want to thank the former and current members of both groups, Ulijn Lab and TuttleLab, for their help and the great atmosphere during these years. I would like to particularly thank those I collaborated with to carry out amazing science and to those I will always call friends: Daniel A. Cannon (the *A* is for awesome), because he is always there to help; Charalampos (Babis) Pappas, the wee genius; Pim (Pimino) W. J. M. Frederix, because everything started with him; Greg M. Anderson (also known as Gregmander), for the nice switching off chats and laughs through the cluster; Maria Paola Conte (Mari P), for the company while touring Europe; Ines Maria Pimentel Moreira, she is crazy but nice; Enrico (Kiko) Angioni, the parties are not the same without him; Florimond (Flo) Cumine, for his dog laugh; and Joe Cameron, although sometimes he forgets I am Spanish and speaks to me in French.

I would also like to thank the SMARTNET and its members for the good scientific exchanges and, especially the ESRs and ERs for the nice time, every six months, for one week: Matija, Serhii, Tomasz, Vania, Jorge, Kars, Philip, Nishant, Marco, Laura and Ana Catarina.

I should not forget the other people who contributed to make my life in Glasgow awesome: Davide, Ana, Andreia, Nuno, and Laura, which, with those previously mentioned, are my Scottish Family; Sarunas, the nicest flatmate ever; Xavi; and of course all the people in the football email list, because there is nothing like kicking the ball (and some legs) to switch off from work.

I would also like to mention those, from my home university, who contributed to make me the chemist and the person I am today. Those lecturers who convinced me that there were good opportunities abroad, Jorge and Rosa, and those who thought I could aim for the best, Marini and Leo; and of course to my friends there, The Magic

Five, Jago, Liher, Chasco and Edu, because someone said that university friends are for ever, and I hope that is true.

My thanks also extend to my friends from Santander, because true friends are those who, no matter for how long you do not meet them, every time you are with them is like time did not pass: Guzman, Javier, Javi, Laura, Viktor, Cesar, Eva and Ines.

Keeping the best and most important for the last, I would like to thank my family, because sometimes, when you are not sure you can do it, you need people who know you can and convince you with their trust you can: my grandmother, Rosa; my aunt, Pili; my brother, Aitor; and specially, thanks to my parents, Chuspi and Sonia, who pull and pull and never let me down.

Finally I would like to thank Paula, my future wife, because she is, for lack of a better word, perfect, and makes me better.

Table of contents

Abstract

Nanomaterials based on aromatic peptide amphiphiles are interesting new materials with potential applications in the areas of biomedicine and nanotechnology. These natural based materials take advantage of the properties of the peptides, as it is the ability to form the final structures spontaneously without any external stimulus by self-assembly, or the high number of functionality available due to the 20 natural building blocks, amino acids, and their possible combinations. Although it is known that the functionality of these nanostructures is highly dependent on both, the chemical groups and the topology of the nanostructure, and both vary with the amino acids side chains, the relationship between nanostructure shape and peptide chain composition is still unknown. Understanding this is necessary to be able to design Fmoc-peptide nanostructures on demand.

In this thesis a combination between these experimental techniques and computational methods, molecular dynamic (MD) simulations, is used to elucidate the self-assembly motifs for a set of model systems composed of Fmoc-dipeptides. The interpretation of the experimental spectroscopic characterization is improved by using enzymatic self-assembly under thermodynamic control Fmoc-dipeptide and side-by-side comparison of nanostructures using dynamic peptide libraries (DPLs). This approach allowed to resolve which features increase the self-assembly tendency of these molecules.

Both MD and DPL approaches depend on the premise that gels can be at thermodynamic equilibrium, which is not clear in the literature. It has been argued that they represent metastable states, where crystals are suggested to represent the actual thermodynamically favoured structures. Hence, the study starts with a model proposed to demonstrate that nanofibrous gels can represent the thermodynamically favoured structure. This is achieved by using a packing model where self-assembling molecules are represented by prisms with faces of different nature, solvophilic and solvophobic to mimic the amphiphilicity of these molecules as a key feature. This approach gives rise to a combination of solvophobic and solvophilic interactions

where a level of solvent exposure is favourable. The model depends on parameters which can be related with features of the system and demonstrates that the amphiphilicity is key to allow 1D objects, fibres, to be more stable than 3D objects, crystals; and hence, that MD simulations and DPLs can be applied for their study.

For MD simulations, the CHARMM force field is used because it has been applied and validated to a wide variety of peptide-based systems. However, this force field does not include parameters for the Fmoc moiety. Therefore, the second steps for this study was to develop an Fmoc parameterization for the CHARMM force field, in order to be able to run all atoms self-assembling Fmoc-peptides simulations, to improve the understanding of these nanostructures and their formation. The parameterization is based in the CHARMM protocol adapted due to the amphiphilic nature of the Fmoc moiety.

Experimentally, in order to get more valuable information from the experimental characterization, the study of different Fmoc-dipeptides nanostructures with specific changes in their peptide chain are compared in order to understand how these specific changes affect the self-assembled structure: phenylalanine/leucine substitution to understand how the aromatic side chain affects; and amide/methyl ester C-terminus substitution, to understand the role of the possible extra hydrogen bonds of the amide group. Furthermore, DPLs are also applied to rationalize the influence of these changes in the self-assembling tendency of Fmoc-dipeptides.

Then, the experimental information is used to develop a model for Fmoc-TF-NH_2 fibre and simulate it. The analysis of the model in addition with correlation of these data with the experimental insights, allows the refinement of the model. The resulting new model is validated by comparing the simulation analysis with the previous model and, again, correlating the computational results with experimental. Finally, the new model is applied to gain understanding of the experimental observed phenomena of fibres evolving to twisted ribbons. The simulations using the developed fibre model demonstrate that those twisted ribbons are formed by lateral aggregation of the fibres. The useful information obtained using the model, supports its validity.

In conclusion, in this thesis the thermodynamic nature of gels is demonstrated to be able to use MD simulations and DPLs for the molecular level study of Fmoc-dipeptide nanostructures. A parameterization of the Fmoc is also developed to allow the implementation of the MD simulations for these systems. Then, standard characterization of Fmoc-dipeptide nanostructures is combined with DPLs to gain intermolecular

interaction information of these systems to then use this information for an iterative model development of a fibre model, which is correlated and validated with experimental observations. This demonstrates the synergistic effect of combining computational with experimental methods to gain understanding of supramolecular nanostructures at a level which is not accessible with any other technique.

Abbreviations

APA	aromatic peptide amphiphile
BE	binding energy
γ	interaction energy penalty
CD	circular dichroism
CHARMM	Chemistry at HARvard Macromolecular Mechanics
CG	coarse grained
Cryo-TEM	cryogenic transmission electron microscopy
Crys	crystal
CS	crystal state
DCL	dynamic combinatorial libraries
DFT	density functional theory
DPL	dynamic peptide libraries
EM	electron microscopy
F, Phe	phenylalanine
f_x	fraction of faces of type x
FCS	final charges set
FEP	free energy perturbation
Fmoc	fluorenyl-9-methoxycarbonyl
FT-IR	Fourier transformed infrared
G, Gly	glycine
GS	gel state
H-bonds	hydrogen bonds
HT	high tension
Hx	hexagonal based prism
kb	solvophobic faces buried
ks	solvophobic faces exposed to the solvent
L, Leu	leucine
lb	solvophilic faces buried
LMWG	low molecular weight gelator
ls	solvophilic faces exposed to the solvent
MD	molecular dynamics
MM	molecular mechanics
NAMD	Not Another Molecular Dynamics program

NDI	naphthalenediimide
NH_2	amidated C-terminus
OH	carboxylic acid C-terminus
OMe	methyl ester C-terminus
PA	peptide amphiphile
PBC	periodic boundary conditions
QM	quantum mechanics
RDF	radial distribution function
S, Ser	serine
solv	solvation
SOS	simple overlap sampling
SPPS	solid phase peptide synthesis
Sq	squared based prism
T, Thr	threonine
TEM	transmission electron microscopy
Tr	triangular based prism
vdW	van der Waals
VMD	visual molecular dynamics
Y, Tyr	tyrosine
WAXS	wide angle x-ray scattering

Publications

Directly related with this thesis:

Chapter 4:

- **Sasselli, I. R.**; Halling, P. J.; Ulijn, R. V.; Tuttle, T. Supramolecular Fibers in Gels Can Be at Thermodynamic Equilibrium: A Simple Packing Model Reveals Preferential Fibril Formation Upon Crystallization. *ACS Nano*, **2016**, 10, 2661-2668.

Chapter 5:

- **Sasselli, I. R.**; Ulijn, R. V.; Tuttle, T. CHARMM Force Field Protocol For Self-Assembling Peptide Amphiphiles: The Fmoc moiety. *Phys. Chem. Chem. Phys.*, **2016**, 18, 4659-4667.

Chapter 6:

- **Sasselli, I. R.**; Pappas, C.; Wang, T.; Tuttle, T.; Ulijn, R. V. Using Dynamic Peptide Libraries to Assess the Influence of the Aromatic Side Chain and C-terminus Substitution in Self-Assembling Fmoc-Dipeptides. *Langmuir* (Submitted Manuscript).

Chapter 7:

- **Sasselli, I. R.**; Ulijn, R. V.; Tuttle, T. Iterative Fiber Model Development: Fmoc-TF-NH_2 (Manuscript in Progress).

Not directly relevant for these thesis:

- Fleming, S.; Frederix, P. W. J. M.; **Sasselli, I. R.**; Hunt, N. T.; Ulijn, R. V.; Tuttle, T. Assessing the Utility of Infrared Spectroscopy as a Structural Diagnostic Tool for beta-Sheets in Self-Assembling Aromatic Peptide Amphiphiles. *Langmuir,* **2013**, *29*, 9510-9515.
- Pappas, C.; **Sasselli, I. R.**; Ulijn, R. V. Biocatalytic Pathway Selection in Transient Tripeptide Nanostructures. *Angew. Chem. Int. Ed.* **2015**, 54, 8119–8123.
- Moreira, I. M.; **Sasselli, I. R.**; Cannon, D. A.; Hughes, M.; Lamprou, D. A.; Tuttle, T.; Ulijn, R. V. Enzimatically Activated Emulsions Stabilised by Interfacial Nanofibre Networks. *Soft Matter*, **2016**, 12, 2623-2631.

Conferences

- ScotCHEM 2013 (St Andrews), Poster: Assessing the Utility of Infrared Spectroscopy as a Structural Diagnostic Tool for beta-Sheets in Self-Assembling Aromatic Peptide Amphiphiles.
- ScotCHEM 2014 (Edinburgh), Poster: Parameterization of the Fmoc Moiety for the CHARM Force Field.
- ERC 2014 (Berlin), Poster: Modelling the Self-Assembly of Fmoc-Dipeptides.
- ScotCHEM 2015 (Glasgow), Poster: Can Gels Be At Thermodynamic Equilibrium? A Simple Packing Model for LMWG.
- Nanopeptides 2015 (Glasgow), Poster: Modelling the Self-Assembly of Fmoc-Dipeptides.
- Active and Adaptive Materials Symposium 2015 (New York), Poster: Can Gels Be At Thermodynamic Equilibrium? A Simple Packing Model Reveals Preferential Fibril Formation Upon Crystallization.

1 Thesis Layout

The motivation of this research is to improve the understanding of Fmoc-dipeptide nanostructures with detailed intermolecular information by taking advantage of the synergistic effect of combining experimental and computational methods.

This thesis is split into ten chapters. The first chapter introduces the motivation and the layout of the thesis. The following two chapters comprise a literature review: Chapter 2 is an introduction to peptide based low molecular weight gelators (LMWGs); and Chapter 3 is focused on how these systems have been studied previously using experimental and computational methods.

Each of the next four chapters (Chapters 4 – 7) presents research results which involves different aspects of the investigations of these nanostructures.

Chapter 4 deals with a question that has been a longstanding disagreement in the literature and which has significant implications for the methodologies used in the following chapters of the thesis (dynamic peptide libraries and molecular dynamics): *Can Gels Be At Thermodynamic Equilibrium?* In this chapter, a simple packing model for LMWGs is introduced to demonstrate that it is possible for one-dimensional fibres (and by extension, gels) to represent the global thermodynamic equilibrium for certain amphiphilic molecules. This is a pre-requisite for the techniques used in Chapters 5, 6 and 7.

In Chapter 5, a parameterization of the Fmoc moiety for the CHARMM force field is introduced. This is an essential step in order to perform MD simulations of systems containing the Fmoc moiety because, although this force field is well developed and validated for peptides, it does not include parameters for this group. This chapter uses a new procedure for the development and validation of these parameters, which is proposed as an alternative to the standard CHARMM parameterization protocol for self-assembling peptide amphiphiles.

Chapter 6 includes a complete side-by-side characterization of four Fmoc-dipeptides that form nanostructures with a gel macroscopic appearance: Fmoc-TF-NH_2, Fmoc-TL-NH_2, Fmoc-TF-OMe and Fmoc-TL-OMe. This chapter aims to show

the differences that arise in nanostructures due to the changes in the amino acid sequence, specifically the inclusion of an aromatic side chain (F/L substitution) and the changes in the C-terminus (NH_2/OMe substitution). As well as the comparison between the characterization data of the four different molecules, dynamic combinatorial chemistry is used to obtain insights into the importance of the different forces that enhance self-assembly, focusing on the relative rate of formation between the Fmoc-dipeptides.

Chapter 7 presents the development of a detailed molecular model for the fibres formed by the system that was found to be most stable in the comparative experiments presented in Chapter 6, Fmoc-TF-NH_2. To do so, the experimental information covered in Chapter 6 is used to develop a preliminary model. This model is simulated using the Fmoc parameterization presented in Chapter 5. The stability of the simulations and the interactions between the molecules through it are then analysed. The information obtained is used to improve the understanding and interpretation of the experimental information (Chapter 6). The combination of the improved experimental interpretation and the interactions analysis of the first model are used to refine the model. The analysis of the new model simulation is then compared with the first model results and with the experimental information obtained in the previous chapter to validate the model. Finally, the model is applied to understand the evolution of the Fmoc-TF- NH_2 nanostructures from fibres to ribbons.

Chapter 8 highlights the research and conclusions contained in this thesis and finishes with an overall conclusion about the synergistic effect of combining computational and experimental methods for the study of LMWG nanostructures.

The last two chapters (Chapters 9 – 10) include the References and Appendices, respectively.

2 Peptide Based Nanostructures

Peptide based nanostructures have been of increasing interest through the last decades due to the broad variety of potential applications of these bioinspired materials.[1-14] These materials present several advantages over conventional polymers due to the use of peptides as building blocks. Peptides are naturally occurring molecules that are able to form nanostructures spontaneously, without any external stimulus. As their properties depend on the amino acids composition, they can be easily tuned by modifying their peptide sequence.[15-19]

The peptides that have been studied in this context are typically at least 10 amino acids long.[20] However, shorter peptides have been found to be able to self-assemble into nanostructures when they are functionalised with an aromatic group, due to the extra contribution of π-staking interactions to the self-assembly.[9, 19, 21-45] As such, a number of aromatic dipeptides have been found to form nanostructures and more generally short peptides have been converted to self-assembling molecules by conjugating them to aromatic moieties.[9, 19, 28-42] These molecules, known as aromatic peptide amphiphiles (APAs) represent a promising minimalistic approach for the construction of peptide based nanomaterials. An important type of APA is Fmoc (9-fluorenyl-methyloxycarbonyl) protected dipeptides.

There are now dozens of Fmoc-dipeptides, which are known to self-assemble into nanostructures which give rise to a gel-phase or highly viscous solution as their macroscopic appearance.[19, 30-38] The potential functionality of these materials is dictated by the chemical groups present but also by the topology formed. In addition, although it is known that the supramolecular structure of self-assembling Fmoc-dipeptides depends on the dipeptide sequence, little is currently known about how molecular structure relates to supramolecular properties.

Therefore, it is important to understand how aromatic groups and amino acid sequence tunes the supramolecular structure and hence functionality of Fmoc-dipeptide based materials in order to be able to design materials for specific purposes. However, the structural studies of Fmoc-dipeptides have some inherent difficulties. Although, structure analysis techniques are well developed for proteins and oligopeptides, they cannot be directly extrapolated to small peptides.[46] Furthermore,

as the macroscopic appearance of these materials is usually neither liquid nor solid, but contains elements of both (in supramolecular gels), it is not easy to resolve the intermolecular arrangements in the nanostructure. Furthermore, they are frequently highly scattering, which complicates quantitative analysis by spectroscopic methods. Different spectroscopic and microscopic techniques are able to provide some details of the supramolecular structure, but none are able to provide complete information that allows for a full structural characterisation.[19, 30, 46-47] As a result, there is some debate in the literature on the supramolecular organization of these structures.

2.1 Peptide Based Low Molecular Weight Gelators

Peptides are chains formed by amino acids linked *via* peptide bonds (Figure 2.1) which, as natural products, are readily available and biocompatible. This thesis will focus on short peptides, which are composed of less than 5 amino acids.

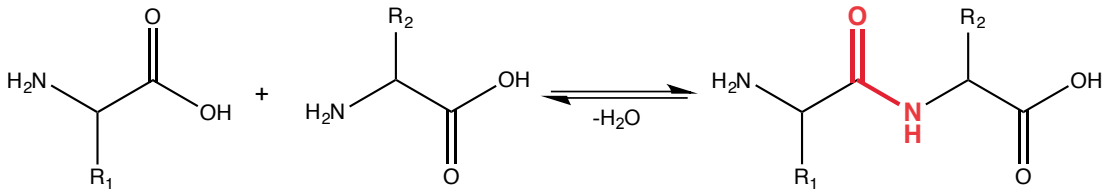


Figure 2.1 Peptide bond condensation scheme between two amino acids with side chains R_1 and R_2 to form a dipeptide. The peptide bond is highlighted in red. The reaction shown does not occur spontaneously and requires the use of catalysts and activating agents to overcome the unfavourable thermodynamics and kinetics of amide bond formation.

There are 20 natural amino acids encoded by the genome, which have a common structure and only differ in their side chain (R_i, Figure 2.1). The side chain provides the amino acids with chemical diversity. The possible sequences for a peptide composed of *n* amino acids are 20^n. This gives 400 possible dipeptides, 8,000 tripeptides and up to 160,000 for tetrapeptides. Peptides are named and presented from their amino terminal group or N-terminus to their carboxylic terminal group or C-terminus.[48-50]

All amino acids are chiral except for glycine (Figure 2.2) and present only the L-configuration in nature. However, artificial D-amino acids have been used to extend the possible topologies for peptide-based nanostructures due to the influence of the chiral centre configuration.[35, 51]

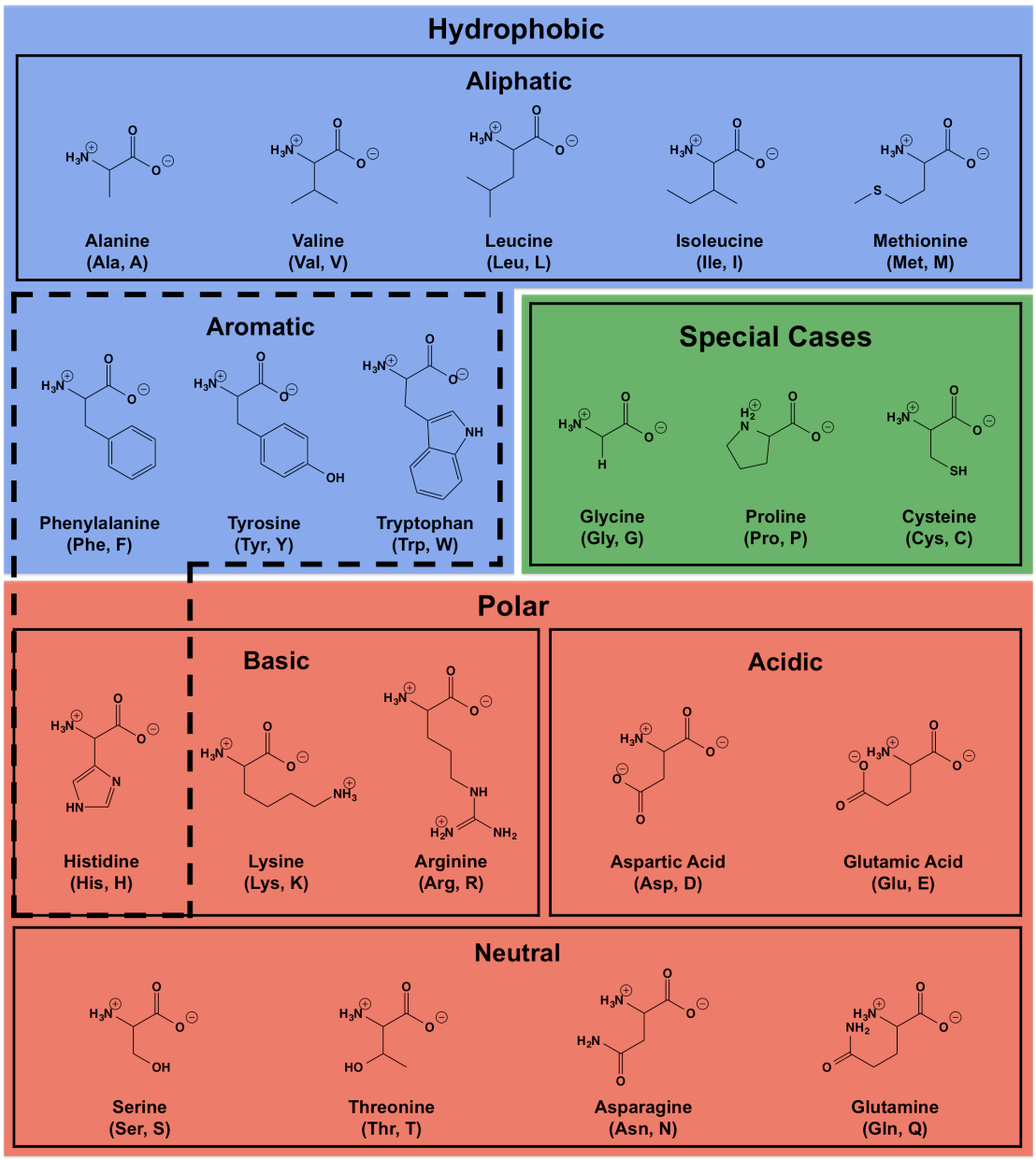


Figure 2.2 The 20 natural amino acids and their classification.

Amino acids can be classified as a function of the chemical nature of their side chain in different ways (Figure 2.2). According to their polarity they can be classed as hydrophobic or hydrophilic. Hydrophobic amino acids can be aromatic or aliphatic while hydrophilic ones can be neutral, acidic or basic, depending on their pH behaviour.

Amino acids are one of the most abundant types of biomolecules in living systems, normally forming proteins, which are macropolymers of amino acids. Proteins owe their functionality to their specific structure. Their structure is typically obtained by intramolecular self-assembly of a long peptide chain, so it is its sequence

which encodes the final structure and, hence, functionality. This idea inspires the use peptides to form nanostructures by taking advantage of intermolecular assembly of multiple shorter chains.[48-50, 52-53]

To be consider a nanostructure, the supramolecular structure formed by these molecules needs to have at least one dimension in the nanoscale (typically 1-100 nm).[54-55] Although in many cases over 90 % of the system is solvent, these structures can modify the bulk properties of the system, for example acting as a nanofibre scaffolding network which confers rigidity to the solution, *i.e.,* making the solution more viscous or gel-like.[55-61] Both the shape of the nanostructure and the entanglement of it affect the properties of the material.[36, 43, 62-63] If the scaffolding is entangled enough it can give rise to a gel. Gels, in which the solvent is water, as is typical for peptide-based nanostructures, are called hydrogels (those formed in organic solvents are called organogels). Due to the small size of these molecules (typically up to about 500 Da for a pentapeptide) and their ability to form supramolecular gels, they are known as low molecular weight gelators (LMWG).

However, it is clear that most amino acid sequences are unable to form nanostructures because self-assembly depends on intermolecular interactions and too short amino acids are usually not able to establish enough interactions.[64] However, it was recently shown that the presence of aromatic side chains allows tripeptides (LFF, KYF) and even dipeptides (FF, FW) to self-assemble into supramolecular structures, probably due to the extra π-stacking interactions.[22-23, 33, 35, 65-67]

Apart from pure peptide building blocks, different artificially modified peptide molecules have been used to create a new generation of LMWGs. There are several examples of peptides modified by the addition of an aliphatic group, called peptide amphiphiles (PAs), which are able to self-assemble into nanostructures.[11, 68-74] These molecules take advantage of a balance between hydrophilicity and hydrophobicity in combination with the peptide chain to form intermolecular interactions to self-assemble into nanostructures. As well as aliphatic chains, aromatic groups can be used to functionalise peptides in order to obtain self-assembling molecules as discussed in detail in the next section.

2.1.1 Aromatic Peptide Amphiphiles

The use of aromatic groups to functionalise peptides is of special interest as it allows especially short peptides to self-assemble due to the ability to tune the

hydrophilic/hydrophobic balance and the introduction of extra π-staking interactions. Different aromatic groups have been used for this purpose (with some examples shown in Figure 2.3).[25, 28, 39-45, 75-87] As an example of gelators made exclusively of biological components, even nucleobases have been used to functionalise peptides.[88-94] These functionalised molecules, as well as taking advantage of the aromatic stacking, also use base pairing to build new topologies.[93-94] The use of azobenzene (Figure 2.3f) introduces a photo-responsive geometry changing group, which allows control over the formation of the nanostructures,[83-85] as discussed in Section 2.2.2.

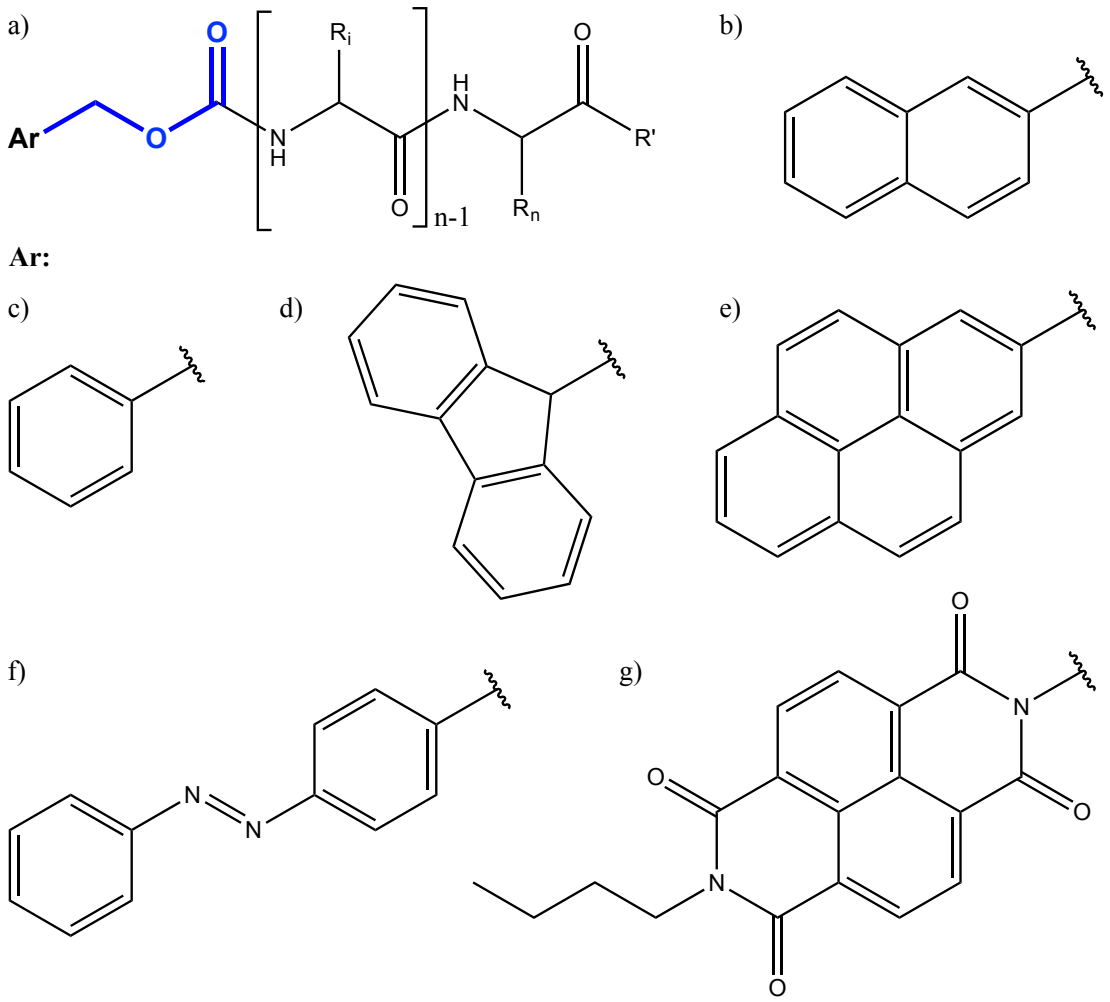


Figure 2.3 (a) General structure of an aromatic peptide amphiphile where *Ar* is the aromatic group, the linker is highlighted in blue (methyloxycarbonyl is used as example), R_i are the side chains of the peptide which has a length of *n* amino acids and *R'* is the C-terminus. (b) Naphthalene, (c) phenyl, (d) fluorenyl, (e) pyrene, (f) azobenzene and (g) NDI are some examples of aromatic groups used in APAs.

The use of some aromatic groups has led to the formation of photoactive supramolecular architectures, which take advantage of the optical properties of these groups.[83-85] Furthermore, the combination of electron donors and acceptors in the structures allows one to tune the self-assembly tendency as it modifies both the strength of the intermolecular interactions and the functionalization of the systems, to build nanowires.[39-40]

It has been found that not only the nature of the aromatic group but also the nature of the linker of the aromatic group with the peptide has a role in the formation of the nanostructure (Figure 2.3).[41] The length of the linker and the possible chemical groups, which affect both its hydrophobicity and the ability to form extra intermolecular interactions, are factors that directly affect the self-assembling behaviour of these molecules.

2.1.2 Fmoc-dipeptide Based Systems

Dipeptides capped at their N-terminus with the Fmoc moiety, known as Fmoc-dipeptides, are still the most common type of APAs. In these molecules fluorenyl is the aromatic moiety and it is attached to the N-terminus of the peptide with the methyloxycarbonyl linker (Figure 2.4). The Fmoc moiety is commonly used as an amino acid protecting group for the N-terminus, as it is used to attach the peptides to the resin in solid phase peptide synthesis (SPPS).[95] Because of this widespread use, Fmoc protected amino acids and some dipeptides are commercially available and relatively cheap. The production of any of the Fmoc-dipeptides involves minimal synthetic effort due to both the availability of Fmoc protected peptides, and, as the Fmoc does not need to be removed afterwards, the simplicity of SPPS to produce these molecules. Furthermore, the methyloxycarbonyl linker appears to be particularly well suited to induce self-assembly when compared with similar linkers.[41]

There are dozens of examples of Fmoc-dipeptides which are able to form nanostructures. These molecules only differ in the amino acid side chains (R_1 and R_2) and in the C-terminus nature (R', Figure 2.4), but these changes are enough to give rise to different types of supramolecular arrangements. Even small changes in the amino acid side chains can provoke big changes in the shape of the nanostructure. For example, such a small change as the threonine/serine (T/S) substitution changes the supramolecular shape from chiral fibres to two-dimensional sheets, which is likely due to the removal of a chiral centre in the side chain.[19]

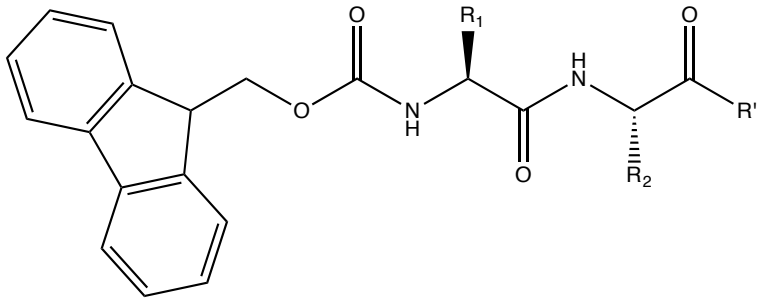


Figure 2.4 General structure for Fmoc-dipeptides. R_1 is the side chain of the first amino acid, R_2 is the side chain of the second amino acid and *R'* is the C-terminus.

There are many examples of Fmoc-dipeptides that are able to self-assemble into nanostructures (Table 2.1). Of the examples known, *ca*. 47 % of these are formed by at least one amino acid with an aromatic side chain (F or Y).

Table 2.1 List of Fmoc-dipeptides which are reported to form nanostructure.

Fmoc-Dipeptide	Hydrophilic Side Chain	Nanostructure Shape	Macroscopic appearance	Ref.
Fmoc-CF-OMe	1st –CH_2-SH	Fibres	Gel	96
Fmoc-SF-OMe	1st –CH_2-OH	Sheets	Viscous	19
Fmoc-SL-OMe	1st –CH_2-OH	Tapes	Gel	19
Fmoc-TF-OMe	1st –CH(-OH)-CH_3	Fibres	Gel	19
Fmoc-TL-OMe	1st –CH(-OH)-CH_3	Fibres	Gel	19
Fmoc-TF-OH	1st –CH(-OH)-CH_3	Fibre	Gel	31
Fmoc-TL-OH	1st –CH(-OH)-CH_3	Fibres	Viscous	31
Fmoc-AD-OH	2nd –CH_2-COOH	not specified	Gel	34
Fmoc-LD-OH	2nd –CH_2-COOH	not specified	Gel	34
Fmoc-ID-OH	2nd –CH_2-COOH	not specified	Gel	34
Fmoc-GS-OH	2nd –CH_2-OH	not specified	Gel	35
Fmoc-GT-OH	2nd –CH(-OH)-CH_3	not specified	Viscous	35
Fmoc-YN-OH	2nd –CH_2-C(O)-NH_2	Fibres	Gel	42

Fmoc-Dipeptide	Hydrophilic Side Chain	Nanostructure Shape	Macroscopic appearance	Ref.
Fmoc-YQ-OH	2^{nd} $-(CH_2)_2$-C(O)-NH_2	Spheres	Solution	42
Fmoc-YS-OH	2^{nd} $-CH_2$-OH	Fibres	Gel	42
Fmoc-YT-OH	2^{nd} –CH(-OH)-CH_3	Fibres	Viscous	42
Fmoc-AA-OH	–	Fibres	Gel	35
Fmoc-AG-OH	–	Fibres	Gel	33
Fmoc-FF-OH	–	Fibres	Gel	30
Fmoc-FG-OH	–	Fibres	Gel	33
Fmoc-FL-OH	–	Fibres	Gel	36
Fmoc-FY-OH	–	Fibres	Gel	36
Fmoc-GA-OH	–	not specified	Gel	35
Fmoc-GF-OH	–	Sheets	Viscous	33
Fmoc-GG-OH	–	Fibres	Gel*	35
Fmoc-LG-OH	–	Fibres	Gel	33
Fmoc-LL-OH	–	Tubes	Gel	31
Fmoc-VL-OH	–	Fibres	Gel	36
Fmoc-VV-OH	–	Fibres	Gel	31
Fmoc-YL-OH	–	Fibres	Gel	36

Clearly, introduction of the Fmoc moiety can render dipeptides able to self-assemble. In some cases, it can even be applied to single amino acids to promote their self-assembly into nanostructures. This has been observed for Fmoc-F and Fmoc-Y that both form unidirectional fibres.[28, 97-99] Other Fmoc protected single amino acids form spheres.[100-101]

Beyond assembly of pure Fmoc-amino acids and peptides, there are also examples of co-assembly of single protected amino acids with Fmoc-dipeptides.[59, 100] In these cases, the Fmoc-amino acid does not need to be able to form nanostructures by itself, but it interacts and modifies, collaborating or disrupting, the Fmoc-dipeptide

nanostructure. These kinds of co-assemblies are useful to tune the properties of the materials using different proportions of the different building blocks.

2.2 Self-Assembly

Self-assembly is a bottom-up approach whereby many molecules spontaneously organise themselves to form nanostructures. Bottom-up approaches refer to processes where simple building blocks are used to build larger structures. By definition, self-assembly processes are spontaneous and do not require an external stimulus. This simplifies the formation of nanostructures relative to the traditional top-down methods which consist in carving away larger materials to give them nano-scale shapes.[54]

As mentioned, the use of peptides as self-assembling building blocks is inspired by nature. Proteins can adopt highly specific structures, typically from one or a few complex peptide chains of typically 100 or more amino acids, with, generally, no external input.[48-50, 102-103] Self-assembly is a spontaneous process where a large number of simpler molecules rearrange to maximize the energetic benefit derived from their intermolecular interactions. During self-assembly, building blocks may form specific well-ordered structures due to repulsion, or lack of interactions with the environment, and interactions with other building blocks. Thermodynamically, solute molecules join together because it is energetically favoured over their interaction with the solvent. Self-assembly is, therefore, an enthalpic driven process, but, as it involves changes in the order of the system, there is also an important entropic contribution.[104-105]

Apart from peptide amphiphiles (PAs) and aromatic peptide amphiphiles (APAs), there are other LMWGs that take advantage of amino acids to form new building blocks for supramolecular nanomaterials.[106-108] These molecules may only contain a small amount of peptide and take advantage of the amphiphilicity added by the synthetic moiety to self-assemble as well as of the ability of dipeptides to form supramolecular arrangements.

Peptides are not the only natural molecules that have inspired the construction of nanostructures. For example, since Ned Seeman's pioneering work in the mid-80s, it is now common to use DNA bases to build nanomaterials.[109-112] DNA double helix structures are formed spontaneously between complementary nucleobases: adenine – thymine and guanine – cytosine. The specificity of these interactions has been

exploited in materials that, as in the case of peptides, use both pure DNA systems and chemically modified nucleobases.[113-117]

A third class of natural molecules, saccharides, have also been employed to design LMWGs.[118-122] As in the cases of peptides and nucleobases, saccharides present chemical groups that give them the ability to form hydrogen bonded networks, which, in combination with the inherent chirality of monosaccharides, make them interesting building blocks for nanomaterials. There are also examples of saccharides which are functionalized with aromatic groups to promote the self-assembly by extra π-stacking interactions.[123-127]

Although natural based self-assembling materials represent an important part of the field, there are also synthetic molecules that are known to self-assemble into nanostructures.[128-134] These molecules very often present chemical groups that may have some similarity to natural building blocks, such as amide bonds.[128-132] This clearly demonstrates the importance of the natural material inspiration even for these synthetic building blocks.

The use of intermolecular interactions to build gels goes beyond the LMWGs. Oligopeptide are commonly used as self-assembling materials. It is common to use specific amino acid sequences that are able to form secondary structure motifs such as α-helix, β-hairpins, collagen and elastin.[135-140] Cyclic peptides have also been exploited to build nanomaterials.[2, 141-142]

Polymers have been exploited to build gel-phase materials.[143-149] These materials can also include chemical groups, such as, amides, esters or aromatic groups, which are able to form intermolecular interactions to form cross-linked materials in solvents. In addition, very specific charge-transfer interactions have been exploited to develop polymer-based hydrogels.[150-152] It is the crosslinking of these macromolecules in solution that forms the scaffolding to confer rigidity and solid like properties to the solution. This is similar to what happens in proteins, where different parts of the peptide chain interact to form a final well-defined structure.

2.2.1 Supramolecular Interactions

Supramolecular interactions are essential for self-assembly to occur.[49-50, 52-53, 55] These interactions are fundamentally the same as those seen in the folding of proteins to form their structures. In the case of the minimalistic nanostructures discussed here, these interactions usually occur between groups in different molecules or building

blocks, rather than, as in the case of proteins, between distant parts of the same macromolecule.

The hydrophobic effect plays an important role in self-assembly. Hydrophobic parts of the molecules prefer not to interact with water because water molecules prefer to interact with other solvent molecules. Although it causes effects similar to a repulsive interaction, the hydrophobic effect is not due to any repulsion but rather to a less favourable interaction.[49-50, 52, 55]

This effect drives the aggregation of molecules as the first step of self-assembly. This process was thought to involve a decrease in entropy as it converts randomly arranged molecules in solution with solvent and solute mixed in a system with two differentiated regions. However, it is currently known that this is not totally correct because solvent molecules form well-ordered arrangements to minimize the contact with the hydrophobic parts of amphiphiles. This order is not present in the interaction of water molecules with the hydrophilic parts, and hence, when the solute molecules aggregate to hide their hydrophobic parts this order disappears, increasing the entropy of the system.

The hydrophobic effect plays an important role also in keeping the hydrophobic parts of the amphiphiles buried in the structures formed, but, as molecules get closer due to the aggregation, other intermolecular, or non-covalent, interactions can contribute to form well-ordered structure. The physical properties of the different chemical groups determine the type of interactions they can establish, which will determine their position, directionality and strength.

Electrostatic interactions are the strongest non-covalent interactions. They occur between charges following Coulombs law. The strength of these interactions is proportional to the charges, and the type of these charges determines their nature: negative or attractive for charges of different sign, and positive or repulsive for charges of the same sign (Equation 2.1). These interactions, as they decrease with the first power of the distance ($1/r$), are long-range but stronger between molecules in close proximity.

$$E_{Coulomb(AB)} = \frac{kq_Aq_B}{r_{AB}} \qquad \text{Equation 2.1}$$

Equation 2.1 shows the mathematical expression of the Coulomb potential energy between two particles A and B of charges q_A and q_B, respectively, where r_{AB} is the distance between the charges and k is the Coulomb's law constant.

Electrostatic interactions are not restricted to molecules which present net charges but also partial charges, which arise from the asymmetrical distribution of charges through a polar bond. Therefore, as well as charge – charge interactions, electrostatic interactions can also be dipole – dipole, which, unlike the former, are short range as they decrease with the third power of the distance ($1/r^3$).

Amino acids show net charges in physiological conditions (pH=7-8) due to the pK_as of their chemical groups: pK_a (acid group) = 1.5 – 2.5 and pK_a (amine) = 9 – 11 (although typically these will not be free as they may be involved in the formation of amide bonds). Due to the free N and C termini, peptides in solution are zwitterions (neutral molecules with separation of charges, Figure 2.5 middle) and, hence, they are expected to form strong electrostatic interactions to drive peptide self-assembly. In Fmoc-dipeptides, the amine group is capped and the acid group is sometimes substituted for a methyl ester or for an amide group. When the C-terminus presents the acid group, it is deprotonated at any pH over 2-4. However, during self-assembly the proximity of molecules with the same charge varies the tendency of this group to be protonated, making it to be protonated upon self-assembly at pH values at which it is usually deprotonated. These pK_a values that seem to change depending on their environment are known as apparent pK_a.[153]

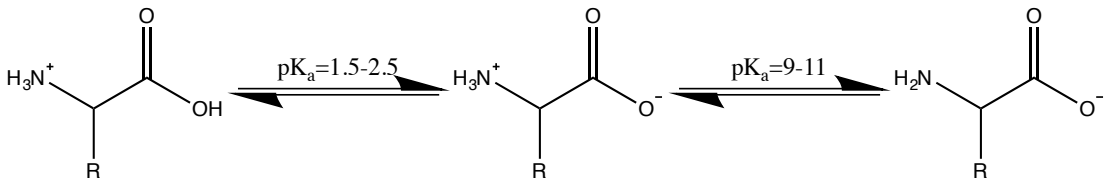


Figure 2.5 Acid/base equilibriums of the acid and amino group of a general amino acid with side chain R.

Hydrogen bonds are mainly electrostatic interactions with some covalent character that appear between an electronegative atom and a hydrogen covalently bound to a electronegative atom (usually F, O or N).[154-155] The hydrogen bond has a strong positive partial charge because its electron is located closer to the electronegative atom avoiding shell repulsion in the areas opposite to it. This partial positive charge allows the second electronegative atom to get closer to the positive charge of the hydrogen core, giving highly linear and strong Coulomb interactions.

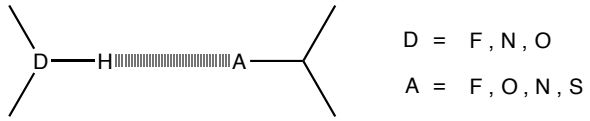


Figure 2.6 Hydrogen bond scheme between a hydrogen bound to donor *D* and acceptor *A*. On the right hand side the atoms that are common donors (*D*) or acceptors (*A*).

The electronegative atoms involved in the hydrogen bonds are known as *hydrogen bond donor (D)*, which is the atom bound to the hydrogen, and *hydrogen bond acceptor (A*; Figure 2.6). These interactions are highly directional, which is illustrated by the fact that a deviation of only 20º in the $D - H \cdots A$ angle from the preferred 180º (Figure 2.7a) would involve a weakening of the interactions by a 10%. Moreover, due to the Coulombic nature, H-bonding strength is also highly dependent on the distance. A typical distance for an $N - H \cdots O = C$, which is a typical hydrogen bond in peptide based systems, is 1.8 Å (Figure 2.7b). This directionality and the distance limitations play an important role in defining the secondary structure of proteins and in other oligopeptides. Therefore, they also have a role to make short peptides to form well-ordered supramolecular arrangements.

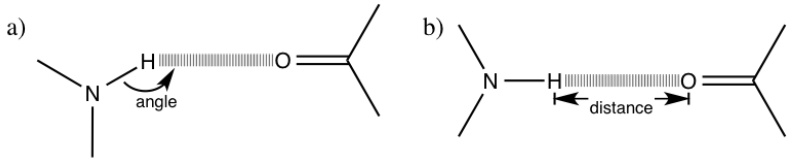


Figure 2.7 Schematic illustration of a hydrogen bond showing (a) angle and (b) distance for an $N - H \cdots O = C$ hydrogen bond.

π-stacking interactions are dispersion interactions between aromatic groups due to the presence of π-conjugated electrons. These interactions are very short range as they decrease with the sixth power of the distance ($1/r^6$). π-stacking interactions can present different conformations (Figure 2.8).[123]

π-stacking interactions play an important role in the self-assembly of short peptides as is evidenced by the high number of F and Y containing LMWG. Also in APAs the role of these interactions is key to promote the formation of nanostructures. Aromatic groups have actually a double effect in self-assembly, as well as the π-stacking interactions, their hydrophobic nature promotes aggregation of these groups through the hydrophobic effect explained above. The fact that a peptide as small as

FF, or even just the amino acid F, is able to self-assemble into nanostructures demonstrates the importance of the role of the aromatic groups.[22, 156]

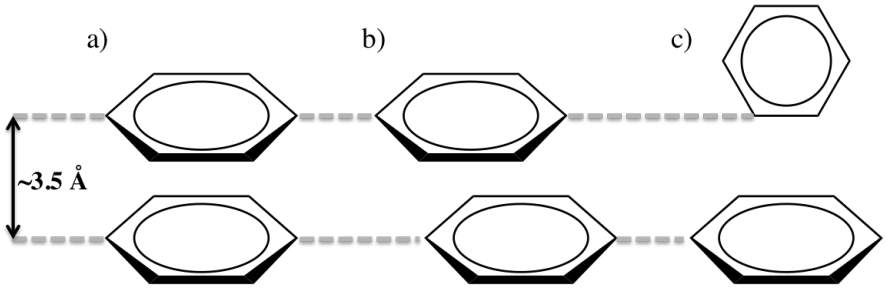


Figure 2.8 π-stacked aromatic rings in (a) parallel, (b) displaced parallel and (c) T-shaped conformation.

In conclusion, electrostatic interactions between net charges are the strongest and longest range intermolecular interactions. However, although the presence of charges is clear for peptides in solution, not only due to the terminal amine and acid groups, but also some side chains (lysine, aspartic acid, etc.), it is not as unambiguous for self-assembled systems because of the acidity/basicity changes of these groups provoked by the proximity of charges of the same sign.[153] Therefore, other interactions have to be considered in the self-assembly process.

Hydrogen bonds are the main interactions inducing supramolecular order due to their strength (4 – 120 KJ/mol) and directionality. π-stacking interactions are usually weaker (0 – 50 KJ/mol) but the presence of large aromatic groups, like fluorene, pyrene or naphthalene, in combination with the formation of extended π-stacked structures, enhances the strength of this type of interactions, making them of similar relevance to hydrogen bonds.[30] These extended stacks are also favoured by the hydrophobic effect in the formation of the nanostructure. Therefore, some structures can show some preference for π-stacking interactions, more so than for hydrogen bonds.

Dipole – dipole interactions are of less importance than the other interactions mentioned (5 – 50 KJ/mol). However, they still have a relevant influence to form well-ordered arrangements of the polar parts of the molecules. Moreover, if the molecules involved in the self-assembly present net charges, electrostatic interactions become more important due to the strength of the ion – ion (100 – 300 KJ/mol) and ion – dipole interactions (50 – 200 KJ/mol). The relative number of each type of interaction will determine their role in the self-assembly process.

2.2.2 On-Demand Self-Assembly

The properties of the self-assembled materials depend on both the building blocks used and the route of self-assembly. This happens because not only the shape of the nanostructure determine the mechanical properties of the gel, but also the topology and interactions between the nanostructures.

Once the building blocks are in solution, self-assembly starts and neither the nucleation nor the growth can be easily controlled. Thus, rather than thermodynamic structures, many self-assembly systems in reality represent kinetically trapped aggregates – nanostructures full of defects where the route dictates the degree of order of the nanostructure. In some cases this has been turned into an opportunity to control a material's properties by kinetic routes. Although it is true that the kinetic control of self-assembly has been used to form different nanostructures from the same building blocks,[36] this still needs to happen in a controlled way in order to obtain reproducible materials.

Therefore, due to the importance of the formation route, a significant effort has been made to control the formation of nanostructures. The control of self-assembly is usually carried out by coupling a step before the self-assembly. This step involves the creation of self-assembling molecules from non-active precursors (Figure 2.9). The control of the kinetics on the previous step allows a tuning the kinetic/thermodynamic control of the formation of the nanostructures.

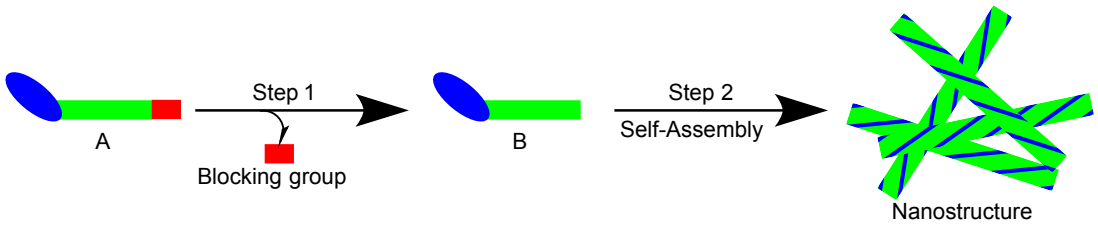


Figure 2.9 Self-assembly (Step 2) coupled to a previous step (Step 1). In Step 1 the blocking group is removed to convert the precursor A in the self-assembling molecule B.

The methods to control the self-assembly can be classified as a function of the nature of the pre-step. Self-assembly can be controlled by changing the temperature, the pH, the ionic strength, the use of light, or enzymes.

The control of self-assembly by *changes in the temperature* is relatively simple. It does not involve the transformation of precursors in self-assembling molecules but the formation and rupture of intermolecular interactions as a function

of the temperature. At high temperature, the intermolecular interactions break and the molecules do not form supramolecular arrangements. When the temperature decreases, the intermolecular interactions form again and the molecules can self-assemble. As different interactions break at different temperatures based on their strength (*e.g.*, π-stacking interactions break at lower temperatures than hydrogen bonds), this method can be used to favour some interactions, leading the system to different polymorphs.[157] Temperature can also be used to overcome kinetic barriers and transform systems trapped in a metastable state into the thermodynamically favoured product.[132, 158-159] There are systems which show the opposite behaviour, self-assembling upon the increase of temperature.[160]

Using *pH changes* is a common and easy way to control self-assembly. It typically affects the whole solution homogeneously. The pH value determines the protonated/deprotonated state of pH sensitive chemical groups, which determines if a molecule is a self-assembling or a non-self-assembling molecule due to the presence of charges. [22, 26, 30, 33, 66, 161] Fmoc-dipeptides with acid C-terminal groups self-assemble upon decreasing the pH below the pK_a (or an apparent pK_a) due to the conversion of negatively charged precursors in neutral self-assembling building blocks (Figure 2.10).

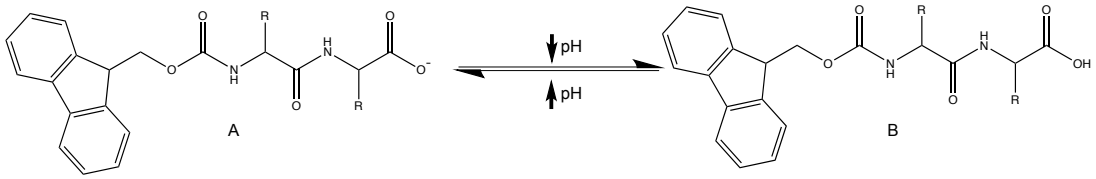


Figure 2.10 pH equilibrium for an Fmoc-dipeptide, which is (A) deprotonated and not able to self-assemble at high pH due to the negative charges repulsion, and (B) protonated and able to self-assemble at low pH.

Peptides which contain amino acids with pH sensitive side chains can also be used to control self-assembly using the pronated/deprotonated transition of these groups.[162-163] This is not only due the repulsion of charges of the same sign but it can be due to the attraction of side chains with complementary charges at certain pH.

Variations on the ionic strength of the system can also be used to control the self-assembly as it masks the effect of charges, decreasing the influence of pH and modifying the strength of the hydrophobic effect.[164-167] The ionic strength variation using salts affects the nanostructures formation following the Hofmeister trends.[167-168]

Light can also be used to convert precursors in self-assembling molecules. This is done by using precursors with a photoactive blocking group, which does not allow the molecules to self-assemble, and which can be cleaved upon the irradiation with light (Figure 2.11).[169-170] There are also systems that use moieties, such as azobenzene (Figure 2.3f), which cannot properly pack to self-assemble, but this steric hindrance depends on the isomeric state of a photoactive group. In these systems, the irradiation promotes the isomerization to a geometry that is able to self-assemble.[83-85] This method, as well as time control, like the others, enables spatial control, as self-assembly starts where the light is irradiated.

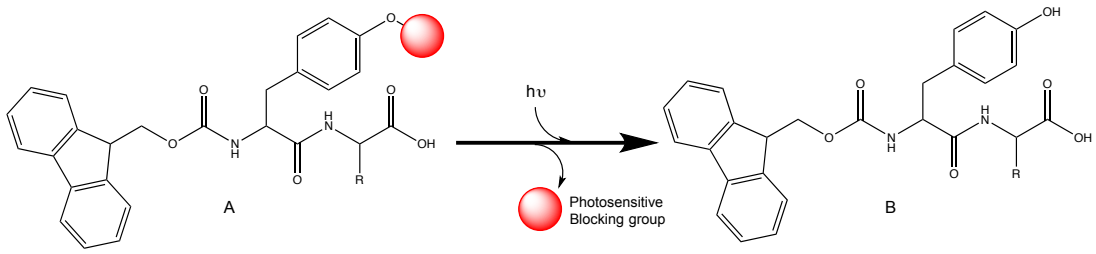


Figure 2.11 Synthesis reaction of a self-assembling molecule (B) from a non-self-assembling precursor (A) by removing a blocking group (red) with light (hυ).

Enzymatically controlled self-assembly has been very popular over the last decade. Enzymes allow the control of the nanostructure formation under constant temperature and pH, and hence, allow the control of the self-assembly under physiological conditions. This has opened a wide range of potential applications in biological systems.[36, 171-172]

The control of the assembly by enzymes can be done either by catalysing the formation of a self-assembling molecule from non-assembling precursors (Figure 2.12a) or by removing a blocking group which avoids the self-assembly of a molecule (Figure 2.12b and c). In the first approach the enzyme catalyses the formation of the bond between both precursors. In the case of Fmoc-dipeptides this is usually the reaction of an Fmoc protected amino acid with another amino acid, which is usually protected at its C-terminus with either an ester or an amide group to avoid side reactions.[19, 31, 172-173] In the second approach, the enzyme catalyses the cleavage of the bond between the self-assembling molecule and the blocking group.[31, 36, 98, 174]

These two approaches are different due to the thermodynamics of each reaction. The formation of the amide bond is slightly endothermic (4 KJ/mol)[172] and

hence, the precursors state is more stable. The reaction is driven forward due to the stabilization of the products upon self-assembly (Figure 2.13). The coupling of the two processes makes the system reversible and ensures the thermodynamic control in the formation of the nanostructure.

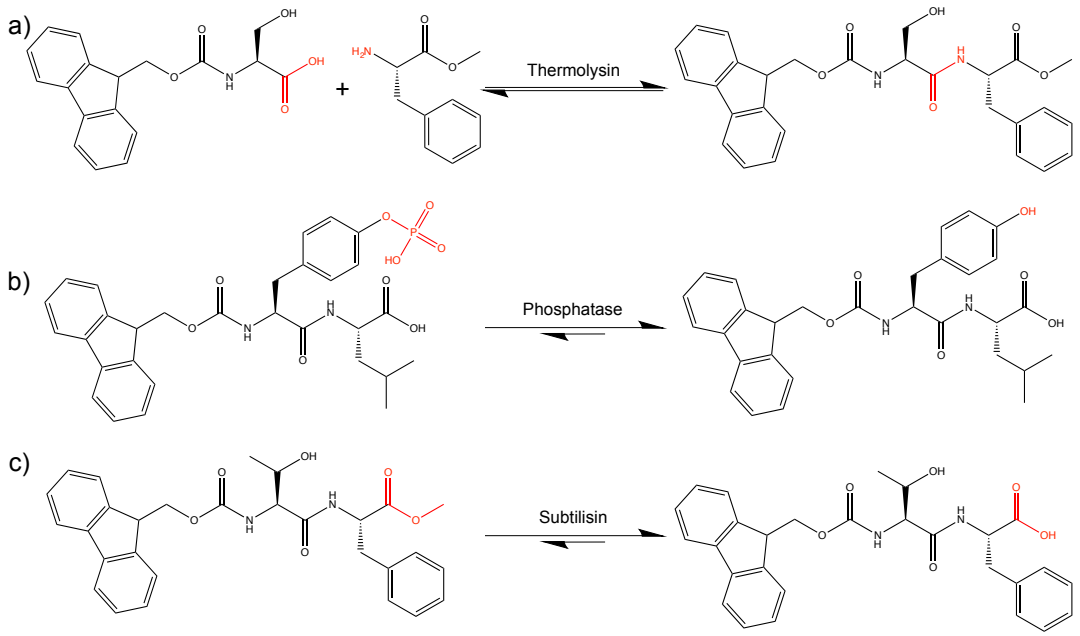


Figure 2.12 Enzymatically catalysed reaction of the synthesis of self-assembling molecules by: (a) the formation of an amide bond; (b) removing the phosphate group; and (c) hydrolysing the methyl ester C-terminus. The chemical groups involved in the reactions are highlighted in red.

However, the reaction in the second approach is usually exothermic, as in the case of the release of the phosphate group (-21 KJ/mol) (Figure 2.12b)[174] or the hydrolysis of a methyl ester C-terminus (-33 KJ/mol) (Figure 2.12c).[31, 36, 175] These reactions are not reversible and the control of the self-assembly is carried out by controlling the rate of production of the self-assembling molecule.

Although enzymes are known to be highly specific, they actually present different levels of specificity and some are able to catalyse reactions involving molecules different to their natural substrate, as long as they are similar, which is also one of the advantages of working with natural products. However, some specific changes in these molecules can make the enzymes not able to catalyse their reactions, as is the case for D-amino acids.

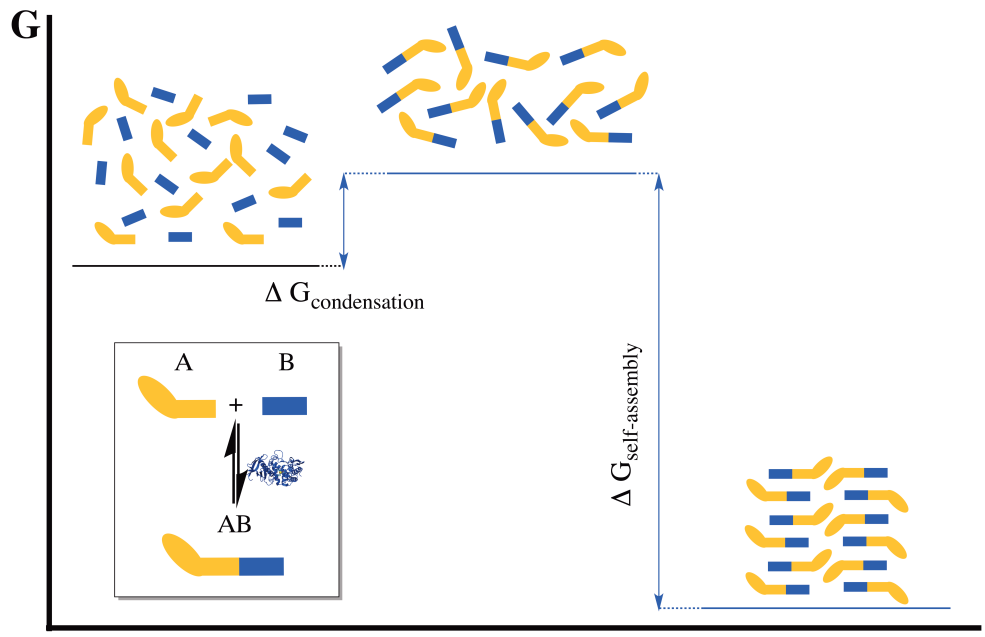


Figure 2.13 Free energy diagram for enzymatic driven self-assembly through amide bond condensation. The inset shows the condensation reaction of the precursor A (orange) with B (blue) to form the self-assembling building blocks AB in the presence of an enzyme.

Enzymes present several advantages as self-assembly controlling tools:

- The *specificity* of the enzymes ensures that they catalyse the formation of self-assembling molecules even in a mixture of components, minimizing the probabilities of having side reactions.
- Self-assembly is *time controlled*, as it starts upon the addition of the enzyme. This type of experimental control is also possible with other control methods but is not as well defined in the case of pH and temperature.
- *Spatial control* of self-assembly. As self-assembling molecules are formed at the enzymes, a high concentration of these molecules meet close to the enzymes, driving the formation of the nanostructures in these areas. This feature has been used to grow nanostructures on surfaces by simply fixing the enzymes to those surfaces.[172]
- The *variation of the enzyme concentration* can be used to control different features of the nanostructure formation. The higher the concentration of the enzyme, the faster the formation of self-assembling molecules is, which speeds up the nanostructures formation. Also, the more disperse the enzyme are in the media, the higher number of active points and the more entangled the nanostructures are. Although it could be expected that at higher concentrations the

enzymes will be more disperse, there are some enzymes that aggregate in clusters, the size of which depends on the enzyme concentration. Therefore, the entanglement is not necessarily proportional to the enzyme concentration.[176]

- The *reversibility* of some enzymes is interesting in terms of being able to drive the self-assembly towards the formation of the thermodynamically most stable structure. In addition with the spatial control, this allows the formation of highly ordered structures.[19, 36, 42, 172, 177]

3 Understanding Peptide Nanostructures

As mentioned in Chapter 2, peptide based nanostructures are of special interest due to their ability to self-assemble into well-ordered structures. Short peptides, like proteins, form these structures through intermolecular interactions. The intermolecular arrangements formed are highly dependent on the structure of the building blocks, and the final structure influences the functionality of the nanomaterial as much as the functional groups present on them. Therefore, short peptide based nanomaterials are highly tuneable. Furthermore, nanostructures based on peptides conjugated with non-peptidic components, such as PAs and APAs, increase the range of structures which can be formed further and, hence, the scope of possible materials that may be formed.

Examples are known of PAs which form supramolecular arrangements with a rigid structure containing groups which are known to be exposed to the solvent and that can be used to functionalize the material. These are usually not minimalistic approaches and require large building blocks (≥ 10 amino acids) so the changes made to functionalize the units represent a low percentage change. This is the case of Stupp's PA which is typically composed of an 11-amino acids long peptide, of which 8 play a structural and 3 a functional role, and an aliphatic chain composed of 15 carbons[11, 178] This PA presents a robust structural part which is able to keep the same fibre structure while the hydrophilic terminus can be functionalized with different groups. Alternatively, an 11-amino acid long peptide in which the functional part is formed by the last 4 amino acids (~20% of the total length), is attached to a 12-carbon long aliphatic chain, contains 4 cysteine residues to be able to enhance the rigidity of the supramolecular arrangement with disulphide bonds. However, PAs also suffer significant structural changes when a short peptide is employed giving rise to more diverse nanostructures with small peptide sequence changes.[71, 179]

Despite these design approaches, the relationship between the primary structure of the building blocks and the supramolecular arrangements formed for APAs and PAs based on short peptides is far from clear. Therefore, the design of peptide nanomaterials from first principles is still challenging and new materials are discovered by serendipity and trial-and-error rather than *de novo* design.[180] The fact

that most of these materials show changes in the supramolecular structure when small changes in the primary structure are introduced, presents opportunities as well as challenges. As the changes in the supramolecular structure cannot be readily predicted or designed, it is difficult for systems to be functionalized for given purposes.

Therefore, in order to be able to design new peptide based materials for specific applications, it is important to understand how the changes in the peptide sequence influence intermolecular arrangements and hence the final nanostructure formed.

3.1 Experimental Structural Studies

In order to understand the interactions and the supramolecular arrangements formed through these interactions different experimental techniques have been used. Most of these techniques, like Fourier transformed infrared (FT-IR) and Circular dichroism (CD), are known to be sensitive to the supramolecular arrangements of amino acids in proteins.[181-185] Therefore, as APAs and PAs supramolecular structures are based on similar interactions as proteins, these techniques can also be expected to be sensitive to the changes in the supramolecular arrangements of their short peptides. However, the interpretation of these techniques for proteins cannot be directly extrapolated for short peptides.[46]

The inclusion of synthetic moieties, like the Fmoc, adds some extra opportunities for the characterization. For example, upon self-assembly changes in the Fmoc orientation and environment may be observed through fluorescence. The inherent chirality of the amino acids typically gives rise to chiral structures, with chiral organization of the fluorenyl groups conveniently observed by changes in its ellipticity through CD, although both fluorescence and CD suffer from difficulties for the interpretation. The inclusion of a carbamate group in the Fmoc-dipeptides is not present in the backbone of proteins and gives rise to specific bands in FTIR.[41, 46]

Further difficulties are found due to the gel nature of may of these materials, which complicates the implementation of a number of experimental techniques which are more suited to solutions (spectroscopy) or solids (XRD). Indeed, some solid state techniques cannot be applied because they require denser systems. This discards the use of, *e.g.,* wide-angle x-ray scattering (WAXS) directly on the gel state. Although these techniques have been applied on dried gels, it is not clear how comparable they are with the actual gel (≥ 90% of water).

Although these systems are mostly water, some of the spectroscopic techniques typically used in solution are also problematic in their application to gels. This is due to the high scattering of the nanostructures formed, especially when the size of the features formed approximate the wavelength used which, in most of the cases, saturates the absorption signal. This makes it impossible to analyse the changes in the absorption spectra upon the formation of nanostructures.

However, despite the difficulties, with careful application FT-IR, fluorescence and CD have been demonstrated to provide insights into the structural features of these materials.

3.1.1 Fourier Transformed Infrared (FT-IR)

Infrared spectroscopy provides information about the vibrations of bonds within molecules. The infrared signal is proportional to the variation in the polarity in a specific vibration. Consequently, only those vibrations which involve a change in polarity are active in IR and, hence, produce an absorption peak in the spectrum. These vibrations are commonly affected by intermolecular interactions and these variations are in turn reflected in the FT-IR spectrum.

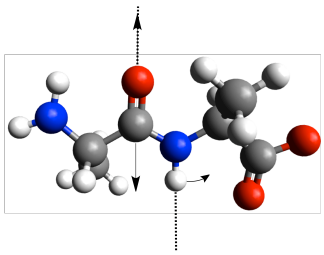


Figure 3.1 Amide I vibrational mode.

The Amide I mode results from the vibration of amide bonds and is known to be highly dependent on the secondary structure of proteins (1600 – 1700 cm^{-1}) and not of the amino acid side chains, which makes it a good diagnostic method for secondary structure.[183-189] It results from movement of the whole amide bond (peptide bond) and hence it is highly affected by the hydrogen bonds of C=O and NH groups which control the secondary structure of proteins. The Amide I vibration mode (Figure 3.1) is principally a C = O stretching vibration with a smaller contribution from the C – C – N deformation, out of phase C – N stretching and the N – H in-plane bending.[183]

This mode presents a single peak for parallel β-sheets but two peaks for antiparallel β-sheets. This is due to the coupling of the dipolar moments in the two dimensions of the antiparallel β-sheet (Figure 3.2): the hydrogen bonding direction ($\nu_{\perp}$) and the peptide chain direction ($\nu_{\parallel}$). In parallel β-sheets the symmetry makes different variations in the dipoles generated by this vibration through the peptide chain to cancel each other making the total coupling negligible. Experimentally the antiparallel peak ($\nu_{\perp}$) appears around 1615 – 1640 cm^{-1} and the parallel ($\nu_{\parallel}$) around 1685 cm^{-1}. However, in the case of short peptides the coupling through the peptide chain becomes negligible. Even for peptides consisting of 7 to 29 amino acids the parallel peak is much weaker than that of the antiparallel.[190-193]

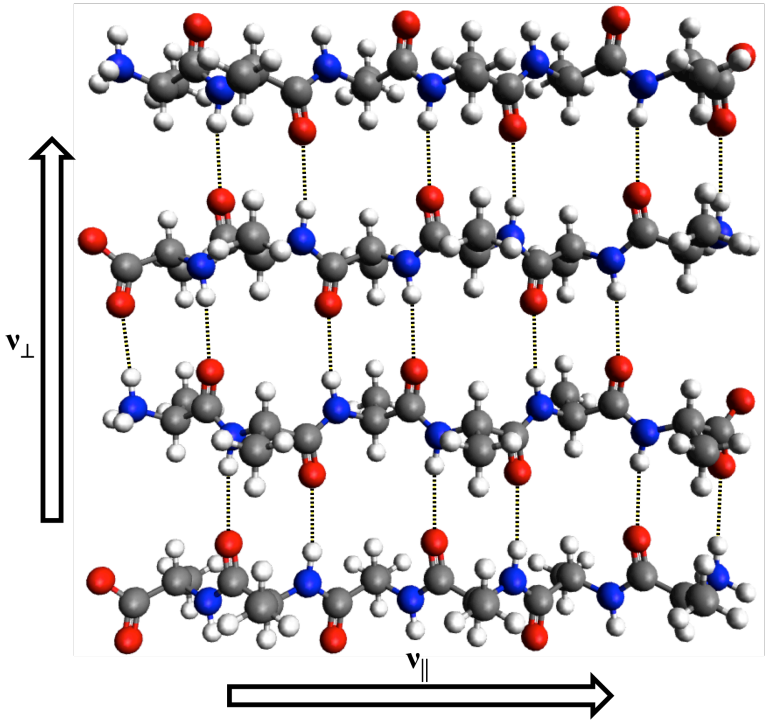


Figure 3.2 Antiparallel β-sheet with the parallel ($\nu_{\parallel}$) and antiparallel ($\nu_{\perp}$) contribution to the Amide I mode coupling.

Therefore, for aromatic peptide amphiphiles, and specifically Fmoc-dipeptides, the dipolar moment of the Amide I vibration is only significantly coupled through the hydrogen bonding direction ($\nu_{\perp}$), because the peptide chain is too short to show relevant coupling in its direction ($\nu_{\parallel}$). However, they usually show more than one peak in the Amide I region (1600 – 1700 cm^{-1}) due to the presence of carbonyl groups of different nature: as well as the normal amide which links both amino acids

to form a dipeptide, there is a carbamate group, which links the dipeptide to the aromatic group. In free Fmoc-dipeptide the amide group vibration appears at 1650 – 1680 cm^{-1} and the carbamate at 1685 – 1730 cm^{-1}.[46, 194] However, when the Fmoc-dipeptides are forming nanostructures, due to the coupling of the dipole moments of the vibrations and the stabilizing effect of the intermolecular hydrogen bonds when they are forming extended stacks, these modes can be found around 1615 – 1670 cm^{-1} and 1670 – 1700 cm^{-1}, respectively.[30, 38, 41, 46]

Although the Amide I mode was demonstrated to be highly shifted by the supramolecular arrangements of Fmoc-dipeptides between parallel and antiparallel arrangements, and even between different kinds of antiparallel arrangements, its correlation with specific intermolecular arrangements is not straightforward and requires costly computational calculations.[46] However, the Amide I mode is useful to address the changes in the supramolecular hydrogen bonded networks formed by different building blocks in Fmoc-dipeptide nanostructures.

3.1.2 Circular Dichroism (CD)

CD measures the difference in absorption between the right-handed and left-handed circularly polarised light.[195] This is commonly used in proteins, where different secondary structures provide well-defined CD spectra.[181-182, 196-197] Therefore, using the deconvolution of the CD spectrum, it can be used to quantitatively address the relative amounts of different secondary structures in a protein.

Although, as in the case of FT-IR, the CD interpretation for proteins cannot be extrapolated to the study of nanostructures formed by short peptides, [198] CD can still be applied in the area of supramolecular chemistry to gain understanding on the supramolecular arrangements of chromophores.[198-199]

In Fmoc-dipeptides, the application of the CD spectra involves studying the ellipticity of its main chromophore, the fluorenyl group. The fluorenyl group does not contain any chiral centre (Figure 3.3) and, therefore, it is not expected to show any CD signal. However, an important contribution of the fluorenyl absorption between 270 and 310 nm has been observed in the CD spectra of different Fmoc-dipeptide nanostructures.[36, 167] This ellipticity, which causes the CD signal of the fluorenyl group, is provoked by the supramolecular arrangement of the fluorenyl group when forming part of the chiral nanostructures formed.[25, 36, 42, 44, 51, 167, 198]

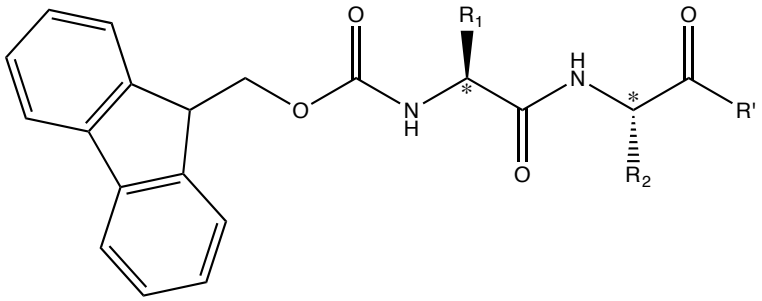


Figure 3.3 Fmoc-dipeptide with its chiral centres marked (*).

As in the case of the FT-IR, although it is difficult to relate absolute CD signals with specific Fmoc-dipeptide supramolecular arrangements, this technique gives valuable information about the supramolecular ellipticity of the fluorenyl groups in the nanostructure, which can be used to compare how this parameter is affected in structures formed by different Fmoc-dipeptides.[42]

3.1.3 Fluorescence Spectroscopy

The use of fluorescence spectroscopy as a characterization technique for Fmoc-dipeptides is not directly extrapolated from how this technique is used in the characterization of the secondary structure of proteins. The application of fluorescence spectroscopy to protein characterization is actually based on adding chromophores, which attach to the proteins and behave in different ways depending in the secondary structure present.[200-203] Nevertheless, the application of fluorescence spectroscopy for the characterization of Fmoc-dipeptides is based in the emission properties of the fluorenyl group, rather than any effects in the dipeptide stacks, even though both are affected by each other.

The fluorenyl group is known to absorb in the UV region and to emit fluorescence with a maximum around 310 – 330 nm.[19, 30-31, 41, 123] Due to the high scattering that arises upon the formation of the nanostructures, it is not possible to study the effects of the supramolecular arrangements in the absorption spectrum. However, it is possible to study the emission. Typically, scattering precludes a quantitative analysis, and emphasis is usually on spectral shape obtained using normalized spectra.

A Jablonski diagram can schematically represent the absorption and emission of the fluorenyl group (Figure 3.4). The energy of these levels is shifted when the fluorenyl group forms π-stacking interaction with other fluorenyl groups in the media.[98, 204] The stabilization provoked by the interactions results in the decrease of

the energy of these levels. As the excited state (S_1) is stabilized more than the ground state (S_0), the presence of this type of interactions shifts the fluorescence to higher wavelengths.

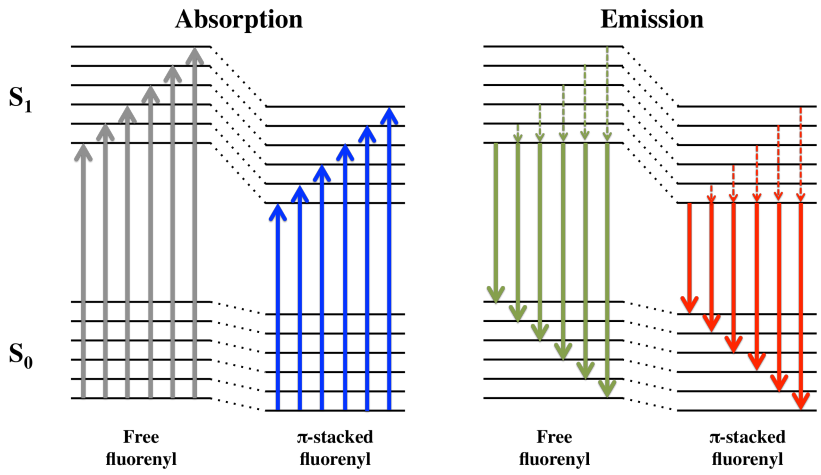


Figure 3.4 Jablonski diagram for the absorption and emission of free and π-stacked fluorenyl groups between the ground (S_0) and the first excited (S_1) electronic states.

Therefore, studying the differences in the emission spectra of nanostructures formed by different Fmoc-dipeptides provides information about the changes in the π-stacking interactions of the different systems.

3.2 Computational Structural Studies

Computational chemistry uses models to mimic experimental systems in order to obtain information which is not accessible with other techniques. In the Oxford dictionary there are different definitions for model. The one that most closely reflects the use of modelling in the context of this thesis is: *A simplified description, especially a mathematical one, of a system or process, to assist calculations and prediction;* and: *A three-dimensional representation of a person or thing or of a proposed structure, typically on a smaller scale than the original*. Computational chemistry or, as it is also called, theoretical chemistry or molecular modelling, applies mathematical models based in three-dimensional representations to mimic some features of the system in order to obtain information of some of its characteristics.

Different computational methods are able to mimic different features of the systems, which limits their applicability and determines what information can be obtained.

Molecular dynamics (MD) methods have been widely used over the last 40 years to model biomolecular systems in order to gain knowledge that is not accessible from other theoretical methods or experimental techniques.[47, 205-207] MD methods simulate the behaviour of molecules representing the atoms as spheres and their interactions and motions using classical mechanics equations rather than quantum mechanical (QM). This simplifies the calculations and allow their application to large systems, such as proteins, DNA, etc. The representation can be atomistic (Figure 3.5a), where each sphere or bead represent a single atom and all atoms are represented, or coarse grained (Figure 3.5c), where each bead represents a group of atoms.[208]

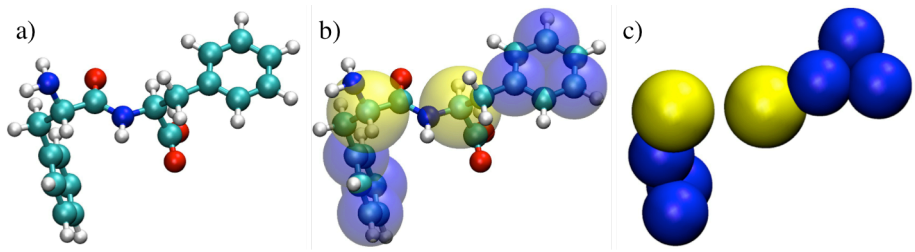


Figure 3.5 FF MD representations: (a) atomistic, (b) mixed atomistic and coarse grained, and (c) coarse grained.

These methods have been widely applied to macromolecular systems, where structure is strongly related with functionality. As structure is dependent on intermolecular interactions, it is also highly dynamic. MD simulations allow studies of how the conformation changes over time under different conditions as well as studies of complexation, for example applied to the modelling of enzyme-substrate complexes. The structure/functionality relationship, the importance of the intermolecular interactions and the high dynamism are also key in supramolecular self-assembling systems. Therefore, MD methods are useful for modelling self-assembling systems.

3.2.1 Molecular Dynamics

MD, as deterministic methods, assume that the future state of the system can be predicted from the previous and use the Newton equation of motion, *i.e.* a classical physics representation, to reproduce the atoms/molecules movement. MD methods use the *Born – Oppenheimer approximation*,[209] which establishes that each atom can be considered as a point mass ignoring quantum dynamical effects. This

approximation considers that the nuclei and electron motion can be decoupled and calculated individually due to their difference in speed (electrons are 10^4 - 10^5 times faster). Nuclei can be considered to be immobile relative to the electron motion and electrons can be assumed to immediately adapt to nuclei movements.

MD simulations are carried out in small time steps (1 – 10 fs) to which the molecular mechanic (MM) potential energy expression is integrated.[210-211] The force is assumed to be constant within the defined time step. The forces on each atom ($\boldsymbol{F}_i$) are calculated using the potential energy variations (dV) based on the position of the atoms ($\boldsymbol{r}_i$), and these forces are used to calculate the acceleration ($\boldsymbol{a}_i$) that each atom, of mass m, suffers (Equation 3.1), originating the motion, *i.e.* the dynamics.

$$\boldsymbol{F}_i = -\frac{dV}{d\boldsymbol{r}_i} = m \cdot \boldsymbol{a}_i \qquad \text{Equation 3.1}$$

Many different algorithms, such as the Verlet or Beeman methods exist to integrate the equations of motion for the small time steps and carry out a simulation based on the potential energy.[212-213]

MM uses classical mechanic expressions to account for the different contributions of interactions to the final potential energy. The equation and parameters used depend on the force field use.[205, 207, 210-211] A typical force field energy expression as a function of the atoms positions (R) is:

$$\begin{aligned} E(R) = & \sum_{Bonds} \frac{k_d}{2}(d - d_0)^2 + \sum_{Angles} \frac{k_\theta}{2}(\theta - \theta_0)^2 \\ & + \sum_{Dihedrals} \frac{V_n}{2}\left(1 + cos(n\omega - \gamma)\right) \\ & + \sum_{Impropers} \frac{k_\varphi}{2}(\varphi - \varphi_0)^2 + \sum_{Electrostatics} \frac{q_i q_j}{4\pi\varepsilon r_{ij}} \\ & + \sum_{vdw} 4 \in_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6}\right] \end{aligned} \qquad \text{Equation 3.2}$$

The first four terms (Equation 3.2: *bonds*, *angles*, *dihedrals* and *impropers*) are called bonded terms and account for the 1 – 4 interactions or interactions between nuclei that are within four 4 atoms. Except the *dihedrals*, bonded terms are expressed as harmonic potentials where the potential energy changes proportional to a force constant (k_d, k_θ, k_φ) as the *bond* between two nuclei (d), the *angle* between three nuclei (θ) or the *improper dihedral* between four nuclei (φ) deviates from the

equilibrium positions (d_0, θ_0, φ_0). The first two terms, *bonds* and *angles*, are known as hard terms because they do not change significantly throughout the simulation due to the high energy required to undergo large deviations from the reference values. Not all force field include a *improper* dihedral term.[214]

The third term in the force field (Equation 3.2, *dihedrals*) accounts for the energetic penalties involved in a bond rotation, described by the rotation angle ω. V_n is the energy barrier height for the rotation and sets the barrier for the rotation. n is the multiplicity and sets the number of minima in the energy profile of the rotation. γ is the phase factor, which establishes the reference system situating the minima in different values of the rotation.

The last two terms in the equation (Equation 3.2, *vdW* and *electrostatic*) are called non-bonded terms and account for the interactions between atoms that are not connected or further than four nuclei away. These are through space interactions where their strength is modelled as function of the inverse of some power of the distance (r_{ij}). The *electrostatic* term uses Coulomb law expression to account for interactions between charges (q_i, q_j). The energy is inversely proportional to the permittivity of the media (ε).

The *vdW* term accounts for the van der Waals interactions between two atoms using a Lennard – Jones function, also known as a 12 – 6 potential due to the powers used in the different terms (Equation 3.2, *vdW*).[215] This term depends on the collision diameter (σ_{ij}), which is the distance between the two particles to which the potential is zero, and the well depth ($\in_{ij}$), which is the depth of the energy minima. The *vdW* term is composed of two parts: the positive part mimics the Pauli repulsion due to the overlap of orbitals at very small distance; and the negative part accounts for dispersion interactions.

Apart from the motion and potential energy expressions, other algorithms are included in the calculations to simulate different conditions in the simulations.[216-221] This is the case of the ensembles, which are collections of all possible different microscopic states of a system to be considered as a unique macroscopic state. The most common ensembles are the canonical ensemble (NVT) and the isobaric-isothermal ensemble (NPT). Both present thermodynamic states with fixed number of atoms (N) and temperature (T). The former also fixes the volume of the system (V) and the latter the pressure (P). *Periodic boundary conditions* (PBC) are used to solve

the issue that simulations are considered to be small in comparison to the relatively infinite real systems.[216] PBC create the effect of doubling the system periodically in all the directions by making the system interact with itself through opposite walls of the system. These methods allow for the calculation of parameters comparable with experimental data, which, in turn, provide validation for the models used as well as the prediction of properties in similar systems.[222-224] In addition to this, the development of analysis tools has also been important: interactions, surface contact, proximity analysis, etc. These tools simplify the process of obtaining useful information from the simulations.[225-226]

3.2.2 MD Simulations for Self-Assembled Systems

Computational methods have been applied with significant success to reveal both molecular level detail and the mechanism of formation for a range of peptide-based nanostructures.[227-237]

Although CG simulations involve the loss of the atomistic resolution and hence the study of explicit interactions, they have been utilized to illustrate the aggregation of peptides and PAs, taking advantage of the possibility of using substantially bigger systems (in the order of hundreds of molecules), which improves the mimicking of the self-assembling systems. [72, 238-240] CG simulations have provided insights into different self-assembling mechanisms by studying the free energy profiles of these processes;[241] and a systematic CG approach was applied to assess the self-assembly propensity of all possible di- and tripeptides and has successfully described literature examples of self-assembled peptides and predicted the self-assembling tendency of other peptides that could subsequently be verified by experimentation.[23, 242]

Atomistic MD simulations have also been applied to increase the understanding of these materials at the molecular level. Primarily, this has been done to study the preferred conformations of the peptides or PAs in the assembled state.[243-245] Atomistic simulations are limited to smaller sizes than CG and require substantially more computation time, but they have the advantage of atomistic resolution, which is required for the study of the specific non-covalent interactions involved in the self-assembly processes. Free energy profile studies using atomistic MD simulations have also been successfully applied for the study of self-assembling mechanism of PAs.[246-247]

APAs have also been studied using atomistic MD simulations to gain understanding on the preferred conformations of these amphiphiles in the assembled state and on the importance of the different interactions in the self-assembling mechanisms.[5, 45, 248-251] Very often these studies are based on studying the relative stability of different conformations through a simulation to then conclude which would lead to a more stable nanostructure. Therefore, an input structure is required, which is often based on experimental observations and basic knowledge of intermolecular interactions. Some of these studies use simple H-bond networks as starting structures, but these are highly affected by the limited size and lack of periodicity, which makes the structures unstable.[5, 45, 250-251] Furthermore, Fmoc-AA fibre models have been built using different supramolecular arrangements.[248-249] Although the latter example is also affected by the edge effects due to the size limit, these are larger systems than the previous and hence they are less affected. However, due to the bigger size of the system, the input structure has to be more detailed and the results are more influenced by the starting conformation. Therefore, the analysis of the results is not straightforward due to the limitations of the method, which have to be taken into account.

As has been mentioned before, MD methods have been widely applied for the study of self-assembling systems because, due to their dynamic character, the possibility of including a large number of molecules in the system and the inclusion of entropy in the calculations, they are more suitable to mimic the main features of these systems.[46, 252-253] However, although the QM implementation for self-assembling systems as well as the interpretation of the results obtained is tricky, QM methods have also been applied to assembling systems and have demonstrated the ability to provide valuable information. For instance, QM IR prediction has been applied to assess the utility of the Amide I vibrational mode as diagnostic method for Fmoc-dipeptide nanostructures.[46] Furthermore, QM structure optimizations have been used to check the different preferred conformations in self-assembled states for different F based short peptides and their corresponding Fmoc-protected variations. [253]

In conclusion, although computational results are useful for the molecular level study of self-assembling systems, a proper selection of the method, which includes type of calculations, starting structures and size of the system, is key to obtain the information required in a reliable way. In addition, the interpretation of

computational results is complicated and their correlation with experimental results is essential for their applicability.

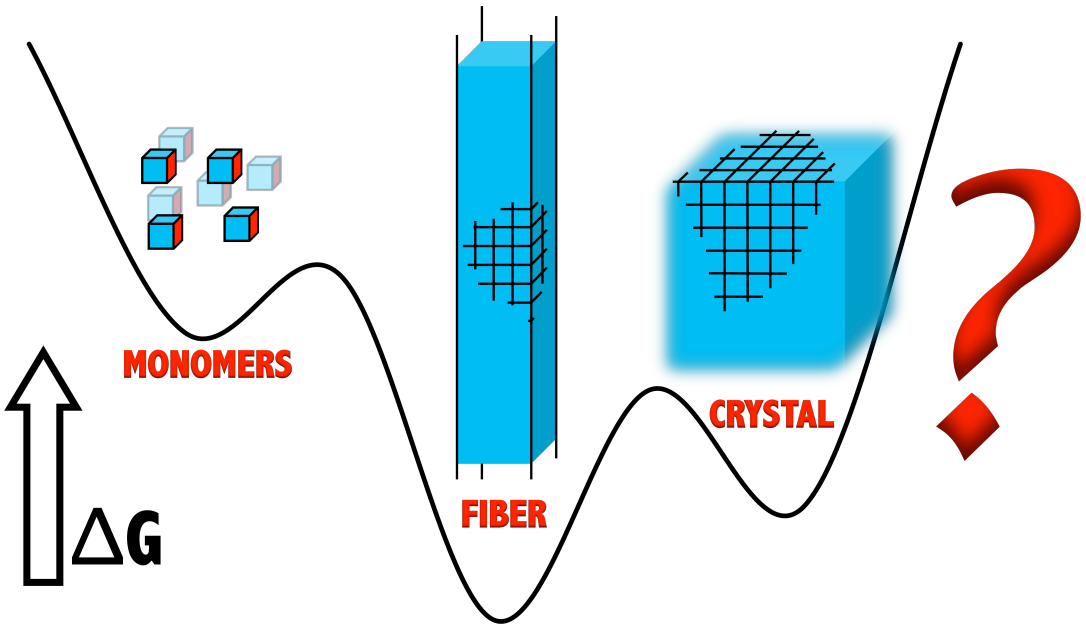
3.3 Computational/Experimental Correlation

Experimental techniques are able to give valuable information about the interactions in peptide nanostructures, but not explicit information of the specific groups involved and hence they currently cannot elucidate their intermolecular arrangements, as they do for proteins. Therefore, the interaction information provided has to be carefully analysed and combined with other techniques and with chemical knowledge to be able to provide some details into the disposition of molecules in the nanostructures. However, this alone still cannot lead to models with detailed interactions.

MD simulations can give details of the intermolecular behaviour of protein-based systems and they have been extrapolated to peptide self-assembly. They have provided information of the relative stability of different dispositions, but their interpretation is also difficult and meaningless if a comparison or correlation with experimental results is not possible. Furthermore, in order to be more effective, they require an input structure, which introduces bias, and to do this accurately would require detailed experimental characterization of the system.

In this thesis a new way of understanding nanostructures is proposed where the information obtained from experimental and computational methods are brought together in a systematic and correlated approach to gain reliable information of nanostructures at intermolecular resolution.

4 The Central Hypothesis: Can Gels Be At Thermodynamic Equilibrium?



4.1 Introduction

The thermodynamic nature of the nanofibrous gel state is not clear in the literature and two different scenarios have been proposed. In the first scenario, the gel state (GS) is composed of a network of fibres that represent a kinetically trapped metastable state, which avoids the transformation of the system to the crystals state (CS).[254-261] The main argument supporting this is that a 3D extended structure (CS) allows a higher number of stabilizing interactions compared to a 1D structure (GS). Depending on the depth of the metastable state and the activation barriers around it, the system could transition to the CS. This gel-crystal transition has been observed for several systems.[257-264]

The alternative scenario is that for certain gelators the 1D structures formed in the GS represent the thermodynamic minimum. For some examples which have clearly defined solvophobic and solvophilic sections, preferential formation of 1D structures, akin to worm-like micelles, is quite intuitive, *e.g.*, Stupp's peptide amphiphiles.[10] The concept of thermodynamically favoured formation of 1D nanostructures has been successfully applied in molecular dynamic simulations and in studies of gels to develop tools to investigate the gelation tendencies as a function of the free energy.[23, 172] Furthermore, self-healing gels in principle depend on the possibility of gels to represent the thermodynamic equilibrium that can be reversibly reached, although it is possible that a local minimum could also represent a gel state that can be reversibly accessed. That is, it has been argued that "thermodynamically stable gels" represent deep local minima surrounded by high activation barriers, which may make it impossible to access the global minimum, the CS.[255-256, 258-261]

As discussed in the literature review (Chapter 2), the functionality of LMWG self-assembling systems is dictated by supramolecular structure. The structure can in turn be tuned by changing the chemical nature of the self-assembling units (or building blocks) as well as the self-assembly pathway.[2, 11, 19, 21, 28-29, 265-266] This pathway dependence clearly suggests that there are both thermodynamic and kinetic aspects to fibril formation and consequent gel formation. Indeed, metastable gels, obtained by controlling the pathway and kinetics of the gelation, have resulted in materials with different nanostructures and properties.[36, 68, 131-132, 158-159, 179, 267] In some cases, different kinetically trapped gels could eventually evolve into the same gel state by using

elevated temperatures to overcome kinetic barriers, which suggests that gels may represent thermodynamic minima that can be accessed reversibly.[132, 158-159]

Simple models have previously been used to describe the gelation and crystallization behaviour of proteins. Dixit *et al*. developed a theoretical model to rationalize the gel-crystal equilibrium for proteins.[268] This model is based on the idea that only in the CS the molecules are organized by well-ordered packing while the GS is an amorphous disordered state. However, it is known that the GS, in the case of most LMWGs, is actually the result of extended well-ordered interactions. Therefore, this kinetic model is not necessarily applicable for LMWGs.

The problem in the case of LMWGs is further complicated due to the similarity and degree of order of the molecular packing in both the CS and the GS, which has been observed experimentally.[257, 259-261, 264] The shape is critical to represent LMWGs and although they have been represented by spheres before,[23, 269] these procedures considered each molecule represented by a number of beads rather than representing each molecule by only one unit. LMWGs are not reasonably represented by spheres because with a spherical potential in the 3D extended structure the CS will always be the most stable.

Furthermore, LMWGs are usually amphiphilic in nature. As such, solvophobic parts of the molecules have a tendency to aggregate while solvophilic parts are preferentially solvent-exposed, thus giving rise to the formation of supramolecular nanostructures. It is clear that not all amphiphiles are able to gelate and there must be a balance between the solvophobic and solvophilic parts to allow the formation of the nanostructures that remain stabilized within the solvent environment.[25, 158, 255, 257, 260] Homogeneous spheres cannot represent the amphiphilicity of a LMWG and thus they cannot represent the balance between solvophobic/solvophilic interactions.

Therefore to develop a model for these molecules, the amphiphilicity of the LMWG, which is clearly critical to formation of 1D fibres and gelation, and the equilibrium between solvophobic and solvophilic interactions, should be taken into account. The challenge is to develop a simple packing model, which takes into account the amphiphilicity of the units, to determine whether fibres, representing the GS, are able to provide a thermodynamic minimum that is more stable than the CS.

A model is presented in this chapter that represents the self-assembling molecules as prisms which present faces of different nature: solvophobic and

solvophilic. The concept of faces of different nature depending on the chemical groups present on them is commonly applied in crystallography and models have been developed to rationalize the final shapes of crystals using this idea.[270-271] However, these are kinetic models that deal with the different growth rates on the different faces of crystals. In contrast, the model presented here focuses on the potential equilibrium stability of fibres of differing lateral dimension (based on the individual molecular units) with reference to the crystals (*vida infra*).

This model shows how the amphiphilicity of the assembling units introduce the possibility of achieving a thermodynamic minimum with a 1D infinite structure rather than a 3D one, demonstrating that fibres can represent the thermodynamically favoured assembled form for certain LMWGs.

The model is in good agreement with both proposed theories, as it shows that the GS or CS can represent the thermodynamic minimum depending on key characteristics of the system related to the balance of solvophobicity and solvophilicity of the LMWG in a solvent. Finally, the model is extended to other geometries to show its validity and applicability to known experimental examples, where it is used to describe the behaviour of aromatic peptide amphiphiles under changing pH conditions. [153]

The conclusions reached in this chapter have critical implications for the theoretical treatment of gels, which is of special importance in this thesis due to the use of techniques which rely on the possibility that gels and nanostructures represent the thermodynamic minima: thermodynamically driven formation of nanostructures using enzymatic assisted self-assembly and dynamic peptide libraries (Chapter 6), and molecular dynamic simulations (Chapters 5 and 7).

4.2 Using Prisms to Represent LMWG

The amphiphilicity of LMWGs is a key property that allows molecules to self-assemble into nanostructures. The presence of "regions" of different nature – solvophobic and solvophilic – allows these molecules to adopt orientations based on a balance between the interactions with the solvent and interactions with other solute molecules. Here, square based prisms (*Sq*), cubes, are used to represent the amphiphilicity of the self-assembling units by considering equal sized faces of different nature (Figure 4.1).

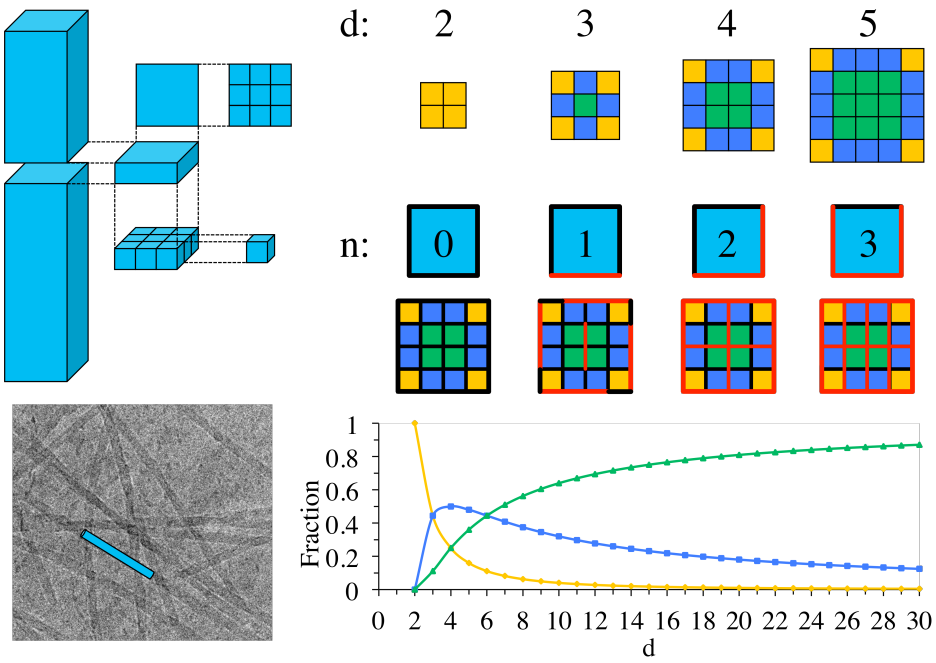


Figure 4.1 Schematic of the model from fibres to the molecular unit (Left) and the cross-section as a function of *n* (showing solvophilic faces in red and solvophobic in black. Lines inside of the fibre are two faces of the same nature, solvophobic or solvophilic, opposed to form the most stable structure) and as a function of *d* (Right). The graph shows the evolution of the fractions of the different units as a function of *d*. Yellow represents corner units (two faces exposed to solvent), blue the side units (one face exposed) and green the fraction of unexposed units.

The model will be subsequently extended to other regular shaped prisms (Section 4.7). The model considers the fibre to be of infinite length in the *z*-direction with the base of the prisms in the *xy*-plane, the cross-section of the fibre. The width of the fibre is a function of the parameter d, which is defined as the number of units per side (Figure 4.1).

4.3 Formulation

The model accounts for the Gibbs energy as the sum of the contributions of the different faces of the prisms that make up the fibre. Clearly, neither the entropic, nor the enthalpic, contributions are strictly additive. However, it is common to approximate them as such. For example, the estimation of solvation Gibbs energies in terms of surface area contributions, common with proteins, uses additivity of entropic contributions.[272] This is a relatively minor approximation within this mesoscale model and should not affect the validity of the general results that the model achieves.

There are two different types of faces: solvophilic *(l,* red in Figure 4.1*)* and solvophobic *(k,* black in Figure 4.1*)*, and these faces can be either exposed to the solvent *(s)* or buried inside the fibre *(b)*. Hence, there are four possible interactions and their contributions to the Gibbs energy are γ_{ls}, γ_{lb}, γ_{ks} and γ_{kb}.

The magnitude, and relative ordering of these parameters is clearly system dependent and different examples are provided below. The favourability of the formation of fibres over crystals requires the interactions between the solvent and exposed faces of the LMWG to be more favourable than the solvent-solvent interactions. While if the solvent-solvent interaction is more favourable, crystals will always be preferred as they eliminate solvent-monomer interactions at the limit of infinite size. Therefore, the relative stability of the solvent-solvent and solvent-fibre interactions is considered within the model with the election of the values used for the γ_{ks} and γ_{ls} parameters. In addition, the model also depends on the solvophobic interactions through the parameter γ_{kb}, which is fundamentally the result of solvent-solvent interactions. However, this dependence does not stop the interaction being well described in terms of Gibbs energy contributions when hydrophobic surfaces contact each other, rather than contacting water. Therefore, the relative stability of the solvent-solvent and solvent-fibre interactions is considered within the model with the election of the values used for the γ_{ks} and γ_{ls} parameters. These parameters do not consider only the direct interaction of the solvent with the self-assembling molecule but also the effect on the solvent, as it is considered implicitly.

To calculate the total contribution to the Gibbs energy of the fibre *(G_{fibre})*, the individual contributions are multiplied by the fraction of faces *(f_{ls}, f_{lb}, f_{ks}* and *f_{kb})* involved in the different interactions (Equation 4.1). By using fractions of the different types of units rather than absolute numbers of units, it is straightforward to consider infinitely large systems in the comparison between fibres and crystals. This is not possible in standard computational methods, which contain fixed number of molecules.

$$G_{fibre} = f_{ls} \cdot \gamma_{ls} + f_{lb} \cdot \gamma_{lb} + f_{ks} \cdot \gamma_{ks} + f_{kb} \cdot \gamma_{kb} \qquad \text{Equation 4.1}$$

These parameters can be related with the total contribution to the Gibbs energy of the crystal state *(G_{crys})* and the solution state *(G_{solv}):*

$$G_{crys} = f_{lb} \cdot \gamma_{lb} + f_{kb} \cdot \gamma_{kb} \qquad \text{Equation 4.2}$$

$$G_{solv} = f_{ls} \cdot \gamma_{ls} + f_{ks} \cdot \gamma_{ks}$$ Equation 4.3

4.4 Parameters Definition

The model derives its parameters from rational considerations:

$n \leq m_{max}$: The number of solvophilic faces in the unit (n) cannot be higher than the maximum number of solvent exposed faces in the cross-section (m_{max}). For example, m_{max} of square prisms is 2 and it already covers the whole fibre surface with solvophilic faces, which has a stabilizing effect (Figure 4.1). However, when the units have one more solvophilic face ($n=3$), the extra faces are buried, which has a destabilizing effect, adding no stabilizing effect to compensate it. This also limits the solubility of the molecules considered in this theory and gives rise to two types of fibres:

• $n < m_{max}$: the fibre exposes both, solvophilic and solvophobic faces to the solvent; *i.e.*, the fibre surface has heterogeneous faces and therefore has an amphiphilic character.

• $n = m_{max}$: the fibre shows a homogenously solvophilic surface but as a result will have solvophilic faces buried in the core for $d>2$ (Figure 4.1).

$n = 0$ is included as a non-amphiphilic reference and $n = 3$ is included to show the effect of having $n > m_{max}$.

$G_{crys} < G_{solv}$: This requirement excludes molecules that are too well solvated to form crystals or fibres.

4.5 Model Development

First, a quantitative formula to describe how exposure of unit faces changes with increased width of a model fibre (d) is developed. The fractions of exposed faces are calculated by dividing the number of units with m faces exposed to the solvent in the cross-section (N_m) by the total number of units on the cross-section (N_{tot}), giving rise to parameters f_m (Equation 4.4). *Square prisms* can have zero (buried, f_0, shown in green), one (fibre side, f_1, blue) or two faces (corner of the fibre, f_2, yellow) exposed to the solvent (Figure 4.1). Both parameters are a function of the parameter d.

$$f_m = \frac{N_m\,(d)}{N_{tot}\,(d)}$$ Equation 4.4

The f_m plots start at $d=2$ because $d=1$ would require a different formulation to represent all the faces in the cross-section exposed to the solvent (Figure 4.1). In the

fraction plot it can be seen how f_2 (the contribution of the corners) is only important at small values of d and decreases rapidly with increasing fibre width (Figure 4.1). Also f_1 (the sides of the fibre) contributes at small values of d (except for the minimum width where the fibre has no side units) and its contribution decreases more slowly, relative to f_2. The contribution of f_0 increases with d and it tends to 1 at infinite values of the d. While the exact interplay of the different parameters is described in detail below, this graph already suggests the possibility of balanced interactions at low d, which would correspond to fibre-like structures.

To calculate the G_{fibre} it is necessary to consider the relationship between f_m and the fraction of faces involving the different interactions (f_{int}). The relationship between these parameters depends on the geometry of the units and on the number of solvophilic faces. The general relationship between these two parameters is described with the equation:

$$f_{int} = \sum_{m=0}^{2} f_m \alpha_{int,m} \qquad \text{Equation 4.5}$$

Where the factor $\alpha_{int,m}$ represents the number of faces that units with m faces exposed to the solvent have involving the different interactions (*ls*, *lb*, *ks* or *kb*). Therefore the general Equation 4.1 is now:

$$G_{fiber} = \sum_{m=0}^{2} \left(f_m \alpha_{ls,m}\right) \gamma_{ls} + \sum_{m=0}^{2} \left(f_m \alpha_{lb,m}\right) \gamma_{lb} + \sum_{m=0}^{2} \left(f_m \alpha_{ks,m}\right) \gamma_{ks} + \sum_{m=0}^{2} \left(f_m \alpha_{kb,m}\right) \gamma_{kb} \qquad \text{Equation 4.6}$$

The factors $\alpha_{i,m}$ can be determined for the case of squares with $n = 1$ (Sq_1) by looking at the disposition of the units in the structure in Figure 4.1 ($\alpha_{int,m}$ values not mentioned are 0):

int = ls: Units with $m = 2$ and with $m = 1$ expose one solvophilic face to the solvent, therefore: $\alpha_{ls,2} = 1$ and $\alpha_{ls,1} = 1$.

int = lb: Only units with $m = 0$ bury their solvophilic face inside of the fibre: $\alpha_{lb,0} = 1$.

int = ks: Only units with $m = 2$ expose one solvophobic face to the solvent: $\alpha_{ks,2} = 1$.

int = kb: Units with $m = 2$ bury 4 solvophobic faces, while $m = 1$ and $m = 0$ bury 5 each: $\alpha_{kb,2} = 4$, $\alpha_{kb,1} = 5$ and $\alpha_{kb,0} = 5$.

The resulting equation is:

$$G_{fiber}(Sq_1) = (f_2 + f_1)\,\gamma_{ls} + f_0\,\gamma_{lb} + f_2\,\gamma_{ks} + (4f_2 + 5f_1 + 5f_0)\,\gamma_{kb} \qquad \text{Equation 4.7}$$

Table 4.1 Expressions for the calculation of the fibre Gibbs energy (G_{fibre}) for *Sq* with the different values of *n*.

Sq₀	$G_{fiber} = (2f_2 + f_1) \cdot \gamma_{ks} + (4f_2 + 5f_1 + 6f_0) \cdot \gamma_{kb}$
Sq₁	$G_{fiber} = (f_2 + f_1) \cdot \gamma_{ls} + f_0 \cdot \gamma_{lb} + f_2 \cdot \gamma_{ks} + (4f_2 + 5f_1 + 5f_0) \cdot \gamma_{kb}$
Sq₂	$G_{fiber} = (2f_2 + f_1) \cdot \gamma_{ls} + (f_1 + 2f_0) \cdot \gamma_{lb} + (4f_2 + 4f_1 + 4f_0) \cdot \gamma_{kb}$
Sq₃	$G_{fiber} = (2f_2 + f_1) \cdot \gamma_{ls} + (f_2 + 2f_1 + 3f_0) \cdot \gamma_{lb} + (3f_2 + 3f_1 + 3f_0) \cdot \gamma_{kb}$

The proposed formulation is applied to derive the G_{fibre} expressions for *Sq* with 0 to 3 solvophilic faces (Table 4.1). $n = 0$ corresponds to a non-amphiphilic molecule; and $n = 3$ is higher than the n_{max} proposed ($n_{max} = m_{max} = 2$) in the parameters definition section. These are used as controls to show how amphiphilicity is required to have fibres at thermodynamic equilibrium.

Finally, taking into account that in solvation all the faces are exposed ($f_6 = 1$) and in the crystal all are buried ($f_0 = 1$) and applying Equation 4.5 to Equation 4.2 and Equation 4.3, the relationship between the parameters for the different types of interactions (γ_i) and of the crystal Gibbs energy (G_{crys}) and of solvation excess Gibbs energy (G_{solv}) can be obtained (Table 4.2). The resulting equations for Sq_n are:

$$G_{crys} = n\,\gamma_{lb} + (6 - n)\,\gamma_{kb} \qquad \text{Equation 4.8}$$

$$G_{solv} = n\,\gamma_{ls} + (6 - n)\,\gamma_{ks} \qquad \text{Equation 4.9}$$

Minimum values for G_{crys} and G_{solv} were calculated (Table 4.2) using the minimum values for γ_{lb} and γ_{ks} mentioned, *i.e.*, 0 and 2 respectively. The equations are applied from these minimum values and as well as a function of *d* they are presented as a function of the parameters γ_{lb} or γ_{ks} (and hence of G_{crys} or G_{solv}) to show

the effect in the ΔG_{fibre} of increasing the destabilizing effect of these interactions. The parameters γ_{lb} and γ_{ks}, can be related with the interactions in a given LMWG, and the difference between G_{crys} and G_{solv} can be related to the solubility of a given molecule in a given solvent. In the case of a molecule which is not soluble in a given solvent, $G_{crys} \ll G_{solv}$, while for a soluble molecule the difference between these two parameters will be minor.

Table 4.2 Expressions and minimum values for the crystal Gibbs energy (G_{crys}) and solvation excess Gibbs energies (G_{solv}) for *Sq* with the different values of *n*.

	G_{crys}	G_{solv}	$G_{crys,min}$	$G_{solv,min}$
Sq₀	$6 \cdot \gamma_{kb}$	$6 \cdot \gamma_{ks}$	6	12
Sq₁	$\gamma_{lb} + 5 \cdot \gamma_{kb}$	$\gamma_{ls} + 5 \cdot \gamma_{ks}$	5	10
Sq₂	$2 \cdot \gamma_{lb} + 4 \cdot \gamma_{kb}$	$2 \cdot \gamma_{ls} + 4 \cdot \gamma_{ks}$	4	8
Sq₃	$3 \cdot \gamma_{lb} + 3 \cdot \gamma_{kb}$	$3 \cdot \gamma_{ls} + 3 \cdot \gamma_{ks}$	3	6

Finally, for ease of comparison between the competing states, the results are presented using the ΔG_{fibre}:

$$\Delta G_{fiber} = G_{fiber} - G_{crys} \qquad \text{Equation 4.9}$$

In this way the results can be interpreted:

- $\Delta G_{fibre} < 0$; $G_{fibre} < G_{crys}$: The fibre represents the thermodynamic minimum.
- $\Delta G_{fibre} > 0$; $G_{fibre} > G_{crys}$: The crystal represents the thermodynamic minimum.

4.6 Results. ΔG_{fibre} Calculation

In order to determine whether a fibre state could represent the thermodynamic minimum, the following energy penalty parameters are defined:

$\gamma_{ls} = 0$: (red face exposed at the surface, Figure 4.1) Solvophilic faces that are solvent exposed are assumed to be the lowest energy penalty as these represent the most stable interactions (*e.g.*, a carboxylate group in water). Since differences in Gibbs energy between states are calculated, it is set to zero as a reference point.

$\gamma_{kb} = 1$: (black, internal, Figure 4.1) The second coefficient, the solvophobic faces buried, is set to 1 as a slightly less stabilizing interaction than solvent exposed solvophilic faces, although still relatively favourable (*e.g.,* aromatic moieties in aqueous solution, which, when buried inside the fibre, can establish π-stacking interactions and liberate solvent).

$\gamma_{ks} \geq 2$: (black, surface exposed, Figure 4.1) The interaction of the solvophobic faces with the solvent is less favourable compared to the two previous types of interactions. This mimics, for example, the exposure of an aliphatic chain to water. The increment of this parameter would be related with increasing the length of the carbon chain.

$\gamma_{lb} > 0$: (red, internal, Figure 4.1) There are two possibilities for interactions of a buried solvophilic face:

- $\gamma_{lb} \geq 2$: solvophilic faces buried in a solvophobic environment – at best comparable to γ_{ks} (*e.g.,* a carboxylate group and the repulsion it generates depending on the pH, and hence, in the fraction of charged groups in the inner part of the fibre).
- $0 < \gamma_{lb} < 2$: solvophilic faces buried in such a way that they interact favourably. This case takes into account the possibility that solvophilic faces can establish interactions in the buried areas with other solvophilic faces (*e.g.,* through hydrogen bonding of amide groups in aqueous solution, or through π-stacking interactions between aromatic molecules in organic solvent. The parameter is varied depending on the strength of the intermolecular interactions). The stability tends towards that of a solvophilic face exposed to the solvent ($\gamma_{ls} = 0$) and is always more stable than the solvophobic face exposed to the solvent ($\gamma_{ks} \geq 2$).

As the parameters γ_{ls} and γ_{kb} are set to 0 and 1 respectively for the G_{fibre} calculations, G_{crys} only depends on the parameter γ_{lb} and G_{solv} only on γ_{ks}, and hence the plots are made with these parameters in the *y*-axes and with the *d* variable in the *x*-axes. The plots also show the evolution of the optimum *d*, defined as the *d* value with the minimum ΔG_{fibre} (d_{min}), in function of the *y*-axes. If the d_{min} is lower than 30, which is the maximum *d* considered in the plots, then the fibre is the thermodynamic favoured assembly.

The results show that for the Sq_0 the ΔG_{fibre} is positive for all the values of γ_{ks} (Figure 4.2a) and it gets closer to 0 as *d* rises (Figure 4.2a and f). However, by adding

amphiphilicity to the units *(n≠0)* the results show values of ΔG_{fibre} below 0 and values of d_{min} below 30. That is, the fibre is the thermodynamic favoured structure.

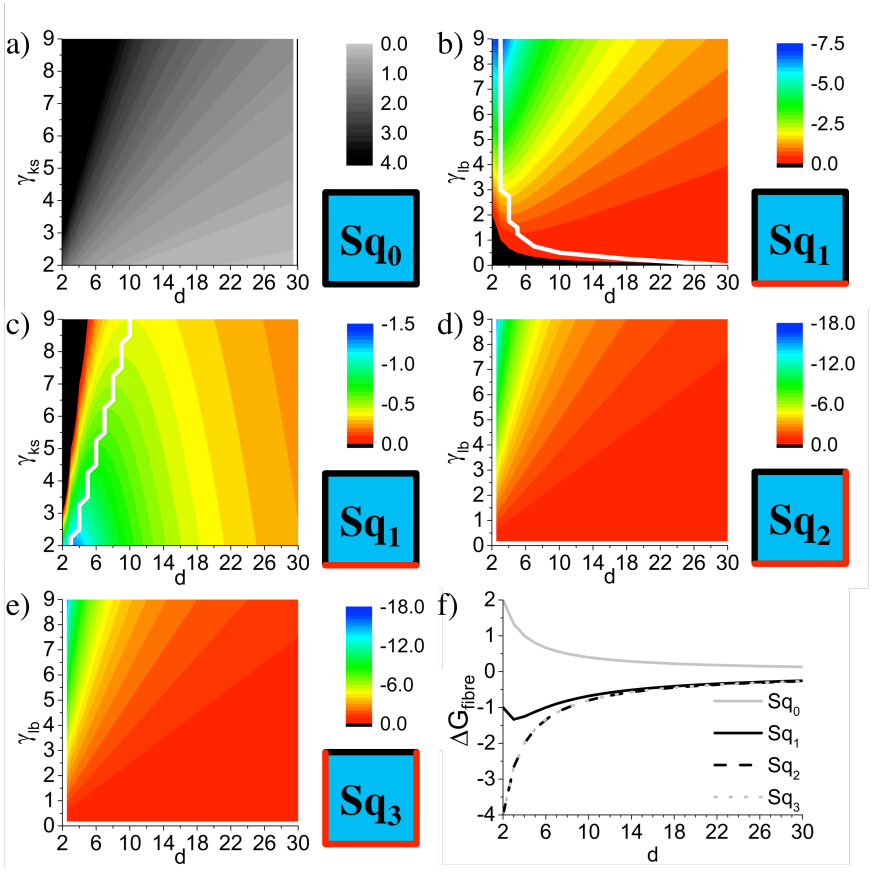


Figure 4.2 ΔG_{fibre} for (a) Sq_0, (b-c) Sq_1, (d) Sq_2 and (e) Sq_3 as a function of d and: (a and c) as a function of γ_{ks} with fixed γ_{lb}; and (b, d-e) as a function of γ_{lb} with fixed γ_{ks}. (f) ΔG_{fibre} for Sq_0, Sq_1, Sq_2 and Sq_3 as a function of d with fixed $\gamma_{ks} = 2$ and $\gamma_{lb} = 2$. d_{min} is represented with a white line (a-e).

Differences can be found between fibres which have solvophobic faces exposed to the solvent, Sq_1 (Figure 4.2b-c), and fibres which do not expose solvophobic faces to the solvent, Sq_2 and Sq_3 (Figure 4.2d-e). In addition, results for the four types of Sq only as a function of d (γ_{lb} and γ_{ks} are set to 2) are shown for a more direct comparison between the four cases (Figure 4.2f). It can be seen how ΔG_{fibre} tends to the minimum at high d's for Sq_0; Sq_1 shows a d_{min} which is different from 2 while Sq_2 and Sq_3, which are overlapped, have a d_{min} which is 2 (favouring the thinnest possible fibre).

Fibres exposing both, solvophilic and solvophobic faces, at the surface have a preferred width (d_{min} value) which is dependent on the interaction parameters (Figure

4.2 b-c, f). When buried solvophilic faces are able to form interactions between themselves $(0 < \gamma_{lb} < 2)$ wider fibres ($4 < d_{min} < 30$) are also stabilized relative to the crystal state (white line in Figure 4.2b). However, as γ_{lb} increases (decreased stability of the solvophilic face when buried) the d_{min} value rapidly decreases and thin fibres ($2 < d_{min} < 4$) are obtained (Figure 4.2b). Conversely, as the energy penalty for exposing solvophobic faces to the solvent increases, the fibres get wider (Figure 4.2c) in order to decrease the fraction of corner units (Figure 4.1). These results suggest that the width of amphiphilic fibres can be tuned with changes in the unit, or molecular structure, of LMWG.

The results for the fibres exposing only solvophilic faces show a constant fibre width (d_{min}) for $\gamma_{lb} \neq 0$. (Figure 4.2d-f). This is not surprising as the minimum fibre width $(d = 2)$ shows no buried solvophilic face and, as there is no solvophobic face exposed to the solvent on these fibres, burying solvophilic faces involves the only destabilizing effect. Overall, this packing model demonstrates that fibres can be at thermodynamic equilibrium under some conditions due to the amphiphilicity of the LMWG.

4.7 Extension to Other Regular Shapes: *Tr* and *Hx*

Square-based prisms may not always be the best way of representing LMWGs but the model can be extended to other prisms: with triangular (*Tr*) and with hexagonal (*Hx*) bases. These are, with *Sq*, the simplest prisms that allow perfect packing in the cross-sections (Figure 4.3). All the faces are considered to have the same area. *Hx* show one main difference with the other two units, namely, that the fibre face units expose 2 faces to the solvent ($m=2$) and the corner units 3 ($m_{max}=3$), while for *Sq* and *Tr* they are 1 and 2 respectively. The formulation for different shapes gives different possibilities to fit the model to real LMWG.

The dependence of the parameters f_m with d varies for the different shapes (Figure 4.4), which already suggest that different shapes can lead to different results in the crystal *vs* fibre competition. However, the fractions for the three shapes have in common the fast decrement of f_{max} and that the three components are important at low *d's*. *Sq* is the only unit which has only corner units at $d=2$ and the fraction of core units increases the fastest for *Hx*.

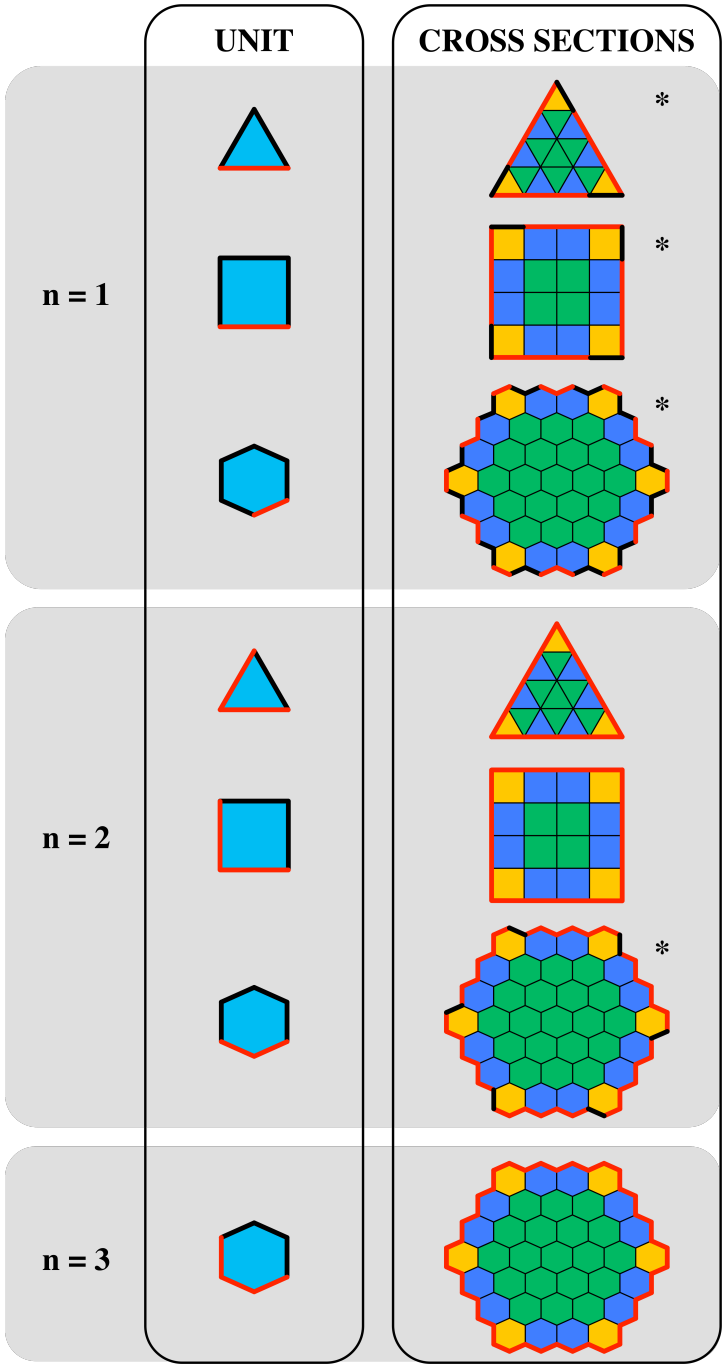


Figure 4.3 Cross-sections of the three shapes with *n=1* and *n=2* for the three and with *n=3* for the hexagons. It shows solvophilic faces (red) and solvophobic faces (black) exposed to the solvent. The cross-sections show buried units in green, units on the fibre faces in blue and units in the fibre corners in yellow (m_{max}). *Cross-sections which propitiate fibres exposing both solvophilic and –phobic faces to the solvent.

The fraction equations for the different shapes can be found in Table 4.3. They show how the number of faces exposed by each type of unit varies as a function of shape and the significant differences derived between *Hx* and the other two units.

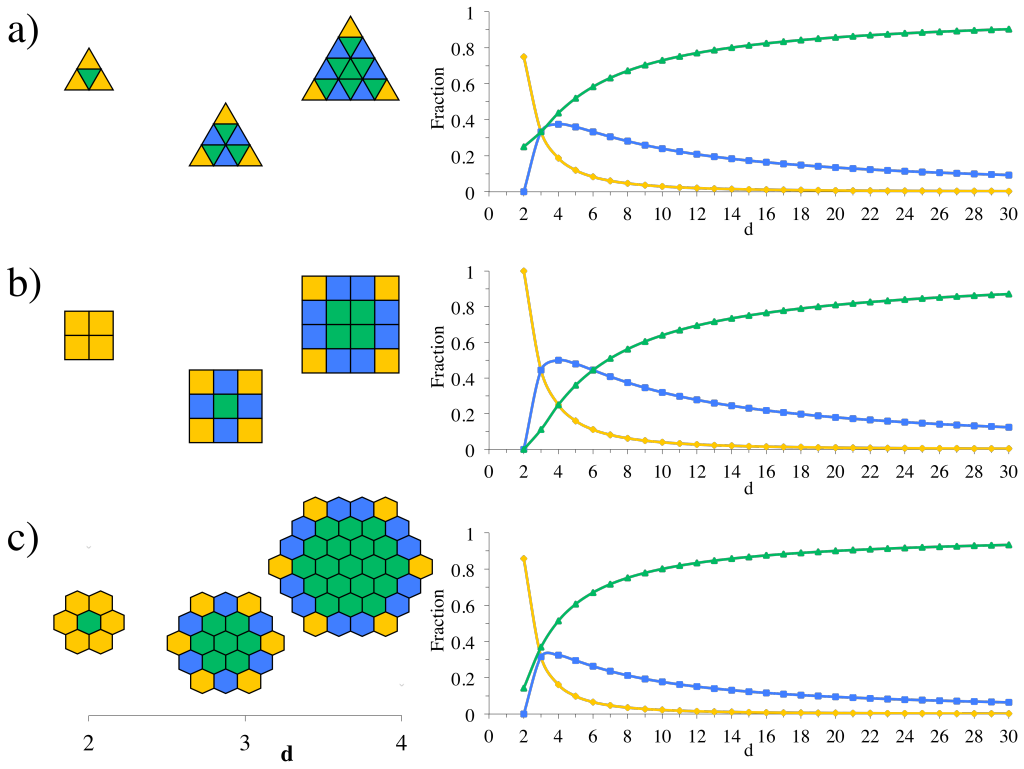


Figure 4.4 Evolution of the fractions of units which can be buried (green), on the faces of the fibre (blue) and on the corners of the fibre (yellow) as a function of d, for: (a) *Tr*, (b) *Sq* and (c) *Hx*. Cross-section examples for d=2-4 (left) and graphic for d=2-30 (right).

Table 4.3 Fractions of units (f_m) with m faces exposed to the solvent for the different regular shapes.

	Triangle (Tr)	**Square (Sq)**	**Hexagon (Hx)**
Buried	$f_0 = \frac{d^2 - 3d + 3}{d^2}$	$f_0 = \frac{(d-2)^2}{d^2}$	$f_0 = \frac{3d^2 - 9d + 7}{3d^2 - 3d + 1}$
Fibre side	$f_1 = \frac{3(d-2)}{d^2}$	$f_1 = \frac{4(d-2)}{d^2}$	$f_2 = \frac{6(d-2)}{3d^2 - 3d + 1}$
Fibre corner	$f_2 = \frac{3}{d^2}$	$f_2 = \frac{4}{d^2}$	$f_3 = \frac{6}{3d^2 - 3d + 1}$

Using the fractions and through the same procedure as described for *Sq* (Section 4.5), related equations can be developed for other regular shaped prisms (Table 4.4).

Table 4.4 Expressions for the calculation of the fibre Gibbs energy (G_{fibre}) for the different regular shapes.

Tr₁	$G_{fiber} = (f_2 + f_1) \cdot \gamma_{ls} + f_0 \cdot \gamma_{lb} + f_2 \cdot \gamma_{ks} + (3f_2 + 4f_1 + 4f_0) \cdot \gamma_{kb}$
Sq₁	$G_{fiber} = (f_2 + f_1) \cdot \gamma_{ls} + f_0 \cdot \gamma_{lb} + f_2 \cdot \gamma_{ks} + (4f_2 + 5f_1 + 5f_0) \cdot \gamma_{kb}$
Hx₁	$G_{fiber} = (f_3 + f_2) \cdot \gamma_{ls} + f_0 \cdot \gamma_{lb} + (2f_3 + f_2) \cdot \gamma_{ks} + (5f_3 + 6f_2 + 7f_0) \cdot \gamma_{kb}$
Tr₂	$G_{fiber} = (2f_2 + f_1) \cdot \gamma_{ls} + (f_1 + 2f_0) \cdot \gamma_{lb} + (3f_2 + 3f_1 + 3f_0) \cdot \gamma_{kb}$
Sq₂	$G_{fiber} = (2f_2 + f_1) \cdot \gamma_{ls} + (f_1 + 2f_0) \cdot \gamma_{lb} + (4f_2 + 4f_1 + 4f_0) \cdot \gamma_{kb}$
Hx₂	$G_{fiber} = (2f_3 + 2f_2) \cdot \gamma_{ls} + 2f_0 \cdot \gamma_{lb} + f_3 \cdot \gamma_{ks} + (5f_3 + 6f_2 + 6f_0) \cdot \gamma_{kb}$
Hx₃	$G_{fiber} = (3f_3 + 2f_2) \cdot \gamma_{ls} + (f_2 + 3f_0) \cdot \gamma_{lb} + (5f_3 + 5f_2 + 5f_0) \cdot \gamma_{kb}$

The G_{crys} and G_{solv} are also defined in a similar way for the two new shapes and their equations can be obtained (Table 4.5). In addition, using the same parameter values as defined in the previous section for the *Sq* (Section 4.6), minimum values can be calculated for both free energies. These values are calculated using the minimum values of the parameters γ_{lb} and γ_{ks}, which are the starting points in the plots below.

Table 4.5 Expressions and minimum values for the crystal Gibbs energies (G_{crys}) and solvation excess Gibbs energies (G_{solv}) for the different regular shapes.

	G_{crys}	G_{solv}	$G_{crys,min}$	$G_{solv,min}$
Tr₁	$\gamma_{lb} + 4 \cdot \gamma_{kb}$	$\gamma_{ls} + 4 \cdot \gamma_{ks}$	4	8
Sq₁	$\gamma_{lb} + 5 \cdot \gamma_{kb}$	$\gamma_{ls} + 5 \cdot \gamma_{ks}$	5	10
Hx₁	$\gamma_{lb} + 7 \cdot \gamma_{kb}$	$\gamma_{ls} + 7 \cdot \gamma_{ks}$	7	14
Tr₂	$2 \cdot \gamma_{lb} + 3 \cdot \gamma_{kb}$	$2 \cdot \gamma_{ls} + 3 \cdot \gamma_{ks}$	3	6
Sq₂	$2 \cdot \gamma_{lb} + 4 \cdot \gamma_{kb}$	$2 \cdot \gamma_{ls} + 4 \cdot \gamma_{ks}$	4	8
Hx₂	$2 \cdot \gamma_{lb} + 6 \cdot \gamma_{kb}$	$2 \cdot \gamma_{ls} + 6 \cdot \gamma_{ks}$	6	12
Hx₃	$3 \cdot \gamma_{lb} + 5 \cdot \gamma_{kb}$	$3 \cdot \gamma_{ls} + 5 \cdot \gamma_{ks}$	5	10

4.7.1 Regular Shapes ΔG_{fibre} Calculations

The ΔG_{fibre} plots for the different regular shapes are presented separately for amphiphilic ($n<m_{max}$) and non-amphiphilic ($n=m_{max}$) fibres. The ΔG_{fibre} plots as a function of γ_{ks} for amphiphilic fibres (Figure 4.5) use a fixed value of γ_{lb} (=2). They show a similar profile for Tr_1 and Sq_1 (Figure 4.5a-b). Both shapes (n=1) present values of ΔG_{fibre} lower than 0 and with d_{min} lower than 10 for a wide range of γ_{ks}. The global minimum in the surface appears at the minimum value of γ_{ks}, which involves the minimum possible destabilization that results from exposing a solvophobic face to solvent. It corresponds with the minimum value of d_{min} and hence the thinnest fibre ($d=3$). As γ_{ks} increases, the thermodynamically favoured fibre is wider because this reduces the fraction of solvophobic faces exposed to the solvent (f_2, Figure 4.4). But even after increasing γ_{ks} by a factor of 4, the fibre width is only 3 times greater.

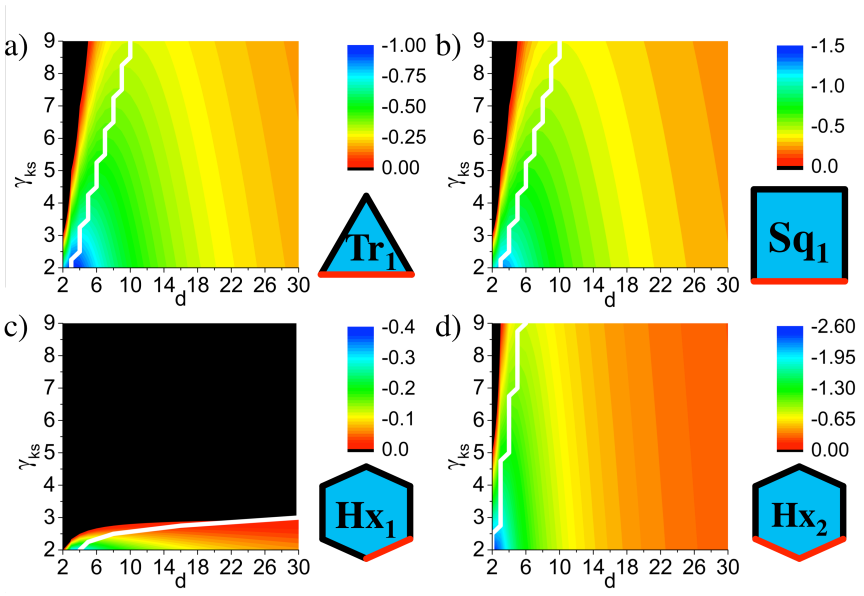


Figure 4.5 ΔG_{fibre} as a function of d and γ_{ks} with fixed γ_{lb} for amphiphilic fibres: (a) Tr_1, (b) Sq_1, (c) Hx_1 and (d) Hx_2. d_{min} is represented with a white line.

The plot for Hx_1 clearly differs from the others (Figure 4.5c). This plot shows that d_{min} rises faster when increasing the value of γ_{ks} than for the other cases. Even for values of γ_{ks} only slightly greater than 3, d_{min} is already over 30, which suggest that the fibre is not the thermodynamic favoured product in this case. This is not surprising because for this shape more than half of the faces exposed to solvent are solvophobic. Thus, a fibre structure can, in this case, only be stable when the loss of stabilization

due to this exposure is very low, as shown in the graph. Nevertheless, when the *Hx* unit has two solvophilic faces (Hx_2), it reduces the number of solvophobic faces exposed to the solvent to almost half and the plot shows a profile more similar to the Tr_1 and Sq_1 (Figure 4.5a-b). Hx_2 actually tends to form thinner fibres than these, due to the fact that N increases faster with d for the *Hx* (Table 4.3) and hence the fraction of core units, which bury the solvophilic faces, as well (Figure 4.4).

The ΔG_{fibre} plots as a function of γ_{lb} (Figure 4.6) take into account that $G_{crys}<G_{solv}$, and hence, the values of γ_{ks} used are those which give values of G_{solv} which are one unit greater than the maximum G_{crys} in the plot range, which corresponds to a γ_{lb} of 9. The values of γ_{ks} for the different prisms are: 3.5 (Tr_1), 3 (Sq_1), 2.4 (Hx_1) and 3.8 (Hx_2).

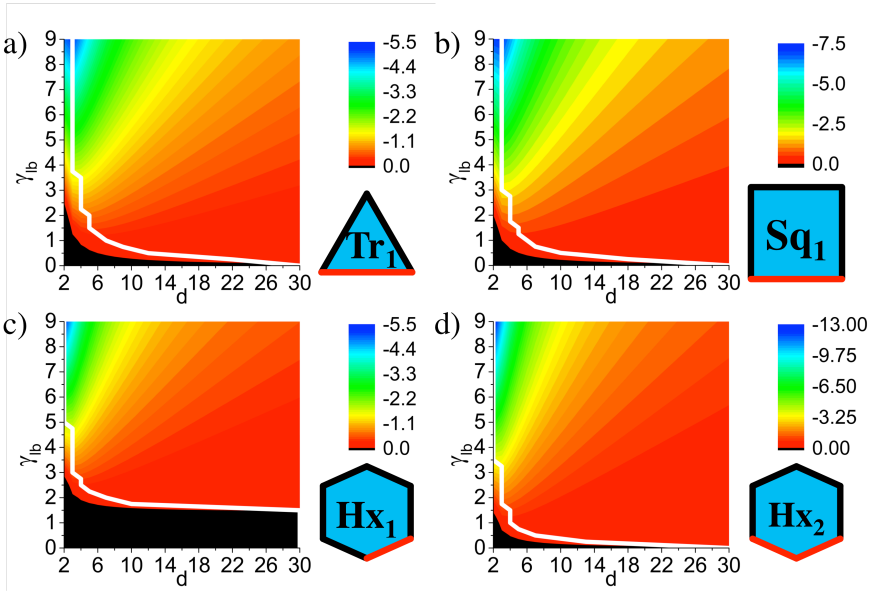


Figure 4.6 ΔG_{fibre} as a function of d and γ_{lb} with fixed γ_{ks} for amphiphilic fibres: (a) Tr_1, (b) Sq_1, (c) Hx_1 and (d) Hx_2. d_{min} is represented with a white line.

The ΔG_{fibre} plots show similar profiles for the four cases (Figure 4.6). All show a minimum with ΔG_{fibre} lower than 0. This is a shallow minimum for low values of γ_{lb} which becomes deeper as the parameter increases. d_{min} also decreases as this parameter rises. These tendencies are due to the increasing destabilization of the buried solvophilic faces with the parameter γ_{lb}, which also enhances the destabilizing effect of increasing d.

In general, the minimum d_{min} corresponds to the maximum value of γ_{lb}, and hence, to the minimum difference between G_{crys} and G_{solv}. This is consistent with the experimental observations which shows that molecules with a limited solubility have a higher tendency to form fibres. Wider fibres are expected as this difference gets higher ($G_{crys} << G_{solv}$), and at the greatest difference, minimum γ_{lb}, the crystal is the thermodynamic minimum for Hx_1.

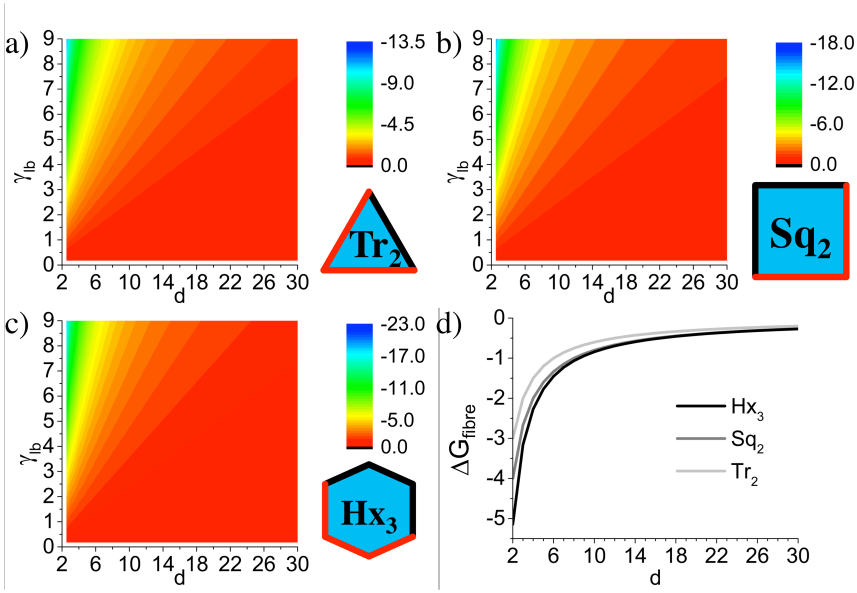


Figure 4.7 ΔG_{fibre} for non-amphiphilic fibres as a function of d and: γ_{lb} with fixed γ_{ks} (a) Tr_2, (b) Sq_2, (c) Hx_3 and (d) with fixed γ_{lb} for Tr_2, Sq_2 and Hx_3. d_{min} is represented with a white line in graphs a-c (which in these cases matches the axes).

For non-amphiphilic fibres (Tr_2, Sq_2 and Hx_3) the trends in the plots (Figure 4.7) are different to those shown previously (Figure 4.6). First of all, as there is no solvophobic face exposed to the solvent there is no dependence with the parameter γ_{ks} and therefore the plots with the fixed γ_{lb} are only shown as a function of d (Figure 4.7d). All the plots show negative ΔG_{fibre} supporting the idea that also non-amphiphilic fibres can be at thermodynamic equilibrium for the different units studied. Furthermore, the d_{min} adopts the minimum possible value in all the cases ($d=2$) because the only destabilizing effect corresponds to burying solvophilic faces, the number of which increases with d.

The fact that similar results are obtained even when changing the shape of the units strongly supports the assumptions to which the model is constructed and provides more possibilities to mimic real systems.

4.8 Extension to 2D Objects

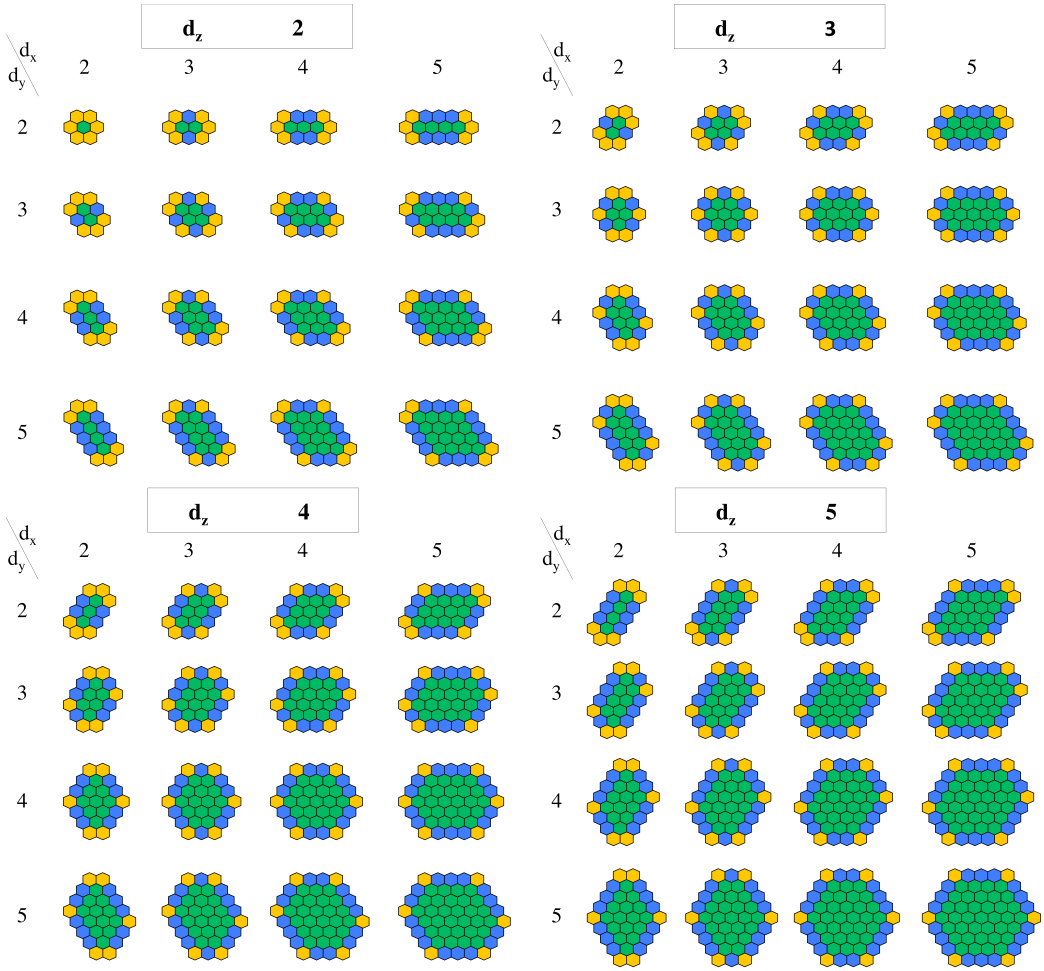


Figure 4.8 Evolution of the fractions of units which can be buried (green), on the faces of the fibre (blue) and on the corners of the fibre (yellow) as a function of three directions (d_x, d_y and d_z) on the cross-section plane (xy-plane).

Experimentally there are other types of nanostructures besides fibres, such as the 2D objects: tapes and sheets. These nanostructures can still be assumed to grow infinitely in the z-direction but they also considerably grow in one direction of the xy-plane. *Hx* units were used to model this (Figure 4.8). It can be seen that the structure can grow in the xy-plane in three different directions (d_x, d_y and d_z) due to the elongation of every two parallel sides of the hexagonal cross-section. By this formulation, if the structure has relatively low values for all these parameters, it is a

1D object, *i.e.*, a fibre, while if the structure grows in only one of these dimensions, the result is a two dimensional object, *e.g.*, tape or ribbon.

The fractions for the *Hx* units were reformulated as a function of the three dimensions (d_x, d_y and d_z) in the *xy*-plane (Table 4.6).

Table 4.6 Fractions of units (f_m) with m faces exposed to the solvent for the *Hx* as a function of d_x, d_y and d_z.

	Hexagon (Hx)
Buried	$f_0 = \frac{(d_x d_y + d_x d_z + d_y d_z) - 3(d_x + d_y + d_z) + 7}{(d_x d_y + d_x d_z + d_y d_z) - (d_x + d_y + d_z) + 1}$
Fibre side	$f_2 = \frac{2(d_x + d_y + d_z - 6)}{(d_x d_y + d_x d_z + d_y d_z) - (d_x + d_y + d_z) + 1}$
Fibre corner	$f_3 = \frac{6}{(d_x d_y + d_x d_z + d_y d_z) - (d_x + d_y + d_z) + 1}$

4.8.1 2D Objects ΔG_{fibre} Calculations

The ΔG_{fibre} was calculated using Hx_2 and Hx_3 as units due to their encouraging results in the 1D formulation and to check if there could be any difference between non-amphiphilic and amphiphilic structures. The ΔG_{fibre} are shown as a function of two of the dimensions in the *xy*-plane: d_x and d_y. As the three directions are equivalent to each other, this is enough to see if the fibre tends to grow or not in just one of these dimensions of the *xy*-plane. Consequently, the two axes are used for two of those dimensions (d_x and d_y) and the results need to be shown with all the energy penalty parameters fixed:

- γ_{ls} and γ_{kb}, like in the previous calculations, are fixed to 0 and 1 respectively.
- γ_{lb} and γ_{ks}, are fixed within each plot (Figure 4.9), but three different pairs of values are showed for each unit:
    - Hx_2: (a) 2:4, (b) 2:2, and (c) 4:2.5
    - Hx_3: (d) 2:4, (e) 3:3, and (f) 6:5

This involves that each unit uses a set of parameters where ($\gamma_{lb}<\gamma_{ks}$), a second set of parameters where ($\gamma_{lb}=\gamma_{ks}$), and a final set of parameters where ($\gamma_{lb}>\gamma_{ks}$)

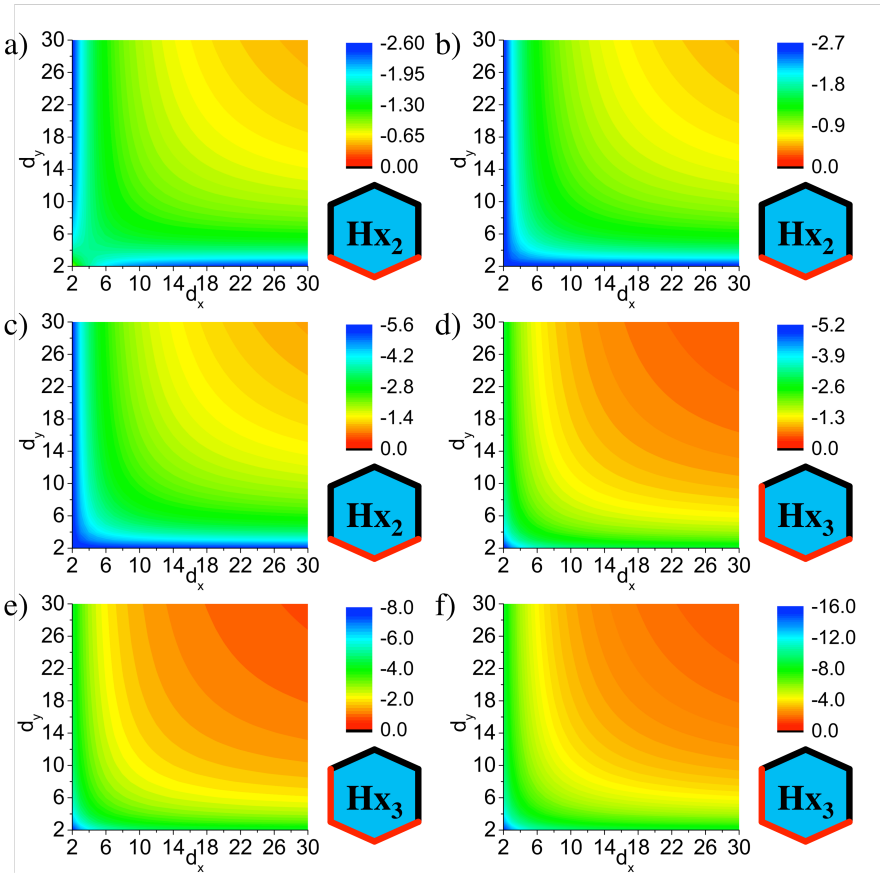


Figure 4.9 ΔG_{fibre} as a function of d_x and d_y for (a-c) Hx_2 and (d-f) Hx_3 with fixed values of γ_{lb} and γ_{ks}: (a) 2:4, (b) 2:2, (c) 4:2.5, (d) 2:4, (e) 3:3 and (f) 6:5.

The results show that for non-amphiphilic fibres, Hx_3, the system shows the minimum excess Gibbs energy for small *d's* with no difference between them (Figure 4.9d-f), suggesting fibres to be the thermodynamic favoured product for these systems independently to the different interaction parameters tried. However, the amphiphilic fibres show different behaviours with different interaction parameters (Figure 4.9a-c). The case where the excess energy of burying a solvophilic face is more favourable than exposing a solvophobic face ($\gamma_{lb}<\gamma_{ks}$) in the structure clearly shows (Figure 4.9a) a preference for a 2D growth ($G_{fibre,min}$ corresponds to $d_x=30$ and $d_y=2$). For the other two cases the preference is not that clear and the G_{fibre} is relatively shallow. For the case where both parameters have the same value ($\gamma_{lb}=\gamma_{ks}$) the actual $\Delta G_{fibre,min}$ (Figure 4.9b) corresponds also to a 2D structure ($d_x=30$ and $d_y=2$) while when the energy penalty of burying solvophilic faces is greater ($\gamma_{lb}>\gamma_{ks}$) the fibres ($d_x= d_y=2$) are the preferred product (Figure 4.9c).

4.9 Comparison with Experimental Observations for Known LMWG

The model has successfully demonstrated that 1D packing can represent the thermodynamic minimum. This is demonstrated by using *Sq* (cubes) and a packing mode which increases equally in all the directions of the *xy*-plane in order to provide a situation where both 1D and 3D packing are possible. The formulation was also extended to other shapes and dimensions and similar results are obtained. This suggests that the formulation is flexible and can be adapted to represent specific LMWGs.

When considering a specific example, it is important to consider what interactions will be competing and the relative strength of these interactions. This is a user-input aspect of the model, the model itself does not prescribe the relative contributions of the different interactions, rather it evaluates the resulting stability of structures based on the user inputs. As such, the model is able to predict if a self-assembling molecule has parameters within the user-defined bounds then it should preferentially form a fibre rather than a crystal (*e.g.*, a hydrophilic interaction on one face that is twice as strong as the possible interactions between the remaining hydrophobic faces results in a fibre). Or alternatively, the model can be used to rationalize the types of structures found by determining what the relative strengths of the interactions that could result in such a structure.

Tang *et al*. demonstrated how the self-assembled structure for N-fluorenyl-9-methyloxycarbonyl-dyphenylananine (Fmoc-FF-OH) varies with the pH.[153] They showed how at high pH values the negatively charged Fmoc-FF-O^- remain in solution, but as the pH decreases and they become protonated they self-assemble into nanostructures, ultimately resulting in the formation of a gel. However, when the pH is reduced further (below 6), the system precipitates due to lateral aggregation (Figure 4.10f). Note that Raeburn *et al*. reported a single apparent pKa for this same system at 8.9, observed upon lowering of the pH using a hydrolysing sugar ester (gluconolactone).[273] However, in the Tang report a heat-cool cycle was used at each incremental pH change in an effort to unlock any kinetic aggregates, and therefore is more likely to represent the equilibrium state.[153]

Prisms with a hexagonal base were used to mimic Fmoc-FF-OH either with one solvophilic face, which represents the COOH terminus (Hx_1, Figure 4.10a), or with two, non-contiguous, solvophilic faces (Hx_{1+1}, Figure 4.10c), representing the

COOH and hydrophilic carbamate (-CO(C=O)-) linker. The pH changes are mimicked by changing the energetic penalties for each type of face (Figure 4.10e). The solvophilic faces buried (γ_{lb}) change from being highly unfavourable (9) due to the charge repulsion of COO^- to be more favourable due to the possibility of forming hydrogen bonds of COOH (2). The γ_{ls} does not involve a penalty (0) when the molecule is charged but it gets less favourable (3.5) when it is neutral. The term for the solvophobic faces exposed to the solvent (γ_{ks}) is kept constant and unfavourable (4.5) but the term for these faces buried (γ_{kb}) changes from involving the same energy penalty as previous due to the proximity of negatively charged groups (4.5) to involve a minimum penalty due to the neutrality which minimizes the repulsion with the *l* faces, and to the possibility of π-stacking interactions with other *k* faces (2). The value of 4.5 was chosen as it is the half of the maximum energetic penalty used in this study, 9. The maximum γ_{ls} (3.5) was chosen to have a value between the minimum γ_{lb} (2) and 4.5. The *x*-axes of the results show only the changes in the parameter γ_{lb} (Figure 4.10b and d).

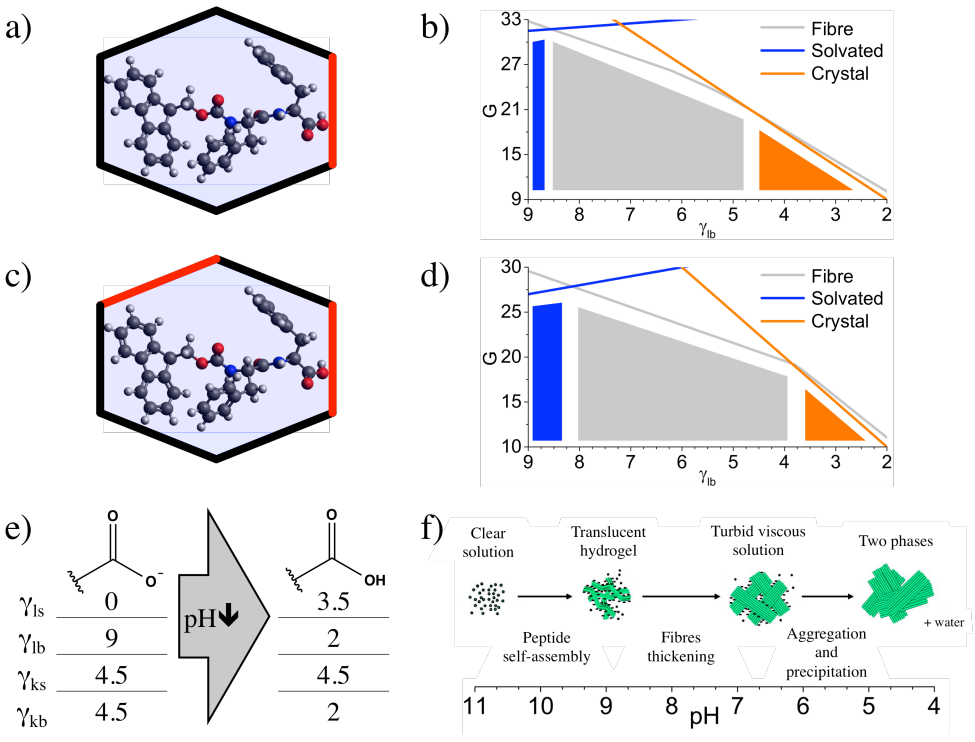


Figure 4.10 Fitting of the (a) Hx_1 and the (b) Hx_{1+1} prisms with the Fmoc-FF-OH molecule (Solvophilic faces in red and solvophobic in black). Results showing the changes of G_{solv} (blue), G_{fibre} (grey) and G_{crys} (orange) for (b) Hx_1 and (d) the Hx_{1+1}. The coloured areas show the most stable state in each region. The plots are presented as a function of γ_{lb}

but the extreme values for the four parameters to reproduce the non-protonated and protonated states are presented in E. (f) Experimental observation for the system by Tang *et al*., adapted from ref [153].

The results for these two prisms are shown as a phase change diagram showing which phase involves the minimum energetic state with coloured areas. This is because the plots used in this chapter until now only show the competition between the crystal and the fibre (Figure 4.11). It has to be taken in consideration that for the Hx_{1+1} both solvophilic faces are treated as equal, but only the side representing the acid group realistically suffers significant parameter changes. However, the Hx_1 does not take into account the changes in the interactions with the carbamate group, which although expected to be less important than those in the acid group, they might still be significant. Therefore, the most accurate representation may reasonably be a combination of both hexagonal based prisms. However, although the two prisms differ in the points where the phase transitions occur, both, Hx_1 and Hx_{1+1} (Figure 4.10b and d) show how the system changes from solution to fibre state and finally to the crystal state as the parameters which mimic the pH drop (Figure 4.10e). Therefore, the model qualitatively reproduces the experimental changes with the pH shown by Tang *et al*. (Figure 4.10f) involving three different self-assembly states depending on the pH of the system.

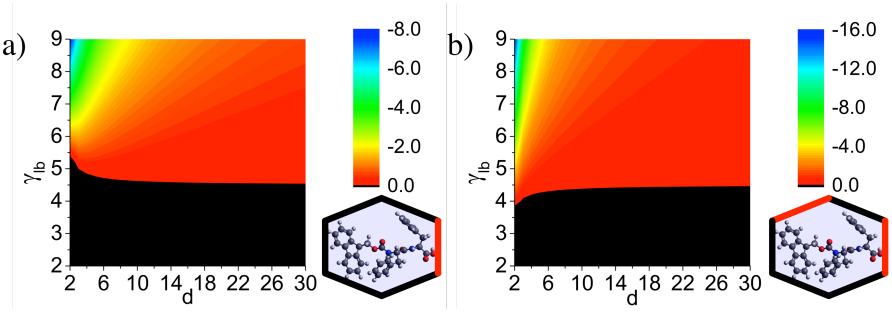


Figure 4.11 ΔG_{fibre} as a function of d and as a function of γ_{lb} for (a) Hx_1 and (b) Hx_{1+1}. The rest of the parameters change as indicated in Figure 4.10.

4.10 Conclusions

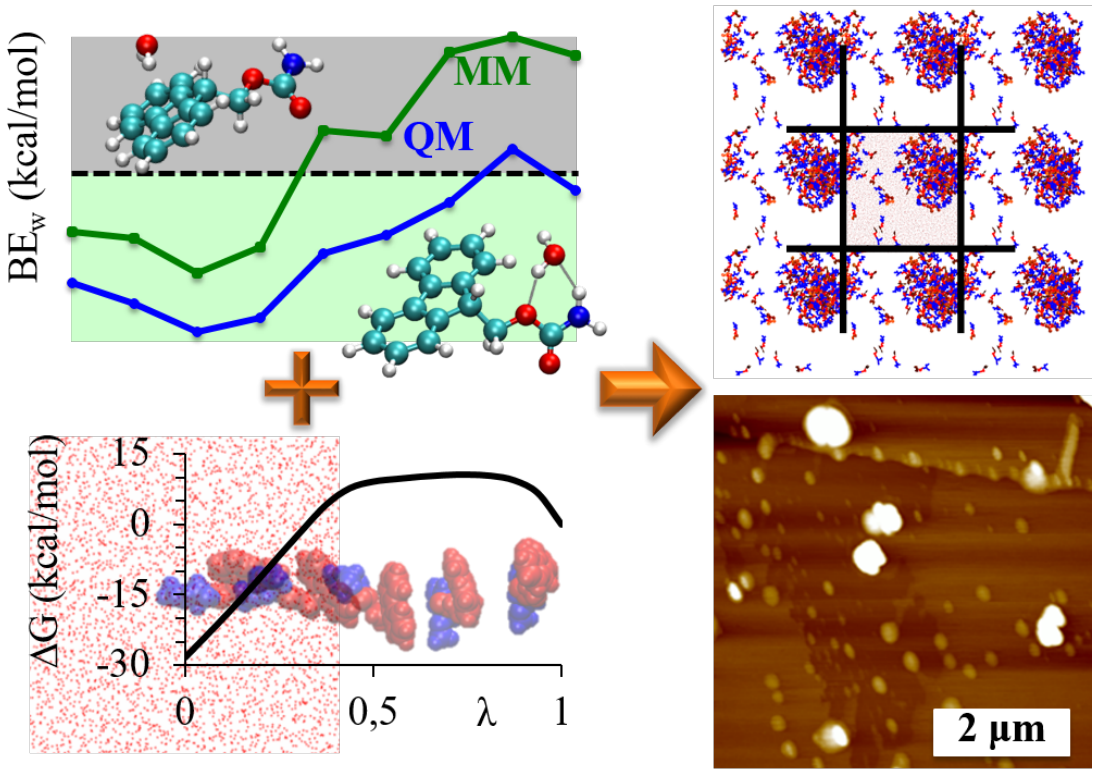
A simple packing model for LMWG has been proposed based on prisms with tuneable faces with different levels of solvophobicity or solvophilicity. This model includes tuneable parameters, which can be related to the characteristics of a given gelator, including the shape; the proportion of solvophilic and solvophilic surface; and

their solubility and crystallization tendency. Within this model, the amphiphilicity of molecules dictates the tendency to form 1D extended structures versus the 3D extended crystalline structure and it is shown that by simply changing the strength of interaction parameters (comparable with *e.g.*, introducing extra H-bonding or stacking interactions to favour or disfavour solvent interactions) a system can be switched from preferential 1D assembly to preferential crystallization. Thus, the model demonstrates that for selected classes of LMWGs the fibre structure can represent the thermodynamic minimum. This observation is critical to this thesis as it supports the use of (equilibrium driven) dynamic peptide libraries and MD simulations for the study of these systems, as these techniques require the systems under study to be at thermodynamic equilibrium.

The formulation has also been successfully applied for other regular shapes, which allows one to easily fit LMWGs into the model, and also for irregular cross-sections which allow the extension of the model to demonstrate that, under some conditions, 2D objects can also be at thermodynamic equilibrium.

The versatility of the model in terms of shapes and cross-sections has allowed the application of the model to reproduce experimental results for a specific LMWG. For example, the pH responsive gelation of Fmoc-FF-OH, which is soluble at high pH, results in gelation at intermediate pH and gives rise to lateral aggregation and eventual precipitation at low pH, can be accurately described by simply changing the surface interaction parameter from unfavourable (repulsion of anionic groups at high pH) to favourable (fibril formation at intermediate pH where surface ionization occurs) to unfavourable (low pH). This demonstrates the potential applications of a simple model to enhance the understanding of the thermodynamics of nanostructures formation.

5 Developing the Methodology: Parameterization of the Fmoc Moiety for the CHARMM Force Field



5.1 Introduction

MD methods have been applied to the study of APAs in order to gain understanding of the preferred conformations of these molecules in the self-assembled state. In addition, they can shed a light on the importance of the different interactions in the self-assembling mechanism.[5, 45, 248-251] In order to study these types of effects at the atomistic level, it is necessary to use well-defined models that are able to provide meaningful information on the particular properties of the systems that are of interest. The definition of these models is not straightforward and this aspect is examined in Chapter 3. The implementation of simulations for these systems has an additional difficulty when using new or unusual chemical moieties that typically are part of APAs, which is that the force fields used for the atomistic simulations of biomolecular systems do not routinely include parameters for the aromatic moieties of interest. Therefore, the first step for the use of MD methods in Fmoc-peptide self-assembly is the parameterization of the Fmoc moiety. As the quality of the results obtained in subsequent simulations – particularly in the case of unbiased, long timescale simulations – will be dependent on the quality of the parameters used, and as such an accurate and consistent parameterization of the Fmoc moiety is critical.

Force field parameters are typically obtained either from experimental results or quantum mechanical (QM) data.[205, 207, 210, 274] The molecules used to derive these parameters and the way they are obtained and optimized are characteristic of a force field and limit its applications. The CHARMM force field was chosen to simulate the self-assembling peptide based systems as it is parameterized and well validated for proteins and peptides.[274-275] The CHARMM force field also includes parameters for many organic molecules (including fluorene and other aromatic groups), lipids, nucleic acids, and some carbohydrates.[275-277] It is common when parameterizing a new molecule in a force field, which includes parameters for such a wide range of molecules to obtain the bonded parameters from similar segments of molecules already parameterized for that force field.[248, 278-279] However, the non-bonded parameters, electrostatic and van der Waals, need to be optimized. As these parameters may have an influence on the torsional potentials, the torsional parameters also need to be optimized.

The CHARMM parameterization protocol evaluates the interactions of the hydrophilic parts of a given molecule with water.[274, 276-277] Nevertheless, to reproduce

the self-assembling behaviour of APAs a balance between the hydrophilic and the hydrophobic nature of the group is essential. Therefore, to parameterize the aromatic moieties used in APAs, a new protocol is required which takes into account these features. In this chapter, a modified protocol to parameterize these aromatic moieties is presented, within the CHARMM force field, and used for the Fmoc moiety parameterization.

This chapter starts with the parameterization of the Fmoc moiety, which includes: the derivation of the bonded parameters from parameterized molecules with similar segments; the charges and van der Waals terms optimization; and the torsional terms optimization. The second part of the chapter is focused on the validation of the parameterization: firstly comparing the partition coefficient of Fmoc-S-OMe obtained computationally with the partition coefficient measured experimentally; and secondly by using the parameters to study a self-assembling system whose final structure is well known.

5.2 Parameterization of Fmoc

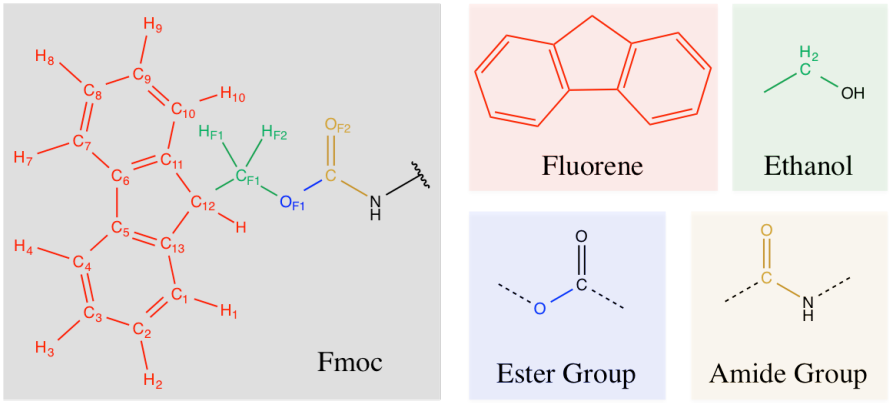


Figure 5.1 Atoms names for the Fmoc moiety (left) and the segments used to obtain some of its parameters (right).

The parameterization was made for the CHARMM force field, which includes parameters for amino acids as well as a wide range of other organic molecules. To parameterize the Fmoc moiety some of the parameters were extrapolated from similar molecules or molecules which present the same chemical groups of the CHARMM parameters library.

The parameters for 9-fluorenylmethyloxycarbonyl, the Fmoc moiety, were obtained as follows. For the fluorenyl group the parameters were taken from the

fluorene molecule (Figure 5.1 red); for the –CH_2– which links the aromatic part of the group to the oxygen, the parameters were taken from ethanol (Figure 5.1 green); the parameters for the following oxygen (–O–) were obtained from standard ester groups (Figure 5.1 blue) and parameters from general amides were used for the carbonyl group (C=O) (Figure 5.1 orange).

The CHARMM force field presents the following expression for the potential energy as a function of the position of the atoms (R):

$$\begin{aligned} \mathrm{E}(R) = & \sum_{bonds} k_b\left(r - r_{0,b}\right)^2 + \sum_{UB} k_{UB}\left(S - S_{0,UB}\right)^2 \\ & + \sum_{angle} k_a\left(\theta - \theta_{0,a}\right)^2 \\ & + \sum_{dihedrals} k_{d,n}\left(1 + cos\left(n\chi - \delta_{d,n}\right)\right) \\ & + \sum_{impropers} k_i\left(\psi - \psi_{0,i}\right)^2 \\ & + \sum_{vdw} \varepsilon_{ij}\left[\left(\frac{R_{min,ij}}{r_{ij}}\right)^{12} - 2\left(\frac{R_{min,ij}}{r_{ij}}\right)^{6}\right] \\ & + \sum_{electrostatic} \frac{q_i q_j}{4\pi\varepsilon_0 r_{ij}} \end{aligned}$$

Equation 5.1

This equation has seven different terms. The force constants (k_b, k_{UB}, k_a, k_i) and reference values ($r_{0,b}$, $S_{0,UB}$, $\theta_{0,a}$, $\psi_{0,i}$) for the bonds stretching, Urey-Bradley terms, angles bending and improper dihedrals as well as the Lenard-Jones parameters (ε_{ij}, $R_{min,ij}$) were directly transferred from the above-mentioned groups in the CHARMM library. The dihedral terms ($k_{d,n}$, n, $\delta_{d,n}$) for the fluorenyl were also taken form the fluorene molecule. Therefore the parameterization effort was focused on the dihedral terms of the *linker* between the aromatic group and the peptide and on the charges (q_i), which cannot be directly transferred from the segments.

5.2.1 QM and MM Binding Energies: Fmoc – Water

To calculate the binding energies of Fmoc with water, the molecule Fmoc-NH_2 (Figure 5.2) was used. This molecule was optimized at the QM level of theory. The parameterization protocol for CHARMM uses the Hartree-Fock method with the 6-31G* basis set for the binding energies calculations.[276] However, the intended application includes the calculation of binding energies between water and the

aromatic group, where dispersion forces may play an important role. As Hartree-Fock is unable to describe dispersion, the QM calculations were carried out using the dispersion-corrected DFT functional, B97-D,[280] with the basis set def2-SVP,[281] in Turbomole.[282-283] The MM binding energies were calculated using the CHARMM force field with the Fmoc parameters in the NAMD program.[284]

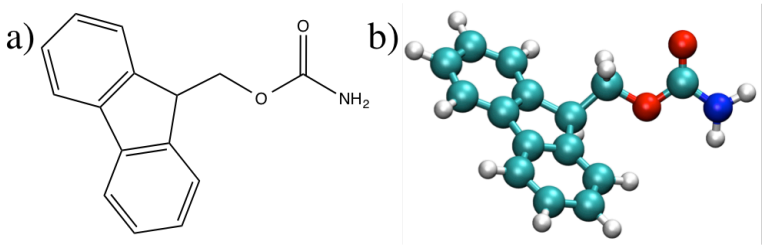


Figure 5.2 Fmoc-NH_2 model molecule in (a) 2D and (b) 3D representations.

The optimized geometry of Fmoc-NH_2 was fixed for all the *Fmoc – water* QM binding energies calculation. The geometry of water was also fixed in these calculations with the TIP3P geometry.[218] Therefore, the binding energies (*BEs*) for the *Fmoc – water* systems are purely interaction energies. The different *Fmoc – water* systems were built in Avogadro.[285] The QM optimized structures were used for the calculation of the MM binding energies. The *Fmoc – water* binding energies (BE_w) are calculated as the energy difference between each *Fmoc – water* system ($E_{water\cdots Fmoc-NH_2}$) and the sum of the internal energies of Fmoc-NH_2 ($E_{Fmoc-NH_2}$) and water (E_{water}) (Equation 5.2).

$$BE_w = E_{water\cdots Fmoc-NH_2} - \left(E_{water} + E_{Fmoc-NH_2}\right) \qquad \text{Equation 5.2}$$

The CHARMM parameterization protocol uses the QM reference binding energies (BE_{QM}) to optimize hydrogen bonds with water.[274, 276-277] However, in this work the binding energies of water with hydrophobic parts, not able to form hydrogen bonds, are also studied. This was done in order to include an additional reference to account for the hydrophilicity/hydrophobicity of the different parts of the moiety. The water – water binding energy ($BE_{w\text{-}w}$) calculated from two water molecules with TIP3P geometry, is used as a reference point to determine the relative strength of the interactions. All BE_w lower than this reference are considered hydrophilic, and those higher than the reference, hydrophobic. Therefore, the Final Charges Set (*FCS*) has to satisfy both conditions: the hydrophobic/hydrophilic behaviour of the different parts of the moiety ($BE_{w\text{-}w}$) and the relative intensities of the interactions (BE_{QM}). The *FCS*

with the final topology and the rest of parameters are presented in the Appendices section (Appendix 2-3).

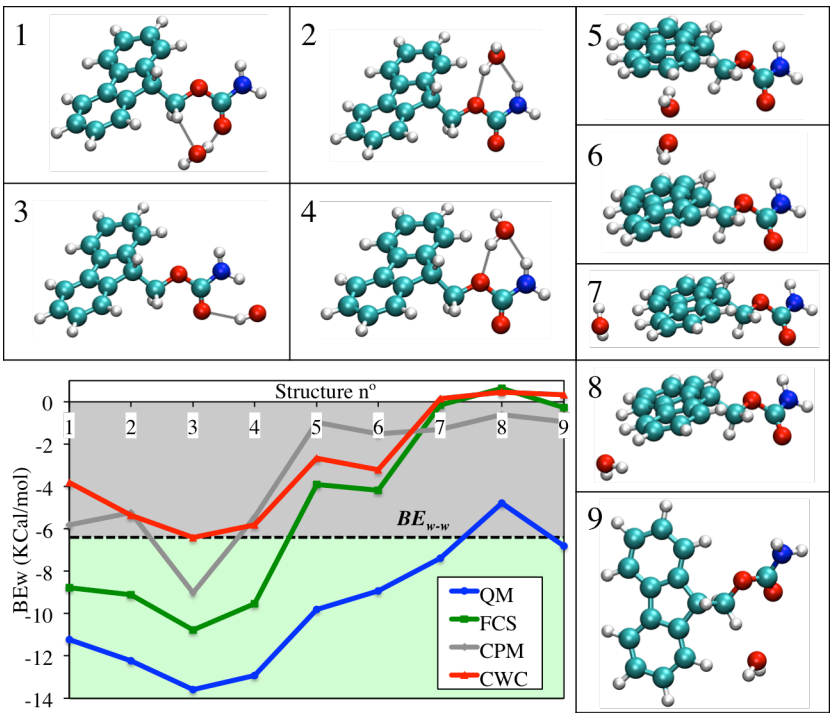


Figure 5.3 Optimized geometries of the *Fmoc – water* systems (1 – 9) and the BE_w graph (bottom left). The graph includes two references: the dashed black line ($BE_{w\text{-}w}$), which separates the hydrophobic (grey) and the hydrophilic (lime) regions; and the BE_{QM} (blue line).

The optimized structures for the *Fmoc – water* are shown in Figure 5.3. The *Fmoc – water* systems were built by situating a water molecule near different parts of the Fmoc-NH_2 molecule. Although up to 15 different structures were built in this way, some of these starting structures optimized to the same final structure. This was the case for a total of three different dispositions of the water molecule on each face of the fluorenyl group which optimized to 5 and 6, respectively, and four disposition on the edge ot this group which evolved to 7 and 8. Structures 1 to 4 involve interactions of the water molecule with the hydrophilic parts of the Fmoc moiety, while in the structures 5 to 9 the water molecule interacts with the hydrophobic parts of the moiety. Therefore the $BE_{w,1\text{-}4}$ are expected to be lower than the $BE_{w\text{-}w}$ (lime area of the graph, Figure 5.3), while the $BE_{w,5\text{-}9}$ must be higher than this reference value (grey area of the graph, Figure 5.3).

The results obtained using the CHARMM protocol method (CPM, Figure 5.3 grey line) are presented to support the method change. The binding energies of the CPM determines that the hydrophobic interactions (5-9) are over 2 kcal/mol. This method underestimates the *BEs* even in the hydrophilic interactions, where three of the four interactions are in the hydrophobic region (Figure 5.3, grey area). The results using the CHARMM website obtained charges (CWC, red) are presented to show that, using these charges, all the intermolecular interactions with water are underestimated, that is three of the four hydrophilic interactions are in the hydrophobic region of the graph (interaction 1 to 4 in the grey area of the plot, Figure 5.3).

Furthermore, the *FCS* successfully reproduces the hydrophilic/hydrophobic behaviour of each part of the Fmoc moiety (Figure 5.3). Furthermore, the *FCS* also reproduces the shape of the BE_{QM} and therefore the relative intensities of the interactions are in good agreement with the QM reference. The interactions 7 to 9, which involve the interaction of the water molecule with the edges of the aromatic group, are very weak or, even, repulsive. However, the relative strength of the interactions is consistent with those calculated at the QM level and the QM method is observed to consistently over stabilise all of the interactions. Therefore, the slightly poorer ability of the force field to reproduce these weak interactions was accepted as a limitation of the force field.

5.2.2 QM and MM Binding Energies: Fmoc – Fmoc Systems

Once the *FCS* was obtained, it was tested against unique interactions that the Fmoc parameters should reproduce to reliably describe the self-assembling behaviour, namely, the *Fmoc – Fmoc* interactions. Various configurations of two Fmoc-NH_2 molecules were built in Avogadro. The systems were optimized in Turbomole using the DFT functional B97-D with the def2-SVP basis set. The optimized geometries were used to calculate the MM energy using the CHARMM force field with the Fmoc parameters in the NAMD program. The binding energies (BE_{dimer}) are defined as the difference in energies between the system with two Fmoc molecules interacting ($E_{Fmoc-NH_2 \cdots Fmoc-NH_2}$) and twice the internal energy of the Fmoc-NH_2 molecule ($E_{Fmoc-NH_2}$) (Equation 5.3). For the *Fmoc – Fmoc* systems the BE_{dimer} are not purely interactions energies but they include the internal energy change. Hence, the *Fmoc –*

Fmoc MM BE presented in Figure 5.4 were calculated using the torsional parameters validated in the following section.

$$BE_{dimer} = E_{Fmoc-NH_2 \cdots Fmoc-NH_2} - \left(2 \cdot E_{Fmoc-NH_2}\right) \qquad \text{Equation 5.3}$$

The MM binding energies calculated using the *FCS* (BE_{dimer}) for these interactions are compared with QM calculated reference binding energies (BE_{QM}). The final geometries and BE are presented in Figure 5.4. Seven different types of π-stacking interactions are studied. Six of the *BE* values are well reproduced as the BE_{dimer} calculated with the *FCS* are within 4.5 kcal/mol from the BE_{QM}.

The only interaction that cannot be reproduced corresponds to the *p_T1* geometry (Figure 5.4). This geometry involves the interactions of the aromatic group of one Fmoc-NH_2 with the hydrophilic part of the other molecule. As in the case of the *Fmoc – water* systems 7 to 9, this is accepted as a limitation of the force field to reproduce some types of interactions. Nonetheless, the ability to successfully reproduce the six π-stacking *Fmoc – Fmoc* interactions is encouraging and suggests the correct balance between the hydrophobic and hydrophilic nature of the moiety has been achieved.

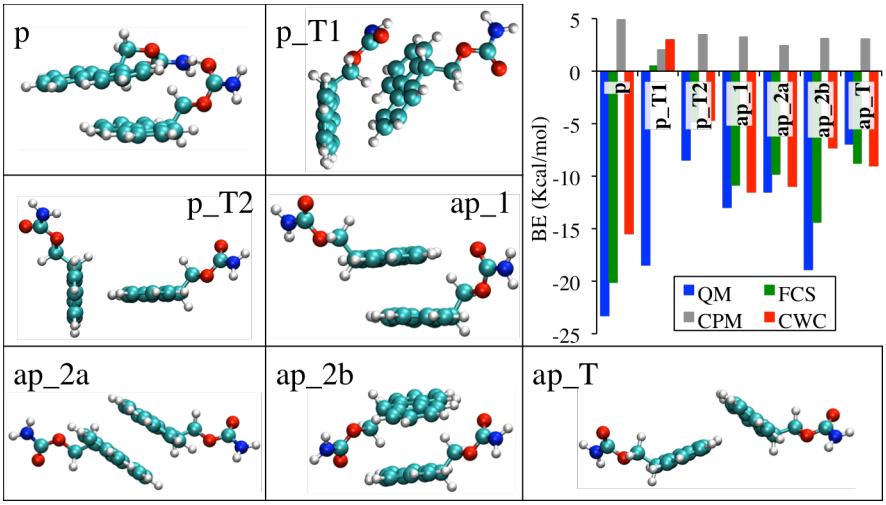


Figure 5.4 Optimized geometries of the *Fmoc – Fmoc* systems and BE_{dimer} graph (top right).

The results obtained using the CPM are included (Figure 5.4) to demonstrate the main reason not to use the CHARMM protocol method for the reference *BEs*. All the *BEs* calculated using the CPM are positive and, hence, repulsive rather than attractive. On the other hand, the CWC is more successful reproducing the *Fmoc –*

Fmoc interactions than the *Fmoc – water*. This charges set gives slightly better results for the structures *ap_1* and *ap_2a* than the FCS. However, the CWC *BEs* are worse than the FCS for the other 5 dispositions, being substantially worse for *p* (by ~5 kcal/mol), for *p_T2* (by ~3 kcal/mol) and for *ap_2b* (by ~7 kcal/mol). This demonstrates the improvement obtained in the optimization of the charges.

5.2.3 Torsion Angle Parameterization

Standard amide and ester dihedral terms were used to build the dihedrals of the linker of the Fmoc. The validity of these terms was evaluated with the *FCS* and compared with reference QM values in order to accurately reproduce the flexibility of the Fmoc moiety. The dihedral angles under study were $C_{13} - C_{12} - C_{F1} - O_{F1}$ (T1), $C_{12} - C_{F1} - O_{F1} - C$ (T2) and $C_{F1} - O_{F1} - C - O_{F2}$ (T3), see Figure 5.1 for atom names. In the case of the torsional potentials, there is no reason not to use the CHARMM protocol suggested method and, hence, it is advisable to keep consistency with the general parameterization of the force field. Therefore, the torsional potential profiles were calculated in Gaussian 09[286] with the B3LYP[287-288] functional and the 6-31G(d,p) basis set.[289-290] The profiles were obtained with rigid scans, calculating the single point energies of structures generated via increments of 10º of rotation around each dihedral. The MM torsional potential profiles were obtained using the same geometries and calculating the energy for the CHARMM force filed with the Fmoc parameters using the NAMD program.

As well as the static MM energies, the QM torsion potentials were compared with the distribution of the three dihedrals during 10 ns MD simulations with a single molecule of Fmoc-NH_2 *(System 1);* and with a single molecule of Fmoc-S-OH *(System 2)*. These simulations reveal the influence of incorporating the parameters of the Fmoc moiety to the parameters of the amino acids in CHARMM. Both systems were simulated in vacuum and in water (TIP3P model). The systems were built in VMD and were minimized with the steep descent technique to avoid high-energy contacts and equilibrated at 298 K before the simulations. The 10 ns simulations were carried out within the NVT ensemble and a 1.0 fs integration time step was applied. Results are shown in Figure 5.5.

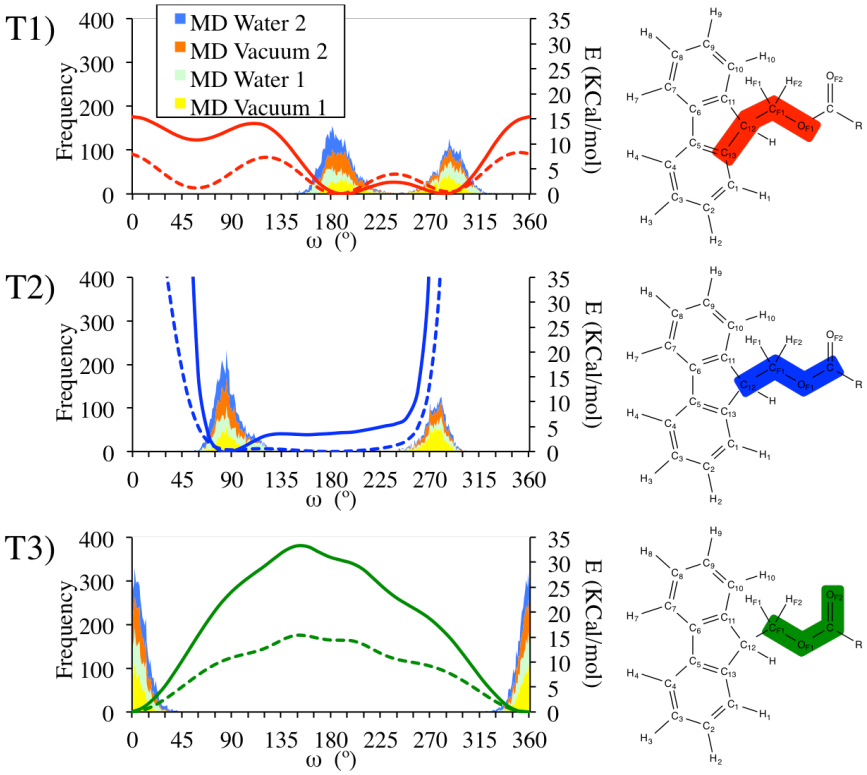


Figure 5.5 Torsion potentials calculated in MM (solid lines) and QM (dashed lines) (Right vertical axis), and torsional angles distribution in MD simulations (Stacked histograms, left vertical axis). The colour-code for the MM and QM results correspond to the images on the right and the colour-code for the histograms is in the legend of the top graph.

The MM potentials successfully reproduce the shape of the QM references (Figure 5.5) although the energy barriers are typically higher relative to the QM gas phase results. The distribution of dihedrals in the simulations maps well on to the QM potentials, with the majority of the torsional space explored around the regions where the potential energy is at a minimum. It can be seen that for the dihedral *T2* both, the MM and the QM barriers are much higher than for the other two dihedrals. This is caused by the proximity of the oxygen in the carbonyl group with the fluorenyl group during the rotation around *T2* for the region between 270º and 45º degrees due to the use of a rigid scan, which cannot relax at each point of the rotation. This also causes a displacement for *T2* in the simulations, which shows a peak between 255º and 300º that is displaced to the right in reference to the energy potentials. This shift is caused by the difference in mobility of the systems used to calculate the different parameters: the histograms are calculated from simulations where the whole molecule can move while the QM and MM plots are calculated by rotating the dihedrals in fixed molecules. Given the agreement between the QM and MM torsional profiles and the

distribution of the torsional angles from the MD simulations, the flexibility in the carbamate region of the Fmoc moiety is considered to be well reproduced.

5.3 Validation via the Partition Coefficient

As an initial validation of the parameterization carried out for the Fmoc moiety, a physical parameter, which can be directly compared with the experimental observable, is calculated. Due to the amphiphilic nature of the molecules for which the Fmoc moiety is parameterized, Fmoc-peptides, and the importance of reproducing this for the self-assembly simulations, the partition coefficient between octanol and water was chosen for this purpose.

The molecule chosen for this validation was Fmoc-S-OMe. This serine is capped at the N-terminus with the Fmoc moiety and at the C-terminus as a methyl ester. Neither of these groups nor the serine side chain contains acidic hydrogen and therefore it is not necessary to take into account any ionization effects, which simplifies the comparison of experimental and computational results. The serine side chain was chosen due to its hydrophilicity, which facilitates the experimental determination of the log P.

The partition coefficient (K_{ow}) is related with the logP and the concentration of the solute in octanol ($[Solute]_{oc}$) and in water ($[Solute]_{w}$) through:

$$logP = log\, K_{ow} = log \frac{[Solute]_{oc}}{[Solute]_{w}} \qquad \text{Equation 5.4}$$

The partition free energy (ΔG_{ow}) is related to the partition coefficient (K_{ow}) using Equation 5.5:

$$\Delta G_{ow} = -2.303\, RT\ log\, K_{ow} \qquad \text{Equation 5.5}$$

The theoretical partition free energy can be calculated from the solvation free energy data of a given molecule in water (ΔG_{w}) and in octanol (ΔG_{oc}) as shown in Equation 5.6:

$$\Delta G_{ow} = \Delta G_{oc} - \Delta G_{w} \qquad \text{Equation 5.6}$$

5.3.1 Experimental Determination

The experimental determination of the Fmoc-S-OMe K_{ow} was carried out using the shake-flask method. [291-293] The concentrations of Fmoc-S-OMe in octanol and water were determined by fluorescence spectroscopy. The emission at 304 nm was calibrated in both solvents (Figure 5.6) using samples of known concentrations.

Emission was measured with a Jasco FP-6500 spectrofluorometer at 304 nm, using medium response with an excitation bandwidth of 3 nm and an emission bandwidth of 3 nm. Samples were measured with 1 cm pathlength cuvettes.

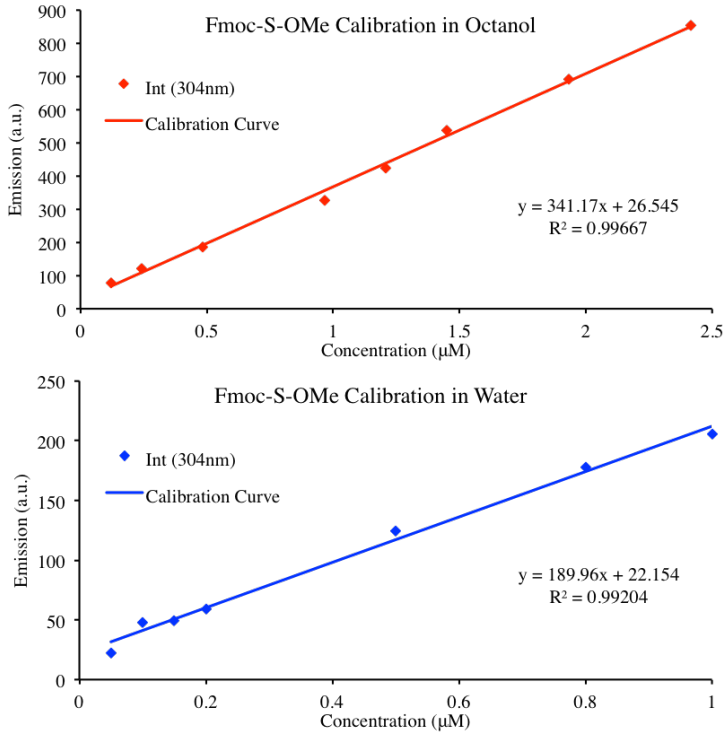


Figure 5.6 Calibration of the emission intensity of Fmoc-S-OMe in octanol (red) and in water (blue) at different concentrations. The range of concentrations was estimated from previous experiments.

Samples for the partition coefficient determination used 2 mL of aqueous Fmoc-S-OMe stock solution (2.01 µM). After mixing them with 2 mL of octanol the mixture was shaken vigorously during 30 s and left resting for 2 hours. 1 mL of each phase was taken to measure fluorescence. Averaged results of 9 samples were used for the final determination of the K_{ow} (Table 5.1).

The final log P, or log K_{ow}, value is -1.4±0.1. The confidence interval was calculated using Equation 5.7:

$$Conf.Interval = \frac{s \cdot t(95\%)}{\sqrt{n}} \qquad \text{Equation 5.7}$$

Where *s* is the standard deviation, *t(95%)* is the confidence coefficient for a 95% of confidence level and *n* is the number of samples used.

Table 5.1. Experimental results for the determination of the Fmoc-S-OMe partition coefficient (K_{ow}).

	Intensity at 304 nm		Concentration (μM)		K_{ow}
	water	**octanol**	**water (x10^{-2})**	**octanol**	
1	40	706	9.49	1.99	21.0
2	36	692	7.26	1.95	26.9
3	35	713	6.85	2.01	29.3
4	39	704	8.78	1.99	22.7
5	35	705	6.65	1.99	29.9
6	42	770	10.4	2.18	21.0
7	35	678	6.52	1.91	29.3
8	45	701	11.8	1.98	16.8
9	47	726	13.1	2.05	15.6
Average			8.97	2.00	23.6
STD Dev			2.39	0.08	5.5

5.3.2 Computational Determination

The theoretical solvation free energies in water (ΔG_w) and in octanol (ΔG_{oc}) were calculated using alchemical methods, specifically free energy perturbation (FEP).[294-295]

The two systems were built in VMD,[296] both include an Fmoc-S-OMe molecule (amino acids parameters from CHARMM27)[274, 297] which is surrounded by TIP3P water[218] for ΔG_w (Figure 5.7a) and by octanol (CHARMM36)[277] for ΔG_{oc} (Figure 5.7b). The systems were built to be 60 x 60 x 60 Å of each solvent with the Fmoc-S-OMe placed in the centre of each phase. All MD calculations were carried out using the NAMD program and our local version of the CHARMM force field, which includes the Fmoc parameterisation presented in this work. The systems were minimized to avoid bad contacts and then heated up and equilibrated at 298 K and 1 atm (NPT ensemble) for 5 ns fixing the position of Fmoc-S-OMe. A 2 fs time step and periodic boundary conditions in the three spatial coordinates[216] were used as well as a 12 Å cut-off for non-bonded interactions. Langevin dynamics were used for the temperature control and Langevin piston Nose-Hoover algorithm was used to keep the

pressure constant.[298] The density of both systems was calculated after this equilibration step to be 0.943 g/ml for water and 0.805 g/ml for the octanol, which sufficiently approximate the experimental values of 1.000 g/ml and 0.824 g/ml, respectively.

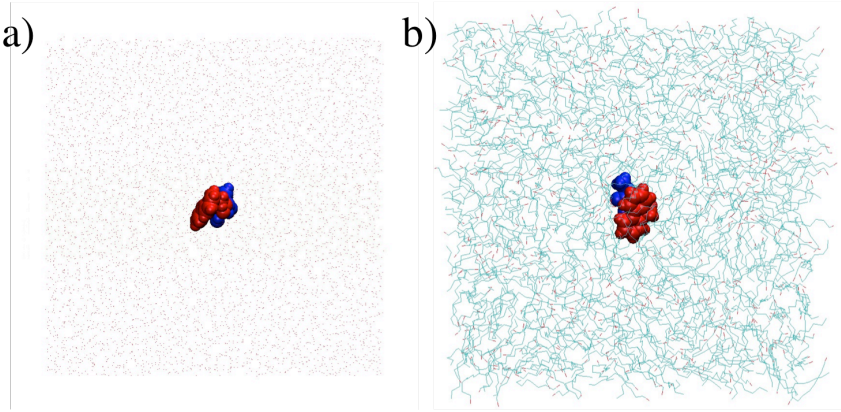


Figure 5.7 Fmoc-S-OMe (Fmoc in red and Ser-OMe in blue) systems in (a) water and (b) octanol. Hydrogens are removed in the solvents for clarity.

The FEP calculations were carried out using the same general MD parameters described before. Reverse FEP is applied, which calculates both the disappearance of Fmoc-S-OMe (from $\lambda=0$, Fmoc-S-OMe in solution, to $\lambda=1$, no Fmoc-S-OMe in solution) and the appearance of Fmoc-S-OMe in each solvent. A soft-core potential[299-300] was applied to avoid end-point problems[301] *($\lambda=0.5$-1)* and simple overlap sampling (SOS)[302] was used to combine both, forwards and reverse simulations. The ParseFEP plugin version 1.9 was used for the error calculation.[303]

Table 5.2 FEP parameters evaluation.

Parameters Set	dλ	Number of windows	Equilibration steps/window	Total steps/window
P-1	0.1	10	2,500	52,500
P-2	0.1	10	30,000	630,000
P-3	0.1	10	500,000	5,500,00
P-4	0.0625	16	250,000	4,250,000

However, prior to the calculation of the final values for the solvation free energies, some of the FEP parameters needed to be optimized. The effect of the

decoupling constant (dλ), which establishes the number of windows in which the calculation is divided, the equilibration steps per window, and the total steps were evaluated (Table 5.2).

To assess the reliability of the results with the different parameters sets the overlap between the forward (solid lines) and backward (dashed lines) simulation was evaluated (Figure 5.8). There is a clear improvement in the overlap as the total number of steps per window is increased from P-1 to P-2. However, when increasing the number of steps per window from P-2 to P-3 (Table 5.2) the improvement in the overlap is minimal (Figure 5.8). However, decreasing the decoupling constant from 0.1 to 0.0625 (Table 5.2) leads to a further improvement as P-4 yields the best overlap between both, the forward and the backward, simulations for the two solvents.

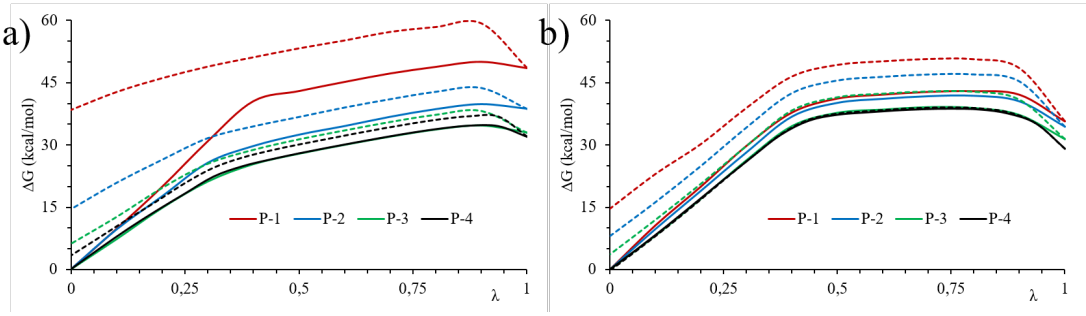


Figure 5.8 Forward (solid lines) and backward (dashed lines) free energy profiles for the solvation of Fmoc-S-OMe (a) in octanol; and (b) in water with different FEP parameters. P-1, P-2, P-3, and P-4 refer to the different parameter sets defined in Table 5.2.

Therefore, the following parameters were used for the decoupling of Fmoc-S-OMe from the solvent and determination of the solvation free energies: a decoupling constant *dλ=0.0625*, giving rise to up to 16 windows in the disappearance and a further 16 windows for the appearance of Fmoc-S-OMe; 250,000 steps (0.5 ns) of equilibration per window; and 4,250,000 (8.5 ns) of simulation per window. The free energy profiles as a function of the decoupling of the Fmoc-S-OMe molecule from each solvent are shown in Figure 5.9. The difference between the starting and the final point gives the solvation free energy of the solute in each solvent: ΔG_{water} is -28.3 ± 0.1 kcal/mol and $\Delta G_{octanol}$ is -29.3 ± 0.3 kcal/mol. This results in a ΔG_{ow} = -1.0 ± 0.2 kcal/mol, which is converted to log P applying Equation 5.5 to yield log P = -0.8 ± 0.2.

The logarithm of the partition coefficients (log P) is used for the comparison between the experimental and the MD-FEP determination. Experimentally, the log P for Fmoc-S-OMe was determined as 1.4 ± 0.1, which compares favourably to the value calculated via MD-FEP (0.8 ± 0.2). The error between the experimental and calculated values ranges from 0.3 to 0.9, which is of a similar magnitude to that obtained for a series of small alkanes reported in the literature.[222] Given the much larger size of the Fmoc-S-OMe molecule (relative to the alkane series) and the amphiphilic nature of the molecules, the level of agreement between the experimental and calculated values is considered to be excellent. Furthermore, the MD-FEP with the developed charges (FCS) shows a significant improvement in comparison with the MD-FEP using the CWC (4.5 ± 0.2, error 2.8 – 3.4), which supports the validity of the parameters optimization process.

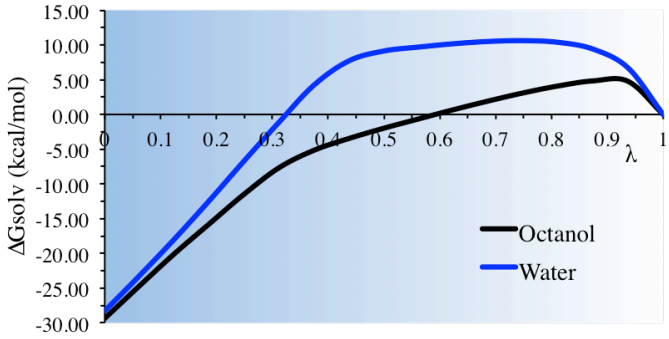


Figure 5.9 Free energy profiles of the decoupling of Fmoc-S-OMe to water and to octanol using MD-FEP.

5.4 Validation via Self-Assembly Simulations

The parameterization of the Fmoc moiety was carried out in order to be able to implement MD simulations of Fmoc-peptides to allow greater insights into the self-assembling process of these molecules as well as of the final structures formed. Therefore, it is critical to validate the Fmoc parameters in a simulation of self-assembling Fmoc containing molecules.

Fmoc-S-OH only contains one amino acid, serine (S), and it is known to form spherical aggregates in aqueous solution.[100] It was chosen due to its simplicity (only one amino acid) and the simplicity of the structures formed, spheres, which can be easily compared with experimental data. The computational results will be compared with the experimental results published by Abul-Haija *et al*. on the aggregation of

Fmoc-S-OH, specifically, the particle size distribution of the AFM image presented in this publication (Figure 5.10).[100]

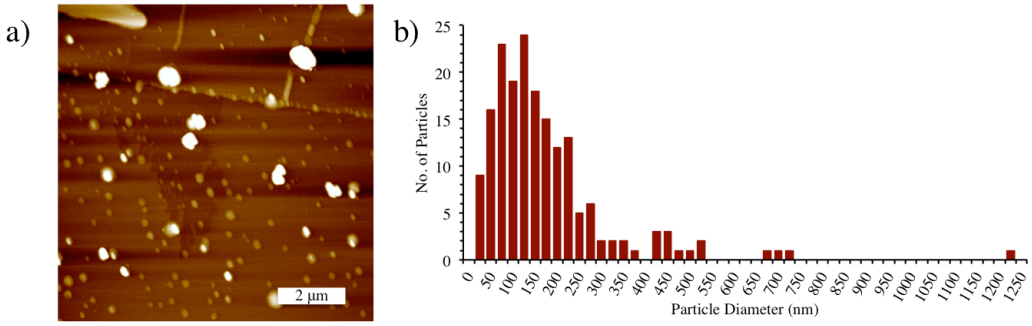


Figure 5.10 (a) Fmoc-S-OH spheres AFM image from reference [100] and (b) particle size distribution analysis.

5.4.1 MD Simulations

The Fmoc-S-OH self-assembling simulation was constructed with 120 randomly distributed Fmoc-S-OH molecules and solvated using VMD[296] (Figure 5.11b, 0 ns) with TIP3P water.[218] Following this, the system was minimized with the steepest descent technique to avoid bad contacts in the starting structure and then gradually (5 K every 1 ps) heated from 0 to 298 K over 60 ps at 1 atm. Following this, the system was simulated for 300 ns in the NPT ensemble (1 atm, 298 K). All other MD parameters were the same as those described in the partition coefficient section. The system, after the heating phase, has a size of ~83 x 84 x 83 Å and, hence, the concentration of Fmoc-S-OH is 0.34 M.

5.4.2 Results and Comparison with Experiments

The 300 ns simulation was analysed using the radial distribution function (RDF) using the C_{12} (see Figure 5.1 for atom names) of the aromatic group in the Fmoc moiety to measure the aggregation and proximity of the aromatic groups at different stages of the simulation (Figure 5.11a). The radial distribution functions for the first 100 ns are calculated by taking snapshots of the systems every 0.01 ns (*i.e.*, 1000 frames per 10 ns), whereas for the distributions from 100 ns – 300 ns are based on snapshots of the system taken every 0.1 ns (*i.e.* 1000 frames per 100 ns). The lower sampling frequency in the latter stages of the simulation was due to the relative stability of the system after 100 ns. A proximity analysis (Figure 5.11c) is also presented, which accounts for the number of groups within 5.5 Å through the

simulation. In the case, of Fmoc – Fmoc proximity, a distance of 5.5 Å between the centres of two aromatic groups is considered as a limiting distance for the possible presence of π-stacking interactions.

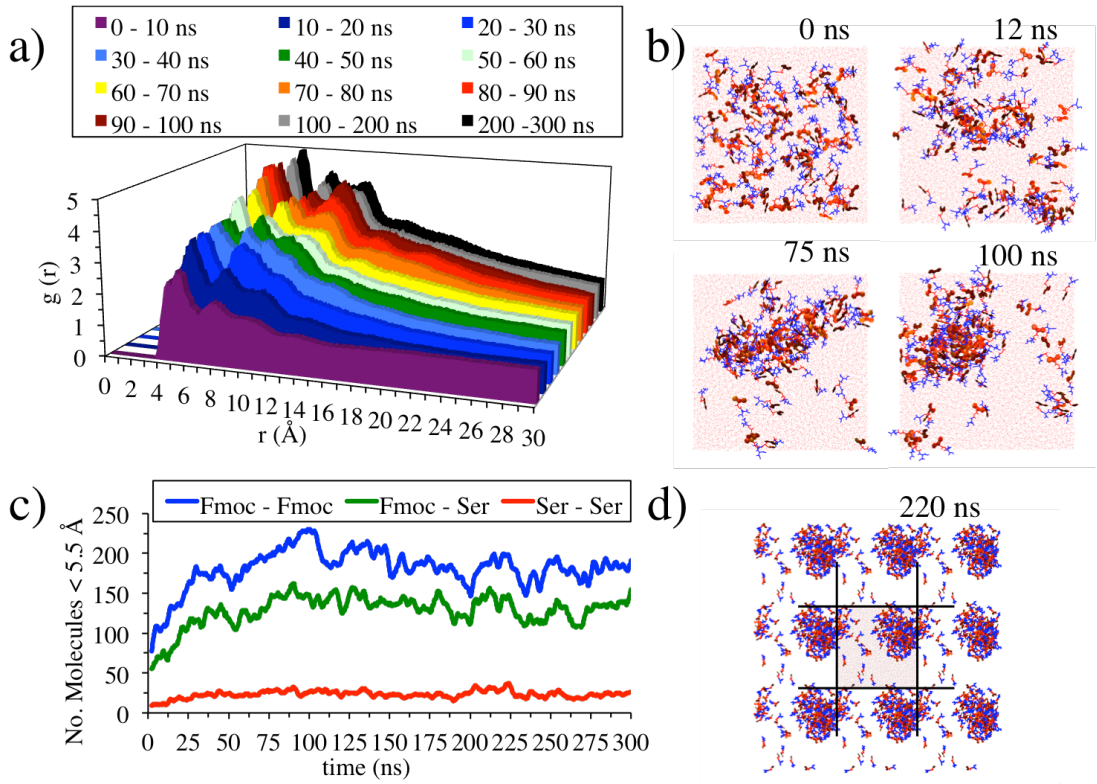


Figure 5.11 Results of the Fmoc-S-OH simulation: (a) RDF analysis, (c) proximity analysis and (b and d) snapshots of the simulation (Fmoc in red and S-OH in blue). (d) The last snapshot shows also the periodic images in the *xy*-plane for clarity.

The RDF analysis (Figure 5.11a) shows three distinctly different sizes of aggregates through the simulation: *peak 1* (3.5 – 7 Å), *peak 2* (7 – 9 Å) and *peak 3* (9 – 12 Å). *Peak 1* corresponds to Fmoc groups in direct contact. This peak increases rapidly at the beginning of the simulation (20 – 50 ns) and represents an early stage of the process where molecules form small aggregates. This initial peak reaches a plateau while *peak 2* and *peak 3* increase due to the junction of the small aggregates (*e.g.*, Figure 5.11b, 12 ns). After ~50 ns, these latter peaks also begin to plateau and *peak 1* starts to increase further, which corresponds to the formation of a single large aggregate (Figure 5.11b, 75 ns). After ~80 ns the variations in the radial distribution function are minimal and the overall shape and magnitude of the distribution remains consistent through the 100 – 200 ns and 200 – 300 ns plots, suggesting that the system is equilibrated following the formation of a spherical aggregate (Figure 5.11b, 100ns).

A fourth and a fifth peak around 12 – 16 Å and 16 – 20 Å, respectively, is also observed in the equilibrated parts of the graph.

The higher *peak 3* between 80 ns and 100 ns in comparison with the 100 – 300 ns results could be due to an elongation of the aggregated spheres to an ellipse-like aggregate. However, as the simulation continues and is averaged over longer timescales, this elongation effect is no longer present.

The proximity analysis (Figure 5.11c) shows a rapid increase in the Fmoc – Fmoc proximity as well as the Fmoc – Ser proximity, which is likely due to the arrangement achieved through the Fmoc π-stacking interactions. The proximity analysis reveals how the molecules aggregate quickly in the first 25 ns to relatively stable structures, until ~50 ns, where the aggregation is seen to increase further. This is in good agreement with the RDF analysis, which shows an early aggregation step at the same time of the simulation. Once the small aggregates are formed they coalesce to form larger aggregates (50 – 100 ns). Therefore, the Fmoc – Fmoc interactions remain constant while *peak 2* of the RDF increases (Figure 5.11a). This suggests that this increase is not due to Fmoc interactions, which are mostly buried in the small aggregates, but once these small aggregates join, the Fmoc groups of the aggregates interact inside the larger aggregate (as *peak 1* in the RDF at 50 – 100 ns). The highest level of Fmoc – Fmoc aggregation is reached around 100 ns. This high number of interactions slightly decreases, in a process of equilibrating the aggregate, and remains relatively unchanged, apart from normal fluctuations, at ~180 Fmoc - Fmoc interactions (~1.5 interactions/molecule) from 110 – 300 ns forming an aggregate around 40 Å.

The AFM image of Fmoc-S-OH published by Abul-Haija *et al*.[100] shows many small aggregates (50 – 250 nm) and only a few larger ones (450 – 550, 700 – 750 and 1250 nm), but not aggregates of intermediate size (Figure 5.10). While care should be taken with the interpretation of AFM images because of possible drying effects influencing the size distribution, the mechanism observed in the simulation reveals how these larger aggregates are formed when many small aggregates coalesce, which is consistent with this experimental observation. Furthermore, the simulation is also in good agreement with the shape of the aggregates, which are mostly spherical, with a limited number showing a degree of ellipticity. However, in addition to this data, the simulation of the Fmoc-S-OH system also provides insight into which interactions are

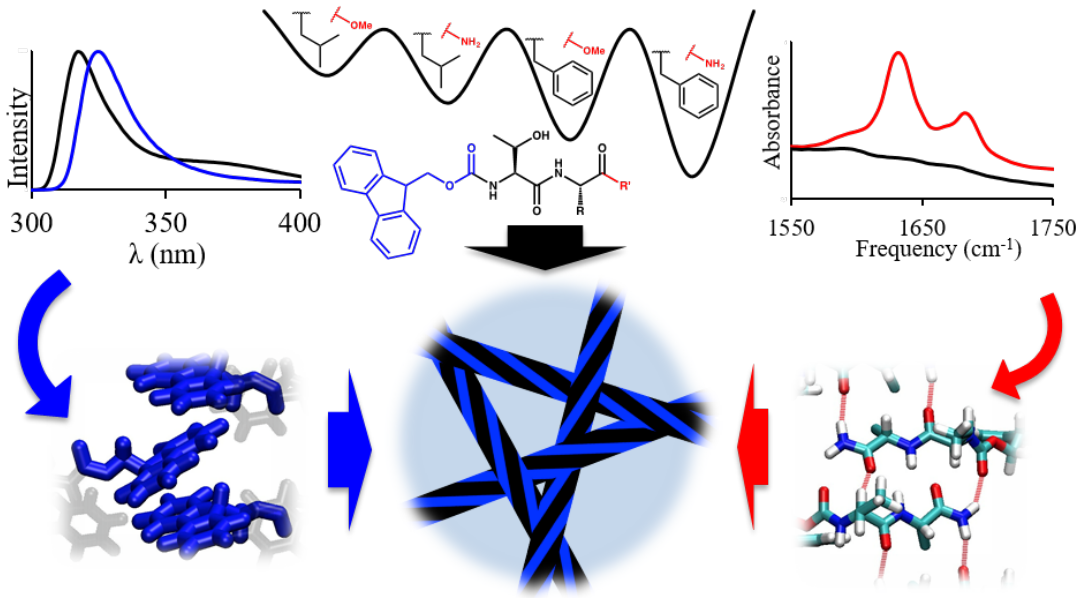
driving the different stages of the process. That is, the initial formation of the small aggregates is driven by the π-stacking interactions of the Fmoc moieties, followed by the H-bonding interactions between the Ser-OH moieties which allow the small aggregates to come together to finally be stabilized through reordering of the π-stacking between the Fmoc moieties. Therefore, in addition to the simulation providing good agreement with the experimental results it also is able to contribute unique insights into the molecular level information, which is otherwise inaccessible experimentally.

5.5 Conclusions

A set of parameters for the Fmoc moiety has been derived and validated for the CHARMM force field. Therefore, it is now possible to run reliable simulations for self-assembling Fmoc-dipeptide based systems as argued in Chapter 4 and which are detailed in Chapter 7.

This set of parameters was obtained by adapting the general CHARMM parameterization protocol to one specifically for amphiphilic molecules. The parameters are shown to be able to reproduce intermolecular interactions by reproducing QM binding energies between the moiety and water and between two Fmoc moieties, and to reproduce the flexibility of the Fmoc group by comparing dihedral distributions in MD simulations and their MM energy profiles with the corresponding QM dihedral scans. Furthermore, the Fmoc parameterization presented in this chapter has successfully reproduced thermodynamic parameters directly related to the self-assembling behaviour and experimental results involving self-assembling of the moiety linked to an amino acid. The validity of the parameters supports the modifications made in the protocol for this type of molecule and suggest a new procedure for the future parameterization of other amphiphilic moieties.

6 The Experimental Test: Dynamics Peptide Libraries to Understand Fmoc-Dipeptide Nanostructures



6.1 Introduction

The influence of peptide sequence in Fmoc-dipeptides has been shown to be important even for small variations in the amino acid side chain (Figure 6.1, black). Hughes *et al*. reported the surprising transition from sheets to tubes, and from fibres to twisted ribbons in Fmoc-dipeptide methyl esters when changing the first amino acid from serine (S) to threonine (T), which involves the addition of a methyl group and creation of an extra chiral centre in the amino acid side chain.[19] However, even for more significant changes, such as the L/F[19, 42] or G (glycine)/F[26] substitution effect, the influence and relative importance of the interactions that result in the various structures formed are evident, but the underlying changes in supramolecular interactions are not well understood.

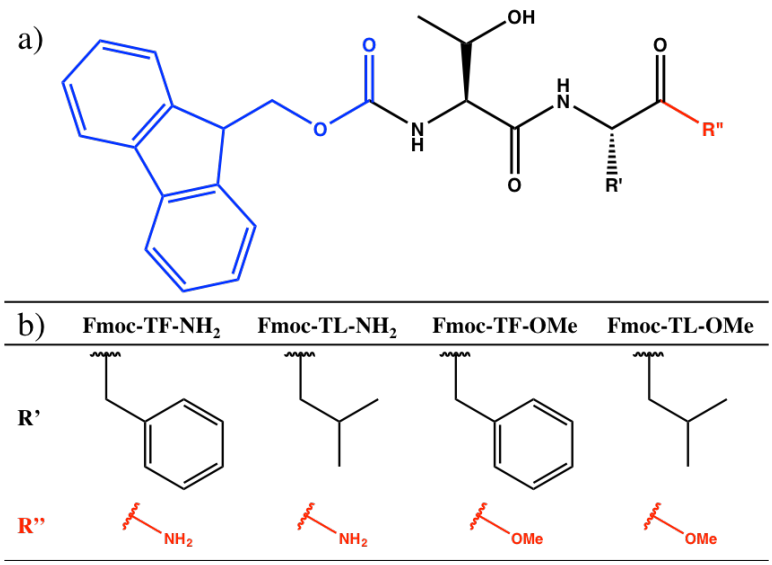


Figure 6.1 (a) General structure of the Fmoc-dipeptides of this study. The Fmoc group is shown in blue, the dipeptide in black and the C-terminus group in red. (b) Amino acid side chains (R') and C-terminus (R") for the four Fmoc-dipeptides under study.

The modification of the C-terminus (Figure 6.1, red) has also been found to have an influence on the self-assembled structure of Fmoc-dipeptides due to both, the formation of further interactions and the modification of the hydrophobic/hydrophilic balance of the building block.[28] Ryan *et al*. demonstrated the importance of this contribution to the self-assembling tendency of fluorinated Fmoc-F by modifying the C-terminus.[304] In this case, those compounds with either the carboxylate or the

amidated C-terminus were able to form gels, while those containing a methyl ester C-terminus could not.

Although the influence of the peptide sequence and C-terminus in the nanostructure has been extensively demonstrated, it is still not clear how these modifications affect the structure at the intermolecular level. One key issue is the difficulty of the interpretation of spectroscopic data, which are established for proteins and oligopeptides, but do not directly map onto those for such small peptides, as discussed in Chapter 3. It is now clear that simple extrapolation is not accurate.[46] Furthermore, the comparison with crystal structures obtained from dried systems is problematic as it is doubtful whether the structure obtained in this state is directly comparable with the gel state (where the system is ≥ 90% water).[264]

As demonstrated in Chapter 4, the relative strength of the interactions between the building blocks, and between the building blocks and the solvent, determine the thermodynamic tendency of a low molecular weight gelator (LMWG) to self-assemble into a nanostructure. However, it is known that many self-assembled systems are kinetically, rather than thermodynamically, driven.[254, 258-260] That is, the nanostructure is due to the formation of a kinetically trapped state and not the formation of the thermodynamically preferred interactions. Often, even in systems where the self-assembly is thermodynamically driven, there are kinetic effects which can lead to structures that differ from the thermodynamically favoured one.[132, 158-159] Therefore, it is important to carefully (and ideally reversibly) control the self-assembly process to ensure thermodynamic control over the formation of the nanostructure.

The control of the self-assembly process has been the objective of different studies.[36, 68, 131-132, 158-159, 179, 267] Thermodynamic control is usually ensured by annealing (ideally through slow heat-cool cycles)[153] or by observing long (sometimes up to weeks) assembly times to ensure thermodynamic minima are reached.[305] However, an alternative way to achieve thermodynamic control is to produce the self-assembling building blocks *in situ* using fully reversible reactions.

Self-assembling Fmoc-dipeptides have previously been synthesized from the non-assembling precursors of an Fmoc-protected amino acid and a C-terminus protected amino acid through an enzyme catalysed direct condensation reaction. [19, 31, 172-173] The coupling of two processes, the condensation, which involves a small

energetic penalty, and the self-assembly, which is thermodynamically favoured, makes the system reversible, and hence, results in thermodynamic control of the process.[172]

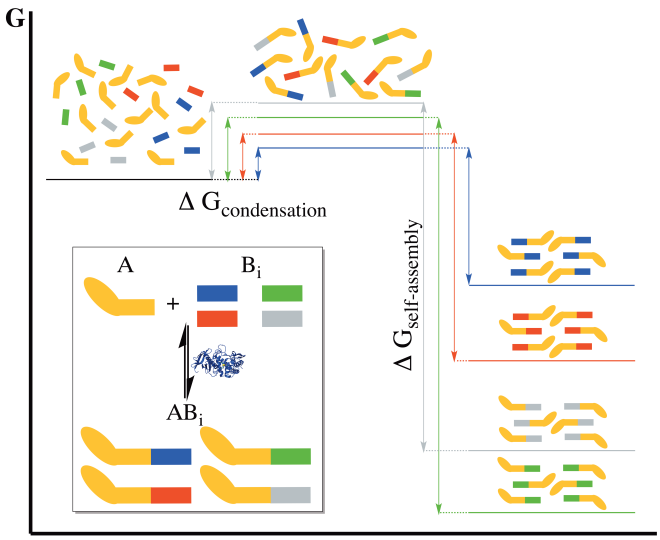


Figure 6.2 Energetic diagram for DPLs. The inset shows the condensation reaction of the precursor A (orange) with four possible precursors B_i (blue, green, red or grey) to form the self-assembling building blocks AB_i in the presence of an enzyme.

The thermodynamically driven formation of nanostructures has been exploited in recent years to carry out competing experiments involving many molecules. These system are known as dynamic combinatorial libraries (DCLs) or, in this case, dynamic peptide libraries (DPLs). In these experiments, one precursor A (an amino acid derivative, such as an Fmoc-amino acid) is placed in a reaction mixture with different precursors B_i (for example, a range of amino acid amides) that can form different self-assembling building blocks AB_i upon enzymatic condensation of A and B (forming an amide bond) (Figure 6.2). Due to the reversibility of the system where the self-assembly is the step which involves the highest change in energy, and hence, is the driving force, the competitive condensation occurs in percentages which are proportional to the self-assembly tendency.[39-40, 51, 109, 172, 306-309]

The fact that DPLs produce a relative ordering of the thermodynamic stability of systems implies that, for the closely related systems studied in this work, this type of information can be used to provide insight into the intermolecular interactions that lead to these changes in stability (*i.e.*, the features that favour self-assembly). Moreover, this insight can be combined with spectroscopic characterisation to allow

the systematic analysis of how the differences between C-terminal ester and amide, as well as the F/L substitution, influence the types of nanostructures formed.

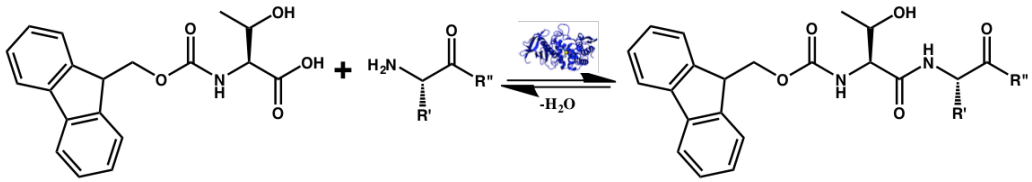


Figure 6.3 Scheme of the condensation reaction between Fmoc-T and a C-protected amino acid (R'-R") *via* reverse hydrolysis in the presence of thermolysin to form the Fmoc-dipeptide, Fmoc-TR'-R".

In this chapter, four Fmoc-dipeptides that form nanostructures, which results in a gel-phase macroscopic appearance, are studied: Fmoc-TF-NH_2, Fmoc-TF-OMe, Fmoc-TL-NH_2 and Fmoc-TL-OMe (Figure 6.1). These materials are formed enzymatically by the condensation of two precursors: Fmoc-T and R'-R" (R'=F or L; R"=NH_2 or OMe; Figure 6.3).[19, 172] All four molecules have the Fmoc group and the first amino acid (T) in common, and the only variations are the phenylalanine/leucine substitution (F/L) and the C-terminus substitution (amidated, NH_2/methyl ester, OMe). In this way, the question of how the amino acid sequence affects the structures is decomposed to two independent questions: what is the role of aromatic *vs* aliphatic amino acid side chains? What is the effect of the C-terminus substitution?

6.2 Materials and Methods

Enzymatic Gel Formation and DPL. Gels were formed catalytically using thermolysin from bacillus thermoproteolyticus rokko (Sigma) in phosphate buffer (100 mM, pH=8) by mixing Fmoc-T (Aldrich) at 20 mM with F-NH_2, L-NH_2 (BACHEM), F-OMe or L-OMe (Aldrich) at 40 mM. For the DPL experiment all were mixed at those concentrations.

HPLC. Gels were vortexed before taking the aliquots. Samples of 30 μL were taken at the different time points and diluted in 1 mL MeCN:H_2O (1:1) with 0.1 % of TFA. Measurements were carried on a Dionex P680 HPLC with a Macherey-Nagel C18 column. Flow rate of 1 mL/min and a MeCN:H_2O mixture as eluting solvent. UV absorption at 280 nm, Fmoc absorption, was used for detection.

FT-IR. IR spectra were recorded on a Bruker optics Vertex 70 spectrometer. Samples were placed between two CaF_2 windows with a 5 μm separation. Spectra

were the result of an average over 25 scans with a resolution of 1 cm^{-1}. Samples were measured after 48 h of equilibration and prepared in 1 mL of phosphate buffer 100 mM pH=8 in deuterated water.

Circular Dichroism. CD spectra were measured on a Jasco J600 spectropolarimeter in a cylindrical cell of 0.1 mm pathlength. Spectra were recorded with step resolution of 1 nm, response of 1 s, bandwidth of 1.0 nm and speed of 100 nm/min. CD spectra were measured between 200 and 350 nm, but the region between 200 and 230 was discarded because the High Tension (HT) voltage signal reaches values close to the maximum and, therefore, the CD signal is unreliable in that region. Samples were prepared and directly added to the cell – the enzyme is added with a pipette when the samples were still liquid and so the gel is formed *in situ*.

Fluorescence. Emission spectra were recorded on a Jasco FP-6500 spectrofluorometer at low response with a 5 nm bandwidth, a 0.5 nm data pitch and a scanning speed of 500 nm/min. Spectra were measured between 300 nm and 600 nm with an excitation wavelength of 285 nm.

TEM. All carbon coated grids were purchased from Electron Microscopy Sciences (EMS) and glow discharged in a 0.39 mbar air atmosphere for 30 s using PELCO easiGlow (Ted Pella, Inc.) before use. A drop of 5 μl of sample solution was applied to the glow-discharged continuous carbon coated grid and incubated for 1 min. Excess solution was removed by blotting the grid with a piece of filter paper, followed by two rounds of staining for 30 sec with 5 μl 2% (w/v) uranyl acetate solution. After blotting excess stain solution the grid was left for air-dry. The negatively stained sample was imaged in an FEI TITAN Halo TEM operating at 300 kV. Cryo-EM grids were prepared in an FEI Vitrobot at 21 °C with the relative humidity set to 100% and the blotting force set to 0 N. 3 μl of sample was pipetted onto a freshly glow-discharged lacey carbon grid. The sample solution was incubated on an EM grid for 15 s, blotted for 4.5 s before being plunged into liquid ethane that was pre-cooled by liquid nitrogen. The cryo-EM grids were then transferred to and stored in liquid nitrogen. The cryo-EM grids were transferred in liquid nitrogen into a Gatan 626 cryo-specimen holder and then inserted into the microscope. The specimen temperature was maintained at −170 °C during data collection. Cryo-EM imaging was performed in an FEI TITAN Halo TEM operating at 300 kV. Both negatively stained

and cryo-EM images were recorded in the low-dose mode (20 $e^{-}/Å^{2}$) on an FEI CETA 16M camera (4,096 × 4,096 pixels).

6.3 Results and Discussion

The four gelators are able to form self-consistent gels within 4 hours after the addition of the enzyme even though the yields of Fmoc-dipeptide formation are relatively low at the concentrations tested (Figure 6.4). It can be seen that the conversion rate and final yield of formation differs between the four molecules already in the pure systems: Fmoc-TF-OMe (~88%), Fmoc-TF-NH_2 (~81%), Fmoc-TL-OMe (~70%) and Fmoc-TL-NH_2 (50%). The latter may not have reached equilibrium within the 48 h of the experiment. It has to be taken into account that at %Y = 100 the gels would be 1 wt%, therefore all these species form gels bellow this concentration. The molecules containing F show the highest rates, which is consistent with the known preference of thermolysin for phenylalanine at the amine end of the forming peptide bond.[173] The macroscopic appearance of the four gels differs in their levels of transparency (Figure 6.4 insets). The images show that both the inclusion of the aromatic residue F and the methyl ester C-terminus result in more opaque gels: Fmoc-TL-NH_2 (Figure 6.4b) is the most transparent gel and Fmoc-TF-OMe is the most opaque (Figure 6.4c).

The experiment was then repeated in a competitive DPL setup, involving the same concentrations as in the previous systems (20 mM Fmoc-T and 40 mM of each C-protected amino acid: F-NH_2, L-NH_2, F-OMe, and L-OMe). To ensure thermodynamic equilibrium was reached, the experiment was extended for 45 days (Figure 6.4e) and the gel was vortexed each time when aliquots were sampled. The yields observed for the DPL experiment (Figure 6.4e): Fmoc-TF-NH_2 (~70%), Fmoc-TF-OMe (~18%), Fmoc-TL-NH_2 (~2%) and Fmoc-TL-OMe (~1%); differ significantly to those obtained in the isolated systems, which suggests that the percentage yields obtained from the pure systems cannot always be used to assess the relatively stabilities of the nanostructures, as has been done previously.[19]

Although the difference between Fmoc-TL-OMe and Fmoc-TL-NH_2 is relatively small, the higher propensity of the latter one to self-assembly was confirmed with another experiment in which at time 0 there were only Fmoc-T and L-OMe, and L-NH_2 was added after 48 h (Figure 6.4f). This experiment showed how

Fmoc-TL-OMe is produced in the absence of other amino acid, but once L-NH_2 is added Fmoc-TL-OMe is hydrolysed and Fmoc-TL-NH_2 is the favoured product.

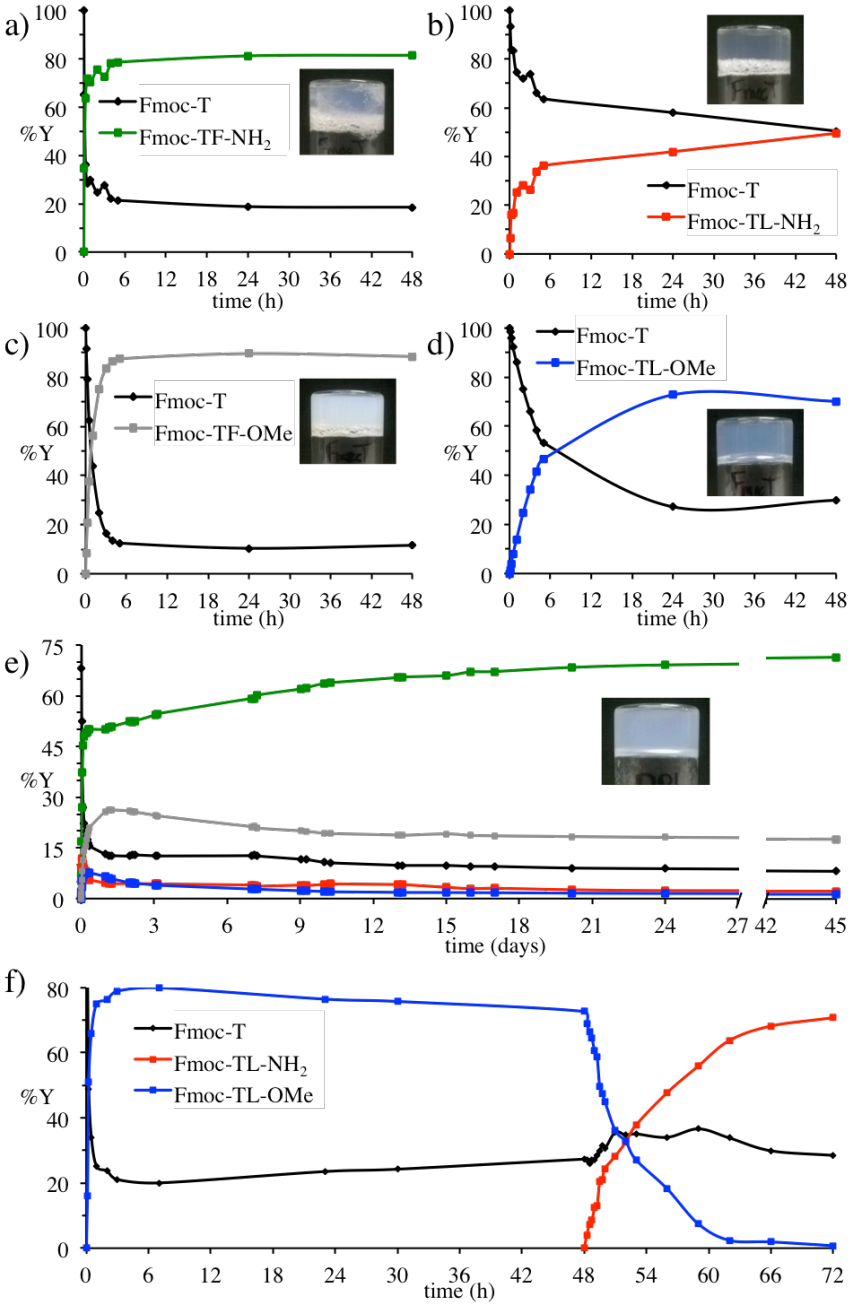


Figure 6.4 Conversion of the Fmoc-dipeptides with time for (a) Fmoc-TF-NH_2; (b) Fmoc-TL-NH_2; (c) Fmoc-TF-OMe; (d) Fmoc-TL-OMe; (e) for the DPL; and for the competition between Fmoc-TL-NH_2 and Fmoc-TL-OMe. Inset into graphs a-e is the picture of the related gel after 6 hours of the addition of the enzyme.

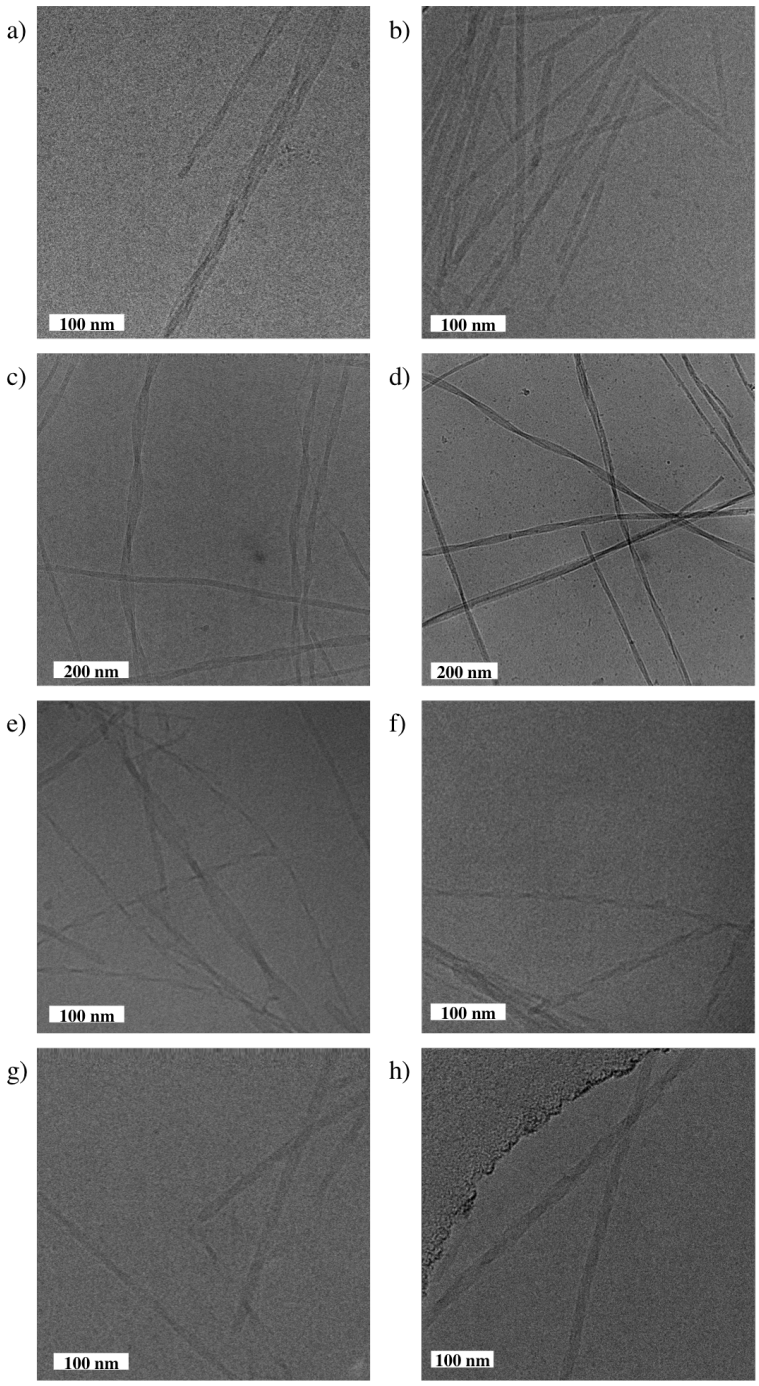


Figure 6.5 Cryo-TEM images for (a-b) Fmoc-TF-NH_2; (c-d) Fmoc-TF-OMe; (e-f) Fmoc-TL-NH_2; and (g-h) Fmoc-TL-OMe.

The DPL results (Figure 6.4e) show Fmoc-TF-NH_2 has the highest yield, which is four times higher than that of the second most stable product (Fmoc-TF-OMe). The preference for the TF dipeptides over TL indicates that the presence of the aromatic group is the main differentiating effect driving the self-assembly, which can either be due to the extra hydrophobicity, or to the extra π-stacking interactions, or, most likely, to a combination of both effects. The last two Fmoc-dipeptides (Fmoc-

TL-NH_2 and Fmoc-TL-OMe) are only formed in negligible amounts in the competition experiment. The secondary effect that governs the relative ranking of the systems with the same peptide unit is the presence of the amide group, which is likely to be due to the extra hydrogen bonds that this group is able to form upon self-assembly.

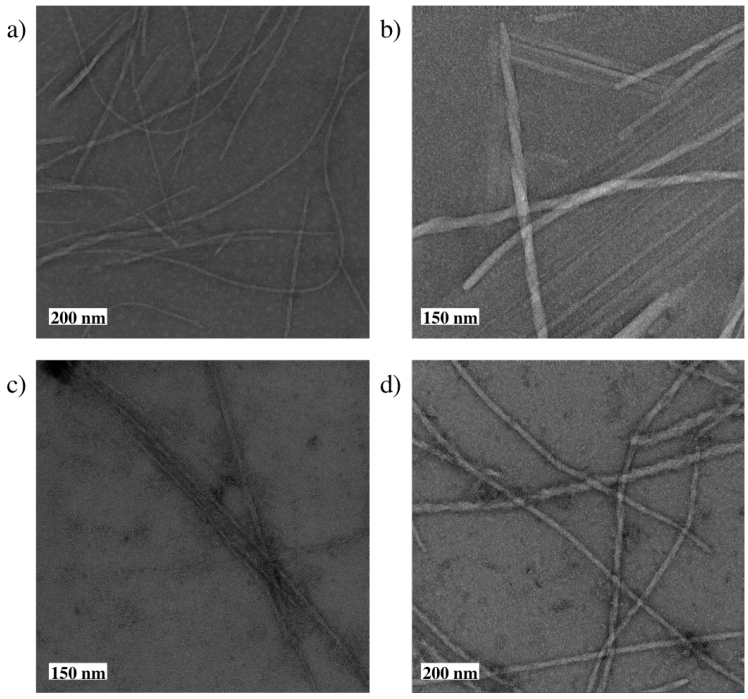


Figure 6.6 TEM images for (a) Fmoc-TF-NH_2: (b) Fmoc-TF-OMe; (c) Fmoc-TL-NH_2; and (d) Fmoc-TL-OMe.

Having established the relative stability of the Fmoc-dipeptides by DPL, they were subsequently characterized separately. The nanostructures formed by the four Fmoc-dipeptides are visually similar in that they have a twisted ribbon appearance with roughly equivalent widths (~20 nm in Figure 6.5a; ~15-50 nm in Figure 6.5c, ~10-30 nm in Figure 6.5e; and ~20 nm in Figure 6.5g). One image also shows a fibre (~10 nm) which evolves into a twisted ribbon (Figure 6.5c middle right). Some images show longitudinal patterns in the ribbons, which suggest that the final ribbons are formed by the fibres lateral aggregation (~10 nm in Figure 6.5a; ~20 nm in Figure 6.5b; ~20 nm in Figure 6.5d; ~10 nm normal TEM Figure 6.6a and d). The similarity between these structures is interesting because, as mentioned, for other systems, even small changes in the peptide resulted in completely different nanostructures.[19, 26, 42] In

the systems studied here, the changes in supramolecular organization and consequent structure formation appear to be more subtle.

The minimal effects on nanoscopic structure suggest that the changes in the self-assembly tendencies observed in the DPL may be explained by different intermolecular interactions, which provides a unique opportunity to obtain information about these structures at the intermolecular level.

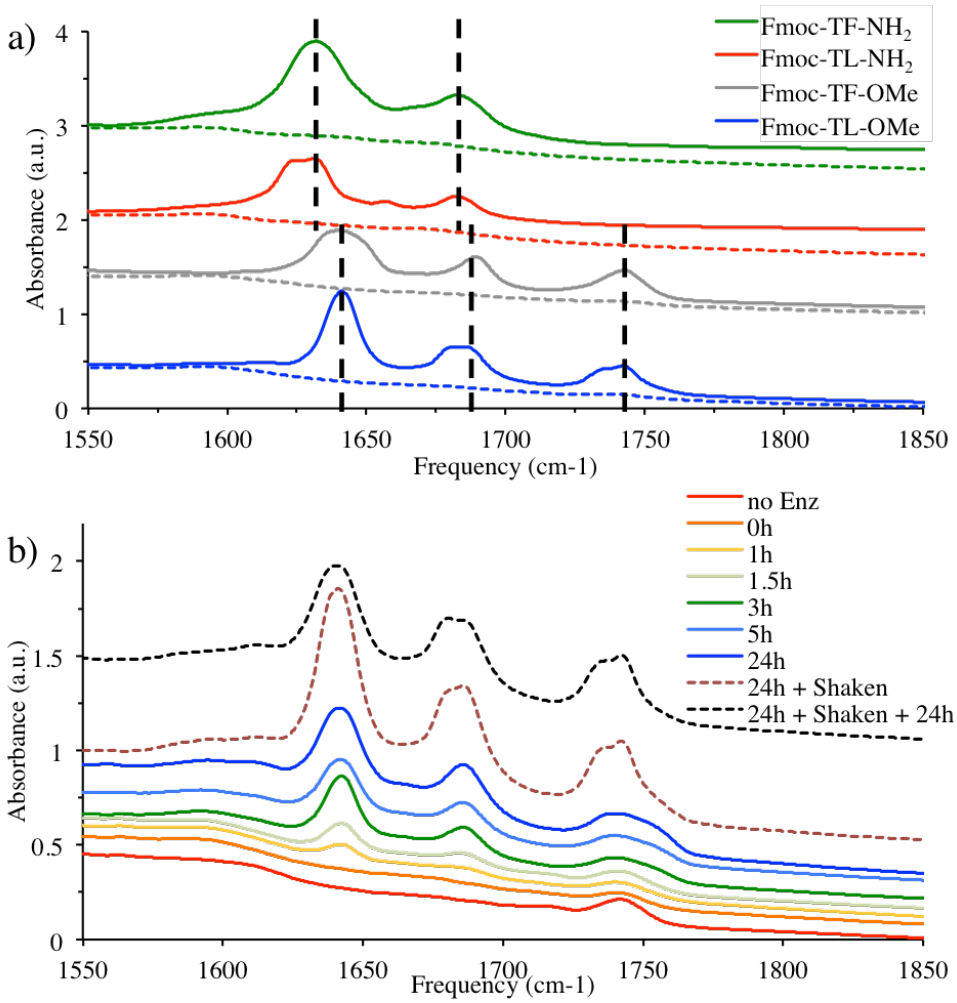


Figure 6.7 (a) FT-IR of the four Fmoc-dipeptide gels (solid lines) and their precursors before the addition of the enzyme (dashed lines). (b) FT-IR spectra evolution with time for the system Fmoc-T + L-OMe with thermolysin added at time 0h. After the formation, the gel was vortexed and two more spectra were taken then and 24 h after (dashed lines).

To understand the intermolecular arrangements in peptides, it is useful to investigate H-bond networks by analysing the FT-IR spectra in the region 1600-1800 cm^{-1}. As mentioned, this interpretation cannot be straightforwardly extrapolated from proteins or oligopeptides to small peptide systems.[46] Hence, the positions of the IR

peaks do not give direct conformational information. However, the presence of clear peaks in this region can still be used to assess the presence of dominant hydrogen bonded motifs, as it is the case for these systems (Figure 6.7a); and the comparison of these peaks for closely related systems can be used to study variations on these motifs through the relative shifts between the related systems.

All the systems show the peaks for the amide (1620-1650 cm^{-1}) and the carbamate (1675-1690 cm^{-1}) vibrations (Figure 6.7a), which indicate the presence of extended hydrogen bonding networks.[30, 38, 41, 46] Beyond distinction between carbamate and amide carbonyls, the results do not provide a clear resolution of which groups or residues are interacting through hydrogen bonds, it is clear that there are differences in the networks between the molecules with different C-termini. As can be expected, the –OMe gelators have an extra peak around 1745 cm^{-1} due to the vibration of the carbonyl in the terminal ester group (Figure 6.7a grey and blue), but also the other peaks show differences depending on the C-terminus group. The amide peak for the amidated Fmoc-dipeptides is around 1630 cm^{-1} and around 1640 cm^{-1} for the methyl ester molecules; while the carbamate is around 1682 cm^{-1} for the NH_2 and around 1687 cm^{-1} for the OMe. Therefore, the amide peak is shifted ~10 cm^{-1} and the carbamate ~5 cm^{-1} higher for the nanostructures containing the methyl ester group. This shift to lower frequencies indicates a better coupling of the carbonyl vibrations in the amidated Fmoc-dipeptide. The better coupling in the NH_2 containing nanostructures can be because the vibration of this terminal group is closer in frequency to the other carbonyl containing vibrations. A more ordered network due to the extra hydrogen bond donor could also explain the better coupling. However, it is most likely due to a combination of both effects and without further information it is difficult to discriminate between the two.

Therefore, the differences in the FT-IR, and hence, in the hydrogen bonds cannot explain the preference observed in the DPL for F containing molecules. However, the information obtained does offer an insight into the secondary preference, *i.e.,* the higher stability of the nanostructures formed by amidated Fmoc-dipeptides through the ability to form a larger H-bond network.

The FT-IR spectra obtained from gels containing L (Figure 6.7a, red and blue lines) show broadened peaks compared to the F peptides that appear as double peaks: the amide vibration peak of the Fmoc-TL-NH_2 (red line, Figure 6.7a) and the

carbamate and methyl ester group vibrations of the Fmoc-TL-OMe (blue line, Figure 6.7a). A time course experiment was carried out to check the nature of these broadened peaks using Fmoc-TL-OMe (Figure 6.7b). In the time evolution of the spectrum for the formation of Fmoc-TL-OMe these extra peaks do not appear, but when the gel is broken by vortexing the sample and reformed again, introducing kinetic trapped states, these peaks arise. These peaks are of similar intensity to those attributed to the supramolecular hydrogen bond networks, which means that they are also hydrogen bonded, albeit in a different arrangement. Therefore, these broadened peaks can be attributed to defects in the structure, most likely caused by the disruption of the gel during the sampling. The thermodynamically driven nature of the DPL approach followed here should ultimately give rise to the correction of suboptimal arrangements over time. This is not observed during 24 hours and it is proposed that the free energy difference between the two assembled states of this Fmoc-peptide-ester is minimal. Another reason for the absence of thermodynamic healing is that enzymatic diffusion is expected to be slow in the gel phase and therefore can greatly reduce the time over which such healing would take place.

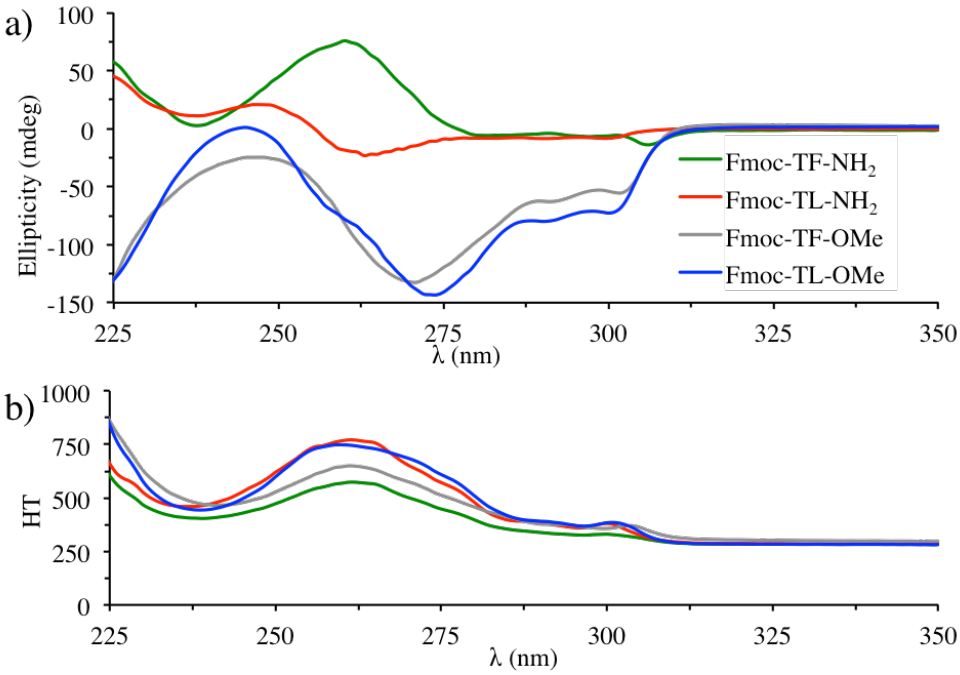


Figure 6.8 (a) Circular Dichroism and (b) the corresponding HT plots.

As discussed in the literature review, the fluorenyl group does not contain any stereo centre, and hence the CD signal shown in the fluorenyl absorption region (270-

310 nm) is due to the supramolecular ellipticity originating from the formation of a chiral nanostructure (Figure 6.8a), as it was indeed observed by cryo-TEM.[25, 36, 44, 51, 198] The CD spectra of the four systems all show a degree of supramolecular ellipticity and, in addition, this ellipticity is higher in the systems with a methyl ester C-terminus. The ellipticity shows dependence of the C-terminus substitution, suggesting a correlation between the hydrogen bonding and ellipticity of the structure. Although this correlation could be expected based on the model presented in 2008 by Smith *et al.*,[30] more recent studies have argued that the highly ordered networks predicted in this model are not realistic.[248-249] However, these studies have not been able to rationalize the existence of the supramolecular ellipticity, which causes the observed CD intensity.

The CD signal that is observed between 250 and 275 nm is due to the phenylalanine, which explains the differences between the spectra of the two amidated Fmoc-dipeptides. The difference between the F and L containing methyl ester Fmoc-dipeptides is not clear because it is masked by the high fluorenyl signal. As F is not present in all the molecules, it is not a relevant diagnostic for the comparison of the four systems.

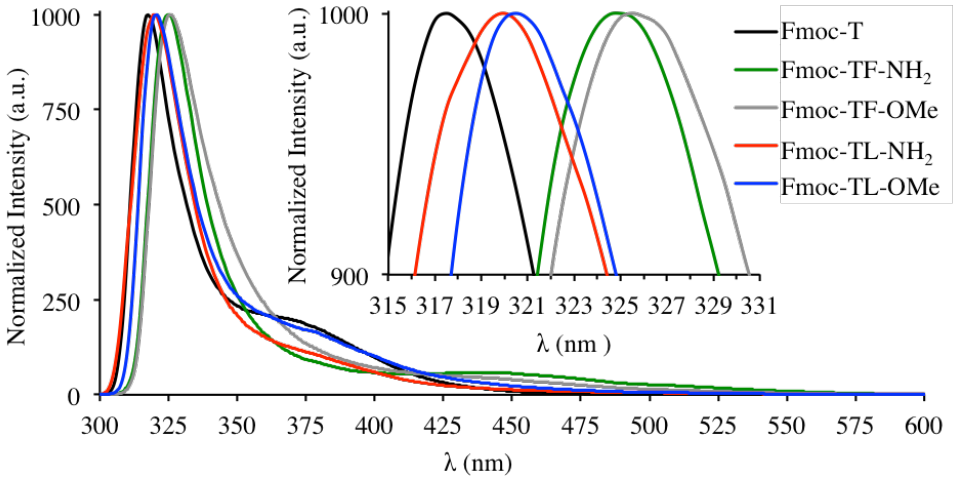


Figure 6.9 Fluorescence results for the four Fmoc-dipeptides. All spectra are normalized to the maximum intensity for clarity. The inset shows a magnified view of the peaks around the λ_{max}.

The fluorescence λ_{max} red shift has been previously related to the formation of π-stacking interactions.[19, 26, 30, 38, 41, 98, 100, 123, 174, 204, 310] In this study, to address and compare the four gels, the λ_{max} for each system is measured and compared to the λ_{max}

of the Fmoc-T precursor (317.5 nm) as a reference (Figure 6.9). It can be seen that the maximum red shift is shown by Fmoc-TF-OMe, 8nm, and the second, by Fmoc-TF-NH_2, 7.5 nm. The red shift for Fmoc-TL-OMe and Fmoc-TL-NH_2 is clearly lower, 3 nm and 2.5 nm, respectively. This suggest that the presence of F strengthens the Fmoc – Fmoc π-stacking, or introduces extra phenyl – Fmoc interactions. The improvement of the Fmoc π-stacking due to the presence of F has been addressed before.[19, 26] The second structural feature which shows an effect on the π-stacking is the presence of the methyl ester C-terminus. Although this effect is small (0.5 nm), it is consistent in the systems and opposite to the effect of the higher hydrophobicity of the environment compared to the amidated Fmoc-dipeptides.

The fluorenyl emission spectra shows other peaks which are commonly used to asses changes in the supramolecular structure, which are the peaks at 365 nm and at 450 nm (Figure 6.9). The first peak is due to the fluorescence of the excimer and the latter one has previously been ascribed to phosphorescence of the excimer. Structurally, the presence of the 365 nm peak is attributed to the presence of micelle Fmoc arrangements,[19, 26, 30, 38, 100, 174] which is supported in this study by the presence of this peak for Fmoc-T (Figure 6.9 black line). Although some of the emissions of the gels also show this peak (especially Fmoc-TL-OMe, Figure 6.9 blue line), which is likely due to the fact that the conversion of this system only reaches ~70%, and hence, there is still 30% of free Fmoc-T molecules which can form micelles. The structural meaning of the triplet emission is still not clear, however it only appears upon the formation of a nanostructure (although not all the Fmoc based nanostructures show this peak).[19, 26, 30, 38, 100, 174]

The fluorescence results can explain the preference, observed in the DPL experiments, for F-containing molecules over L peptides to self-assemble. That is, the improvement of the π-stacking interactions between the Fmoc moieties is a key component in stabilizing these systems, relative to the T-containing molecules. Nevertheless, the preference could also be related to the higher hydrophobicity of the F amino acid. The secondary tendency observed in the fluorescence, the slight preference for OMe molecules, directly contrasts the observed ordering in the DPL. However, the clear difference between amidated and methyl ester molecules observed in the FT-IR data is consistent with the DPL results and overall it can be considered a more robust measure of the influence of the terminal group. Although the higher

hydrophobicity of F (log P=1.25) compared with L (log P=0.91)[311] is expected to give rise to a higher self-assembling tendency, the results show that in the C-terminus substitution the trends are opposite to the hydrophobicity increment. Hence the other intermolecular interactions play a significant role, where the improvement in the π-stacking is the main contributor governing the relative stability of the self-assembled structures, however, the hydrogen bonds are relevant as the extra amide hydrogen bonds compensate the small improvement that the methyl ester group contributes to the π-stacking interaction.

6.4 Conclusions

In this chapter, the effect of F/L and the NH_2/OMe substitutions in self-assembling Fmoc-dipeptides have been used to investigate how the aromatic side chain and the C-terminus substitution influence the self-assembly tendency. The use of catalytically driven self-assembly by reversible amide condensation ensures that the thermodynamically favoured supramolecular structure is obtained from the library of potential structures that could be formed. DPLs were applied to rank the self-assembling tendency based on the differences in the primary structure features. This ranking was then used to gain more insight into the results obtained from cryo-TEM and traditional spectroscopic characterization methods.

The DPL revealed that Fmoc-dipeptides containing F and an amidated C-terminus have the highest tendency to self-assemble into nanostructures, with the presence of the aromatic side chain being the dominant effect.

The cryo-TEM images showed similar types of twisted ribbons nanostructures for the four Fmoc-dipeptides, which are suspected to form by lateral aggregation of thin fibres. The similarity between the nanostructures formed by molecules with different primary structure is rare and is a distinct advantage in terms of being able to analyse the intermolecular level structures rather than effects on the nanoscale structure influenced by the possible differences in thermodynamic preference for 2D or 1D objects.

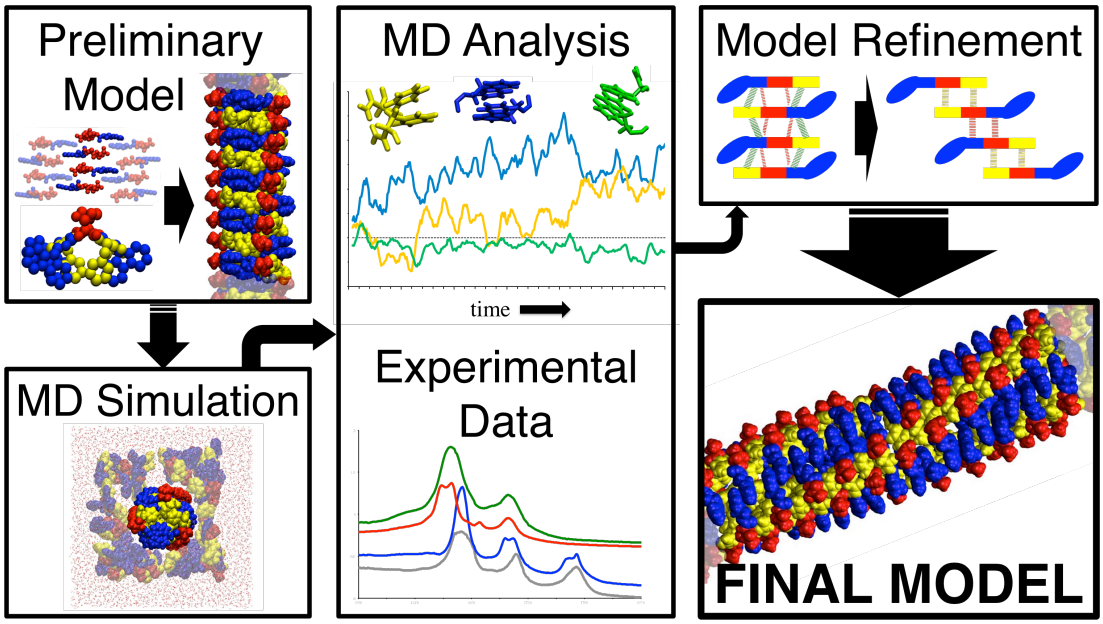
Therefore, to better understand the supramolecular structure and the DPL results, the different gels were characterized using spectroscopic techniques. The FT-IR showed the presence of H-bond networks that varied with the C-terminus substitution. The CD showed similar trends for the supramolecular ellipticity, indicating a correlation between H-bonding and supramolecular ellipticity, which is in

good agreement with a model for Fmoc-FF based on β-sheet like stacks.[30] The H-bonding differences with the C-terminus can explain the secondary preference, shown in the DPL, of amidated molecules to self-assemble, due to the ability to form extra hydrogen bonds.

The fluorescence evidences better Fmoc π-stacking in molecules containing F, which explains why this feature favours the self-assembly, as shown in the DPL. The fact that the improvement of the π-stacking interactions is more important than the hydrogen bonding in the DPL might suggest that the former are more important than the latter. In addition, the higher hydrophobicity of F is also expected to aid the self-assembly. The π-stacking is slightly improved by the methyl ester terminal group, but the additional stability provided by the NH_2 terminal group through the extra H-bonds available overrides this effect. This demonstrates the importance of both, the π-stacking interactions and hydrogen bonds in the formation of the supramolecular structure and how small changes in chemical structure can have significant effects on the stability of self-assembling systems.

In conclusion, spectroscopic data can be interpreted more accurately when comparing results of closely related molecules as opposed to studying the systems separately. The relative shifts in closely related systems act as sensitive antennae, which, when combined with the results from the DPL studies, provide insights into the interactions occurring at the molecular level of these extended nanostructures.

7 The Final Proof: Iterative Model Development for Fmoc-Dipeptide Nanostructures: Fmoc-TF-NH_2



7.1 Introduction

Although experimental methods can provide information about how changes in the Fmoc-dipeptide structure (*i.e.*, the amino acids side chains and C-terminus moiety) affects the final nanostructure and self-assembly tendency,[25, 30, 36, 38, 41-42, 44, 46, 51, 98, 167, 198, 204] they cannot provide detailed intermolecular information, due to the averaged nature of the techniques.[46, 198] Therefore, it is still difficult to establish the real disposition of the molecules when forming the nanostructures and hence, both the exact role of each part of the molecule in the interactions in addition to the actual orientation of the chemical groups in the final nanostructure are unknown. Moreover, as this relationship between the Fmoc-dipeptide chemical structure and the final nanostructure remain poorly understood (and disputed) it is currently not possible to carry out predictions or design to develop materials for given purposes.

Self-assembly is an exergonically driven process, to which intermolecular interactions are key in defining structures, but with critical contributions from the entropy, due to the changes in the order of the system that occurs with particular entropic contributions due to water expulsion as a result of hydrophobic interactions.[104-105] The structural changes in Low Molecular Weight Gelators (LMWG), like Fmoc-dipeptides, are a function of the self-assembling properties of the molecules. Molecular dynamic (MD) simulations have been widely applied to biomolecules in order to understand the process and structural changes that happen at the intermolecular level and are difficult to observe directly using experimental methods.[47, 205-207] Therefore, due to the similarities between the systems and processes previously studied with MD, it is reasonable to use this method to access the detailed information for the self-assembly of APAs into nanostructures.

APA based nanostructures have been studied before using MD simulations, which further supports the validity of these methods for these kind of systems.[5, 45, 248-251] These studies use MD simulations to test and compare the relative stability of different types of starting structures. Some of these studies used a small number of molecules in different arrangements and compared the stability of the arrangements using structure based parameters to define each conformation and observing the evolution of these parameters as a function of time. [5, 45, 250-251] For example, using the distance and angle between the centres of mass of the peptides, in Fmoc-peptides self-assembly, to define β-sheet-like arrangements.[251] The interpretation of results of this

type of systems is not straightforward, because in such small systems (3 – 6 molecules) there are other effects, resulting from the limited size of the model, which have to be taken into account. For example, some of these studies overestimate the stability of the parallel arrangements as this arrangement is formed by one stack of Fmoc-peptides and for such a small system this is the only arrangement that show Fmoc π-stacking.

Other MD studies of Fmoc protected peptide self-assembly use simulations of entire fibres.[5, 248-251] These systems are less affected by the limited size of the simulation and compare the relative stability of different starting points, *i.e.*, different arrangements of Fmoc-dipeptides in the fibres. However, building a fibre as a starting point introduces inherent bias. Structural detail is required and there are a number of possibilities, each of which could represent a local minimum on the potential energy surface in which the structure becomes trapped. Therefore, it is possible to get erroneous results due to choosing the wrong starting structure arbitrarily and using a method that is unable to effectively map the potential energy surface. Clearly the use of an arbitrary starting structure has drawbacks in these types of simulations, as such, it is important to use experimental information to constrain the choice of starting structures and guide the computational studies of APAs self-assembly.

An additional problem when studying APAs self-assembly in MD is that, although the amino acids and peptides are parameterized and highly validated in different force fields, the aromatic moieties are usually not. This was also the case of the Fmoc moiety. Previous studies used parameters based on building the Fmoc moiety from different similar segments, which is commonly done. However, these 'additive' approaches do not include proper evaluation and adaptation of the charges to mimic the intermolecular interactions – both with other solute molecules and also with the solvent, which are critical in self-assembly.[248-249] That is, in addition to the starting structure, the validity of the force field used will also strongly influence the quality of the results obtained.

The CHARMM force field is chosen to study Fmoc-dipeptides self-assembly. This force field has been widely applied and validated for the study of peptide based systems and based on the work described in Chapter 5, it also includes parameters for the Fmoc moiety.

In this study, a model is developed based on the experimental information detailed in Chapter 6 for the Fmoc-dipeptide which was shown to be the thermodynamically most favoured self-assembling molecule of that study: Fmoc-TF-NH_2. This preliminary model is simulated using the modified CHARMM force field (*i.e.*, our in house developed version that contains the parameters of the Fmoc moiety). The information provided by the analysis of the simulation of the preliminary model is used to refine it and develop a new model, which is then subjected to a new MD simulation to test the stability of the new model. The analysis of this final simulation is compared with that of the first model and with the experimental data (Chapter 6), in order to confirm the stability of the composed structure and its ability to conform to the experimentally known facts of the system. Therefore, the final model includes validated and detailed information of the Fmoc-dipeptide intermolecular arrangements and their interactions in an Fmoc-TF-NH_2 fibre. This refinement approach reduces the possibility of producing wrong results due to an incorrect starting structure. Finally, the final model is applied to improve the understanding on an experimental observation of these systems in the cryo-TEM images of Chapter 6 (Figure 6.5), which is the formation of twisted ribbons from fibres.

7.2 Methods

Structures: The models were built in Avogadro and the final system for simulation was built using the genbox GROMACS plugin[312] and VMD.[296] The systems were solvated in VMD using TIP3P water.[218]

The fibre position and size are modified inside the box to take advantage of the periodic boundary conditions (PBC) through the *z*-direction. The final number of Fmoc-TF-NH_2 molecules that compose the fibre (27), which determines the fibre size, was adjusted to fit in the PBC box employed and to ensure an effectively infinite fibre along the *z*-direction of the box. Additional, randomly distributed Fmoc-TF-NH_2 molecules (33) are added to the simulation media. The total number of molecules is 60. These extra molecules help to minimize the effect of the limited size of the system in destabilizing the simulations in the *xy*-directions, as it was observed in previous systems, which included only the fibre molecules (*i.e.*, with no randomly distributed molecules), while the use of the PBC to extend the periodicity of the fibre achieves the same result in the *z*-direction. The extra molecules also allow the investigation of

the role of the Fmoc-dipeptides that remain in solution, *i.e.*, more accurately representing the dynamic nature of a self-assembling system.

Simulations: Calculations were carried out in NAMD[284] using the CHARMM force field[274-275, 277] and the parameterization of the Fmoc moiety presented in Chapter 10 (Appendix 2-3). All the systems were minimized using the steepest descent algorithm for 1,000 steps, in order to avoid bad contacts at the beginning of the simulations. Following this, the systems were gradually heated (0 – 298 K) at a rate of 5 K/ps and equilibrated at 298 K, for a total of 240 ps. Systems with empty space within the fibre (Model 2 and bilayer, see below) were heated and equilibrated for a total of 1.5 ns constraining the C=O and the N-H groups (responsible for H-bonding) in the *z*-coordinate. Finally, the production simulations were carried out on the systems for 150 ns, unless otherwise stated, within the isothermal – isobaric (NPT) ensemble at 298 K and 1 atm. The temperature and the pressure were kept constant using Langevin dynamics and the Langevin Nose-Hoover algorithm, respectively.[298] A 1 fs time step was used to integrate Newton's equation of motion and a 12 Å cut-off was applied for the non-bonded interactions. Periodic boundary conditions (PBC) were applied in the three spatial coordinates.[216]

Analyses: Hydrogen bond analyses were carried out using the VMD package with a 3.5 Å cut-off distance and a 30º of cut-off angle. Backbone hydrogen bonds were measured every 0.5 ns between the different residues. π-stacking interaction were measured by counting the number of aromatic groups closer than 4.5 Å. The potential Fmoc – Fmoc, F – F and Fmoc – F π-stacking interactions were measured every 0.5 ns.

7.3 Results And Discussions

7.3.1 Preliminary Model (Model 1) Development

As mentioned, the preliminary model is based on the structural information obtained in the previous chapter and previous publications,[28] which showed that both hydrogen bonds and π-stacking interactions are important in the formation of these chiral nanostructures. With this information and the idea of maximizing attractive interactions and minimizing repulsive forces between the Fmoc-TF-NH_2 molecules (Figure 7.1a-b) different arrangements were built (Figure 7.1). Firstly, the stacks of molecules were built by maximizing the number of hydrogen bonds (Figure 7.1c-d)

and two different options were found to be possible at this stage depending on the relative orientation of the peptide chain:

- Antiparallel (*ap*): where the Fmoc groups point to alternated sides (Figure 7.1c).
- Parallel (*p*): where all the Fmoc groups point to the same side (Figure 7.1d).

Both these options have been previously proposed in the literature for different Fmoc-peptide and Fmoc-amino acid structures.[30, 304]

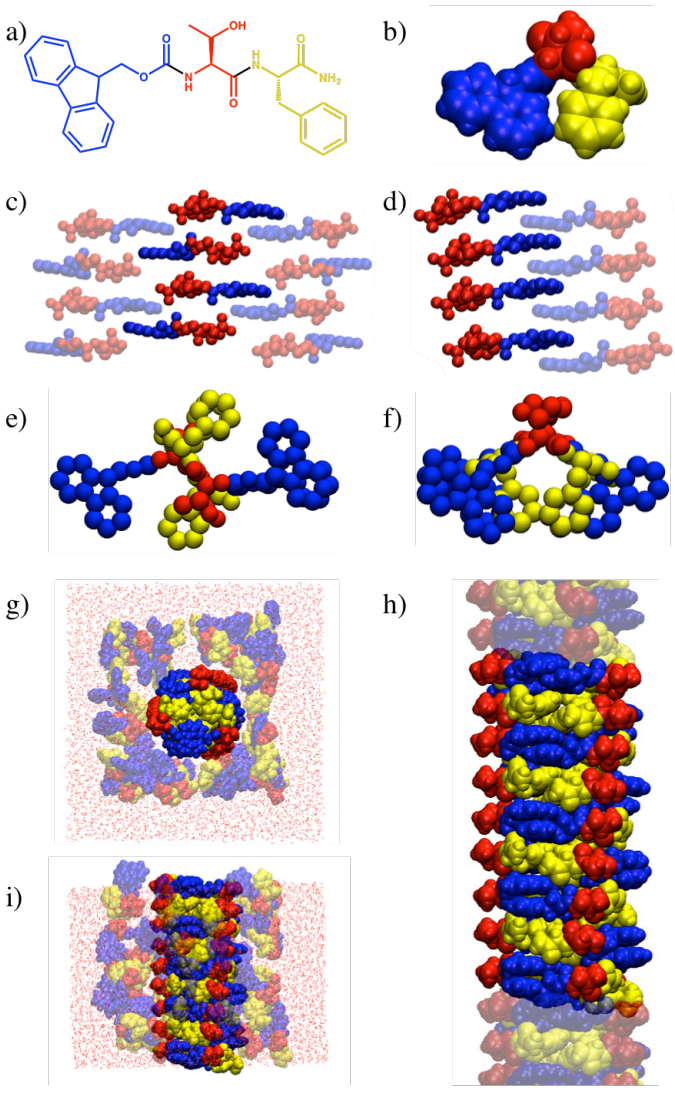


Figure 7.1 (a) Fmoc-TF-NH_2 structure and (b) 3D van der Waals representation. Front view of (c) antiparallel; and (d) parallel arrangements. Top view of (e) antiparallel 1 and (f) antiparallel 2 arrangements. Fibre Model 1: in water box (semi-transparent water) with the extra randomly distributed molecules (semi-transparent), (g) top; (i) side view; and (h) fibre model with periodic images (transparent) along the *z*-direction. (a-b, g-h) Fmoc in blue, T in red and F in yellow. (c-d) Fmoc groups in blue and peptide chain in red.

The *ap* stack was found to be more amenable to extending the structure in two directions (Figure 7.1c), while the *p* configuration only allows extended structures in one direction (Figure 7.1d). However, the *ap* configuration, two potential conformations that differ in the vertical orientation of the Fmoc-dipeptides and changes the position of the side chains:

- Antiparallel 1 (*ap1*): where all the molecules have the same vertical orientation and the side chains point to alternate faces of the stack (Figure 7.1e).
- Antiparallel 2 (*ap2*): where the molecules have alternating vertical orientations with every second side chains pointing to the same face of the stack (Figure 7.1f).

The nature of the building block used has to be taken into account to decide between these two orientations. Fmoc-TF-NH_2 is formed by one amino acid with a hydrophilic side chain (threonine, T, red in Figure 7.1a-b) and another amino acid with a hydrophobic side chain (phenylalanine, F; yellow Figure 7.1a-b). The *ap2* disposition was selected as it situates all the hydrophilic side chains (T) on one face of the stack while all the hydrophobic side chains (F) are situated on the other face (Figure 7.1f). This results in the positioning of the F side chains in the core of the fibre, instead of being stacked with the Fmoc groups. The relative position of the F side chains is the main difference between the model proposed by Smith *et al*. for the Fmoc-FF-OH fibres.[30] Nonetheless, these arrangements are most closely related to those proposed by Smith *et al*.[30] for Fmoc-FF-OH and clearly different from the hydrogen bonded free fibres presented by Mu *et al*. for Fmoc-AA-OH.[248]

Once the fibre is built by connecting three stacks through Fmoc – Fmoc π-stacking interactions, this conformation is shown to be consistent with the importance of the hydrophobic effect in the self-assembly of these molecules to form nanostructures because it situates all the hydrophobic side chains in the core of the fibre, exposing the hydrophilic side chains to the solvent on the surface of the fibre (Figure 7.1g). Furthermore, this conformation allows the F side chains to be close enough to build extra π-stacking interactions both, between the phenyl groups and with the fluorenyl groups of the Fmoc-moieties as also proposed by Smith *et al*. This disposition of the F side chains forming π-stacking interactions would explain the preference shown in the DPL, in Chapter 6. Moreover, the interactions with the Fmoc groups could also explain the better π-stacking interactions evidenced with the

fluorescence spectroscopy in F containing Fmoc-dipeptides. The fibre model and final simulation system is shown in Figure 7.1 (Model 1 in h, simulation system in g and i).

7.3.2 Model 1 Simulation and Analysis

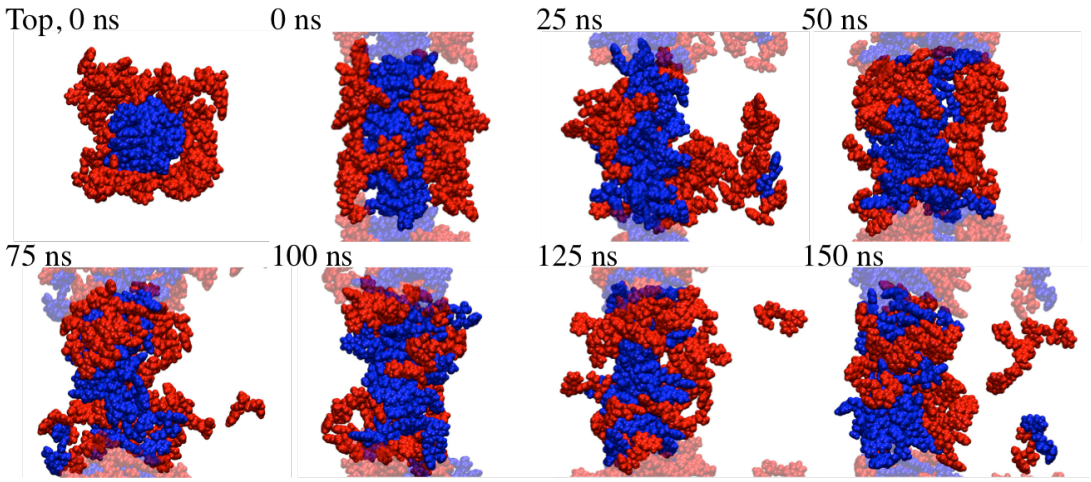


Figure 7.2 Snapshots of the Model 1 simulation. Fmoc-TF-NH_2 molecules in the fibre arrangement are shown in blue and the randomly distributed extra molecules in red. Periodic images through the *z*-direction are semi-transparent. All images are shown form the side with the starting structure also shown from the top (Top, 0 ns).

The snapshots of the simulation show how the fibre has lost its ordered structure after 25 ns, although the overall structure remains elongated and can be considered a uni-directional fibre of infinite length owing to the connections through the PBC (Figure 7.2, blue). The overall aggregated structure is maintained through the simulation due to the additional Fmoc-TF-NH_2 molecules (Figure 7.2, red) used in the simulation. This validates the use of the extra randomly organized molecules for the study of these systems, making the simulations relatively more stable than observed in previously studied systems, which otherwise collapse during simulations of this length.[19] The degree of mixing between the fibre (blue) and the randomly distributed molecules (red) increases through the simulation, but even at the completion of the simulation (150 ns) there are differentiated areas in blue and red, indicating that there are some interactions that work to keep these molecules together throughout the simulation. However, around 50 – 75 ns the fibre breaks down and looks more like a disordered aggregate before rearranging to form like a fibre again at the end of the simulation, albeit with the additional red molecules incorporated into it (150 ns).

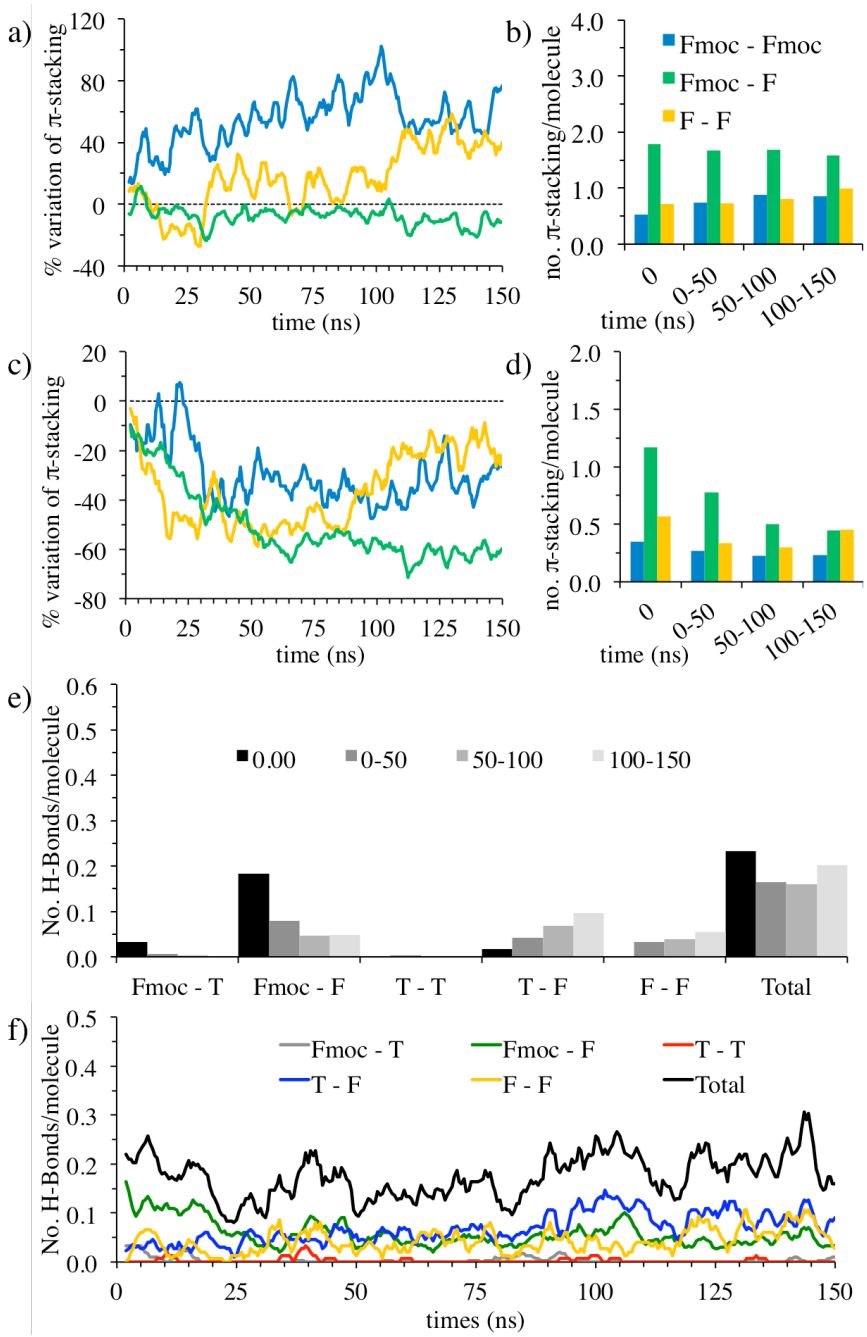


Figure 7.3 Model 1 π-stacking analyses of (a, b) the whole system and (c, d) of the fibre, (a, c) show the percentage of change of the interactions through the simulation; and (b, d) show the average number of interactions during 50 ns intervals and at time 0. Model 1 hydrogen bond analysis showing (e) the average number of hydrogen bonds during 50 ns intervals and at time 0; and (f) the evolution of the hydrogen bonds per residue through the simulation.

The analysis of the π-stacking interactions shows an important variation in the percentage of π-stacking interactions during the time course of the simulations, which indicates that the initial starting structure was not stable (Figure 7.3a). The number of

these interactions per molecule is relatively low even at the beginning of the simulation (Figure 7.3b, Fmoc – Fmoc = ~0.5; F – F = ~0.7; Fmoc – F = ~1.8), which explains the instability of the interactions through the simulation. The overall Fmoc π-stacking interactions increases ~70% (Figure 7.3a) while the interactions between the fibre molecules drop ~40%, which means that the fibre interactions break down in favour of increasing the total number of interactions. This suggests that the Fmoc – Fmoc π-stacking interactions in the fibre are relatively weak, probably due to the lack of periodicity of these interactions (Figure 7.1h shows the Fmoc groups in pairs). The number of F – F interactions initially decreases rapidly suggesting that the stacking of these groups inside the fibre is not stable or potentially the conformation of the peptide's side chain in this orientation is not favourable. After the initial decrease, the F – F interactions increase again, probably after the rearrangement of the F side chains. This provokes an overall final increment as the disordered aggregate is able to incorporate the additional Fmoc-TF-NH_2 molecules leading to a further increase in the number of F – F interactions (yellow line, Figure 7.3a). The Fmoc – F interactions also drop, although at a slower rate, and after this initial drop the total new variations are close to 0 (green line, Figure 7.3a).

The absolute number of interactions is relatively low for both the whole system (Figure 7.3b) and the fibre (Figure 7.3d). The overall increment of the Fmoc – Fmoc (~0.5 – 0.8) and F – F (~0.7 – 1.0) π-stacking even though they decrease in the fibre molecules (~0.35 – 0.25 for Fmoc – Fmoc; ~0.6 – 0.45 for F – F) suggest that these interactions are not well reproduced in the model. A loss of Fmoc – F π-interactions also occurs in the fibre (green bars, Figure 7.3b) analogous to the Fmoc – Fmoc interactions. However, in this case the total number of Fmoc – F interactions is stable across the total system (Figure 7.3b), showing that the loss of interactions in the starting fibre are compensated by the formation of new interactions with the random Fmoc-TF-NH_2 molecules from the starting structure, resulting in the % of variation (Figure 7.3a) remaining close to 0.

The hydrogen bond analysis shows that the count of these interactions is low from the starting structure of the simulation (Figure 7.3e-f). This is due to the linearity of these interactions, which are constantly breaking and forming in the dynamic simulation. However, the evaluation of the changes in the interactions can still give valuable information. It can be seen that at the beginning of the simulation the Fmoc –

F H-bonds are dominant, corresponding to the H-bonds in the starting structure (Figure 7.4, Model 1). This model should also show T – T H-bonds, but they seem to break prior to this simulation, *i.e.*, in the heating up phase, showing that they are highly unstable. During the simulation, the Fmoc – F H-bonds break down while T – F and F – F H-bonds are formed. At the end of the simulation these three H-bonds are of similar importance.

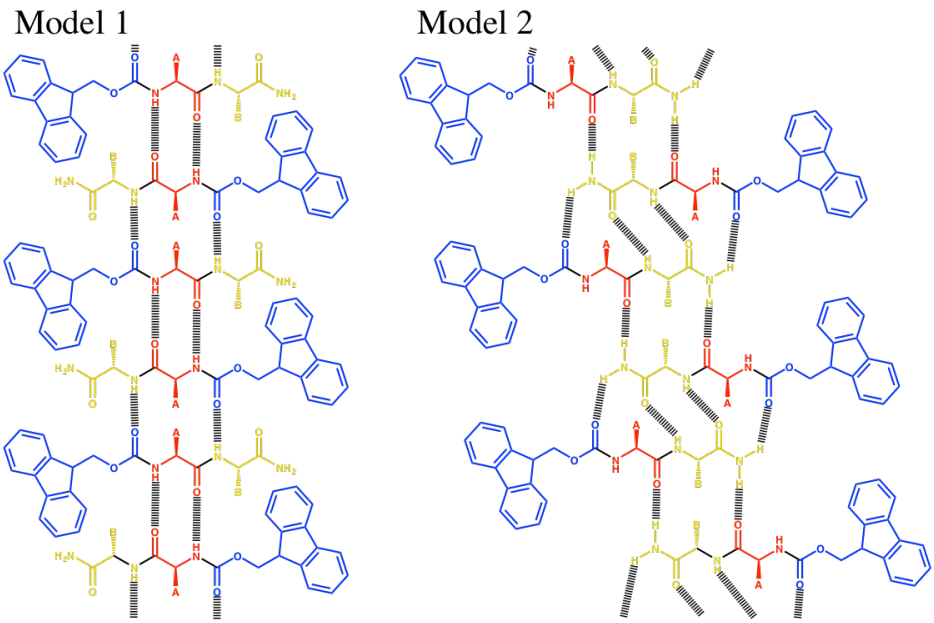


Figure 7.4 Model 1 and Model 2 arrangements showing the hydrogen bonds. Fmoc is shown in blue, T in red and F in yellow.

7.3.3 Refinement of the Model

The change in the H-bonds is used to build the revised version of the model, Model 2 (Figure 7.4). The introduction of the new H-bonds gives rise to a more twisted structure than the Model 1. This correlation between increased H-bonding and a twisted structure was observed experimentally with the correlation between FT-IR and CD signal dependence in the previous chapter (Section 6.3). However, Model 2 also shows more exposed fluorenyl groups (Figure 7.4, blue aromatic group) which need to be taken into account when building the fibre (Figure 7.5).

The Model 2 fibre (Figure 7.5b) shows how the new arrangement improves the periodicity of the Fmoc – Fmoc π-stacking interactions, which was too poor in the previous model. Furthermore, the new arrangement originates a fibre with evident ellipticity, which can explain the CD signal produced by the fluorenyl group shown in

Chapter 6. This is consistent with the arrangement suggested by Smith *et al*[30] but not with that one proposed by Mu *et al*., [248] which does not show any supramolecular ellipticity. The number of Fmoc-TF-NH_2 molecules in this fibre to complete the turn and keep the structure periodic through the *z*-direction is 60 (Figure 7.5b). An additional 60 randomly distributed molecules are added resulting in a system composed of a total of 120 Fmoc-TF-NH_2 molecules (Figure 7.5a and c).

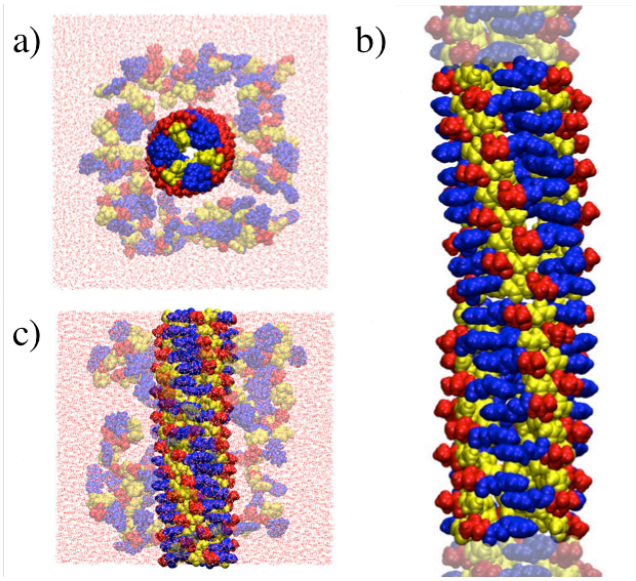


Figure 7.5 Fibre Model 2: in water box (transparent water) with the extra randomly distributed molecules (semi-transparent), (a) top and (c) side view; and (b) fibre model with periodic images (semi-transparent) along the *z*-direction. Fmoc in blue, T in red and F in yellow.

7.3.4 Model 2 Simulation and Analysis

The simulation snapshots of Model 2 show how the fibre shape is kept through the whole simulation (blue, Figure 7.6) and the extra molecules (red, Figure 7.6) fill the areas when an initial fibre molecule dissociates during the simulation. No stages of clear disordered aggregation are observed, suggesting that this structure is more stable than the previous proposal (Model 1, Figure 7.2 50 ns).

The variations of the π-stacking interactions (Figure 7.7a and c) are now smoother than in the previous model (Figure 7.3a-b) and the number of interactions is higher from the start of the simulation: Fmoc – Fmoc 1.45 (0.5 in Model 1), F – F 1.95 (0.7 in Model 1) and Fmoc – F 2.2 (1.8 in Model 1). The percentage of π-stacking lost between the initial fibre molecules is similar to the previous model in the fibre but it is slower (*cf*. Figure 7.7c and Figure 7.3c). Except for the Fmoc – F

interaction, which is relatively constant, showing that the fibre molecules remain close to the fibre even when the arrangement breaks (Figure 7.7c).

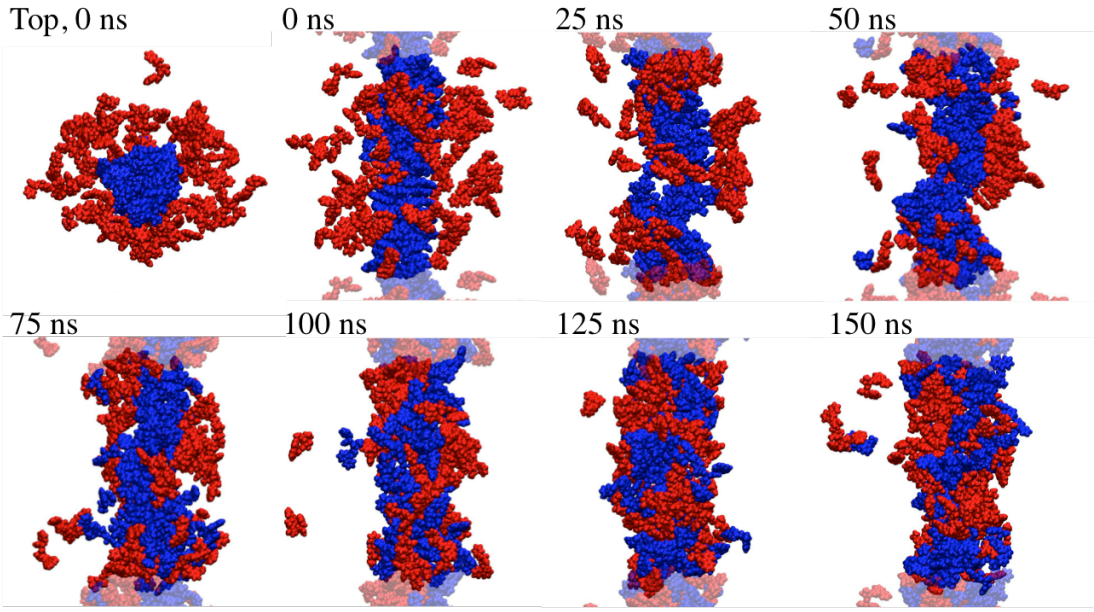


Figure 7.6 Snapshots of the Model 2 simulation. Fmoc-TF-NH2 molecules in the fibre arrangement are shown in blue and the randomly distributed extra molecules in red. Periodic images through the *z*-direction are semi-transparent. All images are shown form the side with the 0 ns image also shown from the top (Top, 0ns).

In Model 1 F – F reached 40 % of lost interactions around 15 ns and Fmoc – Fmoc around 35 ns (Figure 7.3c) while in Model 2 these interactions reach those levels around 40 ns and 120 ns, respectively. The overall change of Fmoc – Fmoc and F – F π-stacking are close to 0, showing that the level of equilibrated interactions corresponds to the level initially considered in the model, and are not higher, as occurred in Model 1. The overall increase in the Fmoc – F π-stacking (Figure 7.7a) is due to the minimal disruption of these interactions in the fibre (Figure 7.7b) and the aggregation of the randomly distributed molecules to the sides of the fibre (Figure 7.6).

The H-bond analysis (Figure 7.7e-f) shows that the ordering of the type of H-bond interactions is conserved from the beginning to the end of the simulation, indicating that the arrangement formed is more stable than that one in Model 1 (*cf.* Figure 7.3e). The randomly distributed (red molecules, Figure 7.6) molecules' aggregation through the simulation does not involve the creation of different arrangements either. The total number of these interactions decreases, especially due

to the overestimation of the F – F and also T – F hydrogen bonds. The F – F are especially overestimated and this can be seen in the Model 2 arrangement (Figure 7.4) to be too optimistic, as the angle required for the H-bonds is too acute.

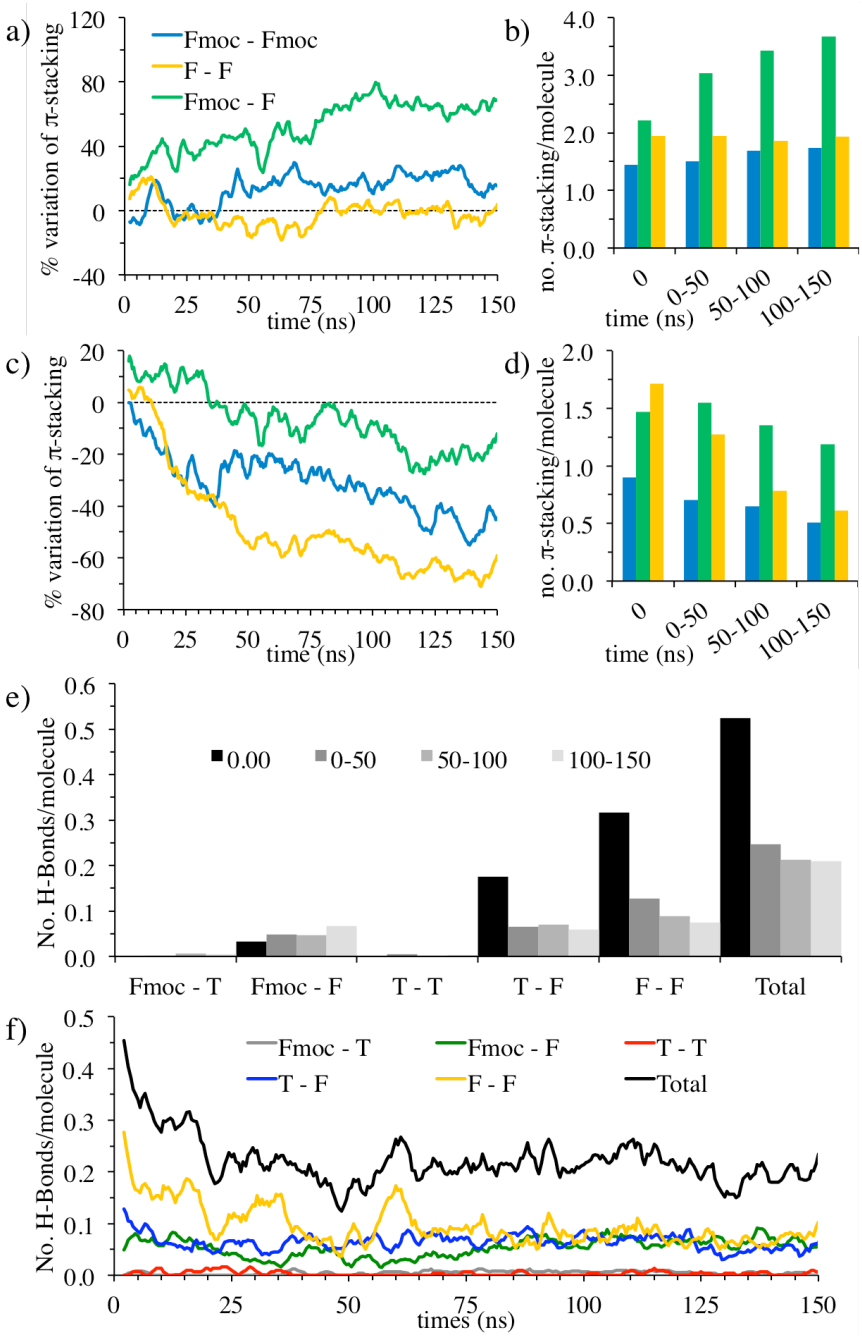


Figure 7.7 Model 2 π-stacking analyses of (a, b) the whole system and (c, d) of the fibre, showing (a, c) the percentage of change of the interactions through the simulation; and (b, d) the average number of interactions during 50 ns time intervals and at time 0. Model 2 hydrogen bond analysis showing (e) the average number of hydrogen bonds during 50 ns

time intervals and at time 0; and (f) the evolution of the hydrogen bonds per residue through the simulation.

Therefore, both starting well-ordered fibre structures (blue) break down through the simulation of the models, due to the limited size of the models, but the systems do not fall apart due to the extra molecules (red) added to the system. However, comparing the analysis for both models it is possible to compare the relative stability of each. The new model (Model 2) shows a greater number of, and more stable, π-stacking interactions. Furthermore, the relative ordering of the H-bonds is the same at the completion of the simulations as it is in the beginning, which is not the case for Model 1. This shows that a more stable H-bonded arrangement is achieved, although some of the interactions (H-bonds between F residues) are overestimated.

7.3.5 Lateral Fibre Aggregation *vs* Bilayer Like Structure in the Formation of Ribbons

The stability of Model 2, and its correlation with the available experimental spectra of the systems, indicates that it may indeed act as a realistic model for the nanostructures resulting from Fmoc-TF-NH_2 self-assembly. This model was therefore used to further study the experimental observation seen in Chapter 6 relating to the formation of ribbons from fibres. Despite the excellent resolution obtained, it is not possible to observe, in the TEM images, how these ribbons are formed at the intermolecular level. However, two possibilities exist: (1) the ribbons are the result of lateral aggregation of fibres; or (2) the ribbons are formed by the assembly of two parallel H-bonded stacks, which involves the breaking of the fibre proposed in this chapter. This is similar to the model proposed for Fmoc-SF-OMe sheets[177] and to that one proposed for Fmoc-FK[NDI] tapes.[32] To study these two scenarios the relative stability of two systems is compared: (1) a system containing two parallel fibres (*2fibres*, Figure 7.8a and c); and (2) a system formed by two stacks, each formed by an open fibre in a bilayer like arrangement (*belt* Figure 7.8e and g). Both systems use the H-bonded arrangement validated in Model 2 as well as the link between H-bonded stacks by Fmoc – Fmoc π-stacking interactions. The hydrophobic F side chains are buried between both layers of the *belt*, where the aromatic phenyl groups can π-stack, as occurs in the fibre Model 2. Therefore, both are formed by the most stable

interactions shown and the only comparison is between the relative stability of the different shapes.

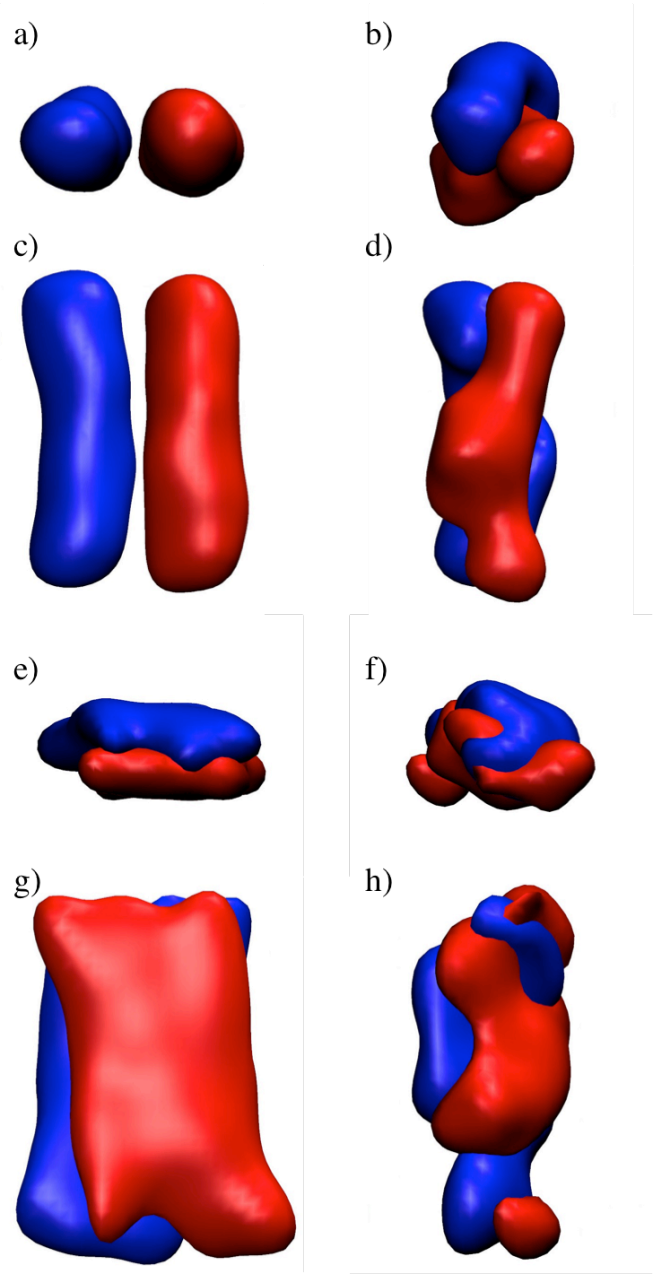


Figure 7.8 Snapshots (a, c, e, g) at 0 ns; and (b, d, f, h) at 100 ns of the systems. (a-d) *2fibres* and (e-h) *belt*. The (a-b, e-f) top and (c-d, g-h) side view of each snapshot are shown. The density surface is represented with one layer/fibre in red and the other in blue.

The simulation shows that in the *2fibres* system after 100 ns (Figure 7.8c-d) both fibres keep their shape and twist around each other, which is in good agreement with what was observed experimentally for these systems in Chapter 6. The *belt* structure is not maintained throughout the simulation. Rather, the two layers deform

during the 100 ns simulation to form an elongated aggregate structure, or fibre-like structure.

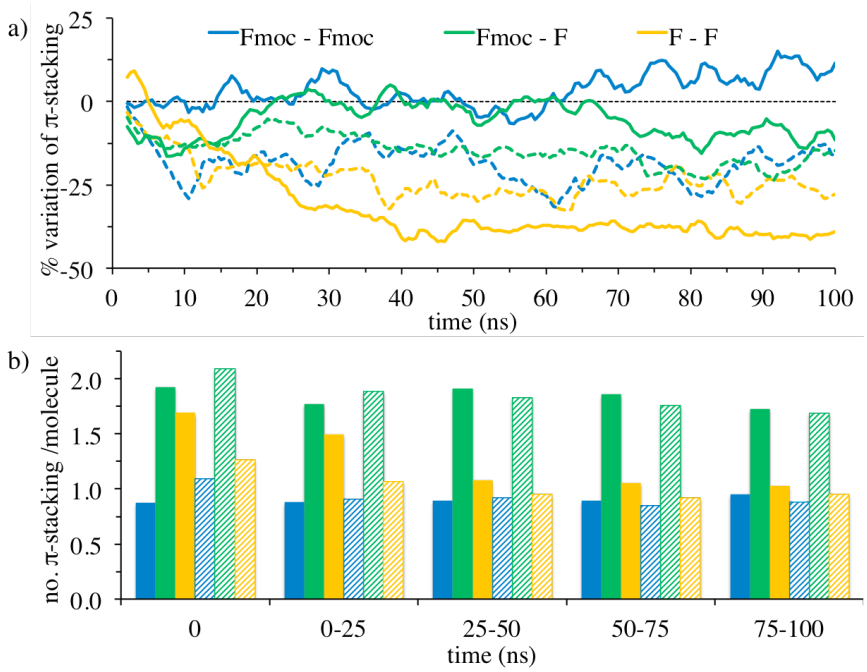


Figure 7.9 *2fibres* (solid lines and stacks) and *bilayer* (dashed lines and striped stacks) π-stacking analyses showing (a) the percentage of change of the interactions through the simulation; and (b) the average number of interactions during 25 ns time intervals and at time 0.

Although visually it is already evident that the *2fibres* system is more stable and agrees better with the experiments, quantitative analysis of the interactions was carried out as in the case of the previous simulations. The π-stacking interactions analysis (Figure 7.9a) shows that Fmoc – Fmoc (blue) and Fmoc – F (green) interactions are more stable in the *2fibres* system than in the system containing only one fibre (*cf*. Figure 7.9a and Figure 7.7a). Moreover, they are more stable in this system than in the *belt* system (Figure 7.9 dashed lines). F – F (yellow) interactions are ~10% less stable in the *2fibres* (Figure 7.9a, solid lines) after 20 ns. However, all the interactions show a steeper decay in the *belt* system (Figure 7.9a, dashed lines), the three interactions ~10% after only 5 ns. The *2fibres* only shows this rate of change in the simulation for the F – F interactions (Figure 7.9 yellow, solid lines). Despite this initial rapid decrease, after 40 ns the F – F interactions are equilibrated and remain unchanged until the end of the simulation. The Fmoc – Fmoc and Fmoc – F, are equilibrated from 0 ns and 25 ns, respectively, even though, they show

fluctuations from 70 ns, due to a rearrangement between these two interactions (positive Fmoc – Fmoc trend matches the negative Fmoc – F, Figure 7.9a blue and green solid lines).

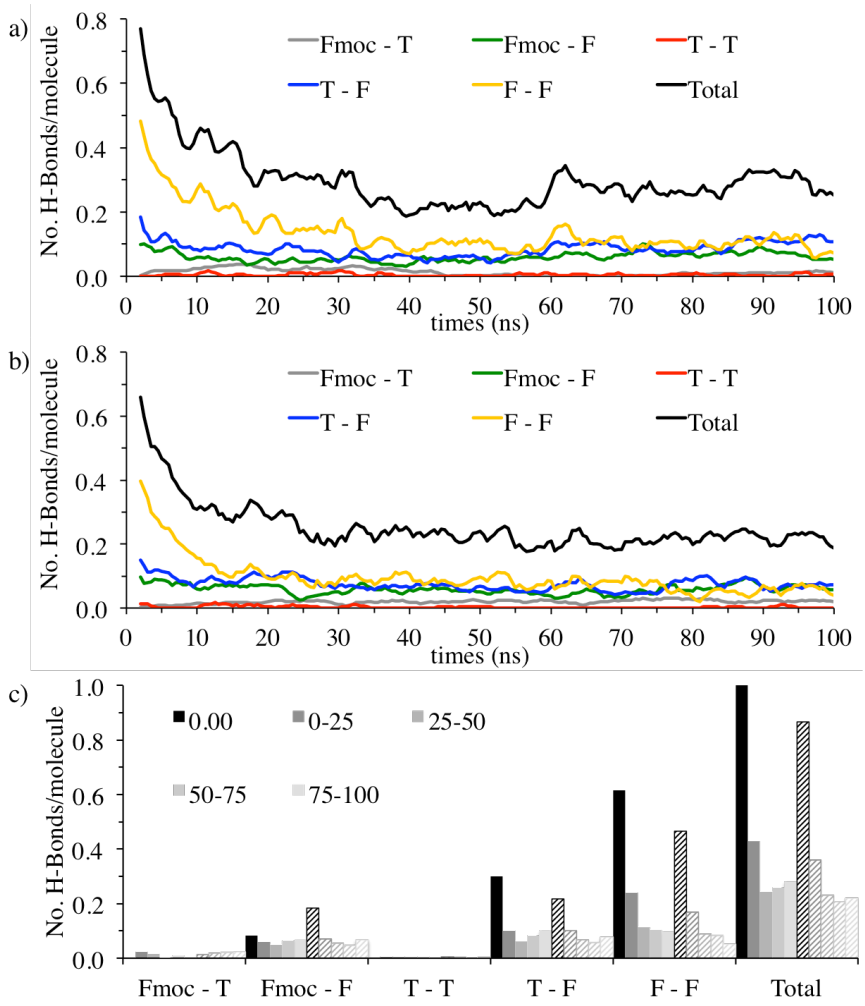


Figure 7.10 *2fibres* (a and solid stacks in c) and *bilayer* (b and striped stacks in c) hydrogen bond analysis showing (a-b) the evolution of the hydrogen bonds per residue through the simulation and (c) the average number of hydrogen bonds during 25 ns time intervals and at time 0.

In absolute numbers, the *belt* structure starts with a slightly higher number of Fmoc – Fmoc and Fmoc – F interactions (Figure 7.9b striped stacks blue and green), but at the end they are slightly lower, which suggest a loss of stability through the simulation even though the starting structure has at least as many interactions as the *2fibres* (Figure 7.9b solid stacks blue and green). The F – F interactions are higher in the *2fibres* (Figure 7.9b yellow solid stacks) and despite the loss mentioned, they continue to be slightly higher even at the end of the simulations.

The trend in the H-bonds for the *2fibres* system (Figure 7.10a) is similar to that one in Model 2 analysis (Figure 7.7e-f), with a small increment in the number of interactions (Total is 0.28 for *2fibres* system and 0.21 in Model 2), which is reasonable as this system contains all the molecules in well-ordered arrangements (*i.e.*, there are no randomly oriented molecules in the *2fibres* system). The *belt* H-bonds show similar trends (Figure 7.10b) for the Fmoc – F and T – F H-bonds, but clear differences in the F – F, which decrease more rapidly for this system than for the *2fibres*. At 15 ns this interaction has a value of 0.1 for the *belt* and 0.2 for the *2fibres*. This, in addition to the averaged total for the last 25 ns, 0.22 vs 0.28 (Figure 7.10c), which evidences the higher stability of the arrangements in the *2fibres* system.

7.4 Conclusions

The inclusion of additional randomly oriented molecules around an ordered structure reduces the problems caused by size limit effects at the sides of a fibre. This allows a more robust study of the simulations due to both, the higher stability of the model and the inclusion of a random effect, which helps to decrease the bias when an erroneous initial structure is chosen.

The interpretation of the analyses is not straightforward and the validation of Model 2 requires comparison of the stability with the results for the previous model. However, the results of Model 1 clearly indicated that better π-stacking and different H-bond networks were required, and it led to the formation of the more stable Model 2. This self-consistent approach for the use of MD simulations was shown to be valid for the study of molecular level interactions in Fmoc-dipeptide nanostructures and can be extrapolated to other nanostructures with adequate parameters. However, this approach cannot be carried out independently of experiment – both the Model 1 formation and the refinement require correlation with the experimental data.

The correlation of H-bonded arrangements and ellipticity has been found to be common in the experimental and the MD results. Furthermore, the need for Fmoc – Fmoc π-stacking and the important role of F in the π-stacking is also corroborated by both methods. Therefore, the correlation between experimental and computational results allows the verification of some structural features, which are otherwise extrapolated from the existing spectra.

The final MD simulation using Model 2 suggests that the twisted ribbons shown in the TEM images in the previous chapter are actually formed by lateral

aggregation of fibres. The simulation shows the twisting of fibres to form a twisted ribbon, although the starting point is a planar ribbon. This is an additional validation to the model, which demonstrates the utility of the model in studying an experimental event.

This model refinement protocol requires less computational effort than usual ways of model development based on checking all the possible conformations in different simulations, for complete fibre models, due to the use of experimental information in the conformational search. Furthermore, it includes the validation of the final model and its correlation with experimental results. It can be concluded, that this protocol is a straightforward approach for getting reliable detailed information of the molecular level structure of self-assembled peptide based nanomaterials, which then results in the ability to obtain useful information on the fibres behaviour at the molecular level.

8 Conclusions

A combination of computational, experimental and theoretical methods, has been applied to the study of Fmoc-dipeptide self-assembled materials. The combination and correlation between the different approaches have provided useful insights into different features of these materials.

The study started with a literature review where the importance of PAs and APAs as potential new materials was shown. These systems have distinct advantages over purely synthetic nanostructures: (a) the bioinspired self-assembling ability, which makes these materials able to form nanostructures spontaneously with no external stimulus; (b) their availability as natural derived products; (c) the high tuneability of these materials due to the dependence of their functionality with the structure; and (d) the dependence of the structure on small chemical changes in the building blocks, which can be rationally constructed from the diverse chemical space provided by twenty amino acids. Chapter 2 finished with the discussion of the importance of controlling the self-assembly process because the formation pathway influences the final structure, and hence, functionality, as much as the chemical structure of the LMWG. Different ways were introduced and the enzymatic controlled self-assembly was shown to present several advantages: specificity, time and spatial control, concentration dependence and – in the case of self-assembly via amide condensation – reversibility.

In Chapter 3 the different methods to study peptide-based nanomaterials were introduced focusing on their application to APA based nanostructures. Firstly, the experimental techniques, which are usually applied for the structural study of these materials, were discussed. The application of these techniques to APAs comes from the well-established use of these techniques for the study of the secondary structure of peptide-based materials such as proteins. However, it is explained that the extrapolation of these techniques to LMWGs is not straightforward and that the information provided for these materials are actually more limited and difficult to interpret. FT-IR spectroscopy, and specifically, the Amide I vibrational mode, evidences the formation of hydrogen bonded networks but does not provide direct information of the arrangements. The CD signal is used to address the supramolecular

induced ellipticity of the aromatic moieties but not of their specific position. Fluorescence spectroscopy is affected by the π-stacking interactions of the aromatic moieties but their relationship with the specific stacks is currently not fully understood.

The use of computational methods is proposed in this chapter to obtain further insights to complement experimental self-assembly approaches. MD methods have been widely applied to study biochemical systems, such as proteins, at the molecular level. Considering the similarity of the molecular structures and interactions in designed self-assembling peptides and protein folding and dynamics, they can be expected to be suitable to study APAs self-assembly. Different examples of their application to APA based systems are presented. In these systems the relative stability of different starting structures is evaluated. However, these studies are affected by the limited size of the systems and the bias from the starting structures chosen.

Computational methods have limitations in their interpretation and validation and the correlation with the experimental characterization is extremely important. The use of experimental techniques to inform the most suitable starting structures for the simulations and the correlation of their analysis with experimental techniques is proposed in this chapter as a synergistic combination for the study of APAs with intermolecular resolution. This information would allow the improvement of the design rules for these materials, and hence, the development of new materials based on specific needs.

Enzymatic driven self-assembly of peptide nanostructures driven by amide condensation has been argued to ensure thermodynamic control in the formation of Fmoc-dipeptide nanostructures, providing the most highly ordered nanostructure with the minimum defects. The use of multiple building blocks to study self-assembly in competition between building blocks (DPLs) provides information on the thermodynamic self-assembly tendency. MD simulations are designed to proceed towards the thermodynamic equilibrium and hence, will provide information of the thermodynamically favoured self-assembled state. Therefore, the validity of the methods used in this thesis for the study of Fmoc-dipeptide nanostructures requires it to be possible for nanofibrous gels to represent thermodynamic equilibrium. However, it can be argued that when the barriers are high enough (4.6 kT), a deep

local minimum can also be studied because the system is not able to access the real thermodynamic minimum.

In order to demonstrate that nanofibres and gels can indeed represent equilibrium, in Chapter 4 a simple packing model was proposed to demonstrate this. Specifically, this is achieved by demonstrating that 1D objects, fibres, can be thermodynamically more stable than 3D objects, crystals. This is a simple packing model for LMWGs that represent each molecule with a prism that has faces of different levels of solvophobicity and solvophilicity. Different prisms were presented which, in addition of tuneable number of solvophilic faces, allows the application of the model to different LMWGs. The energetic penalties can also be adapted to represent the relative strength of the different type of interactions. The model is applied to reproduce the pH gelation dependence of Fmoc-FF-OH, which is in solution at low pH, gelates upon the initial increase of pH and at high pH precipitates. This demonstrates the potential applicability of such a simple model to improve the understanding in LMWGs self-assembly.

The model demonstrates that due to the amphiphilicity of LMWGs, for certain combinations of attractive and repulsive interactions of different faces of a prism, 1D objects, or fibres, can be more stable than 3D objects, crystals. The model was also extended to demonstrate that 2D objects, such as ribbons and tapes, can also be at thermodynamic equilibrium under some conditions.

Once it was demonstrated that MD simulations can be applied to the study of LMWGs self-assembly, the CHARMM force field was selected to carry out MD simulations for Fmoc-dipeptides because it has been frequently used to simulate the self-assembly behaviour of peptide based systems. However, the CHARMM force field does not include parameters for the Fmoc moiety and hence, in Chapter 5 a set of CHARMM parameters for the Fmoc moiety was developed. Taking advantage of the wide range of other molecules, apart from peptides, parameterized in CHARMM, the hard terms and van der Waals contributions were extrapolated from molecules with similar chemical groups. In order to fully parameterize the Fmoc group, the charges and torsions also had to be parameterized.

The CHARMM parameterization protocol was modified to ensure that the parameters can reproduce the amphiphilic behaviour (balance between hydrophilic and hydrophobic interactions) of the moiety by parameterizing the charges against the

interactions of a water molecule with the hydrophilic and hydrophobic parts as a reference. Furthermore, the interactions between two Fmoc moieties were used as an extra reference in order to ensure that the parameters can reproduce π-stacking interactions. Once a set of charges satisfied these requirements, the torsions of the linker between the aromatic fluorenyl group and the peptide were evaluated and corroborated to be able to reproduce the flexibility of the Fmoc moiety.

The final parameters were validated through the comparison of an MD calculated thermodynamic parameter, which is experimentally measureable, the log P, and through a self-assembling simulation of a relatively simple and known system.

Chapter 6 covers the experimental study of a small selection of Fmoc-dipeptides, focused on the effects of the aromatic moiety and C-terminus substitution, carried out by the study of the effect of the F/L and NH_2/OMe substitution, respectively. Therefore, four Fmoc-dipeptide LMWGs were studied: Fmoc-TF-NH_2, Fmoc-TF-OMe, Fmoc-TL-NH_2 and Fmoc-TL-OMe. The self-assembly was controlled enzymatically through direct condensation of Fmoc-T with the relevant amino acid amides or esters, to ensure thermodynamic control of structure formation. The side by side and direct competition (DPL) study of these molecules allowed us to evaluate how the two substitutions affect the H-bonds, comparing the FT-IR spectra in the Amide I region; the ellipticity, through the changes in the CD signal of the fluorenyl group; and the π-stacking, evidenced in the different red shifts of the λ_{max} of the fluorenyl group. The H-bonds and the ellipticity depend on the C-terminus substitution, they were suggested to be correlated, and hence, it could be concluded that the H-bond stack is the responsible of the ellipticity of the nanostructure. The Fmoc π-stacking interactions were improved by the presence of F, suggesting that its phenyl group improves these interactions within the nanostructure. Also OMe groups improve slightly these interactions.

Unlike previously studied Fmoc-peptides that showed dramatic differences in nanoscale morphology, all four APAs studied here were found to form similar nanostructures, which suggest that the changes observed with the different characterization techniques are due to changes in the intermolecular arrangements.

The use of DPLs allowed the comparison of the relative thermodynamic tendencies to self-assembly for the different molecules. These tendencies were related with the substitutions to show that the presence of F is the main effect improving the

self-assembly tendency, and the presence of NH_2 is the secondary effect. The tendencies were related with the interactions characterization to resolve that both π-stacking interactions and H-bonds are important, but the former is dominant.

Finally, the experimental information from Chapter 6 was used to develop a model of a fibre of which was found to be the best self-assembling Fmoc-dipeptide of the ones tested: Fmoc-TF-NH_2. The experimental information was combined with the interactions information presented in Chapter 2 for this type of gelators to develop a preliminary model. This model is formed by three hydrogen bonded networks connected by π-stacking interactions of the fluorenyl group. The arrangement of the molecules made it possible to situate all the hydrophobic side chains, F, buried inside the fibre, and the hydrophilic side chains, T, exposed to the solvent. This was found to be consistent with the better self-assembly tendency of the F containing molecules seen in Chapter 6 due to both, the higher hydrophobicity of F, which would promote the formation of a hydrophobic core, and the possibility of forming π-stacking interactions in the core and close to the Fmoc groups.

The preliminary model was simulated using the Fmoc parameterization developed in Chapter 5 to evaluate its stability. The simulation was carried out using additional randomly distributed molecules, which adds stability to the system and introduces the possibility of dynamic exchange between the molecules in the fibre and in solution. The analysis of the interactions was correlated with the experimental information to refine the model. The hydrogen bonds suggested a more twisted stack, which is in good agreement with the H-bonds/ellipticity correlation found experimentally. The π-stacking interaction showed instability and suggested that the number of Fmoc – Fmoc interactions was low, which agrees with the experimental results for the importance of these interactions.

The model was refined based on this information. The new H-bonded arrangements allowed better Fmoc π-stacking. The new model was found to be more stable and the H-bonds preferences did not change, apart from some overestimation of one of them. Visually, this model kept the fibre shape for a longer time with less exchange of molecules between fibre and solution molecules.

The final model was applied to the study of one experimental observation introduced in Chapter 6, which is the evolution of fibres to twisted ribbons. This was done by comparing systems based on two different possibilities: in the first one, the

ribbons are formed by lateral aggregation of the fibres, using a system with two fibres in parallel; and in the second, the fibres rearrange to form a bilayer type of structure, losing the fibre shape. The simulation using the model shows how the system with two fibres is more stable and the fibres become twisted around each other forming a structure that is similar in morphology to that one observed in the cryo-TEM images.

The combination of both methodologies, computational and experimental, was suggested to have a synergistic effect by combining and correlating the small pieces of information obtained with each technique.

8.1 Future Work

The methods developed in this thesis open new opportunities for the study of Fmoc-dipeptide self-assembly and, in the case of the model developed in Chapter 4, the LMWGs in general.

The implementation of the thermodynamic model (Chapter 4) now depends on the assignation of the interaction parameters and prism type. This information is currently based on the chemical knowledge of the LMWG, which needs to be adapted to the model. In order to improve this model, a more quantitative and standardized way of assigning these parameters and shapes should be developed. It would required the benchmarking of the model with a relevant number of known LWMGs. Ideally it would involve the comparison with an experimental parameter such as the log P, but it should still consider the amphiphilic nature of these molecules and, hence, calculate the parameters based in the chemical groups present on each face once a prism is assigned. Depending on the benchmarking, the final model could be implemented for a more straightforward implementation of the model which could be applied for the study of the self-assembling mechanism or even for prediction of nanostructure shapes and gelation tendencies.

Applying the methodology of comparative spectroscopic results and DPL application (Chapter 6), for the interpretation and understanding of the factors that improve self-assembly, would allow for a better understanding of these structures. Now that the methodology is developed, this can be applied systematically to more Fmoc-dipeptides to obtain general design rules and structure information for this type of LMWGs.

The Fmoc-TF-NH_2 model (Chapter 7) validation is based on qualitative correlation with experimental data. However, the development of a method to

estimate the CD spectra for this type of systems would allow validating the model quantitatively due to the importance of the supramolecular ellipticity induced by the fibre in the model. Also, the fibre model could be applied to study the F/L substitution by running simulations based on the same model but changing the side chain, and for the C-terminus substitution. This is now straightforward due to the development of the Fmoc parameterization for CHARMM. This could be compared with the experimental results in Chapter 6 in order to be able to make a direct comparison between sequence and supramolecular structure. The application of this model should be helpful to better understand the structures formed by Fmoc-dipeptides that have a hydrophilic side chain in the first position and a hydrophobic side chain in the second position, as they could form the same hydrophobic core/hydrophilic surface structure. The knowledge of the Fmoc-TF-NH_2 structures and those to which the model can be extended would also allow a more detailed interpretation of the spectroscopic techniques as their interpretation can be improved through knowing the specific intermolecular arrangements that give rise to the observed spectroscopic data.

The knowledge of the Fmoc-TF-NH_2 allows introducing chemical modifications in order to obtain certain functionalities. For example, the fact that the fluorenyl groups and the phenyl groups show an important level of interaction, in addition to their extended π-stacking disposition, could be exploited to develop donor-acceptor based nanostructures for organic electronics giving rise to materials with anisotropic conductive properties in the direction of the nanostructures. Also, only modifying the F side chain with a proper ratio of donor and acceptor variations could originate this type of materials. In order to do this, the model could be applied to run simulations to show the effects of changing the aromatic groups.

Although the simulations with different aromatic groups would require an extra effort to parameterize these new aromatic groups for the CHARMM force field, the new protocol developed in Chapter 5 facilitates this establishing the way of doing it to obtain a proper reproduction of the behaviour of these groups.

9 References

1. Zhang, S.; Holmes, T.; Lockshin, C.; Rich, A. *Proc. Natl. Acad. Sci. U. S. A.,* **1993,** *90*, 3334

2. Ghadiri, M. R.; Granja, J. R.; Milligan, R. A.; McRee, D. E.; Khazanovich, N. *Nature,* **1993,** *366*, 324.

3. Whitesides, G. M.; Grzybowski, B. *Science,* **2002,** *295*, 2418.

4. Hirst, A. R.; Escuder, B.; Miravet, J. F.; Smith, D. K. *Angew. Chem.,* **2008,** *47*, 8002.

5. Xu, H. X.; Das, A. K.; Horie, M.; Shaik, M.; Smith, A. M.; Luo, Y.; Lu, X.; Collins, R.; Liem, S. Y.; Song, A.; Popelier, P. L. A.; Turner, M. L.; Xiao, P.; Kinloch, I. A.; Ulijn, R. V. *Nanoscale,* **2010,** *2*, 960.

6. Naskar, J.; Palui, G.; Banerjee, A. *J. Phys. Chem. B,* **2009,** *113*, 11787.

7. Zelzer, M.; Ulijn, R. V. *Chem. Soc. Rev.,* **2010,** *39*, 3351.

8. Gao, Y.; Zhao, F.; Wang, Q.; Zhang, Y.; Xu, B. *Chem. Soc. Rev.,* **2010,** *39*, 3425.

9. Zhou, M.; Smith, A. M.; Das, A. K.; Hodson, N. W.; Collins, R. F.; Ulijn, R. V.; Gough, J. E. *Biomaterials,* **2009,** *30*, 2523.

10. Aida, T.; Meijer, E.; Stupp, S. *Science,* **2012,** *335*, 813.

11. Hartgerink, J. D.; Benlash, E.; Stupp, S. L. *Science,* **2001,** *294*, 1684.

12. Adler-Abramovich, L.; Gazit, E. *Chem. Soc. Rev.,* **2014,** *43*, 6881.

13. Gazit, E. *Chem. Soc. Rev.,* **2007,** *36*, 1263.

14. Boekhoven, J.; Stupp, S. I. *Adv. Mater,* **2014,** *26*, 1642.

15. Silva, G. A.; Czeisler, C.; Niece, K. L.; Beniash, E.; Harrington, D. A.; Kessler, J. A.; Stupp, S. I. *Science,* **2004,** *303*, 1352.

16. Zhao, X.; Pan, F.; Xu, H.; Yaseen, M.; Shan, H.; Hauser, C. A. E.; Zhang, S.; Lu, J. R. *Chem. Soc. Rev.,* **2010,** *39*, 3480.

17. Hamley, I. W. *Soft Matter,* **2011,** *7*, 4122.

18. Whitesides, G. M.; Mathias, J. P.; Seto, C. T. *Science,* **1991,** *254*, 1312.

19. Hughes, M.; Frederix, P. W. J. M.; Raeburn, J.; Birchall, L. S.; Sadownik, J.; Coomer, F. C.; Lin, I. H.; Cussen, E. J.; Hunt, N. T.; Tuttle, T.; Webb, S. J.; Adams, D. J.; Ulijn, R. V. *Soft Matter,* **2012,** *8*, 5595.

20. Zhang, S. *Nat. Biotechnol.,* **2003,** *21*, 1171.

21. Smith, A. M.; Collins, R. F.; Ulijn, R. V.; Blanch, E. *J. Raman Spectrosc.,* **2009,** *40*, 1093.

22. Reches, M.; Gazit, E. *Science,* **2003,** *300*, 625.

23. Frederix, P., W. J. M.; Scott, G. G.; Abul-Haija, Y. M.; Kalafatovic, D.; Pappas, C. G.; Javid, N.; Hunt, N. T.; Ulijn, R. V.; Tuttle, T. *Nat. Chem.,* **2015,** *7*, 30.

24. Adams, D. J.; Mullen, L. M.; Berta, M.; Chen, L.; Frith, W. J. *Soft Matter,* **2010,** *6*, 1971.

25. Chen, L.; Revel, S.; Morris, K.; C. Serpell, L.; Adams, D. J. *Langmuir,* **2010,** *26*, 13466.

26. Tang, C.; Ulijn, R. V.; Saiani, A. *Langmuir,* **2011,** *27*, 14438.

27. Adams, D. J.; Topham, P. D. *Soft Matter,* **2010,** *6*, 3707.

28. Fleming, S.; Ulijn, R. V. *Chem. Soc. Rev.,* **2014,** *43*, 8150.

29. Mahler, A.; Reches, M.; Rechter, M.; Cohen, S.; Gazit, E. *Adv. Mater,* **2006,** *18*, 1365.

30. Smith, A. M.; Williams, R. J.; Tang, C.; Coppo, P.; Collins, R. F.; Turner, M. L.; Saiani, A.; Ulijn, R. V. *Adv. Mater,* **2008,** *20*, 37.

31. Das, A. K.; Collins, R.; Ulijn, R. V. *Small,* **2008,** *4*, 279.

32. Shao, H.; Parquette, J. R. *Chem. Commun.,* **2010,** *46*, 4285.

33. Jayawarna, V.; Ali, M.; Jowitt, T. A.; Miller, A. F.; Saiani, A.; Gough, J. E.; Ulijn, R. V. *Adv. Mater,* **2006,** *18*, 611.

34. Vegners, R.; Shestakova, I.; Kalvinsh, I.; Ezzell, R. M.; Janmey, P. A. *J. Pept. Sci.,* **1995,** *1*, 371.

35. Zhang, Y.; Gu, H.; Yang, Z.; Xu, B. *J. Am. Chem. Soc.,* **2003,** *125*, 13680.

36. Hirst, A. R.; Roy, S.; Arora, M.; Das, A. K.; Hodson, N.; Murray, P.; Marshall, S.; Javid, N.; Sefcik, J.; Boekhoven, J.; van Esch, J. H.; Santabarbara, S.; Hunt, N. T.; Ulijn, R. V. *Nat. Chem.,* **2010,** *2*, 1089.

37. Roy, S.; Javid, N.; Frederix, P. W.; Lamprou, D. A.; Urquhart, A. J.; Hunt, N. T.; Halling, P. J.; Ulijn, R. V. *Chem.-Eur. J.,* **2012,** *18*, 11723.

38. Tang, C.; Ulijn, R. V.; Saiani, A. *Eur. Phys. J. E,* **2013,** *36*, 1.

39. Nalluri, S. K. M.; Berdugo, C.; Javid, N.; Frederix, P. W.; Ulijn, R. V. *Angew. Chem.,* **2014**.

40. Nalluri, S. K. M.; Ulijn, R. V. *Chem. Sci.,* **2013,** *4*, 3699.

41. Fleming, S.; Debnath, S.; Frederix, P. W. J. M.; Tuttle, T.; Ulijn, R. V. *Chem. Commun.,* **2013,** *49*, 10587.

42. Hughes, M.; Birchall, L. S.; Zuberi, K.; Aitken, L. A.; Debnath, S.; Javid, N.; Ulijn, R. V. *Soft Matter,* **2012,** *8*, 11565.

43. Yang, Z.; Liang, G.; Ma, M.; Gao, Y.; Xu, B. *J. Mater. Chem.,* **2007,** *17*, 850.

44. Chen, L.; Morris, K.; Laybourn, A.; Elias, D.; Hicks, M. R.; Rodger, A.; Serpell, L.; Adams, D. J. *Langmuir,* **2010,** *26*, 5232.

45. Caruso, M.; Gatto, E.; Placidi, E.; Ballano, G.; Formaggio, F.; Toniolo, C.; Zanuy, D.; Alemán, C.; Venanzi, M. *Soft Matter,* **2014,** *10*, 2508.

46. Fleming, S.; Frederix, P. W. J. M.; Sasselli, I. R.; Hunt, N. T.; Ulijn, R. V.; Tuttle, T. *Langmuir,* **2013,** *29*, 9510.

47. Dokholyan, N. V., *Computational modeling of biological systems*. Springer: 2012; p 360.

48. Branden, C.; Tooze, J., *Introduction to protein structure*. Garland New York: 1991; Vol. 2.

49. Schulz, G. E.; Schirmer, R. H., *Principles of protein structure*. Springer-Verlag KG.: 1979.

50. Ghélis, C., *Protein folding*. Academic Press: 1982.

51. Chronopoulou, L.; Sennato, S.; Bordi, F.; Giannella, D.; Di Nitto, A.; Barbetta, A.; Dentini, M.; Togna, A. R.; Togna, G. I.; Moschini, S. *Soft Matter,* **2014,** *10*, 1944.

52. Castillo, J.; Sasso, L.; Svendsen, W. E., *Self-assembled peptide nanostructures: Advances and applications in nanobiotechnology*. CRC Press: 2012.

53. Fersht, A., *Structure and mechanism in protein science, 1999*. WH Freemann Company, New York.

54. Ozin, G. A.; Arsenault, A. C.; Cademartiri, L., *Nanochemistry: a chemical approach to nanomaterials*. Royal Society of Chemistry: 2009.

55. Steed, J. W.; Atwood, J. L., *Supramolecular chemistry*. John Wiley & Sons: 2009.

56. Graham, T. *Philos. Trans. R. Soc. London,* **1861,** *151*, 183.

57. Jordan Lloyd, D., The problem of gel structure. In *Colloid Chemistry*, Alexander, J., Ed. The Chemical Catalog Co.: New York, 1926; Vol. 1, pp 767.

58. Almdal, K.; Dyre, J. *Polymer Gels and Networks*, 5.

59. Jayawarna, V.; Richardson, S. M.; Hirst, A. R.; Hodson, N. W.; Saiani, A.; Gough, J. E.; Ulijn, R. V. *Acta Biomaterialia,* **2009,** *5*, 934.

60. Jayawarna, V.; Smith, A.; Gough, J. E.; Ulijn, R. V. *Biochem. Soc. Trans.,* **2007,** *35*, 535.

61. Weiss, R. G.; Terech, P., *Molecular Gels*. Springer: The Netherlands, 2006.

62. Cardoso, A. Z.; Alvarez, A. E. A.; Cattoz, B. N.; Griffiths, P. C.; King, S. M.; Frith, W. J.; Adams, D. J. *Faraday Discuss.,* **2014,** *166*, 101.

63. Orbach, R.; Mironi-Harpaz, I.; Adler-Abramovich, L.; Mossou, E.; Mitchell, E. P.; Forsyth, V. T.; Gazit, E.; Seliktar, D. *Langmuir,* **2012,** *28*, 2015.

64. Zhao, Y.; Wang, J.; Deng, L.; Zhou, P.; Wang, S.; Wang, Y.; Xu, H.; Lu, J. R. *Langmuir,* **2013,** *29*, 13457.

65. Reches, M.; Gazit, E. *Nano Lett.,* **2004,** *4*, 581.

66. Reches, M.; Gazit, E. *Physical Biology,* **2006,** *3*, S10.

67. Marchesan, S.; Waddington, L.; Easton, C. D.; Winkler, D. A.; Goodall, L.; Forsythe, J.; Hartley, P. G. *Nanoscale,* **2012,** *4*, 6752.

68. Korevaar, P. A.; Newcomb, C. J.; Meijer, E.; Stupp, S. I. *J. Am. Chem. Soc.,* **2014,** *136*, 8540.

69. Cui, H.; Webber, M. J.; Stupp, S. I. *Peptide Science,* **2010,** *94*, 1.

70. Versluis, F.; Marsden, H. R.; Kros, A. *Chem. Soc. Rev.,* **2010,** *39*, 3434.

71. Cui, H.; Cheetham, A. G.; Pashuck, E. T.; Stupp, S. I. *J. Am. Chem. Soc.,* **2014,** *136*, 12461.

72. Velichko, Y. S.; Stupp, S. I.; de la Cruz, M. O. *J. Phys. Chem. B,* **2008,** *112*, 2326.

73. Newcomb, C. J.; Sur, S.; Ortony, J. H.; Lee, O.-S.; Matson, J. B.; Boekhoven, J.; Yu, J. M.; Schatz, G. C.; Stupp, S. I. *Nat. Commun.,* **2014,** *5*.

74. Paramonov, S. E.; Jun, H.-W.; Hartgerink, J. D. *J. Am. Chem. Soc.,* **2006,** *128*, 7291.

75. Kuang, Y.; Du, X.; Zhou, J.; Xu, B. *Adv. Health. Mat.,* **2014**, n/a.

76. Wang, H.; Yang, C.; Tan, M.; Wang, L.; Kong, D.; Yang, Z. *Soft Matter,* **2011,** *7*, 3897.

77. Zhang, Y.; Yang, Z.; Yuan, F.; Gu, H.; Gao, P.; Xu, B. *J. Am. Chem. Soc.,* **2004,** *126*, 15028.

78. Zeng, G.; Liu, L.; Xia, D.; Li, Q.; Xin, Z.; Wang, J.; Besenbacher, F.; Skrydstrup, T.; Dong, M. *RSC Adv.,* **2014,** *4*, 7516.

79. Kuang, Y.; Xu, B. *Angew. Chem.,* **2013,** *52*, 6944.

80. Kuang, Y.; Shi, J.; Li, J.; Yuan, D.; Alberti, K. A.; Xu, Q.; Xu, B. *Angew. Chem.,* **2014,** *53*, 8104.

81. Yang, Z.; Xu, K.; Guo, Z.; Guo, Z.; Xu, B. *Adv. Mater,* **2007,** *19*, 3152.

82. Yang, Z.; Liang, G.; Guo, Z.; Guo, Z.; Xu, B. *Angew. Chem.,* **2007,** *46*, 8216.

83. Huang, Y.; Qiu, Z.; Xu, Y.; Shi, J.; Lin, H.; Zhang, Y. *Org. Biomol. Chem.,* **2011,** *9*, 2149.

84. Sahoo, J. K.; Nalluri, S. K. M. *Chem. Commun.,* **2014,** *50*, 5462.

85. Shi, J.; Gao, Y.; Yang, Z.; Xu, B. *Beilstein J. Org. Chem.,* **2011,** *7*, 167.

86. Cheng, G.; Castelletto, V.; Moulton, C. M.; Newby, G. E.; Hamley, I. W. *Langmuir,* **2010,** *26*, 4990.

87. Debnath, S.; Shome, A.; Das, D.; Das, P. K. *J. Phys. Chem. B,* **2010,** *114*, 4407.

88. Li, X.; Du, X.; Li, J.; Gao, Y.; Pan, Y.; Shi, J.; Zhou, N.; Xu, B. *Langmuir,* **2012,** *28*, 13512.

89. Du, X.; Zhou, J.; Guvench, O.; Sangiorgi, F. O.; Li, X.; Zhou, N.; Xu, B. *Bioconj Chem.,* **2014,** *25*, 1031.

90. Li, X.; Du, X.; Gao, Y.; Shi, J.; Kuang, Y.; Xu, B. *Soft Matter,* **2012,** *8*, 7402.

91. Li, X.; Kuang, Y.; Shi, J.; Gao, Y.; Lin, H.-C.; Xu, B. *J. Am. Chem. Soc.,* **2011,** *133*, 17513.

92. Wu, D.; Zhou, J.; Shi, J.; Du, X.; Xu, B. *Chem. Commun.,* **2014,** *50*, 1992.

93. Li, X.; Kuang, Y.; Xu, B. *Soft Matter,* **2012,** *8*, 2801.

94. Li, X.; Kuang, Y.; Lin, H. C.; Gao, Y.; Shi, J.; Xu, B. *Angew. Chem.,* **2011,** *50*, 9365.

95. Merrifield, R. B. *J. Am. Chem. Soc.,* **1963,** *85*, 2149.

96. Sadownik, J. W.; Ulijn, R. V. *Chem. Commun.,* **2010,** *46*, 3481.

97. ChandraáKotamarthi, H. *Soft Matter,* **2013,** *9*, 10141.

98. Yang, Z.; Gu, H.; Fu, D.; Gao, P.; Lam, J. K.; Xu, B. *Adv. Mater,* **2004,** *16*, 1440.

99. Sutton, S.; Campbell, N. L.; Cooper, A. I.; Kirkland, M.; Frith, W. J.; Adams, D. J. *Langmuir,* **2009,** *25*, 10285.

100. Abul‐Haija, Y. M.; Roy, S.; Frederix, P. W.; Javid, N.; Jayawarna, V.; Ulijn, R. V. *Small,* **2014,** *10*, 973.

101. Fleming, S.; Debnath, S.; Frederix, P. W.; Hunt, N. T.; Ulijn, R. V. *Biomacromolecules,* **2014,** *15*, 1171.

102. Benyus, J. M., *Biomimicry*. William Morrow New York: 1997.

103. Gross, M., *Travels to the Nanoworld*. Basic Books: 2008.

104. Karplus, M. *Nat. Chem. Biol.,* **2011,** *7*, 401.

105. Palma, C.-A.; Cecchini, M.; Samorì, P. *Chem. Soc. Rev.,* **2012,** *41*, 3713.

106. Singh, N.; Conte, M. P.; Ulijn, R.; Miravet, J. F.; Escuder, B. *Chem. Commun.,* **2015,** *51*, 13213.

107. Rodríguez-Llansola, F.; Miravet, J. F.; Escuder, B. *Chem. Commun.,* **2011,** *47*, 4706.

108. Miravet, J. F.; Escuder, B. *Chem. Commun.,* **2005**, 5796.

109. Sreenivasachary, N.; Lehn, J.-M. *Proc. Natl. Acad. Sci.,* **2005,** *102*, 5938.

110. Parkinson, G. N.; Lee, M. P.; Neidle, S. *Nature,* **2002,** *417*, 876.

111. Zahler, A. M.; Williamson, J. R.; Cech, T. R.; Prescott, D. M. *Nature,* **1991,** *350*, 718.

112. Buerkle, L. E.; von Recum, H. A.; Rowan, S. J. *Chem. Sci.,* **2012,** *3*, 564.

113. Arigon, J.; Prata, C. A.; Grinstaff, M. W.; Barthélémy, P. *Bioconj Chem.,* **2005,** *16*, 864.

114. Iwaura, R.; Yoshida, K.; Masuda, M.; Yase, K.; Shimizu, T. *Chem. Mater.,* **2002,** *14*, 3047.

115. Seela, F.; Pujari, S. S.; Schäfer, A. H. *Tetrahedron,* **2011,** *67*, 7418.

116. Park, S. M.; Kim, B. H. *Soft Matter,* **2008,** *4*, 1995.

117. Du, X.; Li, J.; Gao, Y.; Kuang, Y.; Xu, B. *Chem. Commun.,* **2012,** *48*, 2098.

118. Wang, G.; Cheuk, S.; Williams, K.; Sharma, V.; Dakessian, L.; Thorton, Z. *Carbohydr. Res.,* **2006,** *341*, 705.

119. Matsumoto, S.; Yamaguchi, S.; Ueno, S.; Komatsu, H.; Ikeda, M.; Ishizuka, K.; Iko, Y.; Tabata, K. V.; Aoki, H.; Ito, S. *Chem.-Eur. J.,* **2008,** *14*, 3977.

120. Wang, G.; Cheuk, S.; Yang, H.; Goyal, N.; Reddy, P. N.; Hopkinson, B. *Langmuir,* **2009,** *25*, 8696.

121. Vemula, P. K.; Li, J.; John, G. *J. Am. Chem. Soc.,* **2006,** *128*, 8932.

122. Kiyonaka, S.; Sugiyasu, K.; Shinkai, S.; Hamachi, I. *J. Am. Chem. Soc.,* **2002,** *124*, 10954.

123. Birchall, L. S.; Roy, S.; Jayawarna, V.; Hughes, M.; Irvine, E.; Okorogheye, G. T.; Saudi, N.; Santis, E. D.; Tuttle, T.; Edwards, A. A.; Ulijn, R. V. *Chem. Sci.,* **2011,** *2*, 1349.

124. Chen, Q.; Lv, Y.; Zhang, D.; Zhang, G.; Liu, C.; Zhu, D. *Langmuir,* **2009,** *26*, 3165.

125. Koshi, Y.; Nakata, E.; Yamane, H.; Hamachi, I. *J. Am. Chem. Soc.,* **2006,** *128*, 10413.

126. Wada, A.; Tamaru, S.-i.; Ikeda, M.; Hamachi, I. *J. Am. Chem. Soc.,* **2009,** *131*, 5321.

127. Jung, J. H.; John, G.; Masuda, M.; Yoshida, K.; Shinkai, S.; Shimizu, T. *Langmuir,* **2001,** *17*, 7229.

128. Suzuki, M.; Nakajima, Y.; Yumoto, M.; Kimura, M.; Shirai, H.; Hanabusa, K. *Org. Biomol. Chem.,* **2004,** *2*, 1155.

129. Rodríguez-Llansola, F.; Escuder, B.; Miravet, J. F.; Hermida-Merino, D.; Hamley, I. W.; Cardin, C. J.; Hayes, W. *Chem. Commun.,* **2010,** *46*, 7960.

130. Okesola, B. O.; Smith, D. K. *Chem. Commun.,* **2013,** *49*, 11164.

131. Boekhoven, J.; Poolman, J. M.; Maity, C.; Li, F.; van der Mee, L.; Minkenberg, C. B.; Mendes, E.; van Esch, J. H.; Eelkema, R. *Nat. Chem.,* **2013,** *5*, 433.

132. de Jong, J. J.; Lucas, L. N.; Kellogg, R. M.; van Esch, J. H.; Feringa, B. L. *Science,* **2004,** *304*, 278.

133. Del Guerzo, A.; Olive, A. G.; Reichwagen, J.; Hopf, H.; Desvergne, J.-P. *J. Am. Chem. Soc.,* **2005,** *127*, 17984.

134. Reichwagen, J.; Hopf, H.; Del Guerzo, A.; Belin, C.; Bouas-Laurent, H.; Desvergne, J.-P. *Org. Lett.,* **2005,** *7*, 971.

135. Giano, M. C.; Pochan, D. J.; Schneider, J. P. *Biomaterials,* **2011,** *32*, 6471.

136. Branco, M. C.; Schneider, J. P. *Acta Biomaterialia,* **2009,** *5*, 817.

137. Banwell, E. F.; Abelardo, E. S.; Adams, D. J.; Birchall, M. A.; Corrigan, A.; Donald, A. M.; Kirkland, M.; Serpell, L. C.; Butler, M. F.; Woolfson, D. N. *Nat. Mater.,* **2009,** *8*, 596.

138. Fallas, J. A.; O'Leary, L. E. R.; Hartgerink, J. D. *Chem. Soc. Rev.,* **2010,** *39*, 3510.

139. MacEwan, S. R.; Chilkoti, A. *Peptide Science,* **2010,** *94*, 60.

140. Almine, J. F.; Bax, D. V.; Mithieux, S. M.; Nivison-Smith, L.; Rnjak, J.; Waterhouse, A.; Wise, S. G.; Weiss, A. S. *Chem. Soc. Rev.,* **2010,** *39*, 3371.

141. Montenegro, J.; Ghadiri, M. R.; Granja, J. R. *Acc. Chem. Res.,* **2013,** *46*, 2955.

142. Chapman, R.; Danial, M.; Koh, M. L.; Jolliffe, K. A.; Perrier, S. *Chem. Soc. Rev.,* **2012,** *41*, 6023.

143. Appel, E. A.; del Barrio, J.; Loh, X. J.; Scherman, O. A. *Chem. Soc. Rev.,* **2012,** *41*, 6195.

144. Huang, X.; Terech, P.; Raghavan, S. R.; Weiss, R. G. *J. Am. Chem. Soc.,* **2005,** *127*, 4336.

145. Lutolf, M. P. *Nat. Mater.,* **2009,** *8*, 451.

146. Stuart, M. A. C.; Huck, W. T.; Genzer, J.; Müller, M.; Ober, C.; Stamm, M.; Sukhorukov, G. B.; Szleifer, I.; Tsukruk, V. V.; Urban, M. *Nat. Mater.,* **2010,** *9*, 101.

147. Lee, K. Y.; Mooney, D. J. *Chem. Rev.,* **2001,** *101*, 1869.

148. Wichterle, O.; Lim, D. **1960**.

149. Langer, R.; Tirrell, D. A. *Nature,* **2004,** *428*, 487.

150. Zayed, J. M.; Nouvel, N.; Rauwald, U.; Scherman, O. A. *Chem. Soc. Rev.,* **2010,** *39*, 2806.

151. Zhao, N.; Lloyd, G. O.; Scherman, O. A. *Chem. Commun.,* **2012,** *48*, 3070.

152. Biedermann, F.; Scherman, O. A. *J. Phys. Chem. B,* **2012,** *116*, 2842.

153. Tang, C.; Smith, A. M.; Collins, R. F.; Ulijn, R. V.; Saiani, A. *Langmuir,* **2009,** *25*, 9447.

154. Tuttle, T.; Gräfenstein, J.; Wu, A.; Kraka, E.; Cremer, D. *J. Phys. Chem. B,* **2004,** *108*, 1115.

155. Tuttle, T.; Kraka, E.; Wu, A.; Cremer, D. *J. Am. Chem. Soc.,* **2004,** *126*, 5093.

156. Adler-Abramovich, L.; Vaks, L.; Carny, O.; Trudler, D.; Magno, A.; Caflisch, A.; Frenkel, D.; Gazit, E. *Nat. Chem. Biol.,* **2012,** *8*, 701.

157. Pochan, D. J.; Schneider, J. P.; Kretsinger, J.; Ozbas, B.; Rajagopal, K.; Haines, L. *J. Am. Chem. Soc.,* **2003,** *125*, 11802.

158. Dudukovic, N. A.; Zukoski, C. F. *Soft Matter,* **2014,** *10*, 7849.

159. Hecht, S. *Small,* **2005,** *1*, 26.

160. Kouwer, P. H. J.; Koepf, M.; Le Sage, V. A. A.; Jaspers, M.; van Buul, A. M.; Eksteen-Akeroyd, Z. H.; Woltinge, T.; Schwartz, E.; Kitto, H. J.; Hoogenboom, R.; Picken, S. J.; Nolte, R. J. M.; Mendes, E.; Rowan, A. E. *Nature,* **2013,** *493*, 651.

161. Adams, D. J.; Butler, M. F.; Frith, W. J.; Kirkland, M.; Mullen, L.; Sanderson, P. *Soft Matter,* **2009,** *5*, 1856.

162. Roberts, D.; Rochas, C.; Saiani, A.; Miller, A. *Langmuir,* **2012,** *28*, 16196.

163. Fletcher, N. L.; Lockett, C. V.; Dexter, A. F. *Soft Matter,* **2011,** *7*, 10210.

164. Castelletto, V.; Cheng, G.; Greenland, B. W.; Hamley, I. W.; Harris, P. J. F. *Langmuir,* **2011,** *27*, 2980.

165. Jonker, A. M.; Löwik, D. W. P. M.; van Hest, J. C. M. *Chem. Mater.,* **2012,** *24*, 759.

166. Ozbas, B.; Kretsinger, J.; Rajagopal, K.; Schneider, J. P.; Pochan, D. J. *Macromolecules,* **2004,** *37*, 7331.

167. Roy, S.; Javid, N.; Sefcik, J.; Halling, P. J.; Ulijn, R. V. *Langmuir,* **2012,** *28*, 16664.

168. Hofmeister, F. *Arch. Exp. Pathol. Pharmakol.(Leipzig),* **1888,** *24*, 247.

169. Haines, L. A.; Rajagopal, K.; Ozbas, B.; Salick, D. A.; Pochan, D. J.; Schneider, J. P. *J. Am. Chem. Soc.,* **2005,** *127*, 17025.

170. Tie, Z.-X.; Qin, M.; Zou, D.-W.; Cao, Y.; Wang, W. *Chin. Phys. Lett.,* **2011,** *28*, 028702.

171. Zelzer, M.; Todd, S. J.; Hirst, A. R.; McDonald, T. O.; Ulijn, R. V. *Biomaterials Science,* **2012,** *1*, 11.

172. Williams, R. J.; Smith, A. M.; Collins, R.; Hodson, N.; Das, A. K.; Ulijn, R. V. *Nat. Nanotechnol.,* **2009,** *4*, 19.

173. Toledano, S.; Williams, R. J.; Jayawarna, V.; Ulijn, R. V. *J. Am. Chem. Soc.,* **2006,** *128*, 1070.

174. Sadownik, J. W.; Leckie, J.; Ulijn, R. V. *Chem. Commun.,* **2011,** *47*, 728.

175. Carpenter, F. H. *J. Am. Chem. Soc.,* **1960,** *82*, 1111.

176. Javid, N.; Vogtt, K.; Roy, S.; Hirst, A. R.; Hoell, A.; Hamley, I. W.; Ulijn, R. V.; Sefcik, J. *J. Phys. Chem. Lett.,* **2011,** *2*, 1395.

177. Hughes, M.; Xu, H. X.; Frederix, P. W. J. M.; Smith, A. M.; Hunt, N. T.; Tuttle, T.; Kinloch, I. A.; Ulijn, R. V. *Soft Matter,* **2011,** *7*, 10032.

178. Hartgerink, J. D.; Beniash, E.; Stupp, S. I. *Proc. Natl. Acad. Sci.,* **2002,** *99*, 5133.

179. Pashuck, E. T.; Stupp, S. I. *J. Am. Chem. Soc.,* **2010,** *132*, 8819.

180. van Esch, J. H. *Langmuir,* **2009,** *25*, 8392.

181. Sreerama, N.; Woody, R. W. *Anal. Biochem.,* **2000,** *287*, 252.

182. Johnson, W. C. *Proteins: Struct., Funct., Bioinf.,* **1990,** *7*, 205.

183. Barth, A.; Zscherp, C. *Q. Rev. Biophys.,* **2002,** *35*, 369.

184. Rughani, R. V.; Schneider, J. P. *MRS Bulletin,* **2008,** *33*, 530.

185. Yan, H.; Frielinghaus, H.; Nykanen, A.; Ruokolainen, J.; Saiani, A.; Miller, A. F. *Soft Matter,* **2008,** *4*, 1313.

186. Lee, C.; Cho, M. *J. Phys. Chem. B,* **2004,** *108*, 20397.

187. Hahn, S.; Kim, S.-S.; Lee, C.; Cho, M. *J. Chem. Phys.,* **2005,** *123*, 084905.

188. Ganim, Z.; Chung, H. S.; Smith, A. W.; DeFlores, L. P.; Jones, K. C.; Tokmakoff, A. *Acc. Chem. Res.,* **2008,** *41*, 432.

189. Cheatum, C. M.; Tokmakoff, A.; Knoester, J. *J. Chem. Phys.,* **2004,** *120*, 8201.

190. Rughani, R. V.; Salick, D. A.; Lamm, M. S.; Yucel, T.; Pochan, D. J.; Schneider, J. P. *Biomacromolecules,* **2009,** *10*, 1295.

191. Measey, T. J.; Smith, K. B.; Decatur, S. M.; Zhao, L.; Yang, G.; Schweitzer-Stenner, R. *J. Am. Chem. Soc.,* **2009,** *131*, 18218.

192. Measey, T. J.; Schweitzer-Stenner, R. *J. Am. Chem. Soc.,* **2006,** *128*, 13324.

193. Schneider, J. P.; Pochan, D. J.; Ozbas, B.; Rajagopal, K.; Pakstis, L.; Kretsinger, J. *J. Am. Chem. Soc.,* **2002,** *124*, 15030.

194. Nuansing, W.; Rebollo, A.; Mercero, J. M.; Zuñiga, J.; Bittner, A. M. *J. Raman Spectrosc.,* **2012,** *43*, 1397.

195. Berova, N.; Nakanishi, K., *Circular dichroism: principles and applications*. John Wiley & Sons: 2000.

196. Fasman, G. D., *Circular dichroism and the conformational analysis of biomolecules*. Springer Science & Business Media: 2013.

197. Greenfield, N. J. *TrAC, Trends Anal. Chem.,* **1999,** *18*, 236.

198. Gottarelli, G.; Lena, S.; Masiero, S.; Pieraccini, S.; Spada, G. P. *Chirality,* **2008,** *20*, 471.

199. Wang, Y.; Xu, J.; Wang, Y.; Chen, H. *Chem. Soc. Rev.,* **2013,** *42*, 2930.

200. Groenning, M. *J. Chem. Biol.,* **2010,** *3*, 1.

201. Naiki, H.; Higuchi, K.; Hosokawa, M.; Takeda, T. *Anal. Biochem.,* **1989,** *177*, 244.

202. Biancalana, M.; Makabe, K.; Koide, A.; Koide, S. *J. Mol. Biol.,* **2009,** *385*, 1052.

203. Vassar, P. S.; Culling, C. *Archives of pathology,* **1959,** *68*, 487.

204. Kang, H. K.; Kang, D. E.; Boo, B. H.; Yoo, S. J.; Lee, J. K.; Lim, E. C. *J. Phys. Chem. A,* **2005,** *109*, 6799.

205. Adcock, S. A.; McCammon, J. A. *Chem. Rev.,* **2006,** *106*, 1589.

206. Karplus, M.; Kuriyan, J. *Proc. Natl. Acad. Sci. U. S. A.,* **2005,** *102*, 6679.

207. van Gunsteren, W. F.; Bakowies, D.; Baron, R.; Chandrasekhar, I.; Christen, M.; Daura, X.; Gee, P.; Geerke, D. P.; Glättli, A.; Hünenberger, P. H. *Angew. Chem.,* **2006,** *45*, 4064.

208. Marrink, S. J.; Tieleman, D. P. *Chem. Soc. Rev.,* **2013,** *42*, 6801.

209. Born, M.; Oppenheimer, J. R. *Ann. Physik,* **1927,** *84*, 457.

210. Leach, A. R., *Molecular modelling: principles and applications*. Prentice Hall: Harlow [etc.], 2001.

211. Jensen, F., *Introduction to computational chemistry*. John Wiley & Sons: 2013.

212. Swope, W. C.; Andersen, H. C.; Berens, P. H.; Wilson, K. R. *J. Chem. Phys.,* **1982,** *76*, 637.

213. Beeman, D. *J. Comput. Phys.,* **1976,** *20*, 130.

214. Allinger, N. L.; Chen, K.; Lii, J. H. *J. Comput. Chem.,* **1996,** *17*, 642.

215. Debye, P. *Nachrichten von der Gesellschaft der Wissenschaften zu Göttingen, Mathematisch-Physikalische Klasse,* **1920,** *1920*, 55.

216. de Souza, O. N.; Ornstein, R. L. *Biophys. J.,* **1997,** *72*, 2395.

217. Jorgensen, W. L. *J. Am. Chem. Soc.,* **1981,** *103*, 335.

218. Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.,* **1983,** *79*, 926.

219. Jorgensen, W. L.; Tirado-Rives, J. *Proc. Natl. Acad. Sci. U. S. A.,* **2005,** *102*, 6665.

220. Sen, S.; Nilsson, L. *Biophys. J.,* **1999,** *77*, 1782.

221. Steinbach, P. J.; Brooks, B. R. *Proc. Natl. Acad. Sci.,* **1993,** *90*, 9135.

222. Bhatnagar, N.; Kamath, G.; Chelst, I.; Potoff, J. J. *J. Chem. Phys.,* **2012,** *137*.

223. Park, S.; Khalili-Araghi, F.; Tajkhorshid, E.; Schulten, K. *J. Chem. Phys.,* **2003,** *119*, 3559.

224. Benay, G.; Wipff, G. *J. Phys. Chem. B,* **2013,** *117*, 7399.

225. Gao, J.; Kuczera, K.; Tidor, B.; Karplus, M. *Science,* **1989,** *244*, 1069.

226. Fang, T.-H.; Weng, C.-I. *Nanotechnology,* **2000,** *11*, 148.

227. Guo, C.; Luo, Y.; Zhou, R.; Wei, G. *ACS Nano,* **2012,** *6*, 3907.

228. Khalfa, A.; Treptow, W.; Maigret, B.; Tarek, M. *Chem. Phys.,* **2009,** *358*, 161.

229. Meli, M.; Morra, G.; Colombo, G. *Biophys. J.,* **2008,** *94*, 4414.

230. Solar, M.; Buehler, M. J. *J. Mech. Behav. Biomed. Mat.,* **2013,** *19*, 43.

231. Vácha, R.; Frenkel, D. *Biophys. J.,* **2011,** *101*, 1432.

232. Villa, A.; Peter, C.; van der Vegt, N. F. A. *Phys. Chem. Chem. Phys.,* **2009,** *11*, 2077.

233. Villa, A.; van der Vegt, N. F. A.; Peter, C. *Phys. Chem. Chem. Phys.,* **2009,** *11*, 2068.

234. Zheng, X.; Wang, D.; Shuai, Z. *Nanoscale,* **2013,** *5*, 3681.

235. Wu, C.; Lei, H.; Duan, Y. *Biophys. J.,* **2004,** *87*, 3000.

236. Bellesia, G.; Shea, J.-E. *Biophys. J.,* **2009,** *96*, 875.

237. Hauser, C. A. E.; Deng, R.; Mishra, A.; Loo, Y.; Khoe, U.; Zhuang, F.; Cheong, D. W.; Accardo, A.; Sullivan, M. B.; Riekel, C.; Ying, J. Y.; Hauser, U. A. *Proc. Natl. Acad. Sci. U. S. A.,* **2011,** *108*, 1361.

238. McCullagh, M.; Prytkova, T.; Tonzani, S.; Winter, N. D.; Schatz, G. C. *J. Phys. Chem. B,* **2008,** *112*, 10388.

239. Lee, O.-S.; Cho, V.; Schatz, G. C. *Nano Lett.,* **2012,** *12*, 4907.

240. Zhang, H.; Wang, H.; Xu, G.; Yuan, S. *Colloids Surf., A,* **2013,** *417*, 217.

241. Yu, T.; Schatz, G. C. *J. Phys. Chem. B,* **2013,** *117*, 14059.

242. Frederix, P. W. J. M.; Ulijn, R. V.; Hunt, N. T.; Tuttle, T. *J. Phys. Chem. Lett.,* **2011,** *2*, 2380.

243. Lee, O.-S.; Liu, Y.; Schatz, G. C. *J. Nanopart. Res.,* **2012,** *14*, 1.

244. Lee, O.-S.; Stupp, S. I.; Schatz, G. C. *J. Am. Chem. Soc.,* **2011,** *133*, 3677.

245. Fletcher, J. M.; Harniman, R. L.; Barnes, F. R. H.; Boyle, A. L.; Collins, A.; Mantell, J.; Sharp, T. H.; Antognozzi, M.; Booth, P. J.; Linden, N.; Miles, M. J.; Sessions, R. B.; Verkade, P.; Woolfson, D. N. *Science,* **2013,** *340*, 595.

246. Yu, T.; Lee, O. S.; Schatz, G. C. *J. Phys. Chem. A,* **2013,** *117*, 7453.

247. Yu, T.; Schatz, G. C. *J. Phys. Chem. B,* **2013,** *117*, 9004.

248. Mu, X. J.; Eckes, K. M.; Nguyen, M. M.; Suggs, L. J.; Ren, P. Y. *Biomacromolecules,* **2012,** *13*, 3562.

249. Eckes, K. M.; Mu, X.; Ruehle, M. A.; Ren, P.; Suggs, L. J. *Langmuir,* **2014,** *30*, 5287.

250. Castelletto, V.; Moulton, C.; Cheng, G.; Hamley, I.; Hicks, M. R.; Rodger, A.; López-Pérez, D. E.; Revilla-López, G.; Alemán, C. *Soft Matter,* **2011,** *7*, 11405.

251. Lopez-Perez, D. E.; Revilla-Lopez, G.; Hamley, I. W.; Aleman, C. *Soft Matter,* **2013,** *9*, 11021.

252. Foster, J. S.; Zurek, J. M.; Almeida, N. M.; Hendriksen, W. E.; le Sage, V. A.; Lakshminarayanan, V.; Thompson, A. L.; Banerjee, R.; Eelkema, R.; Mulvana, H. *J. Am. Chem. Soc.,* **2015**.

253. Mayans, E.; Ballano, G.; Casanovas, J.; Díaz, A.; Pérez‐Madrigal, M. M.; Estrany, F.; Puiggalí, J.; Cativiela, C.; Alemán, C. *Chem.-Eur. J.,* **2015,** *21*, 16895.

254. Raeburn, J.; Adams, D. J. *Chem. Commun.,* **2015,** *51*, 5170.

255. Hanabusa, K.; Matsumoto, M.; Kimura, M.; Kakehi, A.; Shirai, H. *J. Colloid Interface Sci.,* **2000,** *224*, 231.

256. Terech, P. *Langmuir,* **2009,** *25*, 8370.

257. Zhu, P.; Yan, X.; Su, Y.; Yang, Y.; Li, J. *Chem.-Eur. J.,* **2010,** *16*, 3176.

258. Moffat, J. R.; Smith, D. K. *Chem. Commun.,* **2008**, 2248.

259. Terech, P.; Sangeetha, N. M.; Maitra, U. *J. Phys. Chem. B,* **2006,** *110*, 15224.

260. Adams, D. J.; Morris, K.; Chen, L.; Serpell, L. C.; Bacsa, J.; Day, G. M. *Soft Matter,* **2010,** *6*, 4144.

261. Wang, Y.; Tang, L.; Yu, J. *Cryst. Growth Des.,* **2008,** *8*, 884.

262. De Rudder, J.; Bergé, B.; Berghmans, H. *Macromol. Chem. Phys.,* **2002,** *203*, 2083.

263. Roy, B.; Bairi, P.; Nandi, A. K. *Soft Matter,* **2012,** *8*, 2366.

264. Houton, K. A.; Morris, K. L.; Chen, L.; Schmidtmann, M.; Jones, J. T. A.; Serpell, L. C.; Lloyd, G. O.; Adams, D. J. *Langmuir,* **2012,** *28*, 9797.

265. Cui, H.; Muraoka, T.; Cheetham, A. G.; Stupp, S. I. *Nano Lett.,* **2009,** *9*, 945.

266. Aggeli, A.; Bell, M.; Boden, N.; Keen, J. N.; Knowles, P. F.; McLeish, T. C. B.; Pitkeathly, M.; Radford, S. E. *Nature,* **1997,** *386*, 259.

267. Tidhar, Y.; Weissman, H.; Wolf, S. G.; Gulino, A.; Rybtchinski, B. *Chem.-Eur. J.,* **2011,** *17*, 6068.

268. Dixit, N. M.; Zukoski, C. F. *Phys. Rev. E,* **2003,** *67*, 061501.

269. Cheng, S.; Stevens, M. J. *Soft Matter,* **2014,** *10*, 510.

270. Hartman, P.; Bennema, P. *J. Cryst. Growth,* **1980,** *49*, 145.

271. Weissbuch, I.; Popovitz-Biro, R.; Lahav, M.; Leiserowitz, L. *Acta Crystallogr., Sect. B: Struct. Sci.,* **1995,** *51*, 115.

272. Privalov, P. L.; Makhatadze, G. I. *J. Mol. Biol.,* **1993,** *232*, 660.

273. Raeburn, J.; Pont, G.; Chen, L.; Cesbron, Y.; Levy, R.; Adams, D. J. *Soft Matter,* **2012,** *8*, 1168.

274. Mackerell, A.; Bashford, D.; Bellot, M.; Dunbrack, R.; Evanseck, J.; Field, M.; Gao, J.; Guo, H.; Ha, S.; Joseph-McCarthy, D.; Kuchnir, L.; Kuczera, K.; Lau, F. T. K.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D. T.; Prodhom, B.; E. Reiher, W. E.; Roux, B.; Schlenkrich, M.; Smith, J. C.; Stote, R.; Straub, J.; Watanabe, M.; Wio´rkiewicz-Kuczera, J.; Yin, D.; Karplus, M. *J. Phys. Chem. B,* **1998,** *102*, 3586.

275. Brooks, B. R.; III, C. L. B.; Jr, A. D. M.; Nilsson, L.; Petrella, R. J.; Roux, B.; Won, Y.; Archontis, G.; Bartels, C.; Boresch, S.; Caflisch, A.; Caves, L.; Cui, Q.; Dinner, A. R.; Feig, M.; Fischer, S.; Gao, J.; Hodoscek, M.; Im, W.; Kuczera, K.; Lazaridis, T.; Ma, J.; Ovchinnikov, V.; Paci, E.; Pastor, R. W.; Post, C. B.; Pu, J. Z.; Schaefer, M.; Tidor, B.; Venable, R. M.; Woodcock, H. L.; Wu, X.; Yang, W.; York, D. M.; Karplus, M. *J. Comput. Chem.,* **2009,** *30*, 1545.

276. MacKerell, A. D.; Banavali, N.; Foloppe, N. *Biopolymers,* **2000,** *56*, 257.

277. Vanommeslaeghe, K.; Hatcher, E.; Acharya, C.; Kundu, S.; Zhong, S.; Shim, J.; Darian, E.; Guvench, O.; Lopes, P.; Vorobyov, I.; Mackerell, A. D. *J. Comput. Chem.,* **2010,** *31*, 671.

278. Prates, E. T.; Souza, P. C.; Pickholz, M.; Skaf, M. S. *Int. J. Quantum Chem.,* **2011,** *111*, 1339.

279. Hansson, A.; Souza, P. C.; Silveira, R. L.; Martinez, L.; Skaf, M. S. *Int. J. Quantum Chem.,* **2011,** *111*, 1346.

280. Grimme, S. *J. Comput. Chem.,* **2006,** *27*, 1787.

281. Schäfer, A.; Horn, H.; Ahlrichs, R. *J. Chem. Phys.,* **1992,** *97*, 2571.

282. *TURBOMOLE*, 6.3.1; TURBOMOLE GmbH: 2012.

283. Ahlrichs, R.; Bär, M.; Häser, M.; Horn, H.; Kölmel, C. *Chem. Phys. Lett.,* **1989,** *162*, 165.

284. Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R. D.; Kalé, L.; Schulten, K. *J. Comput. Chem.,* **2005,** *26*, 1781.

285. Hanwell, M. D.; Curtis, D. E.; Lonie, D. C.; Vandermeersch, T.; Zurek, E.; Hutchison, G. R. *J. Cheminform.,* **2012,** *4*, 17.

286. Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H. P.; Izmaylov, A. F.; Bloino, J.; Zheng, G.; Sonnenberg, J. L.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Montgomery Jr., J. A.; Peralta, J. E.; Ogliaro, F.; Bearpark, M. J.; Heyd, J.; Brothers, E. N.; Kudin, K. N.; Staroverov, V. N.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A. P.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, N. J.; Klene, M.; Knox, J. E.; 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.; Martin, R. L.; Morokuma, K.; Zakrzewski, V. G.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Dapprich, S.; Daniels, A. D.; Farkas, Ö.; Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fox, D. J. *Gaussian 09*, Gaussian, Inc.: Wallingford, CT, USA, 2009.

287. Becke, A. D. *J. Chem. Phys.,* **1993,** *98*, 5648.

288. Lee, C.; Yang, W.; Parr, R. G. *Phys. Rev. B,* **1988,** *37*, 785.

289. Hehre, W. J.; Ditchfield, R.; Pople, J. A. *J. Chem. Phys.,* **1972,** *56*, 2257.

290. Hariharan, P. C.; Pople, J. A. *Theor. Chim. Acta,* **1973,** *28*, 213.

291. Sangster, J. *J. Phys. Chem. Ref. Data,* **1989,** *18*, 1111.

292. Lodge, K. B. *J. Chem. Eng. Data,* **1999,** *44*, 1321.

293. Edelbach, D. J.; Lodge, K. B. *Phys. Chem. Chem. Phys.,* **2000,** *2*, 1763.

294. Zwanzig, R. W. *J. Chem. Phys.,* **1954,** *22*, 1420.

295. Zwanzig, R. W. *J. Chem. Phys.,* **1955,** *23*, 1915.

296. Humphrey, W.; Dalke, A.; Schulten, K. *J. Mol. Graphics,* **1996,** *14*, 33.

297. MacKerell, A. D.; Feig, M.; Brooks, C. L. *J. Comput. Chem.,* **2004,** *25*, 1400.

298. Feller, S. E.; Zhang, Y.; Pastor, R. W.; Brooks, B. R. *J. Chem. Phys.,* **1995,** *103*, 4613.

299. Beutler, T. C.; Mark, A. E.; van Schaik, R. C.; Gerber, P. R.; van Gunsteren, W. F. *Chem. Phys. Lett.,* **1994,** *222*, 529.

300. Pitera, J. W.; van Gunsteren, W. F. *Mol. Simul.,* **2002,** *28*, 45.

301. Chipot, C.; Pohorille, A., *Free energy calculations*. Springer: 2007.

302. Lee, C.; Scott, H. *J. Chem. Phys.,* **1980,** *73*, 4591.

303. Liu, P.; Dehez, F. o.; Cai, W.; Chipot, C. *J. Chem. Theory Comput.,* **2012,** *8*, 2606.

304. Ryan, D. M.; Doran, T. M.; Anderson, S. B.; Nilsson, B. L. *Langmuir,* **2011,** *27*, 4029.

305. Jonkheijm, P.; Schoot, P. v. d.; Schenning, A. P. H. J.; Meijer, E. W. *Science,* **2006,** *313*, 80.

306. Das, A. K.; Hirst, A. R.; Ulijn, R. V. *Faraday Discuss.,* **2009,** *143*, 293.

307. Berdugo, C.; Nalluri, S. K. M.; Javid, N.; Escuder, B.; Miravet, J. F.; Ulijn, R. V. *ACS Appl. Mater. Interfaces,* **2015,** *7*, 25946.

308. Li, J.; Nowak, P.; Otto, S. *J. Am. Chem. Soc.,* **2013,** *135*, 9222.

309. Carnall, J. M.; Waudby, C. A.; Belenguer, A. M.; Stuart, M. C.; Peyralans, J. J.-P.; Otto, S. *Science,* **2010,** *327*, 1502.

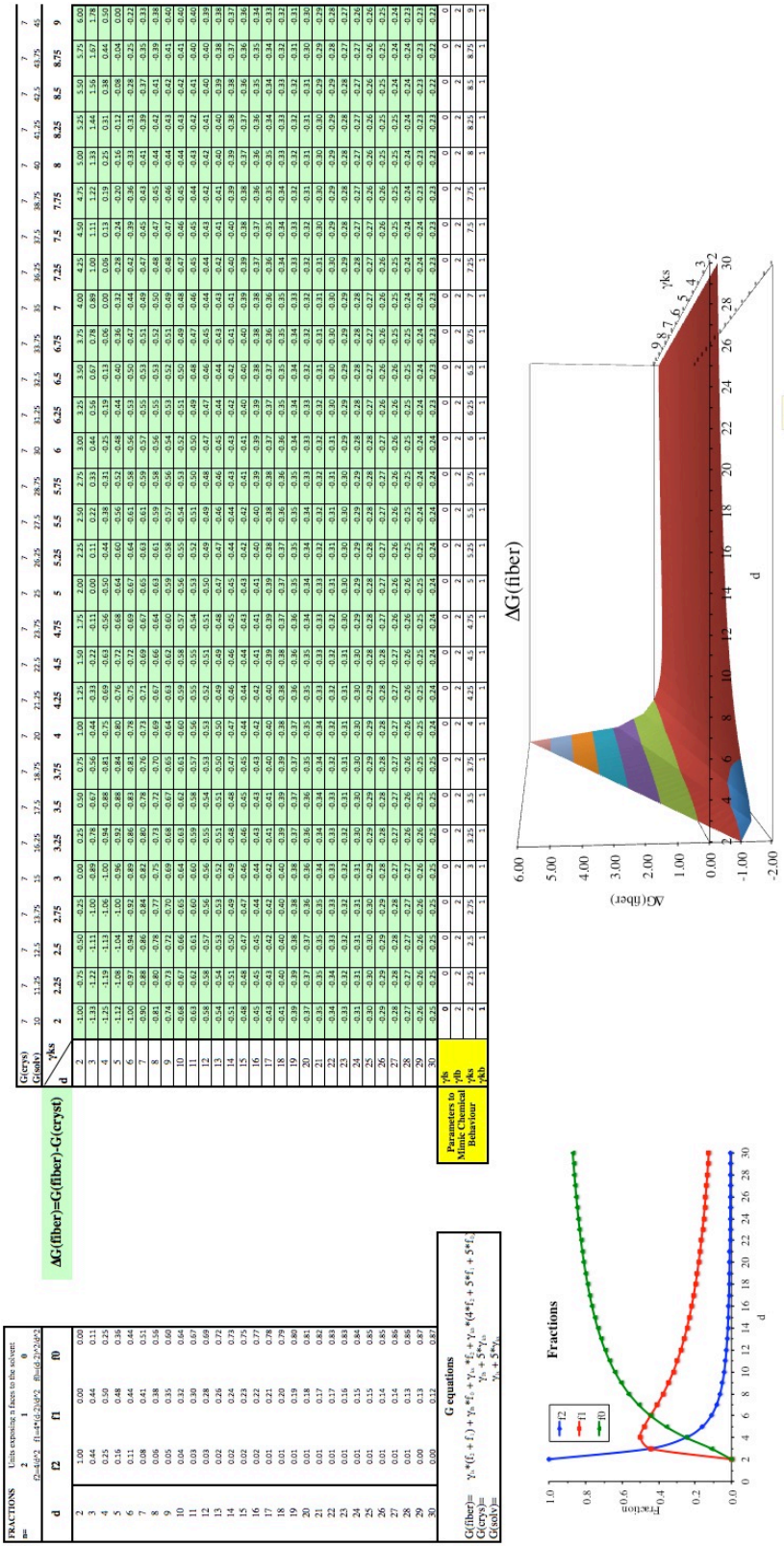
310. Minn, F. L.; Pinion, J. P.; Filipescu, N. *J. Phys. Chem.,* **1971,** *75*, 1794.

311. White, S. H.; Wimley, W. C. *Biochim. Biophys. Acta Rev. Biomemb.,* **1998,** *1376*, 339.

312. Hess, B.; Kutzner, C.; van der Spoel, D.; Lindahl, E. *J. Chem. Theory Comput.,* **2008,** *4*, 435.

10 Appendices

Appendix 1: ΔG_{fibre} Calculation Spreadsheet



Appendix 2: Fmoc CHARMM Topology (NAMD Format)

```
!hydrogens
MASS     1 HGA2     1.00800
MASS     2 HGR61    1.00800
!carbons
MASS     3 CG2O1   12.01100
MASS     4 C2R61   12.01100
MASS     5 CG2R66  12.01100
MASS     6 C2R67   12.01100
MASS     7 C2RC0   12.01100
MASS     8 CG321   12.01100
MASS     9 C3C52   12.01100
!oxygens
MASS    10 OG2D1   15.99940
MASS    11 OG302   15.99940

DEFA FIRS NONE LAST NONE
AUTO ANGLES DIHE

RESI FMO         0.00

GROUP
ATOM C1   C2R61   -0.110 !           H4          H7
ATOM H1   HGR61    0.135 !           |           |
ATOM C2   C2R61   -0.110 !           C4          C7
ATOM H2   HGR61    0.135 !         //  \        /  \\
ATOM C3   C2R61   -0.110 !    H3--C3    C5----C6    C8--H8
ATOM H3   HGR61    0.135 !        |     ||    ||     |
ATOM C4   C2R61   -0.110 !    H2--C2    C13   C11   C9--H9
ATOM H4   HGR61    0.135 !         \\  /  \  /  \  //
ATOM C5   C2R67   -0.100 !           C1    C12   C10
ATOM C6   C2R67   -0.100 !           |    /  \    |
ATOM C7   C2R61   -0.110 !          H1   H21 |  H10
ATOM H7   HGR61    0.135 !                  |
ATOM C8   C2R61   -0.110 !              HF1--CF1--HF2
ATOM H8   HGR61    0.135 !                  |
ATOM C9   C2R61   -0.110 !                 OF1
ATOM H9   HGR61    0.135 !                  |
ATOM C10  C2R61   -0.110 !                 C==OF2
ATOM H10  HGR61    0.135 !                  |
ATOM C11  C2RC0   -0.050 !          TO AMINO ACID RESIDUE NH
ATOM C12  C3C52    0.150
ATOM H21  HGA2     0.090
ATOM C13  C2RC0   -0.050

GROUP
ATOM CF1  CG321    0.240
ATOM HF1  HGA2     0.220
ATOM HF2  HGA2     0.220
ATOM OF1  OG302   -0.920
ATOM C    CG2O1    0.950
```

```
ATOM OF2  OG2D1   -0.850

BOND C1  C2  C2  C3  C3  C4  C4  C5  C5  C6
BOND C6  C7  C7  C8  C8  C9  C9  C10 C10 C11
BOND C11 C12 C12 C13 C13 C1  C5  C13 C6  C11
BOND C1  H1  C2  H2  C3  H3  C4  H4  C7  H7
BOND C8  H8  C9  H9  C10 H10 C12 H21 C12 CF1
BOND CF1 HF1 CF1 HF2 CF1 OF1 OF1 C
DOUBLE C   OF2
BOND C   +N

IC C5   C13  C1    C2    0.0000    0.00    0.00    0.00   0.0000
IC C1   C13  C5    C4    0.0000    0.00    0.00    0.00   0.0000
IC C13  C5   C4    C3    0.0000    0.00    0.00    0.00   0.0000
IC C1   C5   *C13  C12   0.0000    0.00  180.00    0.00   0.0000
IC C4   C13  *C5   C6    0.0000    0.00  180.00    0.00   0.0000
IC C13  C5   C6    C11   0.0000    0.00    5.00    0.00   0.0000
IC C5   C11  *C6   C7    0.0000    0.00  180.00    0.00   0.0000
IC C7   C6   C11   C10   0.0000    0.00    0.00    0.00   0.0000
IC C11  C6   C7    C8    0.0000    0.00    0.00    0.00   0.0000
IC C6   C11  C10   C9    0.0000    0.00    0.00    0.00   0.0000
IC C13  C2   *C1   H1    0.0000    0.00  180.00    0.00   0.0000
IC C1   C3   *C2   H2    0.0000    0.00  180.00    0.00   0.0000
IC C2   C4   *C3   H3    0.0000    0.00  180.00    0.00   0.0000
IC C3   C5   *C4   H4    0.0000    0.00  180.00    0.00   0.0000
IC C6   C8   *C7   H7    0.0000    0.00  180.00    0.00   0.0000
IC C7   C9   *C8   H8    0.0000    0.00  180.00    0.00   0.0000
IC C8   C10  *C9   H9    0.0000    0.00  180.00    0.00   0.0000
IC C9   C11  *C10  H10   0.0000    0.00  180.00    0.00   0.0000
IC C11  C13  *C12  H21   0.0000    0.00  120.00    0.00   0.0000
IC C11  C13  *C12  CF1   0.0000    0.00 -120.00    0.00   0.0000
IC C12  CF1  C11   C10   0.0000    0.00  180.00    0.00   0.0000
IC C12  CF1  C13   C1    0.0000    0.00  180.00    0.00   0.0000
IC CF1  C11  *C12  C13   0.0000    0.00  120.00    0.00   0.0000
IC OF1  CF1  C12   HF1   0.0000    0.00  180.00    0.00   0.0000
IC OF1  HF2  *CF1  C12   0.0000    0.00  180.00    0.00   0.0000
IC OF2  +N   *C    OF2   0.0000    0.00  180.00    0.00   0.0000
IC +N   C    OF1   CF1   0.0000    0.00  180.00    0.00   0.0000
IC C    OF1  CF1   HF1   0.0000    0.00    0.00    0.00   0.0000
IC C    OF1  CF1   HF2   0.0000    0.00  120.00    0.00   0.0000
IC C    OF1  CF1   C12   0.0000    0.00  240.00    0.00   0.0000
```

Appendix 3: Parameters (NAMD Format)

```
BONDS
C2R61 HGR61  340.00    1.0800
C2R61 C2R61  305.00    1.3750
C2R61 C2R67  305.00    1.3750
C2R61 C2RC0  300.00    1.3600
C2R67 C2R67  300.00    1.4900
C2R67 C2RC0  300.00    1.4200
C2RC0 C3C52  305.00    1.5200
C3C52 HGA2   307.00    1.1000
CG321 C3C52  195.00    1.5180
CG321 HGA2   309.00    1.1110
CG321 OG302  320.00    1.4400
CG2O1 OG302  340.00    1.4300
CG2O1 OG2D1  620.00    1.2300
CG2O1 NH1    370.00    1.3450
NH1   H      440.00    0.9970
NH1   CT1    320.00    1.4300
CT1   CD     200.00    1.5220
CT2   CT1    222.50    1.5380
HB    CT1    330.00    1.0800
HA    CT2    309.00    1.1110
OH1   CT2    428.00    1.4200
OH1   H      545.00    0.9600
OB    CD     750.00    1.2200
OH1   CD     230.00    1.4000

ANGLES

CG2O1  OG302  CG321   55.00  109.00
OG2D1  CG2O1  OG302   90.00  125.90
C2RC0  C3C52  CG321   38.00  114.00
C3C52  CG321  OG302   75.70  115.10
C3C52  CG321  HGA2    38.50  115.10
CG321  C3C52  HGA2    38.50  106.80
NH1    CG2O1  OG302   80.00  116.50
NH1    CG2O1  OG2D1   80.00  122.50
CG2O1  NH1    H       34.00  123.00
C2R61  C2R61  C2R61   40.00  120.00   35.00 2.41620
C2R61  C2R61  C2R67   40.00  120.00
C2R61  C2R61  C2RC0   50.00  120.00
C2R61  C2R61  HGR61   30.00  120.00   22.00 2.15250
C2R67  C2R61  HGR61   30.00  120.00
C2RC0  C2R61  HGR61   30.00  120.00   22.00 2.14600
C2R61  C2R67  C2R67   40.00  120.00
C2R61  C2R67  C2RC0   50.00  120.00
C2R67  C2R67  C2RC0   55.00  110.00
C2R61  C2RC0  C2R67   50.00  120.00
C2R61  C2RC0  C3C52   60.00  130.00
C2R67  C2RC0  C3C52  110.00  110.00
OG302  CG321  HGA2    60.00  109.50
```

```
HGA2   CG321  HGA2    35.50  109.00    5.40 1.80200
C2RC0  C3C52  C2RC0   40.00   95.00
C2RC0  C3C52  HGA2    38.00  114.00
HB     CT1    CD      50.00  109.50
H      OH1    CD      55.00  115.00
OB     CD     CT1     70.00  125.00   20.00 2.44200
CT1    CD     OH1     55.00  110.50
OH1    CD     OB      50.00  123.00  210.00 2.26200
NH1    CT1    HB      48.00  108.00
NH1    CT1    CD      50.00  107.00
H      NH1    CT1     35.00  117.00
HA     CT2    CT1     33.43  110.10   22.53 2.17900
OH1    CT2    CT1     75.70  110.10
CG2O1  NH1    CT1     50.00  120.00
H      OH1    CT2     57.50  106.00
HB     CT1    CT2     35.00  111.00
CT2    CT1    CD      52.00  108.00
HA     CT2    HA      35.50  109.00    5.40 1.80200
OH1    CT2    HA      45.90  108.89
NH1    CT1    CT2     70.00  113.50

DIHEDRALS
C2R61  C2RC0  C3C52  CG321   0.5000  3   0.00
C2R67  C2RC0  C3C52  CG321   0.5000  3   0.00
C2RC0  C3C52  CG321  HGA2    0.1950  3   0.00
C2RC0  C3C52  CG321  OG3O2   0.1950  3   0.00
C3C52  CG321  OG3O2  CG2O1   0.0000  3   0.00
HGA2   C3C52  CG321  HGA2    0.1400  3   0.00
HGA2   C3C52  CG321  OG3O2   0.1400  3   0.00
CG321  OG3O2  CG2O1  OG2D1   0.9000  1   0.00
CG321  OG3O2  CG2O1  OG2D1   2.8500  2 180.00
HGA2   CG321  OG3O2  CG2O1   2.0500  2 180.00
OG2D1  CG2O1  C2R61  C2RC0   1.0000  2 180.00
C2R61  C2R61  C2R61  C2R61   3.1000  2 180.00
C2R61  C2R61  C2R61  C2R67   3.1000  2 180.00
C2R61  C2R61  C2R61  C2RC0   3.0000  2 180.00
C2R61  C2R61  C2R61  HGR61   4.2000  2 180.00
C2R67  C2R61  C2R61  HGR61   4.2000  2 180.00
C2RC0  C2R61  C2R61  HGR61   3.0000  2 180.00
HGR61  C2R61  C2R61  HGR61   2.4000  2 180.00
C2R61  C2R61  C2R67  C2R67   3.1000  2 180.00
C2R61  C2R61  C2R67  C2RC0   0.2500  2 180.00
OG3O1  C2R61  C2R67  C2R61   3.1000  2 180.00
OG3O1  C2R61  C2R67  C2R67   3.1000  2 180.00
HGR61  C2R61  C2R67  C2R61   4.2000  2 180.00
HGR61  C2R61  C2R67  C2R67   4.2000  2 180.00
C2R61  C2R61  C2RC0  C2R67   0.2500  2 180.00
C2R61  C2R61  C2RC0  C3C52   0.0000  2 180.00
C2R61  C2R61  C2RC0  C2R67   0.2500  2 180.00
HGR61  C2R61  C2RC0  C3C52   0.0000  2 180.00
C2R61  C2R67  C2R67  C2R61   0.8900  2 180.00
C2R61  C2R67  C2R67  C2RC0   2.0000  2 180.00
```

```
C2RC0  C2R67  C2R67  C2RC0   1.5000  2 180.00
C2R61  C2R67  C2RC0  C2R61   0.0500  2 180.00
C2R61  C2R67  C2RC0  C3C52   6.7500  2 180.00
C2R67  C2R67  C2RC0  C2R61   3.5000  2 180.00
C2R67  C2R67  C2RC0  C3C52   5.0000  3 180.00
C2R61  C2RC0  C3C52  C2RC0   0.9000  3   0.00
C2R61  C2RC0  C3C52  HGA2    0.5000  3 180.00
C2R67  C2RC0  C3C52  C2RC0   0.7500  3 180.00
C2R67  C2RC0  C3C52  HGA2    0.5000  3   0.00
HGR61  C2R61  C2R67  C2RC0   3.0000  2 180.00
HGR61  C2R61  C2RC0  C2R67   3.0000  2 180.00
CG321  OG302  CG2O1  NH1     0.6000  1 180.00
CG321  OG302  CG2O1  NH1     2.0000  2 180.00
OG302  CG2O1  NH1    H       2.5000  2 180.00
OG2D1  CG2O1  NH1    H       2.5000  2 180.00
OG302  CG2O1  NH1    CT1     1.6000  1 180.00
OG2D1  CG2O1  NH1    CT1     2.5000  2 180.00
CG2O1  NH1    CT1    CT2     0.0000  1 180.00
CG2O1  NH1    CT1    CD      0.0000  1 180.00
CG2O1  NH1    CT1    HB      0.0000  1 180.00
H      NH1    CT1    CT2     0.0000  1   0.00
H      NH1    CT1    CD      0.0000  1   0.00
HB     CT1    NH1    H       0.0000  1   0.00
H      OH1    CT2    CT1     1.3000  1   0.00
H      OH1    CT2    CT1     0.3000  2   0.00
H      OH1    CT2    CT1     0.4200  3   0.00
X      CT1    CT2    X       0.2000  3   0.00
X      CD     OH1    X       2.0500  2 180.00
X      CT1    CD     X       0.0000  6 180.00
X      CT2    OH1    X       0.1400  3   0.00

IMPROPERS

CG2O1  X      X      OG2D1    120.00  0     0.00
C2R61  X      X      HGR61     15.00  0     0.00
NH1    X      X         H      20.00  0     0.00
OB     X      X        CD     100.00  0     0.00

NONBONDED nbxmod  5 atom cdiel shift vatom vdistance vswitch -
cutnb 14.0 ctofnb 12.0 ctonnb 10.0 eps 1.0 e14fac 1.0 wmin 1.5

HGA2   0.0  -0.0350  1.3400
HGR61  0.0  -0.0300  1.3582
CG2O1  0.0  -0.1100  2.0000
C2R61  0.0  -0.0700  1.9924
C2R67  0.0  -0.0700  1.9924
C2RC0  0.0  -0.0990  1.8600
CG321  0.0  -0.0560  2.0100   0.0 -0.01 1.90
C3C52  0.0  -0.0600  2.0200   0.0 -0.01 1.90
OG2D1  0.0  -0.1200  1.7000   0.0 -0.12 1.40
OG302  0.0  -0.1000  1.6500
H      0.0  -0.0460  0.2245
```

```
NH1     0.0  -0.2000  1.8500   0.0 -0.20 1.55
HA      0.0  -0.0220  1.3200
HB      0.0  -0.0220  1.3200
OH1     0.0  -0.1521  1.7700
OB      0.0  -0.1200  1.7000   0.0 -0.12 1.40
CT1     0.0  -0.0200  2.2750   0.0 -0.01 1.90
CT2     0.0  -0.0550  2.1750   0.0 -0.01 1.90
CD      0.0  -0.0700  2.0000

END
```