Structure, Dynamics and Applications of Peptide-based Nanomaterials

A thesis submitted to the University of Strathclyde for the degree of Doctor of Philosophy

Pim W. J. M. Frederix

Biomolecular & Chemical Physics

Pure & Applied Chemistry

2013

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.’

Signed:

Date:

Table of Contents

List of abbreviations....10

Acknowledgements....12

Abstract....14

Chapter 1: Introduction and layout of the thesis....16

1.1 Self-assembly and peptides....17

1.2 Layout of the thesis....19

1.3 References....21

Chapter 2: Background: Self-Assembly of Biomolecules....22

2.1 Molecular self-assembly....23

2.1.1 Thermodynamics and key molecular interactions.....23

2.1.2 Self-assembly as a concept in materials science....26

2.1.3 Gelation and definitions....27

2.2 Peptide hydrogels....28

2.2.1 Overview of peptide gelators....28

2.2.2 Preparation of hydrogels: biocatalytic methods....30

2.3 Practical computational approaches to biomolecular self-assembly I: Molecular Dynamics simulations....32

2.3.1 Introduction to computational modelling of self-assembly....32

2.3.2 Principles of Molecular Dynamics....34

2.3.3 Choice of force field....36

2.3.4 System setup....38

2.3.5 Running an MD simulation 41

2.3.6 All-atom Molecular Dynamics for LMW biomolecular self-assembly 42

2.3.7 Coarse-Grain Molecular Dynamics 50

2.4 Practical computational approaches to self-assembly II: Electronic structure methods 55

2.5 Conclusion and outlook 58

2.6 References 59

Chapter 3: Infrared spectroscopy as a structural diagnostic tool for β-sheets in self-assembling aromatic peptide amphiphiles 69

3.1 Abstract 70

3.2 Introduction and objectives 71

3.2.1 IR Spectroscopy and vibrational energy levels 71

3.2.2 IR spectroscopy of proteins and peptides 75

3.2.3 IR spectroscopy in small peptide self-assembly and objectives 78

3.3 Experimental details 82

3.3.1 Formation of gels 82

3.3.2 Fourier transform infrared spectroscopy 82

3.3.3 Vibrational frequency calculations 83

3.4 Results and discussion 85

3.4.1 FTIR spectroscopy and isotope labelling 85

3.4.2 DFT calculations 88

3.4.3 Comparison between experimental and calculated spectra 94

3.4 Conclusion 95

3.5 References 96

Chapter 4: Sequence / structure relations in self-assembling peptide amphiphile nanostructures, studied by all-atom Molecular Dynamics and infrared spectroscopy 100

4.1 Abstract 101

4.2 Introduction and objectives 102

4.2.1 Case study 1: Fmoc-X_1X_2-OMe 103

4.2.2 Case study 2: Core/shell nanostructures for cell culture applications 104

4.2.3 Case study 3: Self-assembly of FF and IF 106

4.3 Methods and materials 107

4.3.1 Hydrogel preparation 107

4.3.2 Fourier transform infrared spectroscopy 110

4.3.3 Atomistic Molecular Dynamics simulations 110

4.3.4 High Pressure Liquid Chromatography 112

4.3.5 Wide-Angle X-ray Scattering 112

4.3.6 Transmission Electron Microscopy 113

4.3.7 AFM 113

4.4 Results and Discussion 114

4.4.1 Case study 1: Fmoc-X_1X_2-OMe 114

4.4.2 Case study 2: Core/shell nanostructures 123

4.4.3 Case study 3: Self-assembly of FF and IF 133

4.5 Discussion and future directions 139

4.6 Conclusion 142

4.7 References 144

Chapter 5: Virtual screening for peptide aggregation: towards predictive tools for peptide self-assembly 149

5.1 Abstract 150

5.2 Introduction and objectives 151

5.2.1 Dipeptides 151

5.2.2 Tripeptides 154

5.3 Coarse-grain force field and simulation details 156

5.4 Dipeptide results and discussion 160

5.4.1 Dipeptide Aggregation Propensities 160

5.4.2 Dipeptide morphological features 162

5.4.3 Dipeptide extended simulations 166

5.5 Tripeptides 169

5.5.1 Scoring method 169

5.5.2 Tripeptide aggregation propensities and evaluation of the AP_H score. 171

5.5.3 Tripeptide structural features 178

5.6 Directions for future research 182

5.7 Conclusion 184

5.6 References 187

Chapter 6: Applications of biomaterials in energy research: Encapsulation of [FeFe]-Hydrogenase Model Compounds in Peptide Hydrogels 191

6.1 Abstract 192

6.2 Introduction and objectives 194

6.2.1 Hydrogen production by hydrogenase enzyme mimics 194

6.2.2 Time-Resolved Infrared Spectroscopy 200

6.3 Experimental details 203
6.3.1 Materials 203
6.3.2 Preparation of gels encapsulating hydrogenase mimics 204
6.3.3 Time-resolved infrared spectroscopy 204
6.3.4 Fourier transform infrared spectroscopy 207
6.3.5 Steady-state UV excitation with IR probe 207
6.3.6 Hydrogenase mimic vibrational frequency calculations 208
6.3.7 Fluorescence spectroscopy 208
6.3.8 Atomic Force Microscopy 209
6.4 Results and discussion 209
6.4.1 FTIR experiments 210
6.4.2 Gel melting experiments 215
6.4.3 TRIR experiments 218
6.4.4 (μ-pdt)$Fe_2(CO)_4(PMe_3)_2$ in a hydrogel with hydrophilic side chains 224
6.4.5 Towards light-driven formation of hydrogen 229
6.5 Conclusion 237
6.6 References 239

Chapter 7: Conclusions and outlook 244

Appendix 1: Typical input parameters for NAMD, GROMACS, Gaussian and Turbomole calculations.....249

Appendix 2: DFT frequency calculation results for additional Fmoc-AA and Fmc-AA conformations.....259

Appendix 3: Additional TEM images of Fmoc-SF-OMe nanostructures.....268

Appendix 4: Simulations results SF, SL, TL and TF.....269

Appendix 5: MD parameters for Fmoc-group.....271

Appendix 6: AP and AP_H scores.....274

Appendix 7: Links to movies of full simulation results.....283

List of publications by the author

Papers are listed in chronological order. Several of these published papers have been used in this thesis, indicated by the chapter number. Papers in grey are not directly relevant to this thesis.

Sep 09 — P.W.J.M. Frederix, C.-H. Yang, G.C. Groenenboom, D.H. Parker, K. Alnama, C.M. Western, A.J. Orr-Ewing, *Photodissociation imaging of diatomic sulfur (S_2)*, J. Phys. Chem. A, 2009, 113, 14995-15005.

Sep 11 Ch. 6 — P.W.J.M. Frederix, R.V. Ulijn, N.T. Hunt, T. Tuttle, *Virtual Screening for Dipeptide Aggregation: Toward Predictive Tools for Peptide Self-Assembly*, J. Phys. Chem. Lett., 2011, 2, 2380-2384.

Sep 11 Ch. 5 — M. Hughes, H. Xu, P.W.J.M. Frederix, A.M. Smith, N.T. Hunt, T. Tuttle, I.A. Kinloch and R.V. Ulijn, *Biocatalytic Self-assembly of 2D Peptide-based Nanostructures*, Soft Matter, 2011, 7, 10032-10038

Jan 12 Ch. 7 — R. Kania, P.W.J.M. Frederix, J.A. Wright, R.V Ulijn, C.J. Pickett, N.T. Hunt, *Solution-phase photochemistry of a [FeFe]hydrogenase model compound: Evidence of photoinduced isomerisation*, J. Chem. Phys., 2012, 136, 044521-9

Mar 12 Ch. 7 — P.W.J.M. Frederix, R. Kania, J.A. Wright, D.A. Lamprou, R.V Ulijn, C.J. Pickett, N.T. Hunt, *Encapsulating [FeFe]-hydrogenase model compounds in peptide hydrogels dramatically modifies stability and photochemistry,* Dalton Trans. 2012, 41, 13112-13119

Apr 12	Ch. 5	M. Hughes, P.W.J.M. Frederix, J. Raeburn, L.S. Birchall, J. Sadownik, F.C. Coomer, I-H. Lin, E.J. Cussen, N.T. Hunt, T. Tuttle, S.J. Webb, D.J. Adams, R.V. Ulijn, *Sequence/structure relationships in aromatic dipeptide hydrogels formed under thermodynamic control by enzyme-assisted self-assembly,* Soft Matter, 2012, 20, 5595-5602
Jun 12		S. Roy, N. Javid, P.W.J.M. Frederix, D.A. Lamprou, A.J. Urquhart, N.T. Hunt, P.J. Halling, R.V. Ulijn, *Dramatic specific ion effect in supramolecular hydrogels*, Chem. Eur. J., 2012, 18, 11723-11731.
Aug 12		K.V. Vidma, P.W.J.M Frederix, D.H. Parker, A.V. Baklanov, *Photodissociation of van der Waals clusters of isoprene with oxygen,C5H8−O2, in the wavelength range 213–277 nm*, J. Chem. Phys. 137, 054305 (2012)
May 13	Ch. 4	P.W.J.M. Frederix, S. Fleming, I. Ramos-Sasselli, N.T. Hunt, R.V. Ulijn, and T. Tuttle, *Assessing the Utility of Infrared Spectroscopy as a Structural Diagnostic Tool for β-Sheets in Self-Assembling Aromatic Peptide Amphiphiles*, Langmuir, 2013, 29, 9510–9515

List of abbreviations

Amino acids:

Single letter code	Three-letter code	Full name	Side chain at pH 7
A	Ala	Alanine	CH_3
C	Cys	Cysteine	CH_2SH
D	Asp	Aspartic acid	CH_2COO^-
E	Glu	Glutamic acid	$C_2H_4COO^-$
F	Phe	Phenylalanine	$CH_2C_6H_5$
G	Gly	Glycine	H
H	His	Histidine	$CH_2C_3H_4N_2^+$
I	Ile	Isoleucine	$CH(CH_3)C_2H_5$
K	Lys	Lysine	$C_4H_8NH_3^+$
L	Leu	Leucine	$C_2H_3(CH_3)_2$
M	Met	Methionine	$C_2H_4SCH_3$
N	Asn	Asparagine	CH_2CONH_2
P	Pro	Proline	cyc-C_3H_6
Q	Gln	Glutamine	$C_2H_4CONH_2$
R	Arg	Arginine	$C_3H_6NHC(NH_2)NH_2^+$
S	Ser	Serine	CH_2OH
T	Thr	Threonine	$CH(OH)CH_3$
V	Val	Valine	$CH_2CH(CH_3)_2$
W	Trp	Tryptophan	$CH_2C_8H_6N$
Y	Tyr	Tyrosine	$CH_2C_6H_4$-p-OH

AA	All-atom
AFM	Atomic Force Microscopy
AP	Aggregation Propensity
ATR	Attenuated Total Reflection
CD	Circular Dichroism
CG	Coarse-grain
CHARMM	Chemistry at HARvard Macromolecular Mechanics
DFT	Density Functional Theory
DLS	Dynamic Light Scattering
FF	Diphenylalanine
Fmc	9-Fluorenylmethylcarbonyl
Fmoc	9-Fluorenylmethyloxycarbonyl
FTIR	Fourier Transform Infrared
HPLC	High Pressure Liquid Chromatography
H-T	Head-to-tail
IR	Infrared
MARTINI	(MARrink's Toolkit INItiative)

MD	Molecular Dynamics
MLCT	Metal-Ligand Charge Transfer
MM	Molecular Mechanics
NAMD	NAnoscale Molecular Dynamics
OPA	Optical Parametric Amplifier
pdt	Propane-dithiolate (-S-C_3H_6-S-)
QM	Quantum Mechanics
REMD	Replica-Exchange Molecular Dynamics
SLS	Static Light Scattering
TDM	Transition Dipole Moment
TEM	Transmission Electron Microscopy
TRIR	Time-Resolved Infrared
UV	Ultraviolet
VCD	Vibrational Circular Dichroism
Vis	Visible (light)
WAXS	Wide Angle X-ray Scattering

Acknowledgements

Written on a beautiful, melancholic Sunday night full of memories, a decent amount of whisky and before all the thesis stress started...

First of all, I would like to thank my PhD supervisors, Neil and Rein. I am especially grateful the have given me trust and flexibility in my project. It was a pleasure (and in my opinion a success) to work with them (it never felt as *for* them). The same goes for Tell, my unofficial third supervisor – a good example how the academic world can be fun too.

However, I also want to take some time to thank everyone who has brought me here, has influenced my life – even though I might never speak to you anymore and you might even have to dig deep to remember me.

Let me start at the beginning: my father Harrie, mother Wilma and brother Rik. Nothing needs to be said there. Love to all the Frederixen and Jans-Bekens, and a special mention goes to the NNC.

Friends through thick and thin: the members of Pimfandischasjo, it is always so great to see you again. San, Leen, Suus, Gil, Jan, Lambros, Martijn, Bazzzz and the Big O: I thank you for your friendship at the EC and afterwards.

My years at Radboud University Nijmegen aren't that far away yet and I do enjoy remembering them. The bainies: such a great group of people, Passio, Hajo-io, Uberpassio, Kwintie, N-bain and Stevie: I admire you and I hope your lives will go to the top N-bainianically. Jaap, Kist, Matthijs, Bram, Mariman, Tempelaars, Frans, Steven, Jarno, Ingrid and other members of the Camelot crew: cheers guys. Boukje, Wieneke and Jilske too – I hope we will keep visiting each other, wherever we go. I should not forget my friend and colleagues from the MLF department: many thanks to Dave Parker and of course Zhichao, Chung-Hsin, Gautam and Ivo, Afric and Zaheed. Kostya: so many thanks; you have been a *fantastische* mentor. I have only stayed at Philips for a short

while, but I won't forget Jos, Anne, Chiara and Samanta (and others!), but most of all Andrei, who I seriously admire as a person. Bas, thank you for being the supervisor I needed.

Thank you Glasgow for bringing me new friends and an awesome experience. I would like to mention the colleagues that have become friends: love to Lynsey, Meghan and Louise for taking me under their wing from the start and giving me a crash course in Glasgwegian. I can't name everyone from the Hunt and Ulijn groups over the years, but thanks for being there and giving my PhD project colour: Daniela, Joy, Chris H., Pier, Marco, Katrin, Danny and Niall. Old Lauren: you're awesome, although you're not Dutch. Thanks John and Bob for the technical support in the lab. My best wishes also for the members of the Tuttlelab: Joe 'Kaiserslautern' C., Zaraschenko, Chris 'Alloa Animal' I., Bhaskar 'You drive me crazy' M., Iván the man, Greg and others: I put my hands up in the air sometimes for you guys. Dan: I'm not a big fan of the term BFF, but I hope it applies.

I would like to give the members of the Strathclyde Volleyball club some credit for enhancing my Glaswegian experience so much as well: Simba, Mr Mike, The Muzzcat, Ali, Bob, Jānis, Kristaps, Robert, RJ, Beth, Steph, Willem, Mirjam, Stefania, Kuba, Lara, Canadian Laura, Scottish Laura, Kermack and of course Alanwelsh, you've all earned a mention for being awesome.

Chris J., Lina, Leyla and my little baboeshka Liudmila: thanks for the homely feelings, I wish you all the best.

All of you, I can only hope my wishes for you will come true.

Finally I should give a tiny wee mention for the most awesome American I have met: Claire. Thank you for your love, patience and even your support with this thesis.

Abstract

Peptide-based nanomaterials have rapidly gained interest over the last 10 years, because of their potential uses in biomedicine, nanotechnology and catalysis. Short peptides can readily self-assemble into ordered structures on the nanometer length scale. The design rules for these constructs are not fully understood yet, which is a key issue in engineering new functional nanomaterials.

This thesis discusses how infrared spectroscopy and computational chemistry can contribute to the understanding of structure and dynamics of peptide-based self-assembled systems. Quantum mechanical calculations will be discussed in the light of determining stacking interactions between aromatic peptide amphiphiles and predicting their infrared absorptions bands. Furthermore, in case studies of various peptides, both IR spectroscopy and all-atom MD are demonstrated to be sensitive to small changes in the supramolecular structure as a consequence of variations in the amino acid side chains of the peptides under study. However, currently all-atom MD is still somewhat limited by computational costs and the specific assignments of the bands in IR spectra are not always clear.

With a lower level of detail in MD simulations, information relevant to the time and length scales of the process of self-assembly can be obtained. The coarse-grain simulation protocol developed here shows good agreement with experiments in terms of predicting a peptide's propensity to aggregate and can reproduce morphological features of self-assembled systems.

Finally, these peptide-based materials are applied in encapsulating an enzyme active site mimic that is of relevance for the cheap, environmentally friendly production of hydrogen as a fuel. Using time-resolved infrared spectroscopy it is shown that the hydrogel environment formed by a short, amphiphilic peptide protected the enzyme

mimic both from degradation by oxygen and UV irradiation. This type of peptide scaffolding for hydrogenase mimics has potential in creating an artificial enzyme for the reversible oxidation of hydrogen.

Chapter 1: Introduction and layout of the thesis

1.1 Self-assembly and peptides

The process in which small components spontaneously form ordered structures is called self-assembly. The importance of this process in biology was realised as early as the 1969 by Kushner, who described the process of self-assembly in the context of virus capsids, phages, microtubules, muscle fibres, etc.[1] and has since then spread as a concept through nanotechnology and materials science. The desire of scientists to mimic Nature's efficiency in creating these structures lies at the origin of the development of a particularly interesting class of self-assembling systems: small peptides. Peptides are short chains of a few amino acid monomers (which are often termed 'the building blocks of life'). In the 1990s, it was first discovered that certain small peptides can interact strongly with each other at sufficiently high concentration in water and form ordered nanostructures in this way, which are different from structures found by precipitation or crystallisation. Zhang *et al*. and Ghadiri and co-workers were among the first to report this type of systems.[2,3] These structures often have nanometer dimensions, and are held together by non-covalent, reversible interactions of, e.g., ionic, hydrogen bonding or dispersive nature. This allows them to self-repair or undergo dynamical changes upon variations in the environment.

Thanks to these properties, the wide variety of chemical functionality and arising morphologies and the envisioned applications in biomedicine and catalysis, the field of peptide self-assembly has been rapidly expanding over the last 10 years as shown by the number of citations per year in Fig. 1.1.

A key paper by Hartgerink *et al*. describing peptide amphiphile nanostructures was published in 2001, but when Reches and Gazit and Silva *et al*. reported on applications in casting metal nanowires[4] and cell culture,[5] respectively, the development in the field was accelerated even further. The first commercial applications as cell culture scaffolds have already materialized, but trials are also ongoing in using peptides as surgical aides in wound healing or mechanical tissue supports.[6]

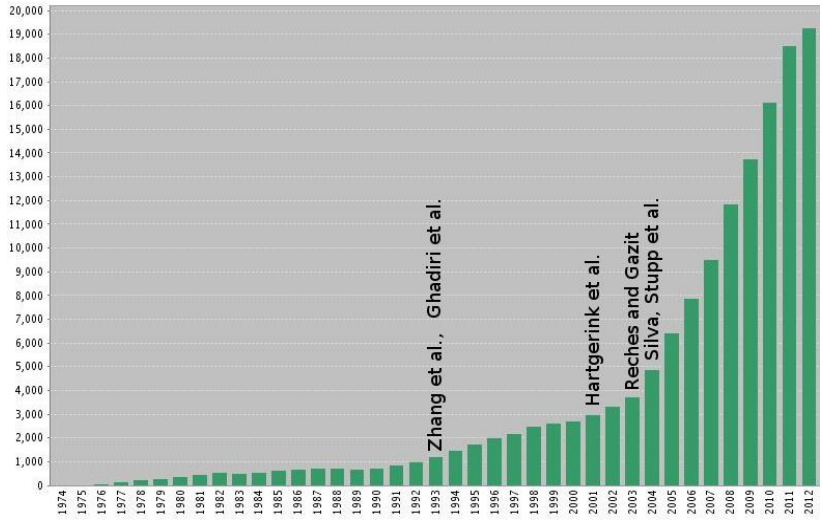


Fig. 1.1 Number of citations of papers containing "peptide" AND "self-assembly" in the Topic per year in the Web of Knowledge database, accessed on 29/04/2013. Key papers are indicated.

Although reports on possible applications and various shapes and sizes of nanostructures are numerous, the molecular architecture of these nanostructures is surprisingly hard to unravel. Microscopy techniques are currently not able to resolve individual molecules for these systems and many scattering techniques (X-rays, neutrons, etc.), usually used for resolving molecular structure, are unsuitable for hydrogels because of their large water content. This thesis aims to shed light on the molecular architecture of peptide hydrogels using infrared (IR) spectroscopy and computational chemistry. IR absorption is sensitive to non-covalent interactions in peptides, which are crucial in determining their supramolecular structure. Computational chemistry can model systems on the molecular scale and can therefore give more valuable insights in the conformation of peptides within a nanostructure. Using this information, work towards applications of peptide nanostructures can then be realised, such as improving the stability and efficiency of catalysts by encapsulating them in fibrous hydrogels.

1.2 Layout of the thesis

In Chapter 2 the detailed background for this thesis is provided, introducing the basic principles of molecular self-assembly. The prerequisites for formation of peptide hydrogels are discussed in particular. In addition, the use of computational methods in the field of self-assembly of small molecules is reviewed in this chapter.

Several experimental (Fourier Transform Infrared (FTIR) spectroscopy, Time-Resolved Infrared (TRIR) spectroscopy) and theoretical methods (Molecular Dynamics (MD) simulations, Density Functional Theory (DFT)) have been used and / or developed further in the work reported in this thesis. It should be noted that other experimental work discussed in this thesis was performed in a collaborative fashion. The author has been responsible for all computational or (time-resolved) infrared experiments. The instances where significant contributions by others have been made are mentioned in the respective chapters.

In the rest of the thesis the results regarding the structure, dynamics and application of peptide nanomaterials are discussed. In Chapter 3, self-assembly is studied at the most detailed molecular level, using quantum mechanical geometry optimizations and frequency calculations to elucidate the actual mode of self-assembly of Fmoc-peptides (Fmoc = 9-fluorenyl methyloxycarbonyl). Moreover, the use of infrared spectroscopy in the field of peptide nanostructures is discussed. Incorrect interpretation of infrared spectra lead to the fact that in spite of years of research, it was still not clear which stacking conformation these self-assembling molecules adopt. The results from new infrared experiments as well as Density Functional Theory (DFT) calculations are reported in order to resolve this issue.

In Chapter 4 the relationship between the amino acid sequence of a peptide and the nanostructure formed upon self-assembly is examined. With a combination of spectroscopic, microscopic and computational methods, it was possible to describe some

design rules with respect to specific nanoscale morphologies. Moreover, details are given regarding the insight into specific molecular arrangements that can be provided by experimental IR evidence and all-atom Molecular Dynamics simulations. .

Chapter 5 concerns the prediction of self-assembly from a given peptide sequence. Using an even coarser approach to computationally modelling self-assembly, a coarse-grain molecular dynamics screening protocol has been developed that can be used to identify interesting candidates for new biomaterials and this has been applied to all combinations of two or three amino acids: 400 dipeptides and 8000 tripeptides.

In Chapter 6 a specific application of peptide nanomaterials in energy research is described. It is shown that encapsulating a hydrogen-producing catalyst, inspired by the active site of the naturally occurring hydrogenase enzyme, significantly improves its stability. (Time-resolved) infrared spectroscopy showed it is more resistant to UV-induced isomerisation and aerobic degradation, even in an aqueous environment.

Finally, the main conclusions from this work and directions for further use of these results are given in Chapter 7, especially in terms of using infrared spectroscopy as an analysis tool, computational chemistry for predicting nanostructure morphologies and applications of hydrogenase mimics and peptide nanomaterials for hydrogen production.

1.3 References

(1) Kushner, D. J. *Bacteriol. Rev.* **1969**, *33*, 302–345.

(2) Zhang, S.; Holmes, T.; Lockshin, C.; Rich, A. *Proc. Natl. Acad. Sci. U.S.A.* **1993**, *90*, 3334 –3338.

(3) Ghadiri, M. R.; Granja, J. R.; Milligan, R. A.; McRee, D. E.; Khazanovich, N. *Nature* **1993**, *366*, 324–327.

(4) Reches, M.; Gazit, E. *Science* **2003**, *300*, 625–627.

(5) Silva, G. A.; Czeisler, C.; Niece, K. L.; Beniash, E.; Harrington, D. A.; Kessler, J. A.; Stupp, S. I. *Science* **2004**, *303*, 1352–1355.

(6) Loh, X. J.; Scherman, O. A. *Polymeric and Self Assembled Hydrogels: From Fundamental Understanding to Applications*; Royal Society of Chemistry, 2012.

Chapter 2: Background: Self-Assembly of Biomolecules

This chapter is aimed at providing an overview of the key idea that connects the various chapters of this thesis: biomolecular self-assembly. First, this concept will be discussed from a thermodynamic point of view, with a brief evaluation of valuable contributions throughout history. A specific focus will be put on nanostructures based on amphiphilic peptides in section 2.2. As computational chemistry plays an important role in this thesis, an extensive overview of biomolecular self-assembly from this point of view is provided in sections 2.3 and 2.4.

2.1 Molecular self-assembly

2.1.1 Thermodynamics and key molecular interactions.

Self-assembly is often mentioned in relation to the formation of complex structures in nature and materials sciences. It is found in chemistry, biology, computer science etc. and can be defined for various types of building blocks on different size scales: on the macroscale, it is linked to evolution and self-replicating structures,[1,2] although it is typically used in the context of molecules. As such, the term is often used interchangeably with 'molecular self-assembly', which encompasses the aggregation of small molecular building blocks into an ordered supramolecular structure. Self-assembly is defined as a spontaneous process, which means in a system with constant temperature, pressure and volume, it follows from the Clausius inequality ($dS - dU/T \geq 0$) the change in Helmholtz free energy (ΔA in Joules) has to be negative:[3]

$$\Delta A \equiv \Delta U - T\Delta S \qquad (2.1)$$

with U the internal energy of the system (J), T temperature (K) and S the entropy of the system (J/K). Typically, the entropy of the self-assembling monomers decreases in the self-assembly process, as the number of accessible states is reduced when flexible molecules are 'fixed' into the nanostructure[4] (although examples of entropy-driven self-assembly exist, see e.g. ref. 5). On the other hand, when apolar moieties aggregate, the

entropy of water increases with a favourable effect on the free energy. In any case, usually other interactions also play a role in lowering the internal energy of the system and adopting a spontaneous assembly process. In molecular self-assembly, these comprise of four types of intermolecular non-covalent interactions (listed by decreasing average strength):

- Ionic interactions (Coulomb type)
- Permanent multipole – permanent multipole interactions (Keesom type, most commonly hydrogen bonds)
- Permanent multipole – induced multipole interactions (Debye type)
- Pure dispersive attractions (London type, induced multipole – induced multipole).

Generally, the combination of the last three terms are denoted as Van der Waals (VdW) interactions. An extensive dissemination of the thermodynamics of self-assembly is outside the scope of this thesis and has recently been provided by Palma *et al.*[6] Moreover, well-developed theories exist that describe the kinetics of both the assembly of monomers to dimers etc. or the formation of fibrous hydrogels (e.g. ref. 7 and references therein). However, it should be noted that the balance between entropy and internal energy is often delicate, and slight changes in building blocks frequently result in dramatic effects on the morphology of the self-assembled structure. Moreover, supramolecular constructs are held together by weak, non-covalent interactions as described above, which means aggregation is often a reversible process. Therefore, the pathway of self-assembly can have a major influence on the product, as illustrated in Fig. 2.1. Knowledge of the potential energy surface of aggregation can be exploited to reach the desired self-assembled state, be it the thermodynamic minimum, or kinetically locked higher energy states.[8,9]

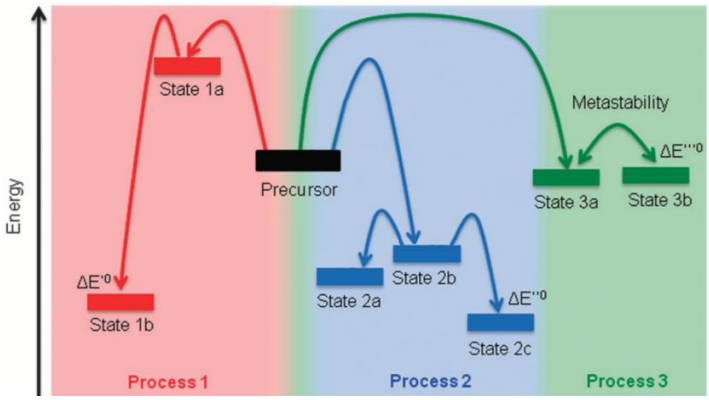


Fig. 2.1. Illustration of 'kinetic' self-assembly on a hypothetical potential energy level diagram: depending on the environment and temperature, self-assembly may proceed to reach the thermodynamically favoured state ΔE^0, or get kinetically locked in various other potential energy wells (e.g. State 2a). Figure adapted from ref. 10.

In aqueous media, one of the main driving forces of self-assembly is the so-called hydrophobic force. This is a complex entropic effect leading to the segregation of polar (water) and non-polar (solute) compounds, driven by the unfavourable breaking of the water structure by an apolar solute. Amphiphilic compounds especially can therefore form ordered structures, with their hydrophobic domains buried in an aggregate and their hydrophilic end exposed to solutions. The most prevalent, well-known example in nature is the phospholipid bilayer that constitutes the cell membrane (see Fig. 2.2). Many other examples are known, including liposomes, the collagen triple helix and actin fibers.

A special case of hydrophobic interactions is the stacking of aromatic moieties, usually called π-π-stacking. Various conformations of the flat aromatic rings provide an extra contribution to the attractive forces between two molecules. Although the quantum mechanical nature of such interactions is still under debate,[11,12] it is a motif that is extensively used in the design of supramolecular structures[13] and novel drugs.[14,15]

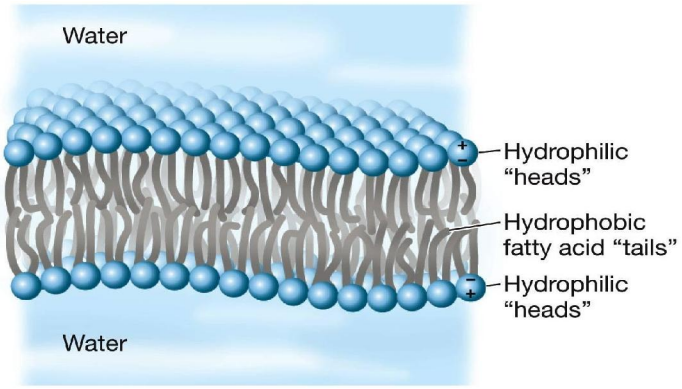


Fig. 2.2. Phospholipid bilayer showing the segregation of hydrophobic and hydrophilic domains. Image taken from ref. 16.

2.1.2 Self-assembly as a concept in materials science

The manufacture of nanomaterials has been identified as a key area for improvement in many applications, from solar cells and electronic circuitry to food and biomedical technology. In general production technology, *top-down* approaches often rely on etching. This technique starts with bulk materials and can be carried out either optically or chemically: in the case of optical etching (widely used in industry) the minimum dimensions of the object are determined by the theoretical diffraction limit of the light used (the wavelength of the light divided by 2) and are therefore usually at least several hundreds of nm. Chemical etching techniques rely on the precision of durable masks and are therefore also limited in their resolution. With the improving understanding of chemistry and physics on the nanoscale, *bottom-up* approaches to construct nanoscale objects are becoming more and more popular. Although Richard Feynman's famous 1959 lecture[17] ("*There's plenty room at the bottom*") mainly described a top-down approach to the field of nanotechnology, it has inspired a lot of researchers such as Eric Drexler to work on and popularize molecular nanotechnology, with atomic resolution 'writing' of molecules and nanostructures as an ultimate goal.[18] Molecular self-assembly is one of the most popular candidates[19] for achieving this: by designing the building blocks to direct bonding to their desired neighbour molecules, one can spontaneously form the desired nanostructures in a reaction vessel, as described above. Other

possibilities include printing techniques such as dip pen nanolithography (DPN), where molecular 'ink' is applied to a substrate with a modified Atomic Force Microscopy (AFM) tip[20] and synthetic bottom-up approaches that form covalent bonds between monomers (e.g. Schafmeister's bispeptides).[21]

2.1.3 Gelation and definitions

In 1861, Thomas Graham (after whom the chemistry department at the University of Strathclyde is named) tried to describe the physical state of a gel: "*While the rigidity of the crystalline structure shuts out external expressions, the softness of the gelatinous colloid partakes of fluidity, and enable the colloid to become a medium for liquid diffusion, like water itself.*"[22] A colloid is simply defined as a substance microscopically dispersed throughout another substance, with at least one dimension between 1 and 1000 nm.[23] However, the definition of a gel has been debated more heavily[24] since Graham's observations, and the current IUPAC definition is "*Non-fluid colloidal network or polymer network that is expanded throughout its whole volume by a fluid,*" which was recently changed from a statement concerning a gel's a finite, usually rather small, yield stress.[23,25] This debate was already envisioned by Jordan-Lloyd in 1926: "*The colloid condition, the 'gel', is one that is easier to recognize than to define,*"[26] and in practice this means that the simple test of vial inversion is often used to determine gelation: when a sample does not exhibit flow under the force of gravity, it is said to be a self-supporting gel (see Fig. 2.3).

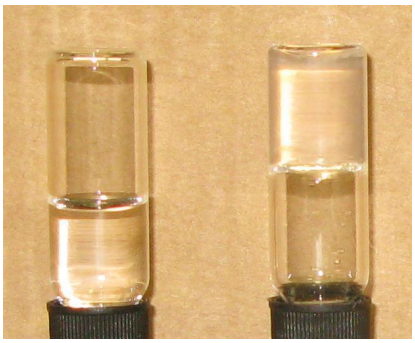


Figure 2.3: Vial inversion test. Left: non-gelling peptide solution. Right: self-supporting hydrogel.

2.2 Peptide hydrogels

2.2.1 Overview of peptide gelators

An important category of self-assembling systems is that of the oligopeptide hydrogels. This field was initially pioneered by Ghadiri[27] and Zhang[28] in the 1990s. They were among the first to *design* peptides that can spontaneously self-assemble into nanostructures. Based on this work, a growing number of researchers have entered the field, with important contributions from ultrashort peptides,[29,30] peptide surfactants,[31,32] and biomimetic systems such as collagen[33] and elastin.[34,35] A whole new range of self-assembling short peptides was developed by capping peptides with (aromatic) functional moieties. This procedure removes charges present at the termini of peptide chains, which reduces charge-charge repulsions between peptides. Moreover, this generally enhances the amphiphilic nature of the molecule, which is favourable for aggregation in an aqueous environment. The variety of building blocks available (20 gene-encoded amino acids, as displayed in Fig. 2.4, and synthetically modified analogues) and the well-developed peptide chemistry have lead to a 'Cambrian explosion' of peptide-based nanostructures with a variety of architectures and different functionalities. Examples of these are the 9-fluorenylmethoxycarbonyl (Fmoc) protected peptides, initially discovered by Vegners *et al*.[36] and further developed by the Ulijn group, and the naphthalene-protected peptides extensively studied by the groups of Xu[37] and Adams.[38–41]

Throughout this thesis, predominantly dipeptides coupled at the N-terminus (the end of the peptide terminated with an amine moiety) with an Fmoc-group were used. A large library of these compounds has been reported, with many different nanostructure morphologies.[8,42–51] It is apparent that minimal changes in the side chains of the amino acids can lead to drastic changes on the properties of the gels. However, a general mode

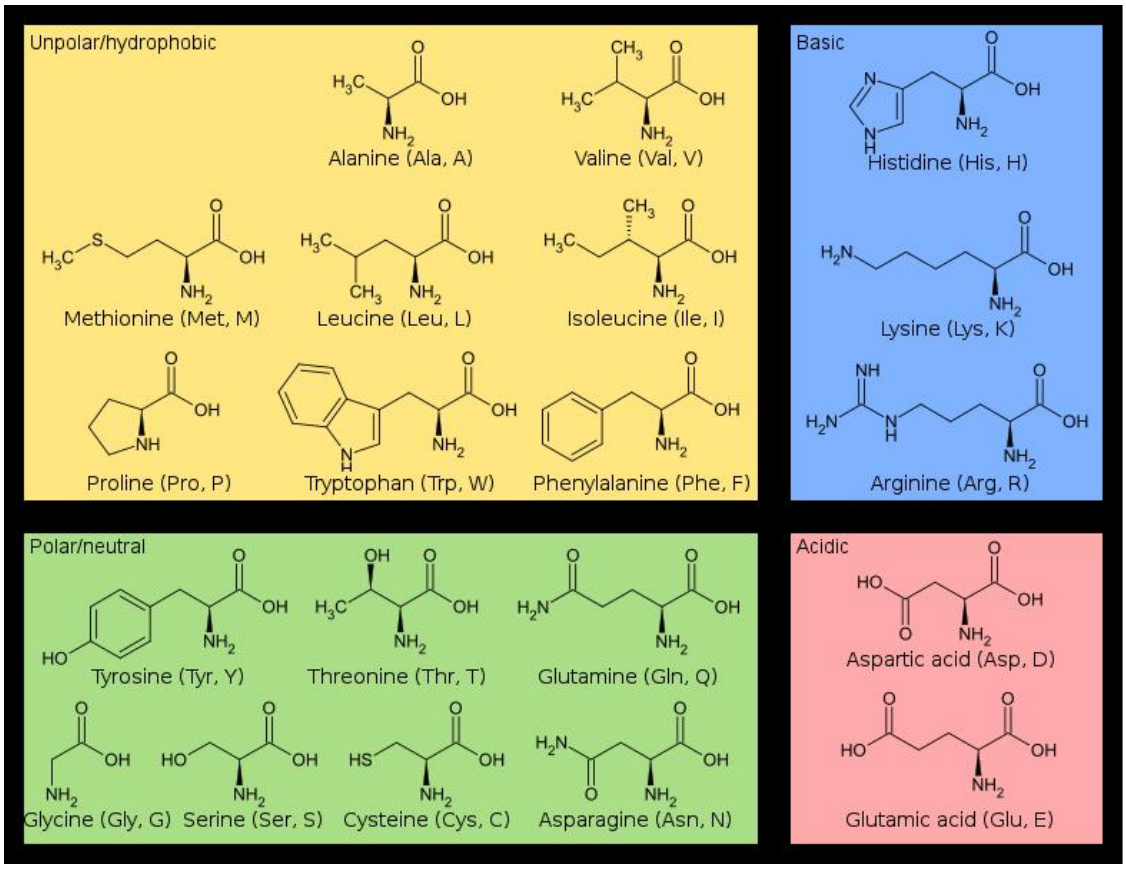


Figure 2.4. All twenty directly gene-encoded amino acids, with three- and one-letter codes.

of assembly for this type of systems has been proposed based on fluorescence and infrared spectroscopy. Results from these techniques indicate the presence of aromatic stacking of the protecting groups and a β-sheet hydrogen bonding pattern for the amino acid backbones. These 'π-β structures'[46] are schematically displayed in Fig. 2.5. The validity of this particular model is further discussed in Chapter 3. These peptide-based materials have been proven to be especially useful in biomedicine[52,53] and drug delivery[54,55] applications as a result of the biocompatibility of amino acid building blocks. Additionally, the gel scaffold has been shown to have suitable properties for templating,[56,57] development of nanoelectronics[29,58,59] and encapsulating enzymes and catalysts.[60–62]

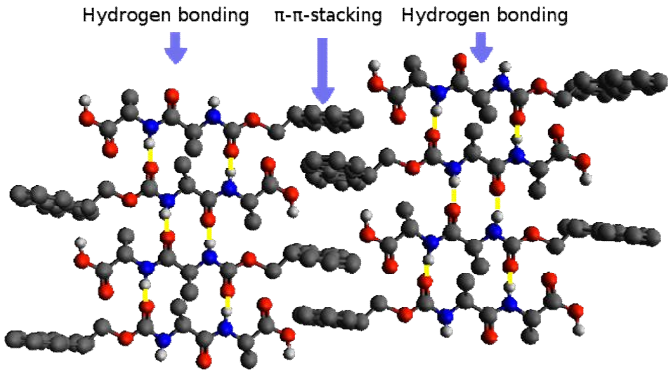


Fig. 2.5. Cartoon ball-and-stick representation of a model antiparallel π-β configuration for eight Fmoc-AA monomers, extendable in two dimensions. Yellow lines indicate the typical hydrogen bonding pattern for an antiparallel β-sheet. Hydrogen atoms connected to carbon not shown for clarity.

2.2.2 Preparation of hydrogels: biocatalytic methods

Typically, peptide hydrogels can be prepared by suspending monomers directly into a buffer solution at the desired final pH or, depending on the solubility of the gelator molecules, by suspension in a basic pH solution, after which the mixture is acidified to below the apparent pKa of the gelator molecule.[63] However, several methods have been developed to provide more control over the self-assembly process, in order to produce more well-defined, monodisperse nanostructures; in the case of hydrogels this leads to materials with reproducible viscoelasticity and fibre network density. They generally involve a trigger mechanism that provides a slow self-assembly process, such as gradual lowering of the pH by hydrolysis of glucono-δ-lactone[64] or light-induced gelation.[65,66] On the other hand, naturally occurring enzymes have been widely used to catalyze the formation of gelator building blocks from non-gelling precursors, as recently reviewed by Zelzer *et al*.[67] The advantages of enzymes as the trigger are their selectivity for their substrates as well as their catalytic efficiency. These traits and the inherent biocompatibility or even natural occurrence of enzymes make enzyme-responsive materials particularly suitable for the development of *in vivo* applications. Examples of enzymatic action to form gelator molecules are displayed in Fig. 2.6.

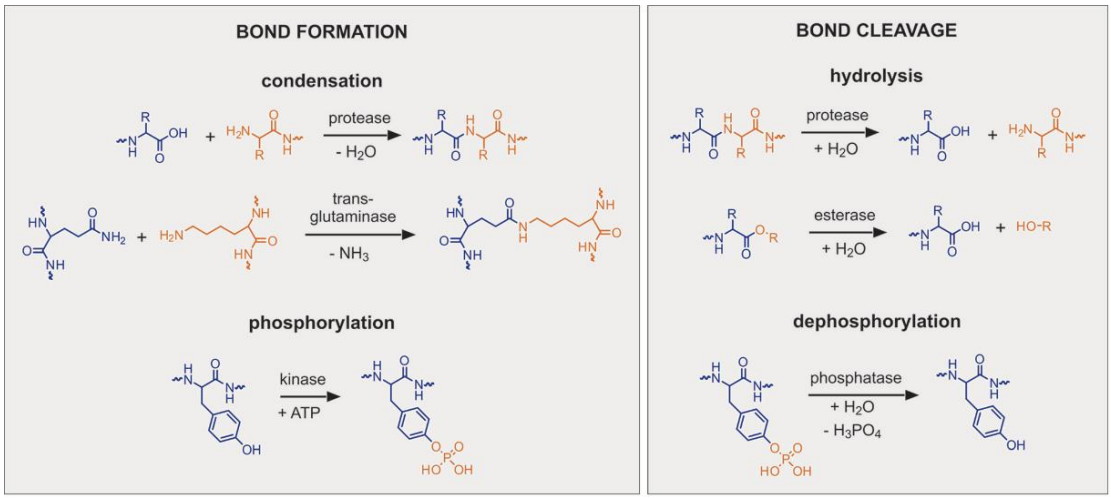


Fig. 2.6. Examples of enzyme-catalysed reactions used to prepare self-assembling building blocks. Figure taken from ref. 67.

The reversible nature of the enzymatic action can lead to the production of well-ordered nanostructures. Several reports have shown that self-assembly takes place under thermodynamic control (i.e. by fully reversible binding of monomers) when the protease thermolysin is used to catalyse the condensation of two precursor molecules.[8,43,45,68] The initial step of condensation between acid and amine is energetically unfavourable (as thermolysin usually operates in the reverse direction), but this is compensated by the free energy change when the self-assembled nanostructure is formed (see Fig. 2.1). This energy profile gives rise to a reversible formation of the assembly, which allows for molecular reorganisation and self-repair, and reproducibly reaching the thermodynamic minimum. This example of self-assembly under thermodynamic control is further discussed in Chapter 4. On the other hand, Hirst *et al*. showed that when gelators are produced by ester hydrolysis of a C-terminally methylated Fmoc-dipeptide using the subtilisin enzyme, kinetic control over the supramolecular assembly can be achieved.[69] In this case, the enzymatic reaction is energetically favourable, and supramolecular order on the nano- and microscale is strongly enhanced by increasing the enzyme concentration.

2.3 Practical computational approaches to biomolecular self-assembly I: Molecular Dynamics simulations

Self-assembly of biomolecules is one of the most important processes in living organisms, as it is responsible for many important tasks such as the formation of actin fibres, gelatinous tissue from collagen proteins and cell membranes. As discussed above, it has been widely used as a design concept in soft matter and pharmaceutical applications. Although many experimental reports are starting to compile the design rules for producing reproducible supramolecular structures, input from computer simulations is thought to be necessary for a complete understanding of the formation of biomolecular nanostructures. Only recently, the computational infrastructure required to perform detailed simulations of systems of a relevant size (10 – 100s of nanometers) has become available to a wide scientific community. An overview of the field so far is given in this section in order to provide a practical guide to obtain useful information from simulations of biomolecular self-assembly. The focus lies on applications in the field of peptide nanostructures. Various computational methods are used in this thesis to model the self-assembly of peptide nanostructures, which emphasizes the need for a review of practical approaches to obtain useful simulation results.

2.3.1 Introduction to computational modelling of self-assembly

Our understanding of the self-assembly of biomolecules and short peptides into nanostructures is continuously improving through the tailoring of building blocks for self-assembly and the empirical discovery of new biomaterials. However, as the toolbox available for designing these is enormous, there is great value in computational methods for virtual screening of new candidate molecules, as well as for advancing our understanding of the specific interactions involved in self-assembly. A good example is the field of computational drug discovery, where molecular modelling has become a crucial part of the search for new active compounds for the treatment of diseases.[70] The approaches used are generally based on 'docking simulations', where computer modelling is used to see if a drug molecule fits in the active site of a target molecule (see,

e.g., ref. 71 for a popular method to perform docking). The difference in the field of self-assembly is that rather than looking at the interactions of one drug molecule with one (large) receptor, there is a need to model 100s to 1000s of small molecules reversibly binding to each other to be able to determine the nanostructure morphology. A recent review by Palma *et al*. describes the computer-aided design of self-assembling systems using insights from thermodynamics.[6]

Currently, theoretical chemistry as a field offers a range of well-developed methods to choose from in calculating the properties of individual or a small group of molecules (electronic structure theory, molecular dynamics simulations etc.). However, only since the recent advances in high performance computing, applying these methods has become not too computationally demanding for the large numbers of molecules that make up a nanostructure. This literature review discusses practical approaches for obtaining meaningful results in various situations regarding self-assembly.

Two principally different methodologies are discussed in this chapter: first, the background of molecular dynamics (MD) simulations is given, which treat interactions between atoms in the system classically solving Newton's equations of motion (see 2.3.2). MD is widely used with the level of detail varying from all-atom representations to coarse grain models, where a group of atoms or molecules is represented by a single particle. A brief section on choosing the most suitable force field is provided in section 2.3.3, followed by a discussion on various ways to set up and run an MD simulation (2.3.4). The two main implementations of MD simulations are discussed: all-atom force fields in 2.3.6 and coarse-grain approaches in 2.3.7. This thesis cannot aim to provide a comprehensive review of all studies of MD simulations on biomolecules, but several key papers relevant to the self-assembly of biomolecules are discussed in these sections. Subsequently, quantum mechanical (QM) methods, which treat the full electronic structure of the molecules in an attempt to solve Schrödinger's equation, but are computationally very expensive, are evaluated for their use in biomolecular self-assembly (see section 2.4).

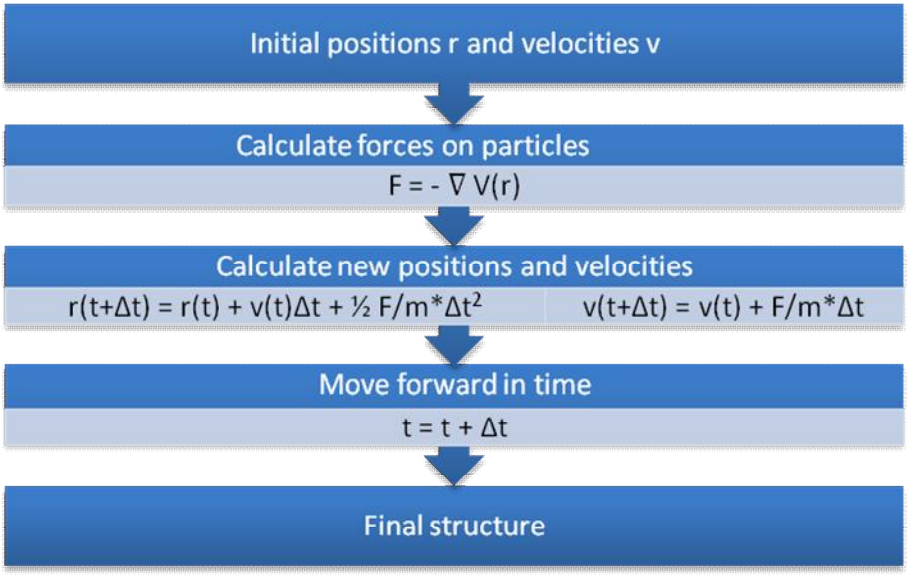


Figure 2.7: Flow diagram of a Molecular Dynamics simulation

2.3.2 Principles of Molecular Dynamics

Molecular dynamics treats all interactions between the particles in the system classically, in a ball-and-spring fashion. Fig. 2.7 sums up how in a typical MD simulation the positions and velocities of all the particles are calculated for every time step: the force F of every particle is determined by the derivative of the potential energy V. Subsequently, Newton's equations of motion are used to determine the new positions (r) and velocities (v) a time step Δt later (usually on the order of femtoseconds). This process is repeated N times until the end of the simulation. It should be noted that in practice the changes in positions and velocities are generally calculated using a more stable algorithm than the Euler integration displayed in the diagram. Examples include the leapfrog algorithm and the velocity Verlet method, which provide accurate positions, velocities and accelerations without compromising precision or energy conservation.[72]

The determination of the potential energy surface is the crucial step in the cycle, and the description of V is called the force field. Although additional terms may be present in various force fields, their mathematical description is generally similar to equation 2.2:

$$V = \sum_{bonds} k_b (b_i - b_0) + \sum_{angles} k_\theta (\theta_i - \theta_0)^2 + \sum_{torsions} k_\psi (1 + \cos(n\varphi - \varphi_0)) ...$$

$$... + \sum_{i}^{N} \sum_{j=i+1}^{N} \left(E_{\min} \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^{6} \right] + \frac{q_i q_j}{4\pi\varepsilon_0 r_{ij}} \right) \qquad (2.2)$$

Where

k_b = bond stretching constant
b_i = bond length
b_0 = reference bond length
k_θ = angle bending constant
θ_i = angle
θ_0 = reference angle
k_φ = torsion angle constant
φ_i = torsion angle
φ_0 = phase angle
n = torsional multiplicity
N = number of particles
E_{min} = Lennard-Jones well depth
σ_{ij} = interparticle separation with $V_{LJ} = 0$
r_{ij} = interparticle distance
$q_{i,j}$ = particle charges
ε_0 = permittivity of free space

The first three terms in Eq. 2.2 are bonded interactions for bonds, angles and torsions / dihedrals. The latter includes both proper dihedrals and improper dihedrals (χ and S, see Fig. 2.8) to account for the planarity of, for example, sp^2 hybridized atoms. The largest part of the computational time required, however, is needed for calculating the non-bonded interactions. In equation 2.2 a typical Lennard-Jones expression (also called 6-12 potential) is given to correct for non-classical effects of attractive dispersion interactions in the long range and repulsive nuclear effect from Pauli exclusion at short interatomic distances. Although the 6-12 potential is not very suitable for describing the interactions at the interface of liquids and solids or liqids and gases, it is suitable for simulations of the condensed phase and generally represents important properties such as oil/water partition coefficients quite well.[73] The last term in equation 2.2 covers coulombic charge-charge interactions, which for most force fields includes the dipole-dipole interactions by using atomic point charges.

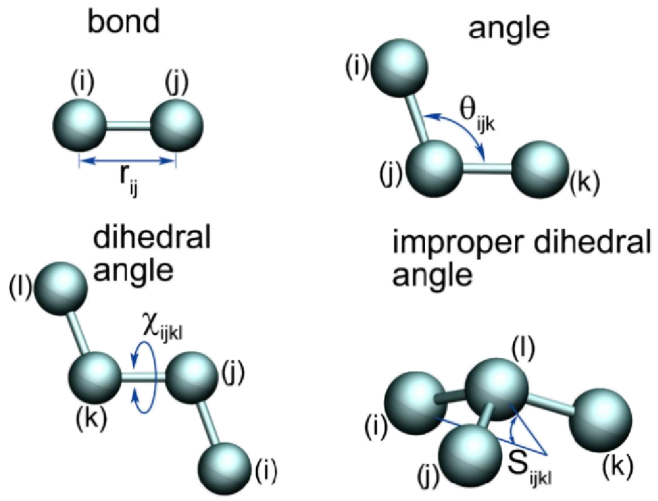


Fig. 2.8. Bonded interactions taken into account in MD simulations. Image obtained from ref. 74.

An excellent review on MD for the modelling of protein folding, including theoretical and historical backgrounds, discussion of software packages and algorithms have been published by Adcock and McCammon in 2006,[75] while Karplus and Kuriyan gave more practical examples on protein folding and enzyme catalysis.[76] This review discusses approaches for self-assembling supramolecular structures made from small biomolecules instead of macromolecules such as proteins.

2.3.3 Choice of force field

As discussed, the force field is the most important factor in an MD simulation as it determines the level of interaction between various molecules. Especially in the case of self-assembly in solvent, the development of force fields is not a trivial task, as the ordered formation of nanostructures often depends on a delicate balance between enthalpic and entropic contributions. Non-covalent interactions including quantum mechanical Van der Waals-forces are relatively important, but notoriously hard to describe (even with a quantum mechanical description).

However, a wide range of force fields has been developed and tested for various applications, some of which are discussed in the next sections. One of the first choices one has to make is the level of detail of the model. Depending on the requirements of the simulation, 'all-atom' molecular dynamics, where every atom is modelled individually as a particle, can be employed. This provides a large amount of information on specific interactions, but is due to its computational cost currently limited to sub-microsecond time scales for systems of 10^5–10^6 atoms on modest computational resources. Systems containing this number of molecules (including solvent) are required in the modelling of self-assembly, due to the large amount of molecules that make up a nanostructure. Moreover, to stay in the range of experimentally relevant concentrations, relatively large amounts of water molecules need to be in the simulation. To access time scales that are relevant for biomolecular self-assembly (total time on the scale of μs[77]) 'united-atom' or coarse-grain methods can be a more suitable option. These methods contain multiple atoms or even molecules per particle ('bead') and by limiting the degrees of freedom of the system, a speedup factor of many orders of magnitude can be achieved, but as a trade-off certain details are lost. Force fields at the coarse grain level are discussed in section 2.3.6, with practical examples from the field of biomolecular assembly.

The best choice of force field obviously depends on the simulation requirements. For all-atom simulations a number of all-atom (AA) force fields particularly suited for biomolecules that have become popular over the last few years are CHARMM (Chemistry at Harvard Molecular Mechanics[78,79]), AMBER (Assisted Model Building and Energy Refinement[80]), GROMOS (Groningen Molecular Simulation[81]) and OPLS-aa (Optimized Potentials for Liquid Simulations – all-atom).[82] Mainly the former two have specifically developed parameters for lipids, proteins, peptides, carbohydrates and other biomolecules. Which force field to choose is nowadays generally a matter of preference or availability of parameters in their respective databases. Moreover, both AMBER and CHARMM have provided automated parameterisation programs (Antechamber[83] and Paramchem,[84,85] respectively) which allow for the straightforward introduction of new molecules in the force field based on existing entries. However, careful checking of the

produced parameters is always required and it is to be hoped that this does not become a pitfall when non-specialist users start using the more and more user-friendly software packages.

2.3.4 System setup

In setting up an MD simulation for a low molecular weight (LMW) self-assembling biomolecule a number of choices have to be made regarding the contents of the simulation. Important factors to consider are the number of building blocks, the size of the simulation box, the starting structure (randomly orientated or predefined), etc.

Especially when computational costs are an important factor, it is common practice to start simulations using a predefined supramolecular structure. For example, in the work by Hughes *et al*., discussed in Chapter 4, the stability of a proposed stack of Fmoc-dipeptide methyl esters in an antiparallel π-β configuration was tested. Another example is the work by Hauser *et al*. who could demonstrate the stability of a hexapeptide nanofibril when starting from a well-oriented stack, but did not observe formation of such a well-ordered structure when starting from well-solvated monomers in a much smaller system.[86] The disadvantage of predefining the initial conformation as a way to setup simulations is the introduction of a structural bias: when a stack is built based on inconclusive experimental data, it is realistic to expect that the proposed structure is stable to a certain extent, but may not represent the particular minimum (either local or global) that is present in the experiment.

When looking at the self-assembly process itself, it is much more attractive to start from a random peptide solution and form nanostructures without the introduction of a structural bias. This is still not a guarantee that the thermodynamic minimum state is reached, as strong binding events can still trap the nanostructure in kinetic minima, but when experimental conditions (temperature, concentrations, ionic strength etc.) are mimicked in the simulation, in principle the final structure should be an accurate

representation of the experimentally observed nanostructure (be it a local or a global minimum state). Moreover, this approach also allows the automation of setting up MD simulations of new systems or screening large numbers of molecules for specific properties, such as the propensity to self-assemble (see also Chapter 5). A good example of the influence of structural bias is given by Mu *et al*. in a computational study of Fmoc-AA self-assembly.[87] Several fibrous starting confirmations were tested for stability, as shown in Fig. 2.10. All simulations gave densely packed, amorphous 1D fibres as a result. However, a Simulated Annealing (SA, see 2.3.5) run, which heats up a fibrous structure to yield random monomer conformations, did not give the same aggregate morphology as for the biased simulations.

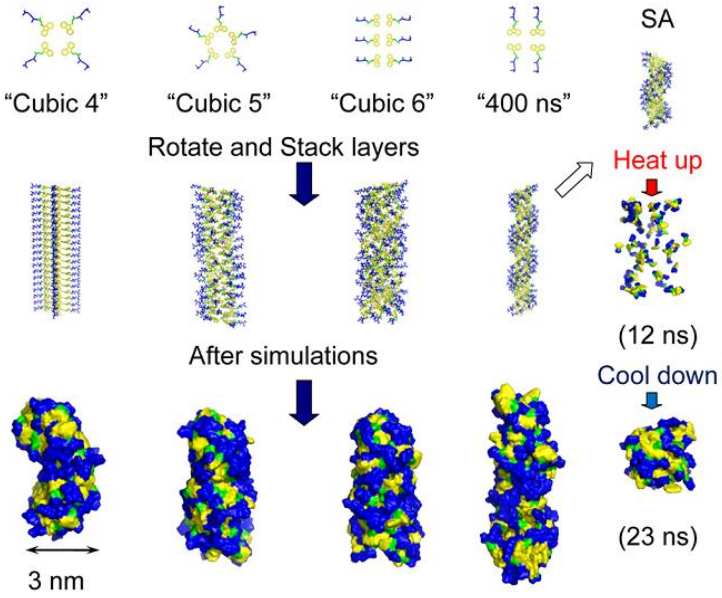


Fig. 2.10 Influence of starting configuration structural bias on the result of MD simulations of Fmoc-AA. The simulated annealing (SA) simulation produced significantly different morphology from the simulations with a predefined fibrous structure. Image obtained from ref. 87.

It is difficult to address the issue of system size on a general basis. Especially in the case of reversibly bound small molecules, it is common to use a system with periodic boundary conditions (PBC). This allows the interactions between particles in the system, and other particles in a virtual neighbouring box, as depicted in Fig. 2.11; without PBC, particle A would only be able to interact with particle B, as other particle are beyond the cutoff distance for interactions. This cutoff is generally about 12–20 Å for MD simulations of biomolecules, and the box size should be greater than double the cutoff distance (for a detailed example, see ref. 88). However, by using PBC, particle A is able to interact with particles C and D across the boundaries of the box. Furthermore, when a particle moves out of the box, it is replaced by the exact same particle at the opposite end of the box. When PBC are used, the simulation resembles an infinite bulk phase and no interactions with simulation walls or vacuum around the system are present.

The use of PBC has consequences for the setup of the simulation: for example, if the stability of a predefined nanostructure needs to be tested, a sufficient solvent layer has to be applied around it to prevent the structure from interacting with itself in the neighbouring periodic boxes as reported by De Souza and Ornstein for biomolecules.[89]

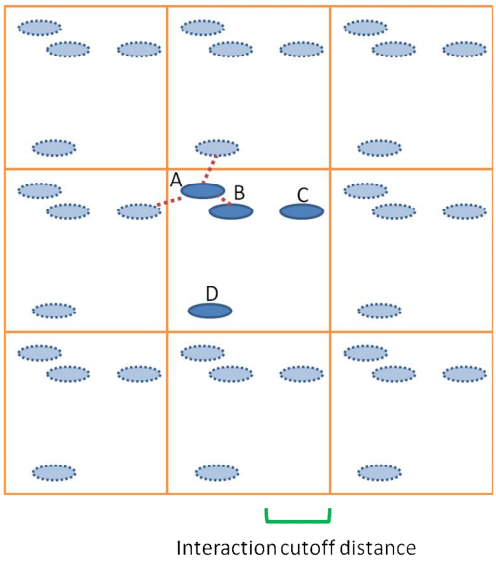


Fig. 2.11 Demonstration of two-dimensional periodic boundary conditions. An example cut-off distance for interactions is displayed as a scale bar. A interacts with B, C and D.

This is a disadvantage as the addition of extra molecules into the system comes with an increased computational cost. However, several methods, including using non-rectangular periodic boxes, can be applied to reduce the necessary amount of solvent molecules.[90] The issue is less directly visible in simulations starting from randomly oriented monomers, but interactions with neighbouring boxes can still occur when a small simulation size is used. Moreover, PBC may artificially induce 1D structures, because of the decrease in surface area that an 'infinite' structure has when it spans the length of the simulation box.

The number of monomers in the box is another issue that needs careful consideration. LMW nanostructures generally still consist of 100s to 1000s of monomers. If these would be present in a simulation box at experimental concentrations, the box size would generally contain many millions of atoms, which is computationally unrealistic. Moreover, a long simulation time would be required to assemble all monomers (experimentally self-assembly can take place on a timescale from minutes up to days, while current state-of-the-art simulations can only reach μs to ms). Therefore, simulations are generally performed at concentrations that are 10 – 100 times higher than experimental. Additionally, it has been shown for many cases that there is a critical size requirement for the systems to be modelled before self-assembly can be observed, analogous to the critical nucleation concentrations in crystallisation and the assembly in polymers.[91] This has been extensively debated for amyloid-type peptides,[92–94] but also been shown for smaller peptides.[95]

2.3.5 Running an MD simulation

After providing the force field parameters and coordinates for the simulation, the simulation needs to be performed on a suitable MD platform. Various software packages are available, with a few of the most popular listed here: GROMACS, NAMD, Desmond, LAMPPS, CHARMM and Discovery Studio. The former four are freely available to academic users and especially GROMACS and NAMD have been implemented with

compatibility with all of the most used force fields. With the release of GROMACS 4.0 in 2008, it was reported that MD simulation in GROMACS are somewhat faster than in Desmond and NAMD, and that GROMACS scales more linearly with the number of processors used in parallel computing.[96] However, the relative performances are comparable nowadays and the platform, like the force field, has mainly become a matter of user preference, unless specialist applications are needed.

One important adaptation of conventional MD simulations that is relevant to biomolecules is replica-exchange MD (REMD).[97] This ensemble runs multiple simulations at the same time, but at multiple different temperatures. By exchanging the temperatures between the parallel runs, small energy barriers can be easily overcome, which allows the system to explore a greater region of the conformational landscape (also called phase space), whilst avoiding becoming trapped in one of the multitude of local energy minima present, especially in simulations of protein folding. A similar method is simulated annealing (SA), in which the systems is gradually heated up, and then cooled down again while the algorithm is looking for a decrease in potential energy and discarding any states with a higher energy.[98,99] The theory of these methods will not be discussed here, as its use in studying peptides has recently been reviewed,[100] but several examples are given as case studies.

2.3.6 All-atom Molecular Dynamics for LMW biomolecular self-assembly

In this section, various case studies where MD simulations were used in the field of peptide self-assembly are discussed. First of all, the assembly of short peptides related to biomaterials science are reviewed, followed by a brief overview of MD simulations of peptides in prebiotic chemistry and at inorganic interfaces.

2.3.6.1 Short peptide self-assembly

Several groups have looked into the simulations of short peptides using atomistic molecular dynamics simulations. These peptides have been shown to have applications in biomaterials science as discussed in section 2.2. An example is the work by Hauser and co-workers on acetylated tri- to hexapeptides with a hydrophilic head group and a tail consisting of aliphatic amino acids, that self-assemble in water to amyloid β (Aβ)-type fibre aggregates.[86] Interestingly, they propose an α-helical intermediate conformation in the formation of the well-studied cross-β sheet[101–104] nanostructure. Results on small systems of 4 monomers in water show that this conformation is adapted after dimerisation for hexapeptide Ac-LIVAGD, around 4 ns into the simulation. Extended simulations were performed on the hexapeptide starting from a proposed fibre structure and showed this conformation was stable in a 20 ns simulation, as shown in Fig. 2.11. However, this tendency to dimerise and subsequently aggregate into larger constructs was not observed for tripeptide Ac-IVD on the time scale of 5 ns. In a second paper, the group presents extended simulations on these peptides, and Ac-NFGAIL, a truncated fragment from human islet amyloid polypeptide (IAPP).[105] In this case, they observed fibre formation for all peptides in 50 ns simulations. These observations are in agreement with the results from Wu *et al*., who have performed similar simulations on the NFGAIL fragment and observe the formation of only partially ordered parallel and antiparallel β-sheet structures (16% of peptides in the simulations was in a β-sheet structure, but 95% of the peptides was part of an aggregate), and hypothesize the disordered aggregates they observe can act as a nucleus for larger amyloid fibrils that cause the formation of plaques in amyloid diseases. Moreover, these studies provide a methodology of determining dynamic secondary structure elements, which is not always straightforward in short peptides.

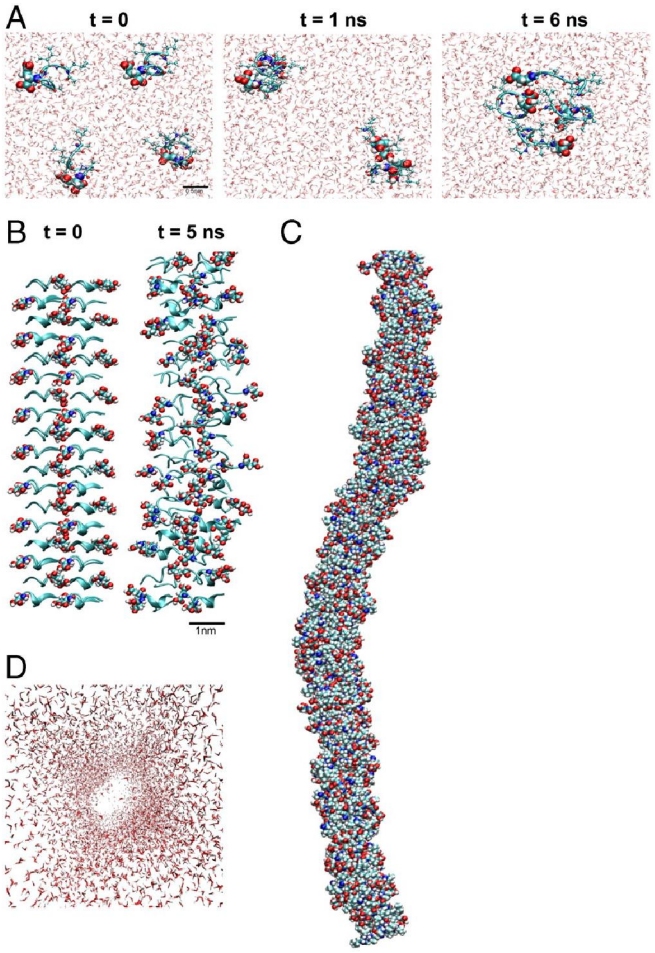


Fig. 2.11. Self-assembled nanostructures of Ac-LIVAGD. Hydrophilic head groups are displayed in VdW surface representation, aliphatic tails as ribbons. (A) Spontaneous assembly of 4 monomers. (B) Snapshots of a simulation of 72 monomers. (C). Final results of a 20 ns simulation on 216 monomers in an initial fibrous conformation. (D) Very few water molecules reside within the fibre, as can be seen from the top view where the fibre is not displayed. Image obtained from ref. 86.

A number of case studies relevant to the work presented on Fmoc-dipeptides in this thesis have been reported. An early study by Smith *et al*. performed molecular mechanics

minimisations* on an Fmoc-FF tubular structure, proposed based on evidence from wide angle X-ray scattering (WAXS) and fluorescence, CD and FTIR spectroscopy.[46] In these simulations, the nanotube was found to be stable, but the lack of dynamic simulations prevents further analysis of the validity of this model. A similar approach was undertaken by Hughes *et al*. who used minimisations to determine the distances between functional moieties, which showed good agreement with the WAXS diffraction patterns of their Fmoc-dipeptide methyl esters nanostructures in a frustrated β-sheet conformation (see refs. 8,106 and chapter 4). Xu *et al*. used MD to test the stability of a predefined Fmoc-LLL tube, which was proposed from evidence in FTIR spectroscopy, transmission electron microscopy (TEM) and WAXS.[59] After minimization and a short (7 ns) simulation in water, they noticed a curvature in the Fmoc-LLL stack, which is in agreement with the experimental curvature of the nanotubes determined by TEM. Both these studies propose structures where π-stacked fluorene rings and a β-sheet hydrogen bonding pattern play an important role in the assembly of the nanotubes.

* When MD simulations are performed without the introduction of temperature, pressure, etc. it is technically a geometry optimisation using a force field, which is called molecular mechanics (MM)

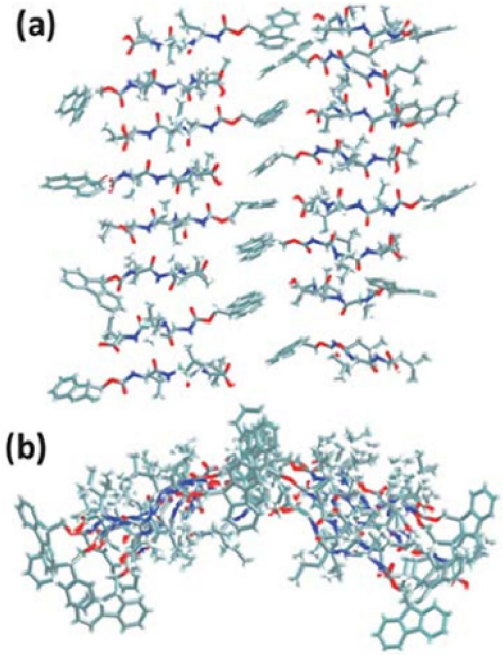


Fig. 2.12 Results of minimisation and equilibration of an Fmoc-LLL stack in π-β conformation (see also section 2.2.1). (a) Side view, (b) top view showing the curvature. Figure adapted from ref. 59.

Another study on Fmoc-peptides has been reported by Mu *et al*.[87] They studied the assembly of Fmoc-AA from various starting conformations, as shown in Fig. 2.10. The simulation trajectories (including a REMD simulation) have been analysed in various ways: they performed a radial distribution function (RDF) analysis, which showed a preferential distance of 4.3–4.5 Å between fluorenyl rings, in agreement with a peak in their wide angle X-ray scattering (WAXS) diffraction pattern of Fmoc-AA fibres. A detailed analysis of H-bonding patterns suggested that Fmoc-groups were not completely buried in the hydrophobic core of the fibre. Moreover, the backbone dihedral angles in their simulation suggested a polyproline II (PPII)-type conformation for the dipeptide backbone, rather than an extended β-sheet structure, which could explain the presence of a 1645 cm^{-1} peak in the IR spectrum of Fmoc-AA hydrogels (see Chapter 3).

A somewhat different type of self-assembling system based on biomolecules are the peptide amphiphiles originally developed by Stupp's group, containing an aliphatic tail and a oligopeptide head group.[31] Recently, MD simulations on these systems were reported by Lee and co-workers: it was found that starting from a cylindrical peptide amphiphile nanostructure based on experiments, a stable fibre formed in a 40 ns simulation of 144 monomers.[107] The aliphatic chain of these peptides is buried in the core of the fibre, while the peptide sequence can form temporary, flexible β-sheet or β-turn conformations. When the peptide chain contains a binding recognition motif (in this case the IKVAV pentapeptide), the flexibility of these peptides plays a role in the adherence to, for example, neurite growth as shown in ref. 108. Moreover, the width of the fibre is in good agreement with experimental data. When the peptide sequence is varied, more β-sheet type structures can be induced in the peptide part of the fibre, while the hydrophobic core remains intact.[109] These studies are a good example of how MD simulation can aid the elucidation of nanostructure architecture, including peptide secondary structure elements.

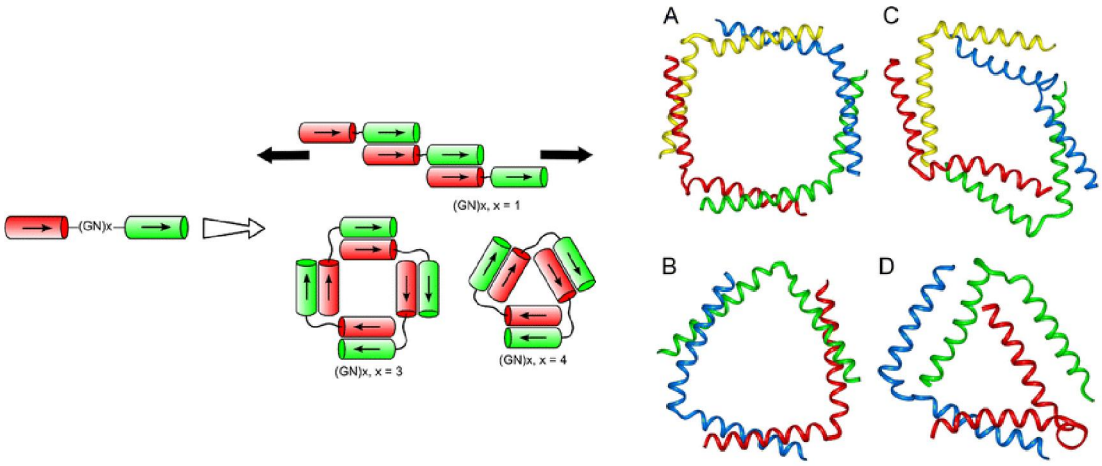


Fig 2.12 Left: cartoon representation on folding possibilities for a de novo *designed coiled-coil structure with different linker lengths. Right: A, B) Starting configurations of a trimer and tetramer of the same coiled-coil dimeric unit, but with different linker length between the two parts of the dimer. C, D) Results of a 100 ns MD simulation. Image adapted from ref. 110.*

A last example is the work by Boyle *et al*. on *de novo* designed coiled-coil structures.[110] These structures can be designed to form dimers in solution, and MD simulations in the AMBER force field investigated the conformational behaviour of dimeric structures with different turn regions between the two parts of the dimer (see Fig. 2.12). With a longer linker/turn region $(GN)_4$, triangles were shown to be stable, while with shorter linker length $(GN)_3$ the angle between the dimers is limited and a diamond type shape was observed after a 100 ns simulation.

2.3.6.2 Prebiotic chemistry and surfaces

Several cases where small peptides and peptide derivatives has been known to form supramolecular structures have been shown above. A set of quite different examples that deserves a note in this review come from the field of prebiotic chemistry. It is believed that biomolecular self-assembly is the driving force behind the emergence of the first life-forms[111] through self-replication and subsequent compartmentalization.[112] Coveney *et al*. recently summarized the role theory, modelling and simulation can play in origins of life studies.[113] They state the importance of kinetics in the 'prebiotic soup' and, relevant to the self-assembly of biomolecules, provide an overview of what simulation methods can provide in terms of molecular information. Quantum mechanical (QM) methods are important for modelling processes as bond-making and breaking, and interactions between organic molecules and inorganic minerals, important in origin of life studies. They discuss the reactivity at mineral surfaces, the reactivity of amino acids in aqueous solution and chiral selectivity in minerals using DFT calculations. However, they also state MD simulations are better suited to studying the dynamical interactions of biomolecules.

Coveney summarises how the adsorption of biomolecules on surfaces can lead to their enhanced reactivity. These interactions are important in everyday life and technology, including interactions of molecules with biominerals (e.g. teeth) or oxidation of metal

nanoparticles. This research field mainly concerns peptide-surface interactions, rather than the peptide-peptide interactions discussed in the previous section, but the same types of interactions need to be considered. The MD modelling of these systems was addressed in a comprehensive review by Slocik and Naik, discussing the interactions of peptides with metal surfaces, nanoparticles and carbon nanotubes.[114] They conclude that generally unstructured peptides with high flexibility are especially suitable for binding to metal surfaces, but the specifics of crystallographic play an important role in surface recognition and the interaction strength.

For simulations at inorganic surfaces, a range of specialist force fields have been developed. For example, the ClayFF describes hydroxide, oxyhydroxide, and clay phases.[115] For metal surfaces, Heinz *et al*. have recently developed the INTERFACE force field, that can be used as an extension to most widely used existing force fields.[116] In previous studies they report on, *inter alia*, the nature of the molecular interactions of peptides on gold and palladium surfaces.[117] As peptides have been shown to influence the properties or architecture of nanoscale metal structures, this is important in the design of functional inorganic or biological nanostructures. Another example showed small

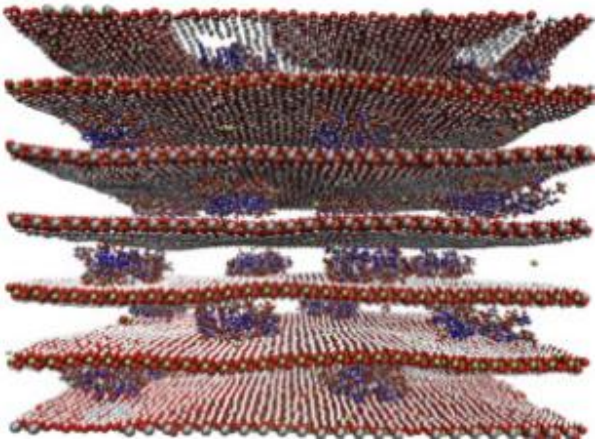


Fig. 2.13. Result of a 30 ns MD simulation of RNA-containing layered double hydroxide models as produced by Swadling et al.[119] *Magnesium, aluminium, oxygen, hydrogen, nitrogen, carbon and phosphorus are displayed as green, silver, red, white, blue, grey and orange spheres respectively.*

nucleic acids that can carry genetic information[118] (RNA, DNA or peptide NA) assemble together with MgAl-layered double hydroxides (LDH) into ordered biomineral nanostructures, as displayed in Fig. 2.13.[119] These force fields are starting to play a more and more important role in the development of catalysts and nanotechnology.

2.3.7 Coarse-Grain Molecular Dynamics

As discussed earlier in this chapter, system size can play a crucial role in modelling the self-assembly of biomolecules. Moreover, the spontaneous formation of nanostructures typically takes place on a longer time scale than is accessible by all-atom simulations. In order to address this problem, a range of coarse-grain (CG) force fields have been developed. The level of coarse graining ranges from qualitative, often solvent-free models to models that include specific chemical details. They are generally parameterised against macroscopic experimental observables such as heat of vaporisation, partition coefficients and, for the more detailed models, AA MD simulations. In this section, first the MARTINI force field is discussed, as it is one of the most popular CG force fields that can be readily used for a wide variety of biomolecules and it is used in Chapter 5 of this thesis. Several other CG models for small biomolecules are discussed in section 2.3.7.2.

2.3.7.1 The MARTINI force field for peptide self-assembly

The MARTINI force field was originally developed for CG simulations of lipids (specifically for lipid membrane studies),[120] but has since its official release in 2007 been extended to peptides and proteins,[121,122] carbohydrates[123] and glycolipids.[124] A recent paper by Marrink and Tieleman[73] gives a comprehensive overview of the design principles, applications and limitations of MARTINI: the philosophy is to develop a residue-based force field, where careful parameterisation of individual beads ensures the transferability of properties to the molecule as a whole. This means the force field can be

used to model a wide variety of biomolecular systems without the need to change the coarse-grain representation and parameters from case to case, which partially explains the popularity of the force field.

To achieve a compromise between chemical accuracy and computation speed, a four-to-one mapping scheme is employed, where four heavy atoms and their associated hydrogens are represented by a single 'bead' (2 – 3:1 for ring systems). For amino acids, this means the entire backbone is represented by a single bead, while one to four beads are used for representing the side chains,[121] which is discussed further in section 3.3.2. This significantly reduces the number of the degrees of freedom and typically speeds up simulations by a factor of approximately 10^3 in comparison with all-atom simulations.

Marrink and Tieleman summarise the applications of MARTINI in many different systems, including the study of lipid membranes, membrane proteins, surfactant self-assembly, polymers and nanoparticles. Several studies have used MARTINI in the context of aggregation of peptides and proteins. Chapter 5 and ref. 95 describe a high-throughput screening of all combinations of di- and tripeptides for their ability to aggregate as a precursor to ordered self-assembly, with a focus on diphenylalanine (FF) nanotubes. An independent study confirmed the formation of hollow FF nanotubes and reported more details on the intermediate structures in the initial stages of tube formation, as displayed in Fig. 2.14.[125] Moreover, they were able to reproduce the results by Yan *et al*. on cationic FF peptides that a reversible transition between tubes and vesicles is possible by diluting or concentrating an FF solution in water.[126] Concentration-dependent studies are generally not accessible at the full-grain (all-atom) level due to the computational costs of simulating large water boxes, while maintaining enough monomers to form a complete nanostructure.

Another study relevant to this thesis is the coarse-grain simulation of Fmoc-AA peptides into hydrogels by Zhang and co-workers.[127] π-π stacking was indicated as the main

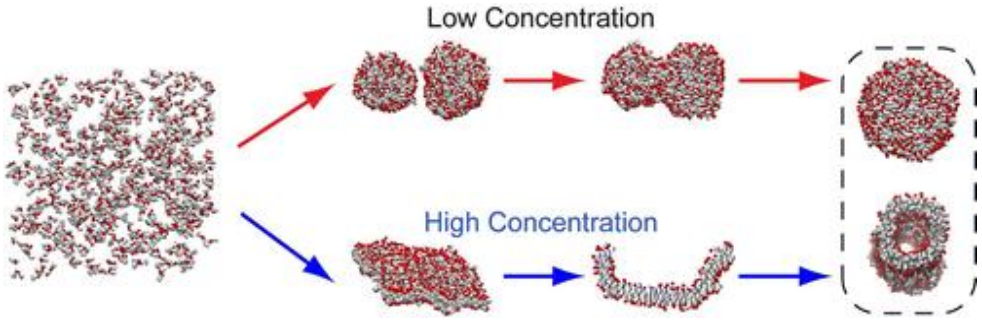


Fig. 2.14. Formation of diphenylalanine nanostructures depending on the peptide concentration modelled using the MARTINI force field. Red and white beads are backbone and side chain beads, respectively. The resulting morphology changes between peptide concentrations. Image obtained from ref. 125.

driving force for this protected peptide with only small apolar side chains. Mainly parallel stacking interactions were observed. This study is in reasonable agreement with the results presented by Mu *et al*.,[87] who performed all-atom simulations on this system as discussed above, and also found that hydrophobic interactions between the Fmoc-rings are the main driving force of assembly.

An interesting comparison between AA and CG simulations is given by Lee *et al*., who previously used atomistic simulations to study the aggregation of peptide amphiphiles as described above, but more recently also used the MARTINI force field to model the same process.[128] They found a good agreement between the structures formed by CG and AA methods. Thanks to the larger time scale accessible, however, they were now able to start from a homogenous peptide amphiphile solution and as such observe the mechanism of formation of the final fibrous structure on a 16 μs time scale (c.f. 40 ns simulations were possible on the AA level).

Other examples from the area of peptide amphiphiles are the nanovesicles formed by the self-assembly of branched amphiphilic peptides[129] and micelles from short block-copolymer oligopeptide fragments.[130] These examples strongly suggest that the MARTINI force field is indeed quite suitable for a range of different sytems, including

micelles, vesicles, amphiphiles and small peptides. However, a range of other coarse-grain methods is also available, as is discussed next.

2.3.7.2 Other coarse-grain methods

In principle, there are many potential ways to model biomolecular self-assembly, as there is a great number of ways to divide up the monomer structure into CG beads. The work by Villa *et al*. is an example of the setup of a CG force field specific for their system under study, in this case self-assembly of diphenylalanine (FF, see Fig. 2.15).[131] They have designed 5 different beads and specified interaction strengths between all the 15 possible combinations of beads based on potential of mean force calculations. Moreover, they have defined the specific force field bonded and non-bonded terms for simulations using this particular CG representation. The obvious disadvantage of this method is the need to parameterise every system from scratch and the speed-up achieved was only 3-fold. On the other hand, accurate results can be obtained regarding larger supramolecular constructs when monomers are parameterised to follow the system-specific target data, which for these CG representations comes from atomistic simulations. In a second paper, they describe the use of implicit solvent, next to the CG representation.[132] The intramolecular covalent interactions in this case need to be redesigned such that the CG molecule captures the conformational sampling observed in the atomistic simulation, even without explicit solvent. This is an interesting step though, as interactions between water particles often represent a large part of the computational cost of these simulations. Treating the water implicity significantly speeds up the calculation further and would allow the treatment of large amounts of molecules. However, Villa's dipeptide simulation did not converge to the experimentally observed nanostructure. For peptides, it is sometimes hard to justify creating a new set of interaction terms and parameters with the open source software and force fields such as GROMACS and MARTINI, which have been extensively tested by large amounts of users and developers.

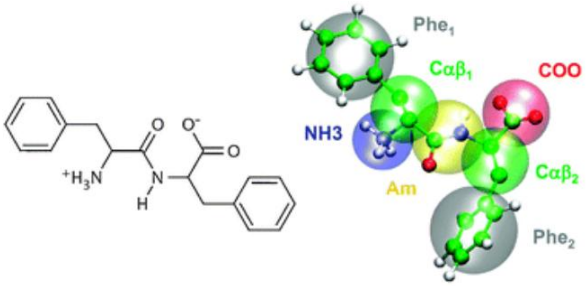


Fig. 2.15 Molecular structure and coarse-grain representation of diphenylalanine by Villa et al.[131]

A completely different level of coarse-graining is discussed by the group of Frenkel: rather than supplying specific molecular detail, they use the shape of biomolecules and the morphology of their aggregates.[133] The model used in this case is that of a rodlike particle with an attractive (hydrophobic) stripe on its side, also called 'patchy particles'. Interestingly, by simply changing the size of the hydrophobic patch, various aggregate structures with experimental analogous could be accessed: micelles, fibers, ribbons etc. This type of approach may be more suitable for protein-protein interactions, in the study of polymers or in systems of colloidal particles, although a recent study has also shown the formation of amyloid fibrils from combinations of 'random coil' and 'β-sheet beads'.[134] These simulations are especially suited for connecting macroscopic observations with the assembly events on a microscale, e.g. concentration-dependent fibril nucleation events.

Several methods are somewhere between these two levels of detail, but most force fields in the area of peptide and protein research use a similar mapping strategy to the MARTINI force field, as this allows separate representation of the peptide backbone and the side chain. For example, force fields for protein and peptides developed by Bond *et al*. and Shih *et al*. use a four-to-one mapping strategy too,[135,136] and have been made compatible with the MARTINI lipid force field. Others have made their own extensions to this easily adaptable force field, for example in the study of photo-switchable self-assembly of azobenzene-containing compounds.[137] As it is the most widely-used CG

force field, it will most likely also be the fastest-developing one, with developments including better representations of long-range interactions, introduction of foldable protein backbones, nucleic acid parameters and the underlying algorithms (for details see ref. 73). Note that Chapter 5 discusses the use of this versatile force field for high-throughput screening of aggregating peptides.

2.4 Practical computational approaches to self-assembly II: Electronic structure methods

Instead of using MD approaches to model self-assembly, as discussed in the previous section, electronic structure methods can be used. In terms of simulating self-assembling systems, quantum mechanical (QM) techniques are generally not the method of choice. Popular methods such as Density Functional Theory (DFT) have reached the points where useful accuracy (e.g. interaction energies within a few kcal/mol) can be achieved at modest computational costs for a large variety of systems. However, a typical DFT calculation is currently limited to approximately 10^3 atoms depending on the code used, due to the computational cost of taking at least all valence electrons in the system into account, even with small basis sets. Methods to extend this to thousands of atoms do exist (e.g. Density Functional-based Tight Binding (DFTB), for a recent implementation see ref. 138), but these are often more suitable for extended semiconductor or metal systems. Moreover, QM optimizations typically try to reach an energy minimum, while MD simulations model the dynamics of a system, allowing for many binding events and an increased sampling the conformational space. However, A recent review by Burke offers a useful perspective and further reading on DFT.[139] When it comes to studying self-assembly, electronic structure calculations have been used for:

i. Generating target data for parameterization of the force fields discussed in the previous sections.
ii. Giving accurate estimates of the binding energy of two monomers.
iii. Predicting spectroscopic properties (IR, UV/Vis, etc.) of self-assembled systems.

When new molecules are introduced in a force field, parametrization is a crucial and not necessarily trivial task, therefore researchers generally release a sigh of relief when the molecule they want to study is already parameterised. In principle, for every term in the force field (bonds, angles and (improper) dihedrals equilibrium values and force constants) the potential energy surface has to be determined using QM values and fitted to the appropriate function (see previous section). Also, determination of the point charges for every atom in the molecule is important, as it often implicitly used to account for dipole-dipole and dispersion interactions. Depending on the force field, charges and dipoles may have to be overestimated by as much as 50% compared to values calculated in the gas phase at QM level.[79] Automated parametrization software is available for AMBER (Antechamber[83]) and CHARMM (ParamChem[84,85]), these packages rely solely on analogy with other, parameterised molecules and thus has to be used with care. A relevant example to small biomolecules is given by Kaminski *et al*., who reported the evaluation and reparametrisation of the OPLS-aa force field based on QM calculations of all the homo-dipeptides.[140]

Comparing the binding of two monomers at MD level with that at QM level can be a valuable step. Not only when comparing the QM binding energy with the MD energy for parametrisation purposes, but in general this procedure can give important information regarding the propensity of a molecule to self-aggregate (the free energy change has to be negative). Moreover, the lowest-energy conformation of the stacked molecules can provide insight into the preferred packing of the monomers. Slight differences in DFT level stacking energy for various supramolecular conformations of gelator Fmoc-AA were observed as is reported in Chapter 3, but the effect of the limited size of the calculated stack was too prominent to draw conclusions on the most favourable stacking.[141] It is recommended that currently the comparison between theoretical and experimental spectroscopic results is a more suitable way to determine correct stacking.

Prediction of light absorption of small assemblies, most prominently in the IR, is an area of research that has been fairly well-developed for proteins. Especially for ubiquitous structural motifs such as α-helices and β-sheets, various methods have been proposed for prediction of infrared spectra, as outlined below. Currently, most calculations of protein absorption spectra concern the amide I band, as it is sensitive to secondary structure. The carbonyl group of the amide moiety is generally modelled as an oscillating dipole, so that the normal mode frequencies and couplings between the various oscillators can be calculated.[142–146] Although results from this Transition Dipole Coupling (TDC) method were initially only meaningful for infinite secondary structure elements, recent adaptations by La Cour-Jansen *et al*. and Kubelka *et al*.[147,148] have provided successful modelling of the amide I band, even for inhomogeneous structures. Zanni's group has recently described the use of excitonic Hamiltonians in the prediction of linear and two-dimensional IR (2DIR) spectroscopy for α-helices[149] and parallel β-sheets[104] in amyloids and proteins. Despite these successes, for more complicated systems it quickly becomes difficult to use these methods and the direct use of quantum chemical calculations is desirable, as recently summarised by Besley.[150] Using a partial Hessian method on a polyglycine peptide, a speed-up of approximately one order of magnitude was achieved with respect to the method by Kubelka *et al*. An important advance of this field is the introduction of DFT-MD and Carr-Parinello MD methods, recently reviewed by Gaigeot for gas and liquid phase calculations on peptides.[151] The combination MD with DFT allows for sampling of the conformational space, which provides an improved agreement with experimental IR spectra, especially for flexible peptides.

Few examples of spectra of self-assembled systems have been reported. Fleming *et al*. have discussed IR as a probe of supramolecular structures for Fmoc-dipeptides based on DFT calculations.[141] They concluded that their electronic structure calculations can aid structural understanding of nanostructures, but definitive differentiation between the various parallel or anti-parallel Fmoc-AA β-sheets still required larger calculations. Setnička *et al*. reported the use of IR and vibrational circular dichroism (VCD) spectroscopies coupled with DFT calculations.[152] They were able to elucidate the

specifically distorted structure of a supramolecular species in a gel state, in this case a potent guanosine-derived hydrogelator. VCD can be applied to resolve supramolecular structure in biopolymers as reviewed by Keiderling,[153] and DFT-based spectral predictions are available in popular computational chemistry packages such as Gaussian.

2.5 Conclusion and outlook

This chapter has provided an introduction to the field of molecular self-assembly, focussing on peptide and peptide derivative nanostructures and ways to model their aggregation. Sections 2.1 reviews the most important driving forces and definitions relevant to self-assembly. Moreover, a very brief overview of the field of peptide hydrogels and routes to their formation was given in section 2.2. These sections have provided the necessary background information for this thesis, in which self-assembling hydrogels are extensively discussed.

The main part of this chapter involves an overview of Molecular Dynamics (MD) simulations and their use in the field of the self-assembly of low molecular weight biomolecules. The principles of MD and crucial choices in setting up a simulation are discussed in sections 2.3.1-5, after which a number of case studies on mimicking the assembly of small biomolecules *in vitro* were discussed.

From these case studies, it becomes clear that MD simulations of self-assembling systems are still in the early stages of development, as only in the last few years computational resources have become available to a wider scientific community to run systems of relevant size and on a timescale suitable for the observation of nanostructure formation. Several results from recent studies on both all-atom and coarse-grain level have shown agreement between experiments and simulation results in terms of morphology and dynamics of assembled structures, including studies on the morphology of peptide amphiphiles, biologically relevant amyloid fragments and the high-throughput screening of peptides for new biomaterials. It seems that the field is about to reach the

point where meaningful results will be accessible even to non-specialists: many force fields and simulation platforms have reached a stage where automated parameterisation and intuitive user interfaces can help to set up an MD simulation of your system in a few hours. As MD simulations provide an intuitive insight into the supramolecular architecture, it is expected that MD simulations will become a standard addition to many research papers in the field of molecular self-assembly within 5–10 years. The advancement of the field will be accelerated by benchmarking of simulations and QM calculations with experimental results. There is a definite need for linking experimental observables with results from theoretical studies. Examples of such methods that are still under development for self-assembling systems include DFT-MD, VCD and time-dependent DFT (TD-DFT[154]) calculations, which can predict properties such as electronic band gaps and such UV/Vis or fluorescence spectra.

Finally, some suggestions for the use of quantum mechanical methods in the context of self-assembly are given in section 2.4. Electronic structure theory was found to be currently mainly important as a benchmark for informing all-atom force fields. However, with the development of multiscale simulations (QM/MM) the gap between the two levels of theory will slowly be bridged.

2.6 References

(1) Whitesides, G. M.; Grzybowski, B. *Science* **2002**, *295*, 2418–2421.

(2) Gross, R.; Dorigo, M. *Proc. IEEE* **2008**, *96*, 1490–1508.

(3) Atkins, P. W.; de Paula, J. *Atkins' physical chemistry.*; Oxford University Press: Oxford, 2006.

(4) Karplus, M. *Nat. Chem. Biol.* **2011**, *7*, 401–404.

(5) Frenkel, D. *Phys. Stat. Mech. Its Appl.* **1999**, *263*, 26–38.

(6) Palma, C.-A.; Cecchini, M.; Samorì, P. *Chem. Soc. Rev.* **2012**, *41*, 3713–3730.

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

(8) 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.

(9) Jayawarna, V.; Roy, S.; Holly, R. J.; Javid, N.; Zelzer, M.; Lamprou, D. A.; Urquhart, A. J.; Frederix, P. W. J. M.; Hunt, N. T.; Oreffo, R. O. C.; Richardson, S. M.; Merry, C. L. R.; Dalby, M. J.; Ulijn, R. V. *In preparation.*

(10) Raeburn, J.; Cardoso, A. Z.; Adams, D. J. *Chem. Soc. Rev.* **2013**.

(11) Hunter, C. A.; Lawson, K. R.; Perkins, J.; Urch, C. J. *J. Chem. Soc. Perkin Trans. 2* **2001**, 651–669.

(12) Bloom, J. W. G.; Wheeler, S. E. *Angew. Chem. Int. Ed.* **2011**, *50*, 7847–7849.

(13) Burattini, S.; Greenland, B. W.; Hayes, W.; Mackay, M. E.; Rowan, S. J.; Colquhoun, H. M. *Chem. Mater.* **2011**, *23*, 6–8.

(14) Babine, R. E.; Bender, S. L. *Chem. Rev.* **1997**, *97*, 1359–1472.

(15) Gonthier, J. F.; Steinmann, S. N.; Roch, L.; Ruggi, A.; Luisier, N.; Severin, K.; Corminboeuf, C. *Chem. Commun.* **2012**, *48*, 9239–9241.

(16) Sadava, D.; Heller, H. C.; Orians, G. H.; Purves, W. K.; Hillis, D. M. *Life: The Science of Biology*; Eighth Edition.; W. H. Freeman, 2007.

(17) Feynman, R. P. *J. Microelectromechanical Syst.* **1992**, *1*, 60–66.

(18) Drexler, K. E. "There's Plenty of Room at the Bottom." *www.metamodern.com* **2009**.

(19) Zhang, S. G. *Nat. Biotechnol.* **2003**, *21*, 1171–1178.

(20) Piner, R. D.; Zhu, J.; Xu, F.; Hong, S.; Mirkin, C. A. *Science* **1999**, *283*, 661–663.

(21) Brown, Z. Z.; Alleva, J.; Schafmeister, C. E. *Pept. Sci.* **2011**, *96*, 578–585.

(22) Graham, T. *Philos. Trans. R. Soc. Lond.* **1861**, *151*, 183–224.

(23) In *IUPAC Compendium of Chemical Terminology*; Nič, M.; Jirát, J.; Košata, B.; Jenkins, A.; McNaught, A., Eds.; IUPAC: Research Triangle Park, NC.

(24) Almdal, K.; Dyre, J. *Polym. Gels Networks* 5–17.

(25) Alemán, J.; Chadwick, A. V.; He, J.; Hess, M.; Horie, K.; Jones, R. G.; Kratochvíl, P.; Meisel, I.; Mita, I.; Moad, G.; Penczek, S.; Stepto, R. F. T. *Pure Appl. Chem.* **2007**, *79*, 1801–1829.

(26) Jordan Lloyd, D. In *Colloid Chemistry*; Alexander, J., Ed.; The Chemical Catalog Co.: New York, 1926; Vol. 1, pp. 767–782.

(27) Ghadiri, M. R.; Granja, J. R.; Milligan, R. A.; McRee, D. E.; Khazanovich, N. *Nature* **1993**, *366*, 324–327.

(28) Zhang, S.; Holmes, T.; Lockshin, C.; Rich, A. *Proc. Natl. Acad. Sci. U. S. A.* **1993**, *90*, 3334 –3338.

(29) Reches, M.; Gazit, E. *Science* **2003**, *300*, 625–627.

(30) Sanchez-de Groot, N.; Parella, T.; Aviles, F.; Vendrell, J.; Ventura, S. *Biophys. J.* **2007**, *92*, 1732–1741.

(31) Hartgerink, J. D.; Benlash, E.; Stupp, S. L. *Science* **2001**, *294*, 1684–1688.

(32) Zhao, X.; Pan, F.; Xu, H.; Yaseen, M.; Shan, H.; Hauser, C. A. E.; Zhang, S.; Lu, J. R. *Chem. Soc. Rev.* **2010**, *39*, 3480–3498.

(33) Fallas, J. A.; O'Leary, L. E. R.; Hartgerink, J. D. *Chem. Soc. Rev.* **2010**, *39*, 3510–3527.

(34) MacEwan, S. R.; Chilkoti, A. *Pept. Sci.* **2010**, *94*, 60–77.

(35) 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*, 3371–3379.

(36) Vegners, R.; Shestakova, I.; Kalvinsh, I.; Ezzell, R. M.; Janmey, P. A. *J. Pept. Sci.* **1995**, *1*.

(37) Yang, Z.; Liang, G.; Ma, M.; Gao, Y.; Xu, B. *J. Mater. Chem.* **2007**, *17*, 850–854.

(38) Chen, L.; Morris, K.; Laybourn, A.; Elias, D.; Hicks, M. R.; Rodger, A.; Serpell, L.; Adams, D. J. *Langmuir* **2010**, *26*, 5232–5242.

(39) Chen, L.; Revel, S.; Morris, K.; C. Serpell, L.; Adams, D. J. *Langmuir* **2010**, *26*, 13466–13471.

(40) Adams, D. J.; Morris, K.; Chen, L.; Serpell, L. C.; Bacsa, J.; Day, G. M. *Soft Matter* **2010**, *6*, 4144–4156.

(41) Morris, K. L.; Chen, L.; Raeburn, J.; Sellick, O. R.; Cotanda, P.; Paul, A.; Griffiths, P. C.; King, S. M.; O'Reilly, R. K.; Serpell, L. C.; Adams, D. J. *Nat. Commun.* **2013**, *4*, 1480.

(42) Xu, X.-D.; Chen, C.-S.; Lu, B.; Cheng, S.-X.; Zhang, X.-Z.; Zhuo, R.-X. *J. Phys. Chem. B* **2010**, *114*, 2365–2372.

(43) Toledano, S.; Williams, R. J.; Jayawarna, V.; Ulijn, R. V. *J. Am. Chem. Soc.* **2006**, *128*, 1070–1071.

(44) Jayawarna, V.; Ali, M.; Jowitt, T. A.; Miller, A. F.; Saiani, A.; Gough, J. E.; Ulijn, R. V. *Adv. Mater.* **2006**, *18*, 611–614.

(45) Das, A. K.; Collins, R.; Ulijn, R. V. *Small* **2008**, *4*, 279–287.

(46) 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*, 37–41.

(47) Das, A. K.; Hirst, A. R.; Ulijn, R. V. *Faraday Discuss.* **2009**, *143*, 293–303.

(48) Debnath, S.; Shome, A.; Das, D.; Das, P. K. *J. Phys. Chem. B* **2010**, *114*, 4407–4415.

(49) Cheng, G.; Castelletto, V.; Moulton, C. M.; Newby, G. E.; Hamley, I. W. *Langmuir* **2010**, *26*, 4990–4998.

(50) Adams, D. J.; Mullen, L. M.; Berta, M.; Chen, L.; Frith, W. J. *Soft Matter* **2010**, *6*, 1971–1980.

(51) Hughes, M.; Birchall, L. S.; Zuberi, K.; Aitken, L. A.; Debnath, S.; Javid, N.; Ulijn, R. V. *Soft Matter* **2012**, *8*, 11565–11574.

(52) Jonker, A. M.; Löwik, D. W. P. M.; van Hest, J. C. M. *Chem. Mater.* **2012**, *24*, 759–773.

(53) Sinthuvanich, C.; Veiga, A. S.; Gupta, K.; Gaspar, D.; Blumenthal, R.; Schneider, J. P. *J. Am. Chem. Soc.* **2012**, *134*, 6210–6217.

(54) Keyes-Baig, C.; Duhamel, J.; Fung, S.-Y.; Bezaire, J.; Chen, P. *J. Am. Chem. Soc.* **2004**, *126*, 7522–7532.

(55) Koutsopoulos, S.; Unsworth, L. D.; Nagai, Y.; Zhang, S. *Proc. Natl. Acad. Sci.* **2009**.

(56) Sharma, N.; Top, A.; Kiick, K. L.; Pochan, D. J. *Angew. Chem. Int. Ed.* **2009**, *48*, 7078–7082.

(57) Chung, W.-J.; Oh, J.-W.; Kwak, K.; Lee, B. Y.; Meyer, J.; Wang, E.; Hexemer, A.; Lee, S.-W. *Nature* **2011**, *478*, 364–368.

(58) Shao, H.; Parquette, J. R. *Chem. Commun.* **2010**, *46*, 4285–4287.

(59) 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*, 960–966.

(60) Wang, Q.; Yang, Z.; Wang, L.; Ma, M.; Xu, B. *Chem Comm* **2007**, 1032–1034.

(61) Gao, Y.; Zhao, F.; Wang, Q.; Zhang, Y.; Xu, B. *Chem. Soc. Rev.* **2010**, *39*, 3425–3433.

(62) Frederix, P. W. J. M.; Kania, R.; Wright, J. A.; Lamprou, D. A.; Ulijn, R. V.; Pickett, C. J.; Hunt, N. T. *Dalton Trans.* **2012**, *41*, 13112–13119.

(63) Tang, C.; Smith, A. M.; Collins, R. F.; Ulijn, R. V.; Saiani, A. *Langmuir* **2009**, *25*, 9447–9453.

(64) Adams, D. J.; Butler, M. F.; Frith, W. J.; Kirkland, M.; Mullen, L.; Sanderson, P. *Soft Matter* **2009**, *5*, 1856–1862.

(65) Haines, L. A.; Rajagopal, K.; Ozbas, B.; Salick, D. A.; Pochan, D. J.; Schneider, J. P. *J. Am. Chem. Soc.* **2005**, *127*, 17025–17029.

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

(67) Zelzer, M.; Todd, S. J.; Hirst, A. R.; McDonald, T. O.; Ulijn, R. V. *Biomater. Sci.* **2012**, *1*, 11–39.

(68) Williams, R. J.; Smith, A. M.; Collins, R.; Hodson, N.; Das, A. K.; Ulijn, R. V. *Nat. Nano.* **2009**, *4*, 19–24.

(69) 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*, 1089–1094.

(70) Luque, F. J.; Barril, X. *Physico-chemical and Computational Approaches to Drug Discovery*; Royal Society of Chemistry, 2012.

(71) Goodsell, D. S.; Morris, G. M.; Olson, A. J. *J. Mol. Recognit. JMR* **1996**, *9*, 1–5.

(72) Leach, A. R. *Molecular modelling: principles and applications*; Prentice Hall: Harlow [etc.], 2001.

(73) Marrink, S. J.; Tieleman, D. P. *Chem. Soc. Rev.* **2013**, *42*, 6801–6822.

(74) MBNExplorer http://www.mbnexplorer.com/users-guide/4-energy-and-force-calculation/43-molecular-mechanics-potential.

(75) Adcock, S. A.; McCammon, J. A. *Chem. Rev.* **2006**, *106*, 1589–1615.

(76) Karplus, M.; Kuriyan, J. *Proc. Natl. Acad. Sci. U. S. A.* **2005**, *102*, 6679–6685.

(77) Ash, W. L.; Zlomislic, M. R.; Oloo, E. O.; Tieleman, D. P. *Biochim. Biophys. Acta BBA - Biomembr.* **2004**, *1666*, 158–189.

(78) Brooks, B. R.; Brooks, C. L.; MacKerell, A. D.; 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*, 1545–1614.

(79) 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*, 671–690.

(80) Salomon-Ferrer, R.; Case, D. A.; Walker, R. C. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2013**, *3*, 198–210.

(81) Reif, M. M.; Hünenberger, P. H.; Oostenbrink, C. *J. Chem. Theory Comput.* **2012**, *8*, 3705–3723.

(82) Jorgensen, W. L.; Tirado-Rives, J. *Proc. Natl. Acad. Sci. U. S. A.* **2005**, *102*, 6665–6670.

(83) Wang, J.; Wang, W.; Kollman, P. A.; Case, D. A. *J. Mol. Graph. Model.* **2006**, *25*, 247–260.

(84) Vanommeslaeghe, K.; MacKerell, A. D. *J. Chem. Inf. Model.* **2012**, *52*, 3144–3154.

(85) Vanommeslaeghe, K.; Raman, E. P.; MacKerell, A. D. *J. Chem. Inf. Model.* **2012**, *52*, 3155–3168.

(86) 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*, 1361–1366.

(87) Mu, X.; Eckes, K. M.; Nguyen, M. M.; Suggs, L. J.; Ren, P. *Biomacromolecules* **2012**, *13*, 3562–3571.

(88) Bhandarkar, M.; Brunner, R.; Chipot, C.; Dalke, A.; Dixit, S.; Grayson, P.; Gullingsrud, J.; Gursoy, A.; Hardy, D.; Hénin, J.; Humphrey, W.; Hurwitz, D.; Krawetz, N.; Kumar, S.; Phillips, J.; Shinozaki, A.; Zheng, G.; Zhu, F. *NAMD User's Guide v2.6*; 2008.

(89) De Souza, O. N.; Ornstein, R. L. *Biophys. J.* **1997**, *72*, 2395–2397.

(90) *Molecular modelling: principles and applications*; Pearson Education, 2001.

(91) Jonkheijm, P.; Schoot, P. van der; Schenning, A. P. H. J.; Meijer, E. W. *Science* **2006**, *313*, 80–83.

(92) Röhrig, U. F.; Laio, A.; Tantalo, N.; Parrinello, M.; Petronzio, R. *Biophys. J.* **2006**, *91*, 3217–3229.

(93) Bellesia, G.; Shea, J.-E. *Biophys. J.* **2009**, *96*, 875–886.

(94) Cheon, M.; Chang, I.; Hall, C. K. *Biophys. J.* **2011**, *101*, 2493–2501.

(95) Frederix, P. W. J. M.; Ulijn, R. V.; Hunt, N. T.; Tuttle, T. *J. Phys. Chem. Lett.* **2011**, *2*, 2380–2384.

(96) Hess, B.; Kutzner, C.; van der Spoel, D.; Lindahl, E. *J. Chem. Theory Comput.* **2008**, *4*, 435–447.

(97) Sugita, Y.; Okamoto, Y. *Chem. Phys. Lett.* **1999**, *314*, 141–151.

(98) Yi, J.-Y.; Bernholc, J.; Salamon, P. *Comput. Phys. Commun.* **1991**, *66*, 177–180.

(99) Villani, V.; Tamburro, A. M. *J. Mol. Struct. THEOCHEM* **1994**, *308*, 141–157.

(100) Ostermeir, K.; Zacharias, M. *Biochim. Biophys. Acta* **2013**, *1834*.

(101) Hall, C. K.; Wagoner, V. A. In *Amyloid, Prions, and Other Protein Aggregates, Part B*; Academic Press, 2006; Vol. 412, pp. 338–365.

(102) Barrow, C. J.; Small, D. H. *Abeta Peptide and Alzheimer's Disease: Celebrating a Century of Research*; Springer, 2007.

(103) Marshall, K. E.; Morris, K. L.; Charlton, D.; O'Reilly, N.; Lewis, L.; Walden, H.; Serpell, L. C. *Biochemistry (Mosc.)* **2011**, *50*, 2061–2071.

(104) Woys, A. M.; Almeida, A. M.; Wang, L.; Chiu, C.-C.; McGovern, M.; de Pablo, J. J.; Skinner, J. L.; Gellman, S. H.; Zanni, M. T. *J. Am. Chem. Soc.* **2012**, *134*, 19118–19128.

(105) Lakshmanan, A.; Hauser, C. A. E. *Int. J. Mol. Sci.* **2011**, *12*, 5736–5746.

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

(107) Lee, O.-S.; Stupp, S. I.; Schatz, G. C. *J. Am. Chem. Soc.* **2011**, *133*, 3677–3683.

(108) Silva, G. A.; Czeisler, C.; Niece, K. L.; Beniash, E.; Harrington, D. A.; Kessler, J. A.; Stupp, S. I. *Science* **2004**, *303*, 1352–1355.

(109) Lee, O.-S.; Liu, Y.; Schatz, G. C. *J. Nanoparticle Res.* **2012**, *14*, 1–7.

(110) Boyle, A. L.; Bromley, E. H. C.; Bartlett, G. J.; Sessions, R. B.; Sharp, T. H.; Williams, C. L.; Curmi, P. M. G.; Forde, N. R.; Linke, H.; Woolfson, D. N. *J. Am. Chem. Soc.* **2012**, *134*, 15457–15467.

(111) Kushner, D. J. *Bacteriol. Rev.* **1969**, *33*, 302–345.

(112) Szostak, J. W.; Bartel, D. P.; Luisi, P. L. *Nature* **2001**, *409*, 387–390.

(113) Coveney, P. V.; Swadling, J. B.; Wattis, J. A. D.; Greenwell, H. C. *Chem. Soc. Rev.* **2012**, *41*, 5430–5446.

(114) Slocik, J. M.; Naik, R. R. *Chem. Soc. Rev.* **2010**, *39*, 3454–3463.

(115) Cygan, R. T.; Liang, J.-J.; Kalinichev, A. G. *J. Phys. Chem. B* **2004**, *108*, 1255–1266.

(116) Heinz, H.; Lin, T.-J.; Kishore Mishra, R.; Emami, F. S. *Langmuir* **2013**, *29*, 1754–1765.

(117) Heinz, H.; Farmer, B. L.; Pandey, R. B.; Slocik, J. M.; Patnaik, S. S.; Pachter, R.; Naik, R. R. *J. Am. Chem. Soc.* **2009**, *131*, 9704–9714.

(118) Nelson, K. E.; Levy, M.; Miller, S. L. *Proc. Natl. Acad. Sci.* **2000**, *97*, 3868–3871.

(119) Swadling, J. B.; Coveney, P. V.; Christopher Greenwell, H. *Geochim. Cosmochim. Acta* **2012**, *83*, 360–378.

(120) Marrink, S. J.; Risselada, H. J.; Yefimov, S.; Tieleman, D. P.; de Vries, A. H. *J. Phys. Chem. B* **2007**, *111*, 7812–7824.

(121) Monticelli, L.; Kandasamy, S. K.; Periole, X.; Larson, R. G.; Tieleman, D. P.; Marrink, S.-J. *J. Chem. Theory Comput.* **2008**, *4*, 819–834.

(122) De Jong, D. H.; Singh, G.; Bennett, W. F. D.; Arnarez, C.; Wassenaar, T. A.; Schäfer, L. V.; Periole, X.; Tieleman, D. P.; Marrink, S. J. *J. Chem. Theory Comput.* **2013**, *9*, 687–697.

(123) López, C. A.; Rzepiela, A. J.; de Vries, A. H.; Dijkhuizen, L.; Hünenberger, P. H.; Marrink, S. J. *J. Chem. Theory Comput.* **2009**, *5*, 3195–3210.

(124) López, C. A.; Sovova, Z.; van Eerden, F. J.; de Vries, A. H.; Marrink, S. J. *J. Chem. Theory Comput.* **2013**, *9*, 1694–1708.

(125) Guo, C.; Luo, Y.; Zhou, R.; Wei, G. *ACS Nano* **2012**, *6*, 3907–3918.

(126) Yan, X.; He, Q.; Wang, K.; Duan, L.; Cui, Y.; Li, J. *Angew. Chem. Int. Ed.* **2007**, *46*, 2431–2434.

(127) Zhang, H.; Wang, H.; Xu, G.; Yuan, S. *Colloids Surfaces Physicochem. Eng. Asp.* **2013**, *417*, 217–223.

(128) Lee, O.-S.; Cho, V.; Schatz, G. C. *Nano Lett.* **2012**, *12*, 4907–4913.

(129) Gudlur, S.; Sukthankar, P.; Gao, J.; Avila, L. A.; Hiromasa, Y.; Chen, J.; Iwamoto, T.; Tomich, J. M. *PLoS ONE* **2012**, *7*, e45374.

(130) Mondal, J.; Yethiraj, A. *J. Chem. Phys.* **2012**, *136*, 084902.

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

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

(133) Vácha, R.; Frenkel, D. *Biophys. J.* **2011**, *101*, 1432–1439.

(134) Bieler, N. S.; Knowles, T. P. J.; Frenkel, D.; Vácha, R. *PLoS Comput Biol* **2012**, *8*, e1002692.

(135) Bond, P. J.; Holyoake, J.; Ivetac, A.; Khalid, S.; Sansom, M. S. P. *J. Struct. Biol.* **2007**, *157*, 593–605.

(136) Shih, A. Y.; Arkhipov, A.; Freddolino, P. L.; Schulten, K. *J. Phys. Chem. B* **2006**, *110*, 3674–3684.

(137) Zheng, X.; Wang, D.; Shuai, Z. *Nanoscale* **2013**, *5*, 3681–3689.

(138) Aradi, B.; Hourahine, B.; Frauenheim, T. *J. Phys. Chem. A* **2007**, *111*, 5678–5684.

(139) Burke, K. *J. Chem. Phys.* **2012**, *136*, 150901–150901–9.

(140) Kaminski, G. A.; Friesner, R. A.; Tirado-Rives, J.; Jorgensen, W. L. *J. Phys. Chem. B* **2001**, *105*, 6474–6487.

(141) Fleming, S.; Frederix, P. W. J. M.; Ramos Sasselli, I.; Hunt, N. T.; Ulijn, R. V.; Tuttle, T. *Langmuir* **2013**, *29*, 9510–9515.

(142) Krimm, S.; Abe, Y. *Proc. Natl. Acad. Sci.* **1972**, *69*, 2788–2792.

(143) Torii, H.; Tasumi, M. *J Chem Phys* **1992**, *96*, 3379–3387.

(144) Chirgadze, Y. N.; Nevskaya, N. A. *Biopolymers* **1976**, *15*, 607–625.

(145) Chirgadze, Y. N.; Nevskaya, N. A. *Biopolymers* **1976**, *15*, 627–636.

(146) Nevskaya, N. A.; Chirgadze, Y. N. *Biopolymers* **1976**, *15*, 637–648.

(147) La Cour Jansen, T.; Dijkstra, A. G.; Watson, T. M.; Hirst, J. D.; Knoester, J. *J Chem Phys* **2006**, *125*, 044312–044312–9.

(148) Kubelka, J.; Kim, J.; Bour, P.; Keiderling, T. A. *Vib Spectrosc* **2006**, *42*, 63–73.

(149) Grechko, M.; Zanni, M. T. *J. Chem. Phys.* **2012**, *137*, 184202–184202–9.

(150) Besley, N. A. *Philos. Trans. R. Soc. Math. Phys. Eng. Sci.* **2007**, *365*, 2799–2812.

(151) Gaigeot, M.-P. *Phys. Chem. Chem. Phys.* **2010**, *12*, 3336–3359.

(152) Setnicka, V.; Novy, J.; Boehm, S.; Sreenivasachary, N.; Urbanova, M.; Volka, K. *Langmuir* **2008**, *24*, 7520–7527.

(153) Keiderling, T. A. In *Encyclopedia of Biophysics*; Roberts, G. C. K., Ed.; European Biophysical Societies' Association, 2013.

(154) Runge, E.; Gross, E. K. U. *Phys. Rev. Lett.* **1984**, *52*, 997–1000.

Chapter 3: Infrared spectroscopy as a structural diagnostic tool for β-sheets in self-assembling aromatic peptide amphiphiles*

* Parts of this chapter have been published as Fleming, S.; Frederix, P. W. J. M.; Ramos Sasselli, I.; Hunt, N. T.; Ulijn, R. V.; Tuttle, T. *Langmuir* **2013**, *29*, 9510–9515. [1]

Declaration of contributions to published article:

The author has been responsible for experimental design, collection and analysis of infrared spectra, data analysis of DFT calculations and manuscript preparation. Synthesis of Fmoc-AA and Flu-AA has been performed by S. Fleming, who also contributed to the manuscript. DFT calculations and extraction of the frequencies were performed by I. Ramos-Sasselli.

3.1 Abstract

β-sheets are a commonly found structural motif in self-assembling aromatic peptide amphiphiles and their characteristic 'amide I' infrared (IR) absorption bands are routinely used support the formation of supramolecular structure. In this chapter, the utility of IR spectroscopy as a structural diagnostic tool for this class of self-assembling systems is assessed. In addition, density functional theory calculations are employed to perform geometry optimizations and IR frequency calculations on small stacks of aromatic peptides to find a stable peptide stacking conformation that is agreement with experimental IR spectra.

Using 9-fluorenyl-methyloxycarbonyl dialanine (Fmoc-AA) and the analogous 9-fluorenyl-methylcarbonyl dialanine (Fmc-AA) as examples, it is demonstrated that the origin of the band around 1680–1695 cm^{-1} in Fourier Transform Infrared (FTIR) spectra, which was previously assigned to an antiparallel β-sheet conformation, is in fact absorption of the stacked carbamate group in Fmoc-peptides. IR spectra from ^{13}C labelled samples support our conclusions.

3.2 Introduction and objectives

Infrared (IR) spectroscopy is well-established as a useful technique to assist in the determination of secondary structure elements in proteins, specifically in the amide I region (1600–1700 cm^{-1}) which is very sensitive to hydrogen bonding patterns found in α-helices and β-sheets.[2–8] The basic theory of IR absorption is discussed in this chapter, followed by a more specific overview of its use in protein and peptide systems, including the physical origin of IR absorptions in the amide I region of the spectrum. Throughout the research reported in this thesis, IR spectroscopy was extensively used as a method to help characterize the supramolecular structure of self-assembling hydrogels composed of Fmoc-dipeptide nanostructures (see Chapters 4 and 6). It is therefore important that the use of IR spectroscopy is assessed for these systems.

3.2.1 IR Spectroscopy and vibrational energy levels

Infrared (IR) spectroscopy is a widely used technique to determine chemical structure, as chemical bonds have specific absorption frequencies depending on their atoms and connectivity. However, the environment of the IR-active group also influences the absorption frequency. This effect is frequently employed in secondary and tertiary structure determination of proteins and peptides. The fundamentals of IR spectroscopy relevant to this thesis will be given in this section.

Infrared light covers the wavelength range from about 0.8–1000 μm. Typically, it is divided in three loosely defined regions: near-infrared from 0.8 up to 2.5 μm; mid-IR from 2.5–20 μm and far-IR from 20 μm to 1 mm. Mid-IR is by far the most used in analytical chemistry, as it overlaps with the vibrational absorption bands of most common chemical bonds. Energies in IR spectroscopy are often expressed in reciprocal length units. The energy of a photon in reciprocal meters (m^{-1}) is directly related to the frequency of the light (v in Hz) by

$$\tilde{\nu} = \frac{\nu}{c} \qquad (3.1)$$

where c is the speed of light in m/s. However, in spectroscopy, it is more common to use reciprocal centimetres, or wavenumbers (cm^{-1}) instead. This unit will be used in this thesis.

For the case of simple diatomics, when bonds are modelled as harmonic oscillators, the energies of the vibrational levels (G_v in cm^{-1}) are quantized as

$$G_v = \omega_e(v + 1/2) \qquad (3.2)$$

where v is the vibrational quantum number ($v \geq 0$) and ω_e the equilibrium vibrational frequency in cm^{-1} given by

$$\omega_e = \frac{1}{2\pi c}\sqrt{\frac{k}{\mu}} \qquad (3.3)$$

which is influenced by the bond force constant k (in N/m) and the reduced mass of the molecule μ (in kg) and is therefore a good probe for molecular structure. Note that this principle lies at the basis of isotope label experiments in IR spectroscopy. This technique is often used in protein spectroscopy, if the environment of a single peptide bond needs to be distinguished from all the other peptide bonds in the system. For example when ^{12}C in a carbonyl bond is replaced by its isotope ^{13}C, the frequency of the band shifts by about 40–43 cm^{-1}.[9]

Importantly, the strength of the bond is also significantly influenced by its non-covalent environment. For example, hydrogen bond donation into an IR-active moiety can shift a

vibrational transition by tens of wavenumbers. This makes IR spectroscopy also a sensitive tool to probe supramolecular structure. Additionally, IR absorption is very sensitive to molecular symmetry and orientation: the direction of the transition dipole moments is correlated with the direction of molecular bonds. Therefore, experiments with polarized light can give a direct insight into the structure of a molecule.

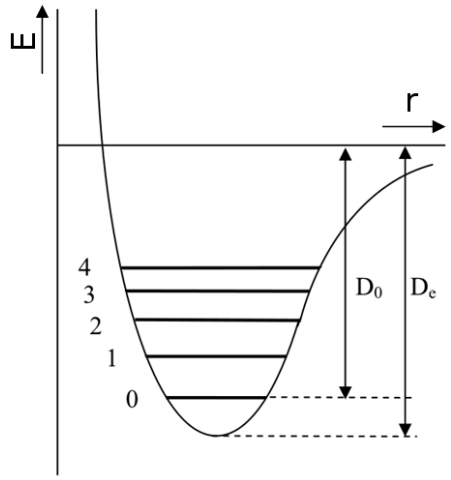


Fig. 3.1: Schematic of a typical potential energy well for a diatomic molecule, with vibrational levels 0-4 indicated. Notice that the spacing between the level decreases with v due to the anharmonicity of the potential well. Horizontal and vertical axes represent interatomic distance and energy, respectively. D_0 and D_e represent dissociation energy and well depth, respectively.

As can be seen from Fig. 3.1, the fact that the potential energy well for chemical bonds is not harmonic has an important influence on the energy of vibrational levels. The Morse potential shown gives a reasonably good dependence of the potential energy on the internuclear distance for a molecular bond: two atoms at infinite distance from each other have an interaction energy of 0, rather than an infinite energy as would be the case for a harmonic potential. From Schrödinger's equation it follows that in a 2^{nd} order approximation the energies of the vibrational levels in wavenumbers are given by

$$G_v = \omega_e(v + 1/2) - \omega_e x_e(v + 1/2)^2 \qquad (3.4)$$

where $\omega_e x_e$ is a small, positive constant that describes the anharmonicity. This has the consequence that the energy spacing between vibrational levels v and v+1 decreases as a function of v. The anharmonicity plays a role in calculations of the transition frequency (see section 3.4.2) and in non-linear IR spectroscopy (see Ch. 6).

When considering a light-induced transition between two vibrational levels, it can be derived from the transition dipole moment matrix elements that, to a first approximation, only transitions with $\Delta v = \pm 1$ are allowed within an electronic state. Overtone transitions ($\Delta v = \pm 2$ or higher) are only allowed when higher order terms are taken into account and are weaker by several orders of magnitude, which significantly simplifies IR spectra. Moreover, vibrational modes are only IR-active when absorption of an IR-photon excites a vibrational mode that is associated with a change in dipole moment ($d\mu/dr \neq 0$) of the molecule. In general, molecules have 3N – 6 vibrational modes (N is the number of atoms, 3N – 5 for linear molecules), although these may not all be IR-active.

Naturally, the nature of IR absorptions will be much more complicated for more complex system. In polyatomic molecules, vibrational modes can contain contributions from large numbers of bonds to form one *normal mode*. Moreover, the coupling of *local* vibrational modes (e.g. a carbonyl stretch in an amide group) between different molecules can lead to changes in the intensity and absorption frequencies in spectra by coupling of the transition dipole moments. The special case where the transition dipole moments of a large number of near-degenerate local modes (often in a one-dimensional stack) couple to form one large IR-active delocalized mode is usually called a (vibrational) exciton (see, e.g., ref. 10). This will be discussed for the case of proteins and peptides below.

3.2.2 IR spectroscopy of proteins and peptides

A well-studied case of strong coupling between transition dipole moments (TDM) is that of protein secondary structure motifs. The formation of α-helices, β-sheets and other structures creates a well-defined pattern of coupled oscillators in close proximity. This is depicted in Fig. 3.2 for anti-parallel β-sheets, which are of particular importance in this chapter. In a general case, spectra of parallel β-sheets often show a single band at approximately 1615–1640 cm^{-1}, while anti-parallel β-sheets have a weaker, but prominent additional component near 1685 cm^{-1} as discussed by Barth and Zscherp.[2] A list of typical frequencies for secondary structure motifs is given in Table 3.1

The exciton model explains that the low frequency component (1615–1640 cm^{-1}) arises from interstrand delocalization and mode coupling ($\nu_{\perp}$, see Figs. 3.2 and 3.3), which shifts the mode to lower wavenumbers as strands become more aligned and the number of strands per sheet increases (see Fig. 3.3).[5,6] In contrast, the frequency of the high wavenumber peak (1680–1695 cm^{-1}), is generally independent of the number of strands

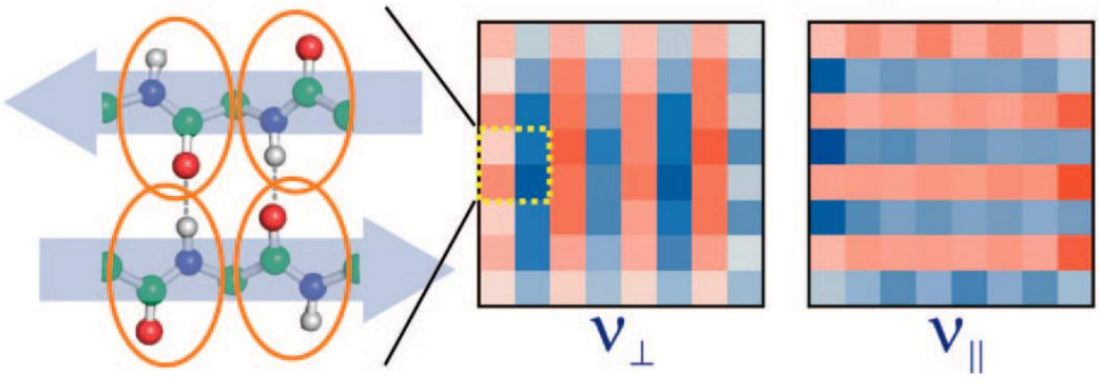


Fig. 3.2. Representation of $\nu_{\perp}$ and $\nu_{\parallel}$ excitons for a model anti-parallel β-sheet. The shading intensity is proportional to the participation ratio of a unit oscillator and the colour denotes its phase; blue and red are 180° out-of-phase. Figure adapted from refs. 7,11.

Table 3.1 Typical infrared absorption frequencies for common protein secondary structure elements in D_2O, obtained from ref. 2.

Protein secondary structure element	Band position in D_2O	
	Average	Extremes
α-helix	1652	1642–1660
β-sheet (par./antipar.)	1630	1615–1638
β-sheet (antipar. only)	1679	1672–1694
Turns	1671	1653–1691
Disordered	1645	1639–1654

and originates from vibrational excitons that run along a particular β-strand ($\nu_{\parallel}$, Figs. 3.2 and 3.8). Note that there is a degree of cooperativity to the intensity of the amide I absorption as strands become more aligned: a single mode can be particularly 'bright' as it is the vector sum of all the TDMs in phase (these modes are shown in Fig. 3.2). However, for symmetry reasons, the excitonic mode that is responsible for the high-frequency peak in antiparallel β-sheets has negligible intensity in parallel β-sheets: the individual local mode TDMs along the strands are out of phase for the parallel case, giving a net TDM value of 0 (in the representation of Fig. 3.2, this would be similar to a chess board-like pattern). The intensity of the amide I band is thus dependent on the presence of secondary structure, as is demonstrated in the results section of this chapter.

Thanks to the different exciton eigenstates, it is theoretically possible to differentiate between infinite parallel and antiparallel β-sheets in IR spectra; the former typically showing a single band at approximately 1615–1640 cm^{-1} and the latter having an additional component near 1685 cm^{-1}. However, it has been well documented that for several proteins with finite β-sheet content these bands can shift tens of wavenumbers, making the determination of secondary structure by linear IR spectroscopy challenging: disorder or twists can severely diminish the delocalization of vibrational modes and

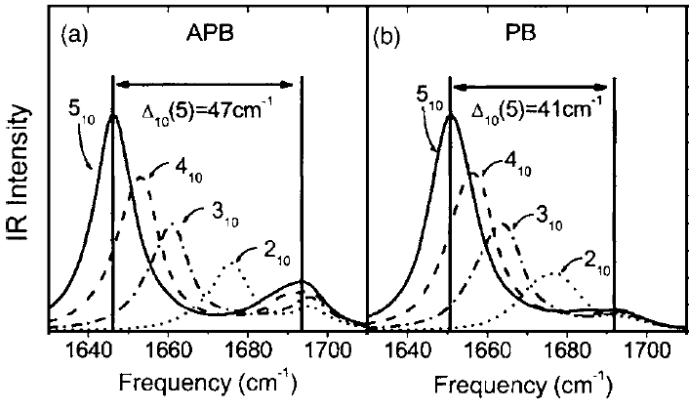


Fig. 3.3 Simulated IR spectra for 2 (denoted 2_{10}) to 5 (denoted 5_{10}) strands of 10 perfectly aligned amide oscillators in (a) anti-parallel β-sheet conformation, (b) parallel β-sheet conformation. Note the shift of the low-frequency component, denoted Δ_{10}, and the presence of a high frequency peak in the antiparallel case. Image obtained from ref. 6.

therefore broaden or shift optical transitions.[2,12–14] Barth and Zscherp have reported average values for several other secondary structures: the absorption of unordered, solvated amide bonds (e.g. random coil conformation in proteins) is generally found around 1645–1650 cm^{-1}, due to the absence of efficient coupling of the transition dipoles.[2] Moreover, from transition dipole coupling theory and their overview of experimental studies it follows that typical α-helical structures absorb only a few wavenumbers higher in frequency than disordered structures. However, spectra from α-helices can sometime be distinguished from randomly oriented peptides by a higher peak optical density (OD), because of the cooperative absorption effect discussed above.

Moreover, the practice of using models meant for the secondary structure determination of proteins may not be valid for interpreting the infrared spectra of aromatic di- and tripeptide hydrogels, which are discussed in this thesis. In fact, the transition dipole moment along the strands is relatively small for short peptide β-sheet strands (≤3 amino acids) and it is conceivable that the high-frequency peak may not be resolvable in spectra of their supramolecular structures, even if the peptides would be in perfect β-sheet

conformation (which is not expected for short, flexible peptides). Therefore, this chapter presents a combined experimental and computational analysis of IR bands in spectra of these short peptide amphiphiles.

3.2.3 IR spectroscopy in small peptide self-assembly and objectives

Various examples of self-assembling peptides that consist of 7 to 29 amino acid residues have been studied using IR spectroscopy.[15–18] Typically, a 1615 cm^{-1} amide I peak and a much weaker 1680 cm^{-1} peak were observed, which is characteristic of proteins with an antiparallel β-sheet structure as discussed above. The assignment β-sheets has also been reported for a class of low molecular weight (LMW) gelators composed from shorter (e.g. di- and tri-) peptides. These peptides are often capped at the N-terminus with an aromatic group to direct self-assembly; most commonly the Fmoc moiety is used (Fmoc, 9-fluorenylmethyloxycarbonyl).[19–28] The assignment of infrared bands that are apparently analogous to those seen in longer peptides, is often used as a key piece of evidence to support an anti-parallel β-sheet structure. While the absorption bands of LMW gelators resemble β-sheets signals in terms of line position, the relative infrared peak intensities for these materials are not always typical of that found for longer peptides, or indeed proteins.[22] Significant variation in the intensity of the 1690 cm^{-1} band has been observed, with this band occasionally being of similar or greater magnitude compared to the lower frequency (1615–1640 cm^{-1}) amide I contribution,[21,26,27,29,30] which is higher than expected for even a perfect anti-parallel β-sheet.[5–7]

The ambiguity surrounding these assignments has led to doubt regarding the presence of antiparallel β-sheet structures in spite of infrared evidence suggesting this conformation; with authors increasingly wary of applying traditional protein secondary structure interpretations.[32–34] In spite of IR evidence for ‘anti-parallel’ β-sheets, parallel β-sheet structures have been proposed based on X-Ray diffraction data,[33] and a recent combined

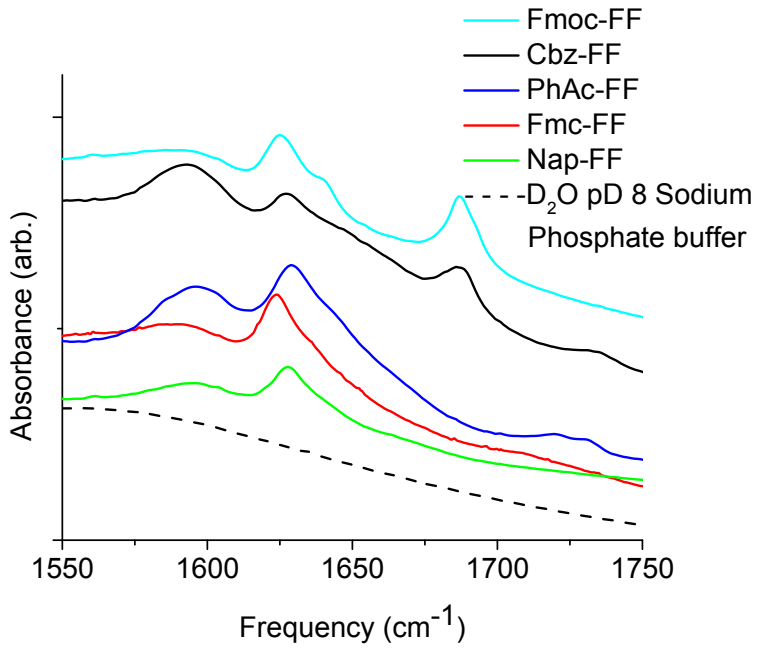


Fig. 3.4. Amide I FTIR spectra of diphenylalanine (FF) peptides protected with various aromatic groups. Spectra are vertically offset for clarity. Fmoc- and carboxybenzyl (Cbz) protection groups contain the carbamate moiety, while phenylacetyl (PhAc), Fmc and naphthalene (Nap) protected peptides do not. A typical spectrum from a 0.1M sodium phosphate buffer at pD 8 in D_2O is given to compare baseline absorption due to water. Figure adapted from ref. 31.

experimental and computational study has suggested a polyproline II conformation for Fmoc-AA,[34] which lacks significant hydrogen bonding between residues altogether. It should be noted that this study has also performed circular dichroism (CD) spectroscopy on Fmoc-AA gels, but suggested that the large differences between gelating peptides found in literature and the discrepancies from well-studied protein data make it difficult to use CD to make specific conclusions regarding the structural arrangement of molecules within a fibril assembly for, for example, Fmoc-AA.

Additionally, we have previously observed the absence of the 1690 cm^{-1} peak, in hydrogels assembled from aromatic dipeptides that lack the carbamate group as

summarised in Fig. 3.4.[31] For phenylacetyl- (PhAc), naphthalene- (Nap) and Fmc-protected diphenylalanine (FF), no high frequency peak was observed, while for carbamate containing Fmoc- and carboxybenzyl (Cbz) diphenylalanine this band was present. The carbamate group is known to absorb IR light in the 1685–1730 cm^{-1} range,[35–39] including a report by Nuansing *et al*. on Fmoc-FG powders,[40] significantly higher than the 1650 cm^{-1} absorption for free amide groups. For these reasons, it is important to assess the diagnostic value of the amide I infrared region for these systems.

In this chapter we study the amide I infrared bands of Fmoc-dialanine (from here on: Fmoc-AA) and Fmc-dialanine (Fmc-AA) hydrogels (Fmc, Fluorenylmethylcarbonyl), both experimentally using FTIR spectroscopy and isotope labelling (see section 3.4.1), and theoretically employing Density Functional Theory (DFT, section 3.4.2). Mu and co-workers recently studied the self-assembly of Fmoc-AA and reported the formation of a fibrous hydrogel at low pH.[34] Their results from IR spectroscopy and Molecular Dynamics (MD) simulations is discussed throughout section 3.4.

The use of DFT for self-assembling peptides has been discussed in Chapter 2. Briefly, it is used here for geometry optimisations of small stacks of Fmoc-AA and Fmc-AA to test the stability of particular parallel and antiparallel conformations. Moreover, DFT frequency calculations allow the comparison between experimental and theoretical IR absorption spectra, with insight in the individual vibrational modes for every peak in the spectrum.

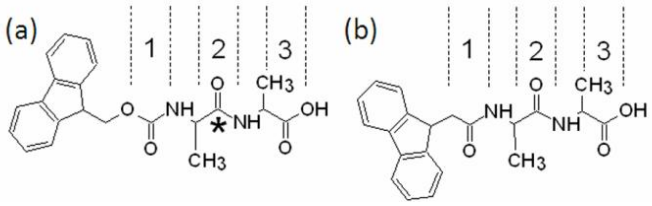


Fig. 3.5. Fmoc-AA and Fmc-AA. (a,b) Chemical structures, with numbered carbonyl groups. The asterisk in (a) indicates the position for the ^{13}C-labelling experiments.

Dialanine was chosen as the model dipeptide as alanine is the smallest chiral amino acid, which helped to facilitate the computationally demanding simulations. Nonetheless, the calculation of amide I modes of a certain peptide is not significantly influenced by the nature of the side chains, so the approach presented here is expected to be valid for all short peptides; the supramolecular structure may differ with different peptides, but this is accounted for in the geometry optimisation and hence the resulting spectrum. Fmc was utilised as a close analogue for the Fmoc moiety, but crucially lacks the carbamate oxygen and is instead linked to dialanine via an amide bond (see Fig. 3.5). By comparing the experimental and computational results, we can discuss the suitability of DFT frequency calculations for predicting IR spectra of this type of self-assembling systems (section 3.4.3).

3.3 Experimental details

All reagents were purchased from commercial sources and were used as supplied, unless stated otherwise in the experimental procedures. Details of synthesis and analysis of Fmc-AA and ^{13}C-isotope labelled Fmoc-AA are available in ref. 1.

3.3.1 Formation of gels (S. Fleming and I. Ramos-Sasselli)

20 mM deuterated gels were formed by dispersing Fmoc-AA or Fmc-AA in deuterium oxide (1 ml), and dissolving via the addition of 0.5 M NaOH deuterium oxide solution (50 μl). 1 M HCl was then added dropwise while the samples underwent repeated sonication and vortexing. The samples were self-supporting upon vial inversion at pH ~4–5 (pH measured in an H_2O sample). Only broad, baseline-type absorptions from HOD or H_2O were observed in the amide I region. Methanol solutions of Fmoc-AA or Fmc-AA were prepared at the same peptide concentration.

3.3.2 Fourier transform infrared spectroscopy

In Fourier transform infrared (FTIR) spectroscopy, the whole spectrum of a broadband source is frequency-modulated in a Michelson interferometer and the light transmitted by the sample is collected at the detector (see Fig. 3.2). The intensity is recorded as a function of the moving mirror's position and the absorption spectrum is obtained by a Fourier transform of the data. This allows for accurate data collection on a seconds timescale in commercially available spectrometers.

FTIR spectra were acquired in a Bruker Vertex 70 spectrometer (see Fig. 3.6) with a spectral resolution of 1 cm^{-1}. Spectra between 800 and 4000 cm^{-1} were obtained by averaging 25 interferograms for each sample. Measurements were performed in a standard IR cuvette (Harrick Scientific) in a transmission geometry, in which the sample

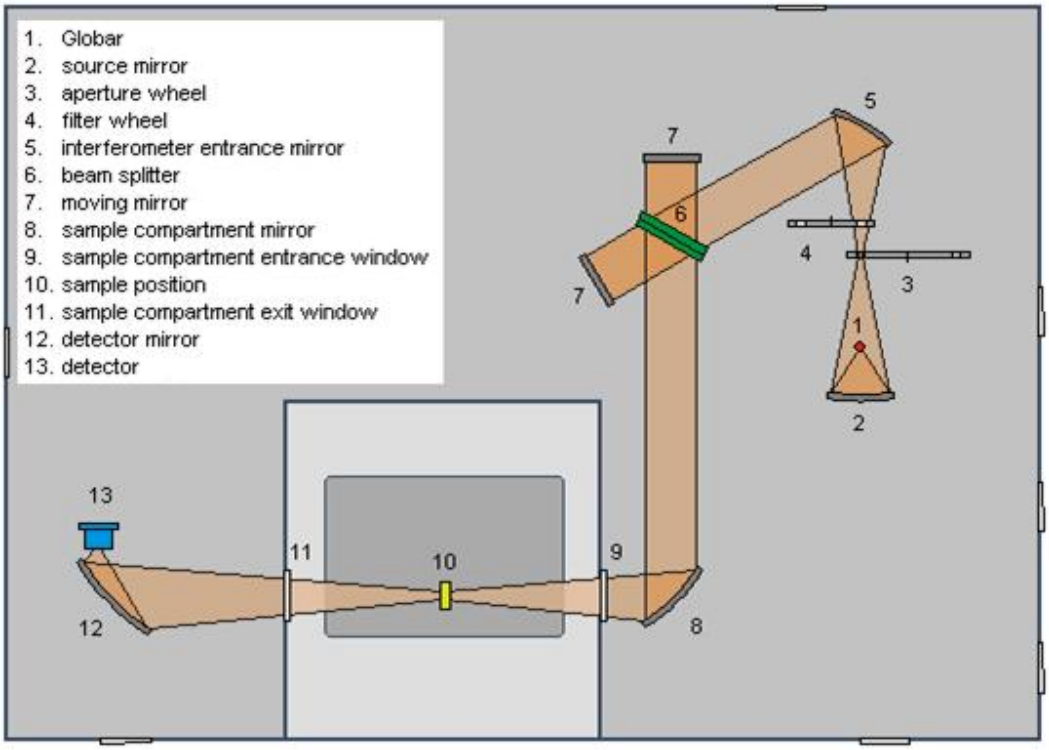


Fig. 3.6 Schematic layout of the Bruker Vertex 70 Fourier Transform Infrared Spectrometer, adapted from the OPUS 6.5 software.

was contained between two CaF_2 windows (thickness, 2 mm) separated by a 25-µm PTFE spacer. D_2O was used as the solvent for the hydrogel. corrected for atmospheric background absorptions from gaseous H_2O and CO_2 by subtracting a scan from an empty sample compartment. Because the broad background absorption from liquid D_2O and HOD in the sample may vary from sample to sample, no attempt was made to subtract this.

3.3.3 Vibrational frequency calculations (I. Ramos-Sasselli, P.W.J.M. Frederix)

Molecules were modelled in their C-terminally protonated form, since spectroscopy has previously shown that within hydrogels, the carboxylic acid groups exist predominantly in this form – a consequence of the "apparent pK_a shift" of the carboxylic acids when in

the self-assembled state.[41] All exchangeable hydrogens in the system were ^{1}H, which has been shown to have a negligible effect on amide I frequencies.[13] As dispersion effects are considered important for geometry optimizations in systems with an aromatic group such as fluorenyl, the gas phase structure optimisations were performed using Grimme's B97-D functional.[42] All calculations were carried out using the def2-SVP basis set,[43] which generally produces good results for main group atoms, while taking advantage of the MARI-J approximation[44] within the Turbomole 6.3.1 program (an example input can be found in Appendix 1.[45] Dimers and tetramers were built using Avogadro[46] from the geometry optimised monomer. It is important to perform the harmonic frequency calculations at the same level of theory, but unfortunately this functional is not available in Turbomole, so the frequency calculations were performed in Gaussian09 (an example input can be found in Appendix 1).[47] No scaling factor was applied to the results to correct for anharmonic effects and basis set truncation. Vibrations were visualised using the TmoleX program.[48]

Vibrational modes located >50% (by magnitude of Cartesian displacement vectors on relevant carbon atoms) on moieties other than the amide(s) or carbamate were omitted from the simulated spectra for clarity. Labels 1 and 2 indicate the number of the carbonyl giving the main contribution to the transition dipole moment (see labelling in Fig. 3.5(a,b)) determined by the largest local displacement of the carbonyl group. Modes with significant (20-50%) contribution from carboxylic acid groups are indicated with a dagger (†), while amide I modes arising from amide groups with their carbonyl group pointing out of the stack are indicated with an asterisk (*). For both of these cases, the calculated amide vibrational frequency is 30–100 cm^{-1} higher than would be expected for an ideally coupled amide mode, due to combination with a higher frequency mode or lack of the cooperative dipole coupling and hydrogen bonding, respectively. For completeness, the full calculation results from 1500-1900 cm^{-1} are given in Appendix 2 (Fig. A2.7).

Binding energies for the dimer and tetramer stacks were calculated according Eq. 3.5:

$$\Delta E_b = E_{n-mer} - n \cdot E_{monomer} \qquad (3.5)$$

where $E_{n\text{-}mer}$ is the energy-minimized di- or tetramer energy and $E_{monomer}$ the minimized monomer energy, both calculated in the gas phase.

3.4 Results and discussion

3.4.1 FTIR spectroscopy and isotope labelling

Opaque hydrogels were observed when the pH of the samples was lowered to ~4–5. The FTIR spectrum of the Fmoc-AA gel in the amide I region (Fig. 3.7(b)) clearly shows the characteristic bands described previously,[34] centred around 1638 cm^{-1} and 1685 cm^{-1}. A full spectrum can be found in Appendix 2, Fig. A2.8. Both bands are of a similar intensity, which is not as expected for antiparallel β-sheets, but is characteristic for Fmoc-dipeptide hydrogels.[41] In contrast, the gel of Fmc-AA shows only a single main peak at 1626 cm^{-1}, although the non-Gaussian shape suggests this is the convolution of two close absorption bands (i.e. amide groups nr. 1 and 2 in Fig. 3.5(b)). A small peak around 1650 cm^{-1} is usually attributed to 'random coil' conformation in protein spectroscopy. In short peptides, this is can be assigned to disordered or unstacked amide groups (as demonstrated in section 3.2.1, and further discussed in sections 4.3.4, 6.4.2 and ref. 49 and 50) indicating imperfect sheet stacking, or partial polyproline II conformation.[34] However, the dominant peaks at 1638 cm^{-1} (Fmoc-AA) and 1626 cm^{-1} (Fmc-AA) are more consistent with β-sheet structure.[2] Note that no bands from the fluorene moiety are present in this spectral region.[51]

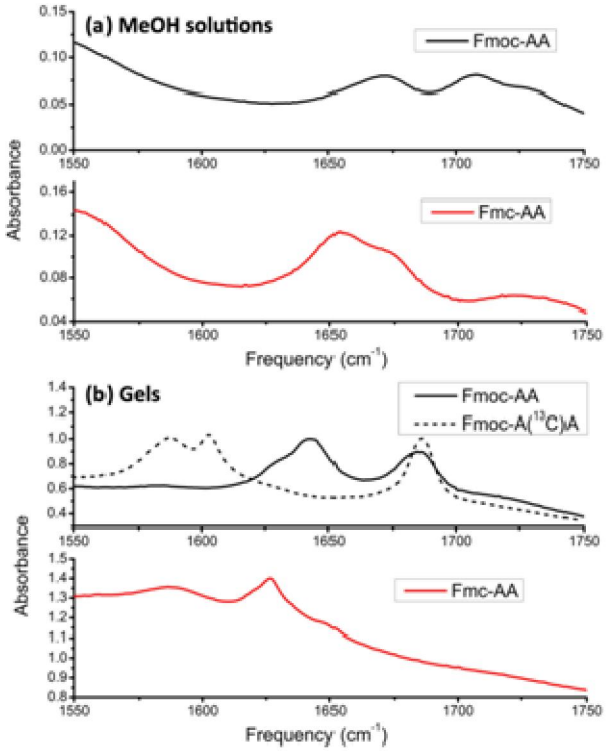


Fig. 3.7. Experimental FTIR spectra of Fmoc-AA and Fmc-AA as (a) 20 mM methanol solutions, (b) 20 mM deuterated hydrogels. Path length 25 μm. A broad, weak MeOH absorption is responsible for the sloped background below 1600 cm^{-1} in (a), a weak D_2O absorption is responsible for the sloped background in (b). Peaks in the gels are strongly redshifted and amplified in intensity compared to the spectra in MeOH.

In both spectra in Fig. 3.7(b), a broad contribution around 1590 cm^{-1} is discernible, which indicates that a fraction of terminal carboxylic acid groups remain deprotonated, in line with other Fmoc-dipeptide gels (refs. [41,50] and melting experiments in Chapter 6). Assuming that the self-assembly of Fmoc-AA and Fmc-AA proceeds in a similar fashion, these results would indicate the 1685 cm^{-1} peak is more likely a consequence of carbamate moiety absorption than the $\nu_{\parallel}$ component of antiparallel β-sheets, as the latter is expected to be of negligible intensity for very short peptide chains. This hypothesis is confirmed by the spectra of Fmoc-AA and Fmc-AA in neat methanol at the same concentration (Fig. 3.7(a)). Despite the absence of self-assembly in this solvent, it is clear that the spectrum of Fmoc-AA shows distinctly separate bands at 1669 and 1708 cm^{-1}

assigned to the respective amide I and carbamate absorptions. Whereas Fmc-AA shows two overlapping amide I bands around 1652 and 1670 cm^{-1} from carbonyl 1 and 2, respectively. Furthermore, in both experimental spectra a band is present at 1730 cm^{-1} due to the carboxylic acid terminal group. Note that a weak MeOH absorption is responsible for the sloped background below 1600 cm^{-1}. The absence of a 1590 cm^{-1} band indicates gelator molecules are in their fully protonated state when solvated in methanol.

It is apparent from Fig. 3.7(a) that the absorption of Fmoc-AA and Fmc-AA in methanol solution has a much lower intensity due to the lack of the cooperative effect (note the difference in scale between Fig. 3.7(a) and (b)). However, the relative intensities of the amide and carbamate bands in the Fmoc-AA hydrogel and solution are similar. This indicates that, in the hydrogel, both the 1638 cm^{-1} amide band and the 1685 cm^{-1} carbamate band of Fmoc-AA are equally enhanced as part of the β-sheet structure, which again suggests that interstrand delocalisation is responsible in each case, rather than $\nu_{\parallel}$ and $\nu_{\perp}$ components.

Conclusive evidence for the assignment of the 1685 cm^{-1} peak to the carbamate group was given by the IR spectrum of isotope-labelled Fmoc-AA* hydrogels, in which a single ^{13}C substitution was made at the amide carbonyl (see Fig. 3.5(a)). This procedure typically redshifts the vibrational frequency of the H-bonded carbonyl groups by 40-43 cm^{-1} (see section 3.1.1 and, e.g., ref. 9). Fig. 3.7(b) shows that, in comparison with unlabelled Fmoc-AA, only the lower frequency band moves from about 1638 to 1597 cm^{-1}, while the position of the 1685 cm^{-1} peak remains unaffected, in agreement with its different chemical nature.

3.4.2 DFT calculations

Vibrational frequencies were calculated for the single molecules as displayed in Fig. 3.8(a). Despite the absence of any β-sheet structure, the monomer of Fmoc-AA shows two distinct peaks with a separation of about 51 cm^{-1}, whilst Fmc-AA demonstrates a much smaller separation of 18 cm^{-1}. Moreover, when looking at the origin of the bands and assigning them to the carbonyl group with the largest contribution to the vibration (for labelling see Fig. 3.5(a,b)), it becomes apparent that the carbamate carbonyl (label 1 in Fmoc-AA spectra) vibrates at a distinctly higher frequency (65 cm^{-1}) than the analogous Fmc-amide group (1 in Fmc-AA spectra). The amide groups present in the middle of both monomers (labelled 2) are predicted to have similar frequencies as expected.

Next, parallel and antiparallel dimers and tetramers of both Fmoc-AA and Fmc-AA were used to mimic the potential supramolecular structures (Fig. 3.9). In addition, two alternative antiparallel arrangements were considered (see Figs. 3.10 and 3.11). The geometries of the structures were optimised (for results see Appendix 2, Fig 2.9-2.17) and their vibrational frequencies were calculated. For the antiparallel and parallel tetramers (Fig. 3.8(b,c)) and other intermolecular conformations (see Appendix 2), both Fmoc-AA and Fmc-AA spectra show multiple normal modes. This introduction of new peaks is expected as the number of oscillators in the system increases; however, experimentally many of the bands would overlap due to the line broadening associated with aqueous solvation. One would expect that the parallel ($\nu_{\parallel}$) and perpendicular ($\nu_{\perp}$) modes should progressively increase in peak height and become narrower in line width due to the coupling and cooperativity associated with these modes in an extended antiparallel β-sheet form (as discussed in 3.2.1). Nonetheless, the additional modes will also be present experimentally (i.e. as a weak, broad component around 1650 cm^{-1}), even in near-perfect sheets.[5]

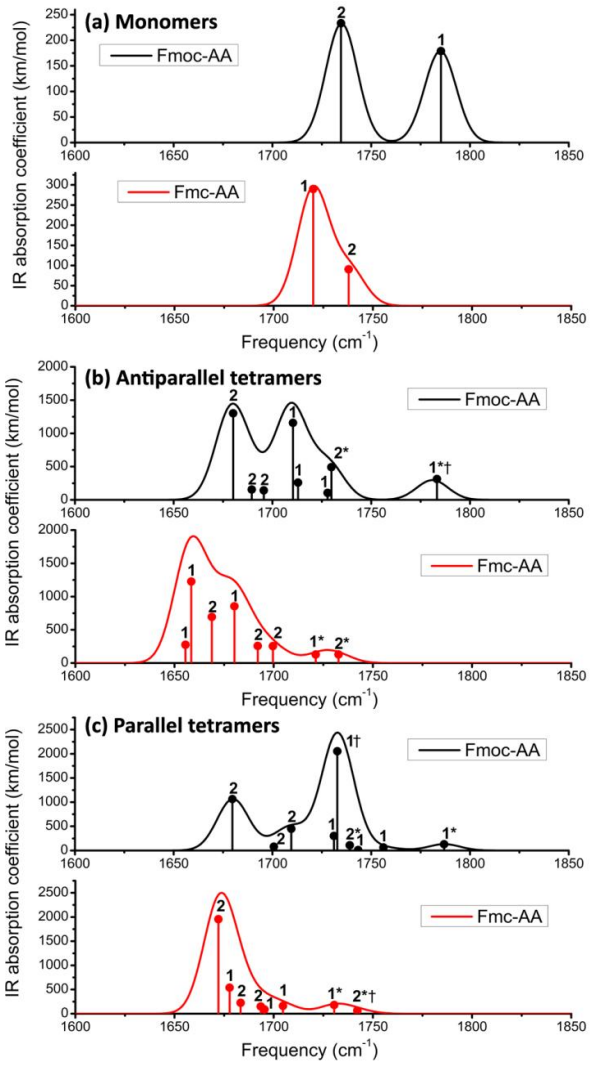


Fig. 3.8. B97-D/def2-SVP calculated amide I spectra in the gas phase for Fmoc-AA and Fmc-AA (a) monomers and (b) antiparallel and (c) parallel tetramers. Spectra are generated by applying an 8 cm^{-1} Gaussian linewidth (half-width at half maximum) to the calculated normal modes (vertical droplines).[52] *Labels 1 and 2 indicate the number of the carbonyl giving the main contribution to the transition dipole moment (see labelling in Fig. 3.5(a,b)). Vibrations localised on -COOH groups (carbonyl 3, generally >1790 cm^{-1}) groups are omitted. The mode indicated with dagger (†) has significant contribution from a -COOH group, while amide I modes arising from amide groups with their carbonyl group pointing out of the stack are indicated with an asterisk (*).*

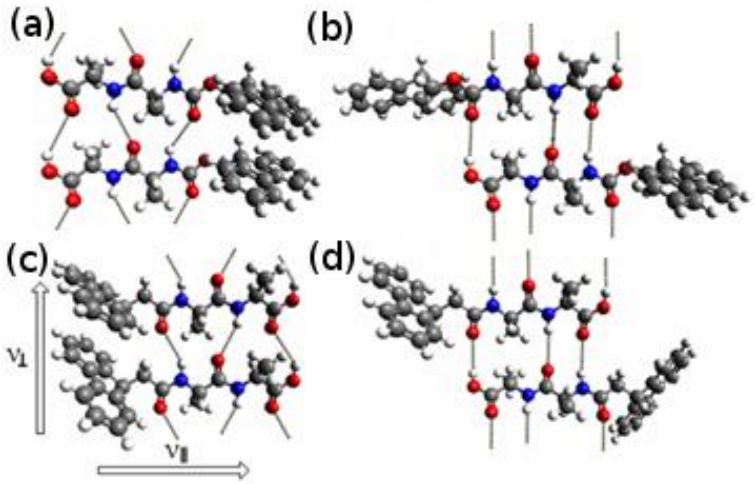


Fig. 3.9. Proposed dimeric molecular models built from B97D/def2-SVP optimised monomers showing the repeated β-sheet hydrogen bonding patterns; (a) parallel Fmoc-AA, (b) antiparallel 'ap1' Fmoc-AA, (c) parallel Fmc-AA, and (d) antiparallel Fmc-AA. Arrows show the direction of parallel ($\nu_\parallel$) *and perpendicular* ($\nu_\perp$) *excitons as discussed in the text. DFT optimised structures can be found in Fig. A2.10, 11, 14 and 15.*

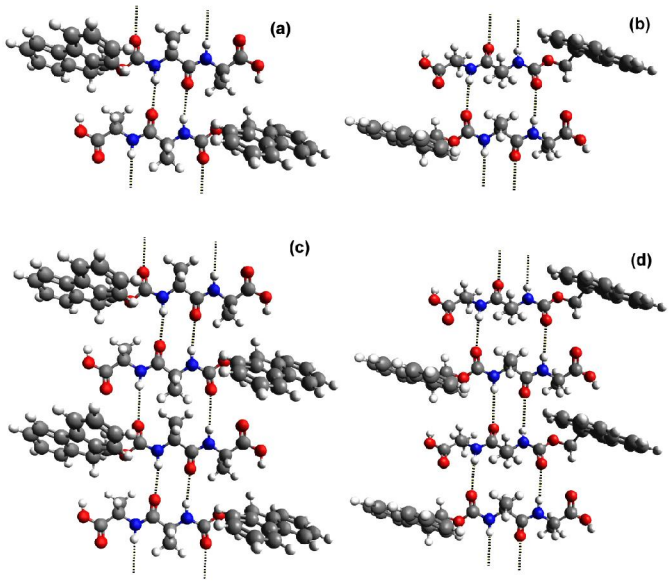


Fig. 3.10. Proposed antiparallel models for Fmoc-AA: (a) ap2a dimer; (b) ap2b dimer; (c) ap2a tetramer; (d) ap2b tetramer. Note that models a and b are different segments of the same structure. DFT-optimised structures can be found in Fig. A2.12, 13, 16 and 17.

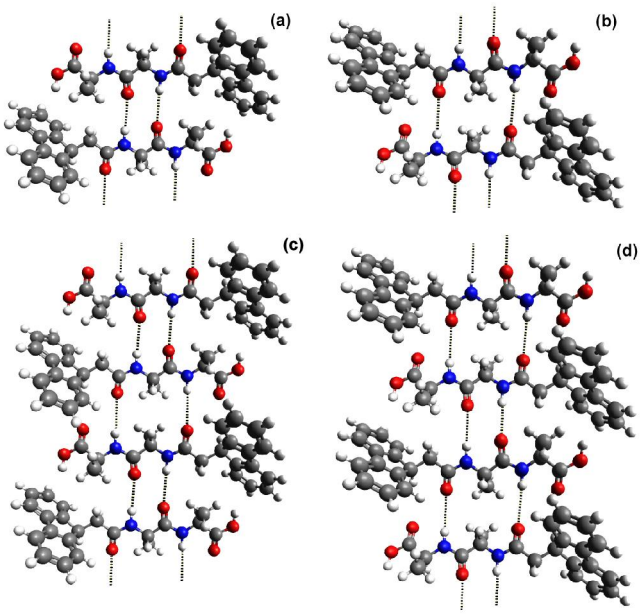


Fig. 3.11. Proposed antiparallel models for Fmc-AA: (a) ap2a dimer; (b) ap2b dimer; (c) ap2a tetramer; (d) ap2b tetramer. Note that models a and b are different segments of the same structure. DFT-optimised structures can be found in Fig. A2.12, 13, 16 and 17.

In our calculated spectra we observe a cooperative increase in the highest IR absorption coefficients with growing stack sizes, due to the delocalisation of the vibrational modes across multiple (often all) amide groups in the stack. Additionally, we see two main groups of Fmoc-AA bands: carbamate and amide groups are predicted to absorb at different frequencies (see Fig. 3.8 and Appendix 2, Fig. A2.1-A2.6). When the line positions for all the tetramer configurations are weighted by the intensity of their modes, the average predicted absorption frequencies are 1739 and 1697 cm^{-1} for Fmoc-AA modes mainly located on carbamate and amide groups, respectively, showing a significant 42 cm^{-1} separation. In contrast, in the calculated Fmc-AA spectra peaks from both amide groups are closer together, although on average modes arising from the amide group labelled 1 absorb at a lower frequency (1678 vs. 1685 cm^{-1}).

It should be noted that some modes, such as the 1781 cm^{-1} band in Fig. 3.8(b) (indicated with †) contain a significant contribution (24%) from a carboxylic acid vibration, which

can shift the vibrations towards higher frequencies by ~30 cm^{-1}. Furthermore, in all calculated spectra, the outermost C=O groups at the periphery of a β-sheet are predicted to absorb IR at higher frequencies due to the lack of H-bonding, as marked by asterisks in Fig. 3.8(b,c) and Fig. A2.1-A2.6. These peaks can be ignored for our purposes as the contribution from the edge of the stack will diminish in larger systems, but could show up in experimental spectra of disordered secondary structures as part of a component around 1650 cm^{-1}, typically assigned to random coil peptides. In summary, although there are clearly size-dependent artefacts in the model systems, these can be easily identified and rationalised. The DFT results for both monomers and supramolecular stacks support the notion that the presence of the 'split' amide I peak in systems based on Fmoc-AA is more dependent upon the presence of a two different chemical moieties (i.e. carbamate and amide), rather than indicative of an antiparallel β-sheet structure.

As well as vibrational calculations, binding energies were computed (see Eq. 3.5) for these structures (Fig. 3.12). β-sheets of Fmc-AA were found to be more stable than Fmoc-AA and the tetramers were more stable (on a per monomer basis) than the corresponding dimers due to cooperative binding effects. In addition, the parallel arrangement was generally found to be more stable than the antiparallel arrangement. However, this is likely to be caused by the π-π stacking of fluorene rings, which provided a contribution to the binding energy that is exclusively present in the parallel configuration due to the limited system size (see Figs. 3.9-11). The ap1 conformation was found to be slight more stable than the ap2 conformations, but these small variations in the relative binding energies of these arrangements are not good indicators for determining the equilibrium distribution of supramolecular structures, due to the neglect of several important features within these simplified models (e.g. solvent and edge effects). Rather, the correlation between the experimental and calculated vibrational spectra is a much more reliable indicator due to the local nature of the factors effecting the shifts in the amide I bands.

In any case, care has to be taken when using quantum mechanical methods for modelling dynamic structures such as LMW gelators. Even when the most energetically favourable nanostructure is found by geometry optimization, subtle influences from entropy and hydrogen-bonding in explicit water, temperature and local minima in the potential energy landscape of aggregation can lead to a variety of different structures being present in experimental situations. Molecular Dynamics simulations may therefore be more suitable for sampling nanostructure morphology, although hybrid methods are already in development that combine the conformational sampling of MD with the accuracy of QM frequency calculations, as recently reviewed by Gaigeot (see also Chapter 2).[53] These methods have been able to provide infrared spectra at finite temperature. However, for these small systems, we show that DFT calculations only can provide a valuable insight into supramolecular structures. Moreover, DFT-MD methods are currently computationally too expensive for extended nanostructures.

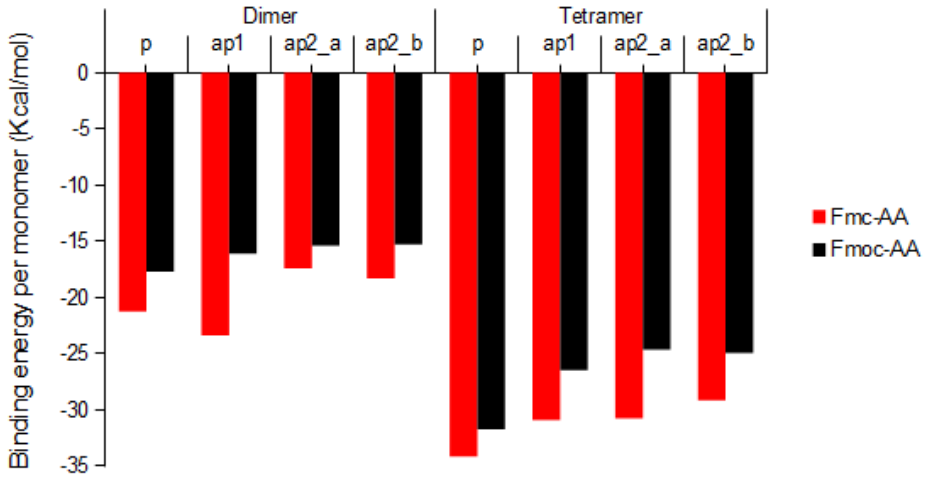


Fig. 3.12. Binding energy per monomer (gas phase, kcal/mol) for the various stacking conformations discussed in the text. For larger stacks, the binding energy is significantly higher. Note that due to edge effects, much larger stacks than tetramers would be necessary to draw firm conclusions as a result of calculated binding energies.

3.4.3 Comparison between experimental and calculated spectra

When the calculated infrared bands (Fig. 3.8) are compared with the experimental results (Fig. 3.7), several key features become apparent. When ignoring size-dependent artefacts in our DFT results such as vibrations located at the edge of the stack, the relative positions of the predicted absorption bands are in good agreement with experimental results. Furthermore, the experimental results consistently show the amide I peak of Fmc-AA at a lower frequency than that seen for the infrared bands of Fmoc-AA. This observation suggests the proximity of the fluorenyl group also influences the amide I vibration, in agreement with the computational spectra: amide group 1 in Fmc-AA absorbs at a lower frequency than amide groups 2 in both Fmc-AA and Fmoc-AA. Although the calculated frequencies are systematically too high, it is apparent that, as the finite computational models become larger, a gradual convergence towards the experimental infrared bands is observed (compare tetramers with dimers and monomers in Fig. 3.8 and A2.1–1.6). It should be noted that our frequency results have not been corrected for anharmonic effects and basis set truncation, because no extensively tested scaling factor is known for the combination of B97-D functional and def2-SVP basis set. The anharmonicity of the vibrational potential energy well usually leads to an overestimation of the frequencies as the difference between vibrational energy level is smaller than expected in the harmonic approximation (see Eq. 3.4). By applying a standard scaling factor (e.g., 0.97 as suggested by Fong and co-workers[54]) the agreement between the calculated and experimental data would be further improved, to within 5–15 cm^{-1} of the experimental result. This suggests accurate spectral predictions can be achieved with DFT calculations in the general case of self-assembling aromatic peptides.

Furthermore, the antiparallel tetramer results (Fig. 3.8(b), A2.5 and A2.6) appear to more closely resemble the experimental infrared spectra (peak spacing and relative intensities agree better with Fig. 3.7 then when experimental data is compared with the analogous parallel tetramer model (Fig. 3.8(c)). Although it is not possible to conclusively determine the stacking mode from these results, this hints at the notion that the antiparallel arrangement is in fact the most accurate depiction of the supramolecular

structure. In any case, this work mainly indicates a more cautious interpretation of infrared results of short peptides will be required in the future, but does not necessarily alter the conclusions of the parallel,[33] antiparallel,[26] or polyproline II models[34] proposed for specific cases. Ultimately, further work and complementary techniques, such as vibrational circular dichroism[13,55,56] or multidimensional IR[6,57,58] will be required to establish if the antiparallel conformation is indeed the prominent one in our case.

3.4 Conclusion

In summary, the use of IR spectroscopy in combination with DFT calculation for determining the supramolecular structure of aromatic peptide amphiphiles has been assessed. It is clear that both methods are sensitive to changes from monomer structures to aggregates as indicated by the changes in the spectra between methanol and gel phase, or between monomers, dimers and tetramers for calculations. Moreover, it is clear that our computed infrared bands vary with the stacking arrangement of these molecules and are thus a useful diagnostic tool to determine different structural features, as is done in Chapters 4 and 6.

Particularly in this chapter, the peak intensities in hydrogels and solution and the peak shifts under isotope labelling demonstrated that for self-assembling Fmoc-containing peptides the presence of two separate peaks in the amide I region of the IR is not indicative of antiparallel β-sheets. Instead, both experimental and computational results show that while the 1620–1640 cm^{-1} band may be assigned to β-sheet-like structures, it is the presence of the carbamate moiety that is responsible for the 1680–1690 cm^{-1} peak observed in these systems rather than a $\nu_{\parallel}$ component. A similar peak pattern can thus be expected for other peptides coupled to an aromatic group via a carbamate linkage, like, for example, carboxybenzyl (Cbz) or *t*-butoxycarbonyl (Boc)-protected peptides. Although the calculated peak patterns and intensities qualitatively agree with experimental results, currently our DFT-based computational models are insufficient in

size to allow us to categorically distinguish between different β-sheet arrangements. As a consequence, neither parallel nor antiparallel β-sheets can be ruled out as potential supramolecular structures on the basis of infrared spectroscopy alone. Work on these systems is currently ongoing to further refine the details that this approach can reveal.

3.5 References

(1) Fleming, S.; Frederix, P. W. J. M.; Ramos Sasselli, I.; Hunt, N. T.; Ulijn, R. V.; Tuttle, T. *Langmuir* **2013**, *29*, 9510–9515.

(2) Barth, A.; Zscherp, C. *Q. Rev. Biophys.* **2002**, *35*, 369–430.

(3) Rughani, R. V.; Schneider, J. P. *MRS Bull. Mater. Res. Soc.* **2008**, *33*, 530–535.

(4) Yan, H.; Frielinghaus, H.; Nykanen, A.; Ruokolainen, J.; Saiani, A.; Miller, A. F. *Soft Matter* **2008**, *4*, 1313–1325.

(5) Lee, C.; Cho, M. *J. Phys. Chem. B* **2004**, *108*, 20397–20407.

(6) Hahn, S.; Kim, S.-S.; Lee, C.; Cho, M. *J. Chem. Phys.* **2005**, *123*, 084905–084905–10.

(7) Ganim, Z.; Chung, H. S.; Smith, A. W.; DeFlores, L. P.; Jones, K. C.; Tokmakoff, A. *Acc. Chem. Res.* **2008**, *41*, 432–441.

(8) Cheatum, C. M.; Tokmakoff, A.; Knoester, J. *J. Chem. Phys.* **2004**, *120*, 8201–8215.

(9) Blume, A.; Huebner, W.; Messner, G. *Biochemistry (Mosc.)* **1988**, *27*, 8239–8249.

(10) Hamm, P.; Zanni, M. *Concepts and Methods of 2D Infrared Spectroscopy*; 1st ed.; Cambridge University Press, 2011.

(11) Chung, H. S.; Tokmakoff, A. *J. Phys. Chem. B* **2006**, *110*, 2888–2898.

(12) Torii, H.; Tasumi, M. *J. Chem. Phys.* **1992**, *96*, 3379–3387.

(13) Kubelka, J.; Keiderling, T. A. *J. Am. Chem. Soc.* **2001**, *123*, 12048–12058.

(14) Torii, H.; Tasumi, M. *J. Raman Spectrosc.* **1998**, *29*, 81–86.

(15) Rughani, R. V.; Salick, D. A.; Lamm, M. S.; Yucel, T.; Pochan, D. J.; Schneider, J. P. *Biomacromolecules* **2009**, *10*, 1295–1304.

(16) Measey, T. J.; Schweitzer-Stenner, R. *J. Am. Chem. Soc.* **2006**, *128*, 13324–13325.

(17) Measey, T. J.; Smith, K. B.; Decatur, S. M.; Zhao, L.; Yang, G.; Schweitzer-Stenner, R. *J. Am. Chem. Soc.* **2009**, *131*, 18218–18219.

(18) Schneider, J. P.; Pochan, D. J.; Ozbas, B.; Rajagopal, K.; Pakstis, L.; Kretsinger, J. *J. Am. Chem. Soc.* **2002**, *124*, 15030–15037.

(19) Shao, H.; Parquette, J. R. *Chem. Commun.* **2010**, *46*, 4285–4287.

(20) Orbach, R.; Mironi-Harpaz, I.; Adler-Abramovich, L.; Mossou, E.; Mitchell, E. P.; Forsyth, V. T.; Gazit, E.; Seliktar, D. *Langmuir* **2012**, *28*, 2015–2022.

(21) 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*, 1089–1094.

(22) 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*, 960–966.

(23) Jayawarna, V.; Richardson, S. M.; Hirst, A. R.; Hodson, N. W.; Saiani, A.; Gough, J. E.; Ulijn, R. V. *Acta Biomater.* **2009**, *5*, 934–943.

(24) Debnath, S.; Shome, A.; Das, D.; Das, P. K. *J. Phys. Chem. B* **2010**, *114*, 4407–4415.

(25) Williams, R. J.; Smith, A. M.; Collins, R.; Hodson, N.; Das, A. K.; Ulijn, R. V. *Nat. Nano.* **2009**, *4*, 19–24.

(26) 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*, 37–41.

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

(28) Xu, X.-D.; Chen, C.-S.; Lu, B.; Cheng, S.-X.; Zhang, X.-Z.; Zhuo, R.-X. *J. Phys. Chem. B* **2010**, *114*, 2365–2372.

(29) Tang, C.; Ulijn, R. V.; Saiani, A. *Langmuir* **2011**, *27*, 14438–14449.

(30) Orbach, R.; Adler-Abramovich, L.; Zigerson, S.; Mironi-Harpaz, I.; Seliktar, D.; Gazit, E. *Biomacromolecules* **2009**, *10*, 2646–2651.

(31) Roy, S.; Jayawarna, V.; Ulijn, R. V. *Soft Matter* (under review).

(32) Cheng, G.; Castelletto, V.; Moulton, C. M.; Newby, G. E.; Hamley, I. W. *Langmuir* **2010**, *26*, 4990–4998.

(33) Ikeda, M.; Tanida, T.; Yoshii, T.; Hamachi, I. *Adv. Mater.* **2011**, *23*, 2819–2822.

(34) Mu, X.; Eckes, K. M.; Nguyen, M. M.; Suggs, L. J.; Ren, P. *Biomacromolecules* **2012**, *13*, 3562–3571.

(35) Paquet, A. *Can. J. Chem.* **1982**, *60*, 976–980.

(36) Isama, K.; Kojima, S.; Nakamura, A. *J. Biomed. Mat. Res.* **1993**, *27*, 539–545.

(37) Egner, B. J.; Bradley, M. *Drug Discov. Today* **1997**, *2*, 102–109.

(38) Jalbout, A. F.; Xin-Hua, L.; Trzaskowski, B.; Raissi, H. *Eclética Quím.* **2006**, *31*, 53–62.

(39) Spegazzini, N.; Siesler, H. W.; Ozaki, Y. *J. Phys. Chem. A* **2011**, *115*, 8832–8844.

(40) Nuansing, W.; Rebollo, A.; Mercero, J. M.; Zuñiga, J.; Bittner, A. M. *J. Raman Spectrosc.* **2012**, *43*, 1397–1406.

(41) Tang, C.; Smith, A. M.; Collins, R. F.; Ulijn, R. V.; Saiani, A. *Langmuir* **2009**, *25*, 9447–9453.

(42) Grimme, S. *J. Comp. Chem.* **2006**, *27*, 1787–1799.

(43) Schäfer, A.; Horn, H.; Ahlrichs, R. *J. Chem. Phys.* **1992**, *97*, 2571–2577.

(44) Sierka, M.; Hogekamp, A.; Ahlrichs, R. *J. Chem. Phys.* **2003**, *118*, 9136–9148.

(45) Ahlrichs, R.; Bär, M.; Häser, M.; Horn, H.; Kölmel, C. *Chem. Phys. Lett.* **1989**, *162*, 165–169.

(46) Hanwell, M. D.; Curtis, D. E.; Lonie, D. C.; Vandermeersch, T.; Zurek, E.; Hutchison, G. R. *J. Cheminformatics* **2012**, *4*, 17.

(47) M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G. A. Petersson, H. Nakatsuji, M. Caricato, X. Li, H. P. Hratchian, A. F. Izmaylov, J. Bloino, G. Zheng, J. L. Sonnenberg, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, J. A. Montgomery, Jr., J.

E. Peralta, F. Ogliaro, M. Bearpark, J. J. Heyd, E. Brothers, K. N. Kudin, V. N. Staroverov, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, N. Rega, J. M. Millam, M. Klene, J. E. Knox, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, R. L. Martin, K. Morokuma, V. G. Zakrzewski, G. A. Voth, P. Salvador, J. J. Dannenberg, S. Dapprich, A. D. Daniels, Ö. Farkas, J. B. Foresman, J. V. Ortiz, J. Cioslowski, *Gaussian 09 Revision A.1*.

(48) Steffen, C.; Thomas, K.; Huniar, U.; Hellweg, A.; Rubner, O.; Schroer, A. *J. Comput. Chem.* **2010**, *31*, 2967–2970.

(49) 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.

(50) Frederix, P. W. J. M.; Kania, R.; Wright, J. A.; Lamprou, D. A.; Ulijn, R. V.; Pickett, C. J.; Hunt, N. T. *Dalton Trans.* **2012**, *41*, 13112–13119.

(51) Stein, S. E. In *NIST Chemistry WebBook, NIST Standard Reference Database Number 69, Eds. P.J. Linstrom and W.G. Mallard*; National Institute of Standards and Technology: Gaithersburg MD, 20899, http://webbook.nist.gov, 2013.

(52) Manor, J.; Feldblum, E. S.; Zanni, M. T.; Arkin, I. T. *J. Phys. Chem. Lett.* **2012**, *3*, 939–944.

(53) Gaigeot, M.-P. *Phys. Chem. Chem. Phys.* **2010**, *12*, 3336–3359.

(54) Fong, A.; Meyer, M.; O'Leary, D. *Molecules* **2013**, *18*, 2281–2296.

(55) Measey, T. J.; Schweitzer-Stenner, R. *J. Am. Chem. Soc.* **2011**, *133*, 1066–1076.

(56) Schweitzer-Stenner, R. *J. Phys. Chem. B* **2012**, *116*, 4141–4153.

(57) Woys, A. M.; Almeida, A. M.; Wang, L.; Chiu, C.-C.; McGovern, M.; de Pablo, J. J.; Skinner, J. L.; Gellman, S. H.; Zanni, M. T. *J. Am. Chem. Soc.* **2012**, *134*, 19118–19128.

(58) Hunt, N. T. *Chem. Soc. Rev.* **2009**, *38*, 1837.

Chapter 4: Sequence / structure relations in self-assembling peptide amphiphile nanostructures, studied by all-atom Molecular Dynamics and infrared spectroscopy*

---

* Parts of this chapter have been published as

Hughes, M.; Xu, H.; Frederix, P. W. J. M.; Smith, A. M.; Hunt, N. T.; Tuttle, T.; Kinloch, I. A.; Ulijn, R. V, *Biocatalytic Self-assembly of 2D Peptide-based Nanostructures*, Soft Matter, 2011, 7, 10032–10038.[1]
and
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., *Sequence/structure relationships in aromatic dipeptide hydrogels formed under thermodynamic control by enzyme-assisted self-assembly,* Soft Matter, 2012, 20, 5595-5602.[2]

Parts of this chapter are being prepared for submission as
Jayawarna, V.; Roy, S.; Holly, R. J.; Javid, N.; Zelzer, M.; Lamprou, D. A.; Urquhart, A. J.; Frederix, P. W. J. M.; Hunt, N. T.; Oreffo, R. O. C.; Richarson, S. M.; Merry, C. L. R.; Dalby, M. J.; Ulijn, R. V.,*Highly tunable gels via non-equilibrium self-assembly enable directive control of stem cell fate.*[3]
and
Abul-Haija, Y. M.; Roy, S.; Frederix, P. W. J. M.; Javid, N.; Jayawarna, V.; Ulijn, R. V. Biocatalytic Self-Assembly of Two-Component Core/Shell Nanofibres. Small (under review).[4]

Declaration of contributions to published articles:

The experimental work on self-assembly of Fmoc-dipeptides summarized in this chapter has been performed by Dr. Meghan Hughes. Experimental work including synthesis of materials for the core/shell nanostructures has been performed by Yousef Abul-Haija, Dr. Sangita Roy, Sanne Bakker, Fiona Divers and Dr. Nadeem Javid, except all data collection and analysis regarding FTIR spectra which has been performed by the author. All molecular modelling discussed was performed by the author.

4.1 Abstract

Molecular Dynamics (MD) simulations are an attractive way to obtain an insight into the architecture of supramolecular structures. Here, an atomistic MD approach to tackle a number of challenges in determining the relationship between sequence and structure in peptide-based nanostructures is discussed. Three case studies are presented in this chapter: (1) Fmoc-X_1X_2-OMe nanostructures (X_1 = S,T, X_2 = F,L); (2) co-assembly systems composed of hydrogelators Fmoc-FF or Fmoc-FY and surfactants Fmoc-X (X = S, T, RGD) and (3) self-assembling unprotected dipeptides FF and IF.

When the results from MD simulations on the systems mentioned above are compared to experimental data it becomes apparent that MD is a valuable tool for explaining nanostructure morphology. A model for the 2D self-assembled state of Fmoc-SF-OMe was obtained from molecular mechanics minimizations, in good agreement with experimental evidence from infrared (IR) and X-ray analysis, although the predicted structures were currently unstable in extended simulations.

Additionally, the core/shell structure proposed for Fmoc-FF/Fmoc-S and Fmoc-FY/Fmoc-S systems was confirmed by MD simulations starting from randomly oriented, solvated peptides. Additionally, the predicted behaviour of these molecules in different C-terminal ionisation states was shown to be consistent with experimental results concerning gelation at different pH values. Amide I IR spectra of these systems provides evidence for a β-sheet structure formed by the gelator, which is coated by the surfactant molecules. However, this propensity to form β-sheet type structures was not apparent from the simulations.

Finally, simulations on FF and IF show aggregation into 1D fibrous structures on a 100 ns timescale in water, where aggregation is dominated by hydrophobic interactions.

4.2 Introduction and objectives

As discussed in Chapter 2, peptide-based nanomaterials have attracted much interest in the fields of biomedicine and materials science. In order to accelerate the development of specific applications of these materials, it is imperative that the relationship between the amino acid sequence and the nanostructure morphology is understood. As the synthetic chemistry of peptides has been well-developed since the description of solid-phase methods by Merrifield[5] and can even be automated using peptide synthesizers,[6] *de novo* peptide hydrogels can then be easily adapted to the needs of new functional systems.

The main research question in this chapter is: How can atomistic Molecular Dynamics simulations and IR spectroscopy assist in explaining and predicting the sequence / structure relations in aromatic amphiphilic peptide hydrogels? To address this question, the results of three different, but related, case studies of self-assembling systems are presented. For each of these systems, experimental results from infrared spectroscopy are discussed, together with critical evidence from previous work. As described in the previous chapter, IR spectroscopy is a technique that has the capacity to provide valuable information on the hydrogen-bonding arrangement within peptide nanostructures. The analysis of more complex cases than Fmoc-AA (Ch. 3) gives a further insight into the use of IR spectroscopy as a tool to study self-assembling biomolecular systems. Subsequently, results from atomistic MD simulations are presented for the case under study.

Following the experimental studies, in section 4.5 an attempt is made to draw together lessons learnt in a critical assessment of the applicability of IR and MD to furthering our understanding of peptide nanostructure systems.

4.2.1 Case study 1: Fmoc-X_1X_2-OMe

Control over the morphology of structures formed by self-assembling aromatic peptide amphiphiles is crucial to their usage in the applications discussed in Chapter 2. By systematically varying the side chains of the amino acids, it is possible to study the relation between sequence and structure. Here, results from IR spectroscopy and MD simulations on a series of four Fmoc-dipeptide methyl esters are presented: Fmoc-SF-OMe, Fmoc-SL-OMe, Fmoc-TF-OMe and Fmoc-TL-OMe, which are referred to as SF, SL, TF and TL in this chapter. Our interest in this system stems from the fact that a dynamic combinatorial library approach has previously identified SF and SL as the thermodynamically most stable assemblies formed in a reaction pot containing Fmoc-S and methyl esters of F, L, Y, V, G and A.[7] An exhaustive study of these systems, using a range of experimental techniques, is discussed here and in refs. 1,2.

The Fmoc-protected dipeptide methyl esters can be produced via the principle of reverse hydrolysis by the thermolysin enzyme (discussed in Chapter 2.2). This method provides a way to perform self-assembly under thermodynamic control, i.e. without getting trapped in kinetic structural minima.[1,2,7] This system is of particular relevance as a case study of the sensitivity of molecular dynamics and infrared spectroscopy for this class of systems, as small changes in the amino acid side chains (e.g. S→T, F→L) incur dramatic changes in the nanostructure morphology.

Computational work on Fmoc-peptide nanostructures was undertaken by Xu *et al*. who proposed a hollow nanotube for Fmoc-LLL molecules based on microscopy and XRD results.[8] They observed a stable 8-by-2 stack of these molecules after a short (7 ns) NPT simulation (constant number of molecules, pressure and temperature) starting from a well-ordered flat initial conformation, which formed a curvature consistent with that of a nanotube with a diameter around the experimental value. Mu *et al.* recently demonstrated stable fibrils of Fmoc-AA in their MD simulations.[9] They observed the formation of densely packed fibres from different starting structures, but concluded that self-assembly

is mainly driven by various π-stacking of the Fmoc-rings, whilst β-sheet interactions don't play a significant role in this small peptide. Their comparison with WAXS data on Fmoc-AA reveals reflections at 26.3 Å and 4.35 Å, which they ascribe to fibril width (in agreement with earlier work on Fmoc-FF by Smith *et al.*[10]) and π-π stacking, respectively. The absence of the typically observed β-sheet peaks at 4.8 Å was explained by a different packing for peptide with small side chains: a detailed analysis of torsion angles of the peptide backbone suggested a large polyproline II-like component next to a small anti-parallel β-sheet element, confirmed by FTIR. However, the FTIR spectrum shows a β-sheet-type absorption of similar intensity to the band attributed to PPII conformation and their simulations show a prominent 4.8–5.0 Å maximum in the radial distribution function between peptide strands. The results presented in Chapter 3 also provide evidence for a β-sheet structure for this particular dipeptide. However, it should be noted that Fmoc-AA is significantly different from the peptides discussed in this case study, as our peptides feature a hydrophilic and strongly hydrophobic amino acid, and are methylated at the C-terminus. Moreover, the enzymatic preparation method here can access a thermodynamically lower state than the pH-controlled method employed by Mu *et al*. However, the application of MD to Fmoc-dipeptides is relevant.

Here, we propose an antiparallel π-β peptide stacking configuration and test its stability using geometry optimisations (energy minimisations) in the CHARMM force field, which has been explicitly developed for simulations of biological and drug-like molecules.[11,12] The resulting structures are compared with the experimental data from wide-angle X-ray scattering (WAXS) and IR spectroscopy.

4.2.2 Case study 2: Core/shell nanostructures for cell culture applications

One area where self-assembling hydrogels have shown great potential is the development of biomaterials as tissue culture media.[13] Specifically, it has been shown that stem cell differentiation can be controlled by encapsulation in hydrogels, although the largest

successes so far have focussed on incorporating biochemical growth factors (such as oligopeptide recognition sequences)[14,15] rather than using the matrix itself as a tool. However, it is well-known that physical stimuli, such as matrix stiffness and topology (i.e. size and shape of microscopic features) also have an important effect on stem cell differentiation.[16] Controlling stem cell fate without complex differentiation media has great advantages for both *in vivo* and *in vitro* applications.

It was previously found that simply using Fmoc-dipeptide gelators creates a scaffold that is too hydrophobic for certain types of cells to grow on.[17] Coating the Fmoc-dipeptide fibres with an Fmoc-protected single amino acid ('surfactant') provides a handle for introducing chemical functionality and a more hydrophilic fibre surfaces into the hydrogels. Various coated nanostructures have been previously tested in our group, including Fmoc-FF / Fmoc-K and Fmoc-FF / Fmoc-D, as lysine and amines have been reported to be potentially more suitable for a cell culture environment.[17–19] However, so far little evidence was provided that supports the proposed core/shell nanostructure. Here, results are presented on hydrogels formed from various coassemblies of gelators (either Fmoc-FF or Fmoc-FY) with protected amino acid 'surfactants' (either Fmoc-S, Fmoc-T, Fmoc-K or Fmoc-RGD, where RGD is the well-known arginine-glycine-aspartic acid cell adhesion motif).[3,4,20,21] It is proposed that these mixed gelator and surfactant systems form core/shell nanostructures with a hydrophobic core and a hydrophilic surface. IR spectroscopy and MD simulations are applied here to aid in clarifying the molecular architecture of these systems.

It should be noted that for the Fmoc-FY / Fmoc-X structures an enzymatic route to hydrogel formation was used. Self-assembly was triggered by alkaline phosphatase, employing a method originally developed by Yang *et al*.[22] and applied by Sadownik *et al.* to Fmoc-FY.[23] This enzyme catalyzes the removal of the phosphate group from a gelator precursor, phosphorylated Fmoc-FY^{P} to change the hydrophobicity of the molecule and start self-assembly. Enzymatic gel formation allows for temporal and/or

spatial control and often a more reproducible production process, as was discussed above and in Ch. 2.2.[24]

Additionally, the systems with co-assembled monomers in this case study provide a good test for the accuracy of MD simulations in simulating nanostructure morphologies. A different approach compared to the case of Fmoc-X_1-X_2-OMe in the previous section was applied to this effect; currently, available computational power still significantly limits the size and length of MD simulations. Therefore, in many cases, it is necessary to start from a pre-defined supramolecular structure to get results on its morphology. However, it is much more attractive to start from a random peptide solution and form nanostructures without introducing bias. This would also significantly reduce the amount of manual work going into setting up MD simulations of new systems or screening large amounts of systems (see also Chapter 5). This 'unbiased' approach was used for the Fmoc-FF / Fmoc-S system discussed here. Additionally, as shown by the work of Tang *et al*.[25] and acidity-induced gelation methods employing glucono-δ-lactone hydrolysis developed by Adams *et al*.,[26] the protonation state of small peptide hydrogelators is crucial to their self-assembly, as gelation can often be triggered by lowering the pH. Therefore, MD simulations of this system in various ionisation states have been performed in section 4.4.2.3.

4.2.3 Case study 3: Self-assembly of FF and IF

In addition to the large library of self-assembling (Fmoc-)protected peptides, a small number of unprotected dipeptides have been experimentally shown to be able to form self-assembled nanostructures in aqueous solution. Diphenylalanine (FF) is the most famous example, first reported by Reches and Gazit as a hollow template for casting gold and silver nanowires.[27] In the same report, they showed that other aromatic dipeptides, FW, WY, WF and WW mainly form amorphous aggregates, although TEM reveals some nanotubes as well for FW. In their study related to amyloid disease, Sanchez-De Groot *et*

al. showed that IF spontaneously forms nanofibres upon aqueous dilution from hexafluoroisopropanol, while similar peptide VF does not.[28] For both FF and IF peptides, crystal structures were reported by Görbitz.[29,30]

In 2009, Tamamis *et al*. performed Replica Exchange MD simulations on 12 FF molecules and observed occasional formation of 5 or 6-membered rings,[31] in agreement with a detail of the XRD structure, but observed many other configurations as well. They concluded hydrophobic interactions dominate aggregation / assembly in aqueous solution, although Görbitz states head-to-tail interactions between the N- and C-termini of dipeptides are more important.[32] More recently, Rissanou *et al*. reported a more detailed analysis of atomistic simulations on FF in water and methanol.[33] They confirmed the formation of nanostructures in water, while they were absent in the organic solvent, and a certain degree of parallel and anti-parallel stacking in 100 ns simulations of 16 monomers. As aggregation is less prominent in methanol solution, they state hydrogen bonding between the termini is the main driving force. However, in coarse-grain simulations of FF performed by Guo *et al*., side-chain interactions were reported to be the largest energetic contribution to stacking.[34] As a third case study, we aim to shed more light on these discrepancies by comparing IR spectra both FF and IF dipeptides and assessing the ability of atomistic MD simulations on larger systems to access the experimentally reported structures, starting from a random, solvated configuration.

4.3 Methods and materials

4.3.1 Hydrogel preparation

4.3.1.1 Preparation of Fmoc-X_1X_2-OMe hydrogels (Dr. M. Hughes)

In a typical experiment, Fmoc-amino acids and amino acid methyl esters HCl salts (both Sigma-Aldrich) were mixed in a 20:80 mM ratio. Addition of 1 mg/mL lyophilised

thermolysin powder (*bacillus Thermoproteolyticus rokko*, Sigma-Aldrich batch 079K1706, mol. wt. 34.6 kDa by amino acid sequence) and subsequent vortexing and sonication would start the formation of the hydrogel by the condensation reaction between the precursor molecules (see also Fig. 2.6). HPLC was used to check reaction conversion after 24 hours of incubation at RT or for monitoring the reaction progression in time.

4.3.1.2 Fmoc-FF / Fmoc-S 'Chemical' gels (Dr. V. Jayawarna)

Cell culture inserts for 12-well and 24 well multiwell plates with 1.0 μm pore size were purchased from Greiner Bio-one, UK. Fmoc-FF and Fmoc-S powders were sterilized under UV light for 45 min prior to use. Samples were prepared by mixing Fmoc-FF and Fmoc-S (Fmoc-FF/S) (1:1 ratio) in 10 mL glass vials and suspending the powders to a total concentration of 20 mM in sterile/distilled H_2O. To a 4 mL Fmoc-FF/S suspension, 130, 140 and 150 μL of 0.5 M sterile NaOH were added dropwise to give rise to opaque, viscous liquids of pH 7.2, 7.8 and 8.4. Final concentrations of Fmoc components in three pre-gelation mixtures (taking the dilution effect in consideration) are 19.37, 19.32 and 19.28 mM respectively.

To produce hydrogels from each of these pre-gelation mixtures, 300 μL were transferred to cell culture inserts in 12-well plates. A volume of 1400 μL of culture medium (αMEM) containing 10% fetal calf serum and antibiotics was added to each well (outside the insert) and incubated for 1.5hrs at 37 °C in a humidified atmosphere with 5% CO_2. Following gelation the medium surrounding the inserts was replaced and 300 μL of new medium were gently added to the surface of the gels. Following overnight culture the pH of the gel stabilized at around pH 7.8 (±0.5).

4.3.1.3 Fmoc-FY / Fmoc-X enzymatic gels (Y. Abul-Haija)

Enzymatically prepared gels for cell culture used 10 mM of each monomer, leading to a total peptide concentration of 20 mM, except from the Fmoc-FY/Fmoc-RGD sample which had a 15:5 mM ratio. 25 units of Alkaline Phosphatase (Sigma, lot no. 050M1581V) were added and immediately after that, the solution was allowed to gel in an incubator at 37°C in 5% CO_2 atmosphere. All characterizations were performed after 24 hours unless otherwise mentioned.

4.3.1.4 Fmoc-FY / Fmoc-S 'chemical' gels (F. Divers)

Fmoc-FY and Fmoc-S were mixed in 100:0, 75:25, 50:50, 25:75 and 0:100 ratios to a total concentration of 20 mM in D_2O and solved by addition of 40 μL of 1M NaOH. A gel was formed by lowering the pH to physiological conditions (7.5 ± 0.4) by addition of 1M HCl. To agree with the enzymatically prepared gels, the solutions were allowed to gel in an incubator at 37°C in 5% CO_2 atmosphere.

4.3.1.5 IF dipeptide gels

IF hydrogels were prepared by making a 100 mg/mL stock solution of the peptide in 1,1,1,3,3,3-hexafluoroisopropanol (HFIP), followed by dilution using deionized water to a final peptide concentration of 1.5% w/v, following the procedure in ref. 28. No gelation was observed without the presence of HFIP; after dissolving the peptide in water at 50°C, crystallisation was observed when the solution was cooled back to room temperature. Note that the FF samples by Reches and Gazit were also obtained by aqueous dilution of a HFIP stock solution.

4.3.2 Fourier transform infrared spectroscopy

FTIR spectra were acquired in a Bruker Vertex 70 spectrometer with a spectral resolution of 1 cm^{-1} (2 cm^{-1} for Fmoc-X_1X_2-OMe and Fmoc-FF/Fmoc-S samples). The spectra were obtained by averaging 25 interferograms for each sample (64 for Fmoc-X_1X_2-OMe samples, 56 for Fmoc-FF/Fmoc-S). Measurements were performed in a standard IR cuvette (Harrick Scientific), in which the sample was contained between two CaF_2 windows (thickness, 2 mm) separated by a 25 μm PTFE spacer (10 μm for Fmoc-X_1X_2-OMe samples). D_2O (Sigma-Aldrich, 99.9 atom% D) was used as the solvent for all the IR measurements. Spectra were corrected for atmospheric background absorptions from gaseous H_2O and CO_2 by subtracting a scan from an empty sample compartment. Because the broad background absorption from liquid D_2O and HOD in the sample may slightly vary from sample to sample, no attempt was made to subtract this.

4.3.3 Atomistic Molecular Dynamics simulations

All atomistic simulations were performed using the open source NAMD package[35] and the CHARMM27 and CHARMM c36a3 force fields.[11,12] Parameters for all amino acids existed, but parameters for Fmoc-groups were created in-house as follows. Existing charges, angles and dihedrals were used where analogous moieties (fluorene, esters and amides) were found. For all dihedral angles in the –CH_2–O–CO–NH– group, equilibrium angles and spring constants were determined by fitting the CHARMM energy profile to the potential energy surface calculated at the density functional theory (DFT) level. This surface was determined by single-point energy calculations of the DFT-optimized structure with the relevant dihedral angle varying in 15° steps, all at the B3LYP/def2-SVP level of theory[36–38] in the TURBOMOLE package.[39] Subsequently, CHARMM charges were optimized by comparison of dimer binding energies of fluorene rings and amino acid side chains at the MM level with those at the DFT level. Dimerization energies for both cases were calculated by equation 4.1:

$$\Delta E_{dimerisation} = E_{dimer} - 2E_{monomer} \qquad (4.1)$$

Where E_{dimer} is the energy-minimized dimer energy and $E_{monomer}$ the minimized monomer energy, both calculated in the gas phase. An overview of all atomic charges and new bonded interaction parameters used can be found in Appendix 4.

In a general atomistic simulation, peptides or peptide stacks, built in the Avogadro software,[40] were solvated in a TIP3P water box using the VMD scripting tools.[41] Periodic boundary conditions were used throughout. Initially, a 10,000 step energy minimization with fixed box volume was performed using the conjugate gradient method, after which the minimization was deemed to be converged (i.e. energy fluctuations are on the order of 10^{-4} kJ/mol). For systems starting with a box of randomly distributed molecules (section 4.4.2 and 4.4.3), a short 50 ps equilibration phase was performed using 1 fs time steps at a temperature of 300 K (damping coefficient 5 ps^{-1}) and a pressure of 1 atmosphere (piston period 100 fs, piston decay 50 fs) using the Langevin coupling scheme to allow for changes in the box volume. These are recommended values for CHARMM calculations and where also used for the production runs. The sensitivity to these parameters was not tested. For structures with a predefined starting structure (e.g. a peptide fibre, Ch. 4.4.1), the system was instead gradually heated from 0 to 300 K in 25 K steps (1 step per 5 ps) to prevent 'explosion' of the nanostructure due to the assignment of high random velocities upon the introduction of kinetic energy (temperature). The margin was increased to 2.0 to allow for larger atom displacements for the equilibration phase only. Afterwards, a production run in the constant number of molecules, pressure and temperature (NPT) ensemble was performed. Detailed parameters for simulations can be found in Appendix 1.

For Fmoc-X_1X_2-OMe nanostructures, proposed molecular structures were solvated by adding a water layer of 0.7 nm on all sides of the stack to avoid interactions across the periodic boundaries. For Fmoc-FF / Fmoc-S and Fmoc-FY / Fmoc-S simulations, cubic

boxes of 125 and 225 nm^3 were filled with 30 or 60 molecules, respectively (details can be found in section 4.4.2.3). For FF and IF dipeptide simulations, 40 dipeptides were solvated in a 150 nm^3 box.

Average stacking distances for comparison with X-ray data were measured between 1) C_α of the top residue to C_α of the bottom residue in the β-sheet stacking direction, divided by 7; 2) nearest carbon atoms of Fmoc-ring pairs in the stacking direction and 3) identical C-atoms on neighbouring Fmoc-rings along the length of the molecule.

Hydrogen bonds were determined by VMD's H-bond tool with a donor-acceptor cut-off distance of 0.34 nm and cut-off angle of ±25°.

4.3.4 High Pressure Liquid Chromatography (HPLC, Dr. M. Hughes)

A Dionex P680 system operating with a Macherey-Nagel 250 Å, 4.6 x 250 nm, C18 column was used for reversed phase HPLC. 10 μL of sample was injected. The mobile phase was comprised of water and acetonitrile (both with 1 wt% TFA) ramped from 20-80 % over 35 minutes at a flow rate of 1 mL min^{-1}. Detection of the peptide amphiphiles was carried out using a UVD170U UV-Vis detector at a 280 nm wavelength

4.3.5 Wide-Angle X-ray Scattering (WAXS, Dr. F. Coomer)

WAXS analysis were performed using a Philips X'Pert diffractometer with a Cu K-alpha source at 1.5406 Å. Gels and buffer solutions were spread on two separate silica substrates as film and allowed to air dry prior to data collection.

4.3.6 Transmission Electron Microscopy (TEM, Dr. L. Tetley)

Carbon-coated copper grids (200 mesh) were glow discharged in air for 30 seconds. A 10 μL volume of gel was transferred onto the support film and blotted down using filter paper. 20 μl of the negative stain was applied and the mixture blotted again using filter paper to remove excess. The dried specimens were then imaged using a LEO 912 energy filtering transmission electron microscope operating at 120kV fitted with 14bit/2K Proscan CCD camera.

4.3.7 AFM (Dr. Dimitrios Lamprou)

To image the hydrogel structure, cell culture media were removed and gels were placed on a mica sheet for 1 min and rinsed twice with 200 μL of distilled H_2O. The images were obtained by scanning the mica surface in air under ambient conditions using a Veeco MultiMode with 4 NanoScope IIID Controller Scanning Probe Microscope (Digital Instruments, Santa Barbara, CA, USA; Veeco software Version 6.14r1) operated intapping mode. The AFM measurements were obtained using a sharp silicon probe (TESP; nominal length (l_{nom}) = 125 μm, width (w_{nom}) = 40 μm, tip radius (R_{nom}) = 8 nm, resonant frequency (υ_{nom}) = 320 kHz, spring constant (k_{nom}) = 42 N m^{-1}; Veeco Instruments SAS, Dourdan, France), and AFM scans were taken at 512 × 512 pixels resolution. Typical scanning parameters were as follows: tapping frequency 308 kHz, integral and proportional gains 0.3 and 0.5, respectively, set point 0.5–0.8 V and scanning speed 1.0 Hz. The images were analyzed using Veeco Image Analysis software Version 6.14r1.

4.4 Results and Discussion

4.4.1 Case study 1: Fmoc-X_1X_2-OMe

4.4.1.1 X-ray scattering and IR spectroscopy

Addition of thermolysin enzyme to the 20 : 80 mM Fmoc-X_1 : X_2-OMe precursor solutions gave relatively high conversion to the dipeptide product (yields after 24 hours: 98, 79, 94 and 75% for SF, SL, TF and TL, respectively), as confirmed by HPLC. This is in good agreement with previous studies that indicated that SF was the thermodynamically most stable product, followed by TF and TL ≈ SL.[7] In all cases, self-assembling hydrogels were formed within 24 hours, although for SF formation of spherulitic structures on longer time scales destabilized the initially formed hydrogel (see ref. 1). TEM revealed different nanostructure morphologies (Fig. 4.1), i.e. SF formed 2D sheets, SL straight ribbons, TF long twisted fibres and TL short, strongly twisted fibrils. More TEM images of the SF sheets can be found in Appendix 3.

Especially the formation of extended 2D sheets in the case of SF is unusual the inherent chirality of amino acid building blocks and the nucleation and growth mechanisms in

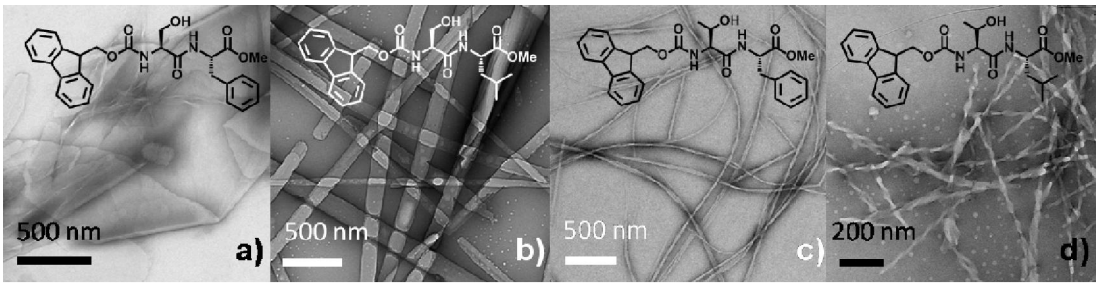


Fig 4.1. TEM images of biocatalytically self-assembled hydrogels (t = 24 h). Chemical structures of gelators are displayed in the top parts of the panels. a) Fmoc-SF-OMe, b) Fmoc-SL-OMe, c) Fmoc-TF-OMe, d) Fmoc-TL-OMe. Figure adapted from ref. 2.

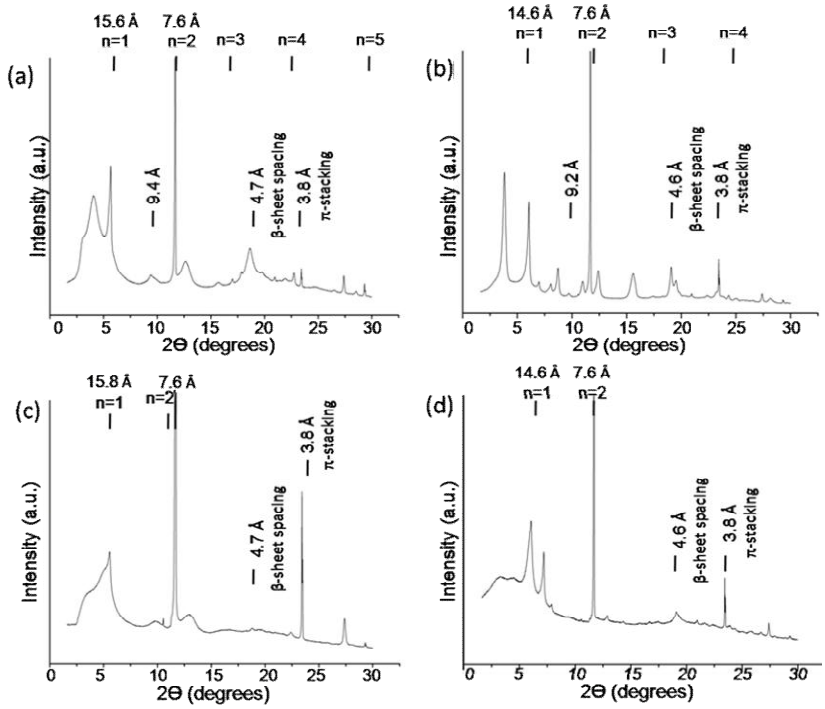


Fig. 4.2. (a) – (d): WAXS data for dried Fmoc-SF-OMe, Fmoc-SL-OMe, Fmoc-TF-OMe and Fmoc-TL-OMe, respectively. The main scattering angles are indicated with their d-spacings in Å. Higher order reflections are indicated by n=1, 2, ... Figure obtained from ref. 2.

nanostructures from low-molecular weight compounds generally favour chiral unidirectional structures.[42] A limited number of peptide-based nanomorphologies extending in two dimensions (i.e. not by lateral association of 1D fibres) have been reported. Cui *et al*. and Shao and Parquette both reported on 2D assembly of peptide amphiphile with alternating hydrophobic and hydrophilic residues,[43,44] which have often been shown to form β-sheet structures thanks to association of their hydrophobic and hydrophilic faces (see e.g. Brack and Orgel, ref. 45). Nam *et al*. observed flat structures by association of achiral peptoid polymers, although this could also be described uni-directional assembly of dimerised long chains.[46]

In order to understand the molecular packing that leads to the observed differences in morphology with the subtle changes in amino acid side chains, wide angle X-ray scattering (WAXS) analysis and IR spectroscopy were undertaken (Figs. 4.2 and 4.3). WAXS can reveal a periodic spacing between planes of atoms (called d-spacing) for (semi-)crystalline samples and the data presented in Fig. 4.2 shows recurring reflections at 3.8 Å, 4.7–4.8 Å and 14.6–15.8 Å with higher order reflections (up to n=5) for the latter. These d-spacings are typically assigned to the distances between π-stacked Fmoc-groups,[10,44,47,48] stacked β-strands[8,48–50] and lateral repeating units (along the length of molecule), respectively. However, the WAXS data were insufficient to determine a (crystalline) packing structure. Moreover, they failed to explain why Fmoc-SF-OMe forms extended two-dimensional sheets.

FTIR data (Fig. 4.3) showed various bands in the 1600–1700 cm^{-1} amide I region. Two bands around 1640 cm^{-1} and 1680 cm^{-1} are typically associated with Fmoc-peptide β-sheet structure (see ref. 51 and Chapter 3). Moreover, SF and SL show a prominent absorption around 1650 cm^{-1}. Generally, a band at this frequency would be assigned to a random coil in proteins (see, e.g., ref. 52), but the fairly narrow width and intensity of the peak, and the absence in the threonine systems, may indicate this is related to a more ordered, but frustrated arrangement of the amide groups, related to the planarity of the SF and SL nanostructures. Note that no bands for the fluorene moiety are expected in this region.[53]

For all dipeptides, a band around 1620 cm^{-1} grows slowly over time (a stable intensity was reached after approximately 24 hours), which is therefore attributed to quaternary structure formation.[1] For this case specifically, this occurs on a similar time scale to the observation of the macroscopic spherulites. This IR mode could therefore arise from amide I groups participating in inter-sheet interactions, or in other words, laterally associating β-sheets. The peak around 1740–1745 cm^{-1} is assigned to absorption of the terminal methyl ester group. These spectra are discussed further in the light of the MD results on this system.

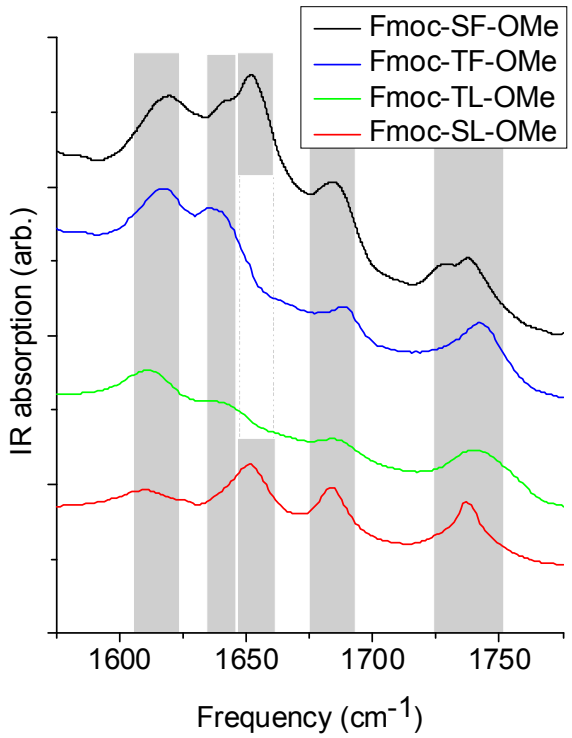


Fig. 4.3 Gel FTIR spectra in the amide I region of SF, TF, TL and SL hydrogels produced by thermolysin-catalysed reverse hydrolysis (t=24h). Shaded regions indicate bands discussed in the main text. Final dipeptide concentration ~ 20 mM, path length 10 µm. Figure adapted from ref. 2.

Together, WAXS and FTIR results suggested the formation of a type of π-β assembly (see chapter 2.2) for all studied dipeptides, with Fmoc-groups stacked as indicated by WAXS and the dipeptide in a β-sheet type assembly as shown in the FTIR spectra. However, more molecular insight is needed to draw a more conclusive picture and explain the differences in morphology,

4.4.1.2 Molecular modelling

In order to explain the unusual two-dimensional self-assembly of Fmoc-SF-OMe, the stability of a stack of gelator molecules was tested. This stack was built in the antiparallel π-β configuration (Fig. 4.4(a,b)), using a distance of ~4.8 Å between peptide

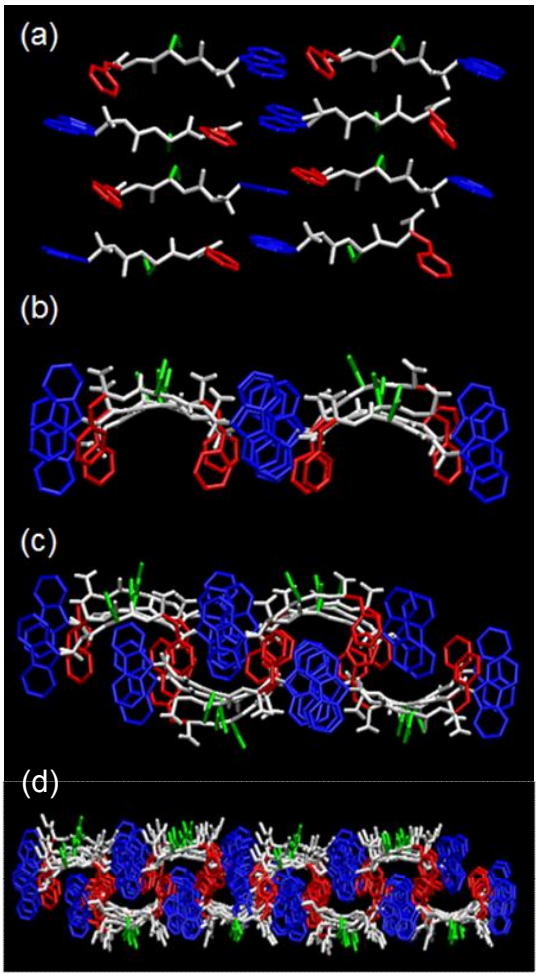


Figure 4.4 – Proposed molecular association models of Fmoc-SF-OMe. Blue: fluorenyl groups. White: peptide backbone. Red: phenylalanine side chains. Green: serine side chains. Hydrogens and water molecules not shown for clarity. (a) 8-mer anti parallel β-sheet arrangement, (b) 8-mer top view, (c) 16-mer top view of two separate 8-mers where their hydrophobic residues interact to from a bilayer, (d) extended 64-mer top view.

backbone C_α's as obtained from the WAXS results. It was noted that thanks to the antiparallel arrangement and typical peptide backbone dihedral angles for a β-sheet, the proposed stacks had a hydrophobic face with all the Phe aromatic side groups, and a

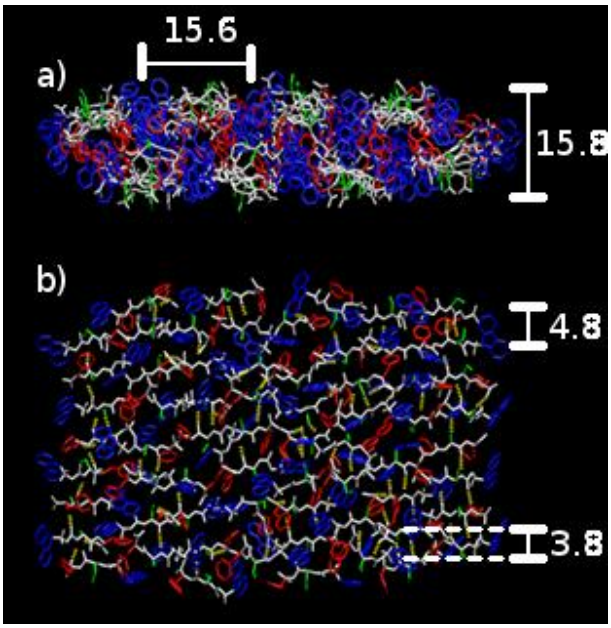


Fig. 4.5. Results from energy minimisations in the CHARMM force field of 64 Fmoc-SF-OMe monomers in an anti-parallel π-β configuration. Legend as in Fig. 4.4. a) Top view, b) side view. Average distances in Å for molecule length (15.6), bilayer width (15.8), C_α-C_α (4.8) and fluorene-fluorene rings (3.8) are indicated. See Table 4.1 and Fig. A4.1 for SL, TF and TL results.

more hydrophilic face with the Ser hydroxyl-groups pointing outwards. The unique size and amphiphilicity of these structures allowed the formation of 'bilayers' as displayed in Fig. 4.4(c), which can extend in two dimensions: both left to right in (c) through π-π stacking and in/out of the paper through hydrogen bonding. This conformation resembles the 2D structure proposed by Nam *et al*., where alternating hydrophobic and charged polymeric peptoids formed similar bilayer structures.[46] From this structure, a 64-mer was built of 8 x 4 x 2 molecules (see Fig. 4.4(d)). This structure was then solvated and energy-minimized for 10,000 steps using the steepest descent algorithm, when energy fluctuations were smaller then 10^{-4} kcal/mol and the structure was considered to be converged. The result of the energy minimization is displayed in Fig. 4.5. The same procedure was carried out for SL, TF and TL, for which results can be found in Appendix 3.

Table 4.1. Comparison of WAXS data (displayed in Fig. 4.2) with molecular mechanics results. All distances are in Å as indicated for the case of Fmoc-SF-OMe in Fig. 4.5.

	π-π		**C_α-C_α**		**Fmoc-Fmoc column**	
	WAXS	model	WAXS	model	WAXS	model
SF	3.80	3.78 ± 0.34	4.77	5.06 ± 0.29	15.56	16.44 ± 0.49
TF	3.80	3.84 ± 0.47	4.72	5.03 ± 0.14	15.83	16.69 ± 0.55
SL	3.80	disordered	4.65	4.85 ± 0.18	14.56	disordered
TL	3.80	3.97 ± 0.42	4.65	4.87 ± 0.19	14.6	15.67 ± 0.54

The proposed structures for SF and TF proved to be stable under minimization, while the SL stack showed more disruption of the network in the minimized structure. The TL stack was mostly stable, although several molecules at the edge of the simulation box were disordered. However, SF, TF and TL clearly exhibited repeating units with average distances of ~16 Å (Fmoc-column to Fmoc-column), and ~4.9 Å (peptide backbone to backbone), which are in excellent agreement with the reflections observed in WAXS (see Table 4.1). The proposed model is also consistent with the spectroscopic results such as the formation of stable β-sheets observed from FTIR as significant hydrogen bonding was observed between the amide and carbamate groups of the peptide backbone. For the hydrophobic residues to be efficiently buried into the bilayer core, the β-sheet configuration has to be twisted, which would explain the magnitude of the 1650 cm^{-1} amide component in the FTIR spectra (see Figure 4.3): instead of the random coil orientation of the peptide backbone as an origin of the peak, it is hypothesized that this absorption originates from a relatively large component of structured, but frustrated amide groups. As discussed in section 4.2.1, self-assembled peptide nanostructures usually adopt a chiral nanostructure, as a direct amplification of their molecular chirality. As both SF and SL form planar structures, each of these systems contain some degree of molecular frustration in the self-assembled state, leading to a blueshift in the amide I band with respect to the β-sheet absorption generally observed in Fmoc-dipeptide nanostructures. This would be consistent with the intensity and fairly narrow width of

this absorption, as a random coil configuration would give broader and less intense peaks, due to the lack of the cooperative dipole coupling effect, as discussed in Ch. 3. However, as TF and TL adopt the natural twist, the continuous H-bonding perpendicular to the peptide backbone is relatively uninterrupted along its axis and therefore the spectra of these systems not exhibit this disordered component. For the nanostructures that show flat structures (SF and, to a certain extent, SL), the minimized model also shows that the methyl ester groups may intra-molecularly hydrogen bond with the serine hydroxyl groups, which could explain the observed shift to 1730 cm^{-1} in the methyl ester peak of SF.

It should be noted that the proposed assignment of the distances in the WAXS data in Table 4.1 is not conclusive. Most notably, the width of the bilayer was found also to be 15–16 Å on average (indicated in Fig. 4.5), which offers an alternative explanation to the reflections of 14.6–15.8 Å observed in Fig. 4.2, especially considering the less bulky Leu side chains could afford denser packing, which is observed in the WAXS data. In general, there may be other arrangements that give the same distances. Moreover, molecular mechanics (MM) minimization is prone to getting stuck in local energy minima, as effects from thermal energy and entropy are excluded. The final structure of a MM calculation is therefore strongly dependent on the starting conformation and, as a consequence, the sheet structure was also stable for Fmoc-TF-OMe and Fmoc-TL-OMe, which don't exhibit planar morphologies.

Initial studies in NPT ensemble were carried out where molecular dynamics simulations were performed of the stack of 64 peptides in a water box for 100 ns. The final snapshots of the simulations are displayed in Fig. A4.2 and show that the stacks do deform on this timescale: SF and TF stay mostly in bilayered sheet conformation, although a twist was noticed. TL quickly collapses to a more cylindrical fibril and SL is most loosely bound, but still shows some of the initial bilayer structure. The limited size of the system may also play an important role in this, as in fact 40 out of 60 monomers are technically 'on the edge' where they are more exposed to water on at least one side than the molecules in

the middle of the stack. Note that these simulations took approximately 44 days on a 24-core computing node, so larger systems were not feasible due to computational costs. Although no quantitative analysis has been performed on these simulation results, they may point towards a different molecular arrangement.

Possible alternatives to the antiparallel π-β structure in terms of packing are offered by the columnar stacked cross-β pattern suggested by Adams *et al.* and Braun and Cardoso.[54,55] They propose a type of parallel π-β structure for naphthalene dipeptides and Fmoc-FF, respectively, where aromatic columns are stabilised by face-to-edge π-π interactions, based on crystallographic data. However, this model fails to explain the supramolecular chirality (for a review of nanostructure chirality, see ref. 56) of the normally achiral Fmoc-group shown for many of this type of systems, as discussed for Fmoc-GL and Fmoc-LL by Smith *et al*.[57] and shown for related Fmoc-dipeptides by Hughes *et al*.[58] Adams also noted that their model is derived from X-ray Diffraction results on crystals, and diffraction patterns of crystals grown from the gel-phase are significantly different from crystals of the same molecules grown from solution and care has to be taken drawing parallels between the two methods. In summary, it is still unclear from experimental work what the most common stacking conformation for aromatic peptide amphiphiles is. With the advance of computational infrastructure and force fields, Molecular Dynamics programs will be able to mimic experimental conditions in terms of timescale, concentration and molecular interactions more and more accurately. Therefore, it is expected that sufficiently large simulations for solving the conformational question will emerge within this decade, independent of experimental techniques.

4.4.2 Case study 2: Core/shell nanostructures

4.4.2.1 Fmoc-FF and Fmoc-S

Fmoc-FF formed strong self-assembling hydrogels, consistent with previous reports.[10,25] AFM results confirm the fibrous morphology for the pure Fmoc-FF hydrogel (see Fig. 4.6(b)). Surfactant Fmoc-S, on the other hand, formed spherical aggregates at this pH. However, when these two are mixed, evidence of micelles disappeared (as demonstrated in Fig. 4.5(b).

FTIR spectroscopy was used to study the supramolecular arrangement of the structures observed using AFM. To obtain more information on the mode of coassembly, the IR spectra for samples with various concentrations of surfactant and gelator molecules were recorded (see Fig. 4.5(c)). In the pure Fmoc-FF spectrum, bands at 1625–1640 and 1687 cm^{-1} are assigned to Fmoc-FF β-sheets in agreement with previous work.[10,25] In the 100% Fmoc-S sample on the other hand, no such structures were observed. Instead, broad bands around 1595 and 1675 cm^{-1} are assigned to free carboxylate and unstructured carbamate absorptions, respectively. Interestingly, no significant changes in the peaks at 1625 and 1687 cm^{-1} were observed upon inclusion of up to 50% of surfactant into the Fmoc-FF hydrogel (samples with more than 50% Fmoc-S did not form self-supporting hydrogels). This indicates Fmoc-S does not significantly intercalate the stack of Fmoc-FF molecules as this would reduce the peak heights by decreasing the cooperative excitonic coupling effect as discussed in Chapter 4. The disappearance of micellar structures can therefore be explained by an Fmoc-S coating of the Fmoc-FF fibres, as proposed in the cartoon in Fig. 4.5(a). This conclusion is further confirmed by evidence from fluorescence spectroscopy and Static Light Scattering (SLS), of which details can be found in ref. 3. However, more detailed molecular insight is needed to be able to conclusively determine the presence of a core/shell type structure rather than a randomly mixed fibre. This gap can be filled by molecular dynamics (see section 4.4.2.2).

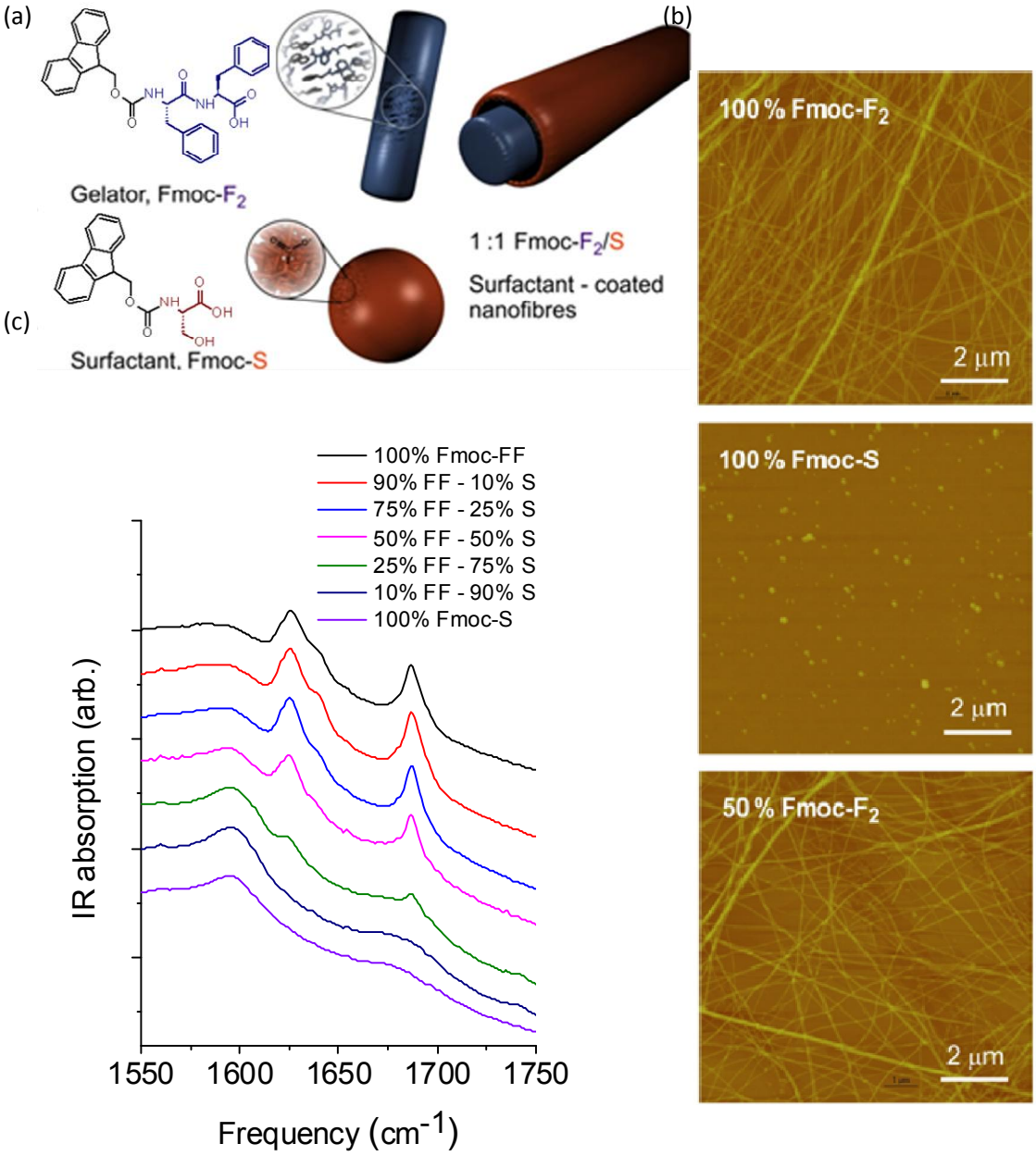


Fig. 4.6. (a) Chemical structures of Fmoc-FF and Fmoc-S and cartoon representations of their nanostructures, courtesy of Dr. M. Zelzer. (b) AFM images of Fmoc-FF, Fmoc-S and Fmoc-FF / Fmoc-S mixed nanostructures, courtesy of Dr. D. A. Lamprou and Dr. V. Jayawarna. (c) Amide I FTIR spectra of samples with different ratios of Fmoc-FF and Fmoc-S at a total peptide concentration of 20 mM. The path length is 25 µm.

4.4.2.2 Fmoc-FY and Fmoc-S, -T or -RGD

FTIR spectroscopy was employed to study the supramolecular structure of mixtures of phosphorylated Fmoc-FY^P with Fmoc-X (X = -S, -T or -RGD) in D_2O in the presence of the enzyme phosphatase. Spectra from before and 24 hours after the addition of the enzyme are shown in Fig. 4.7. No extended amide I or carbamate stacking interactions were present in the pre-gelation components in solution (Fmoc-FY^P, Fmoc-X), implying a lack of β-sheet structure for these pre-gelators (see Fig. 4.6(a)). Mixtures of Fmoc-FY^P and the surfactants in solution gave linear combinations of the spectra. Note that no absorptions from the fluorene rings or phosphate group are present in this region. The sloped baseline in this region arises from water (D_2O and trace HOD) absorptions.[59]

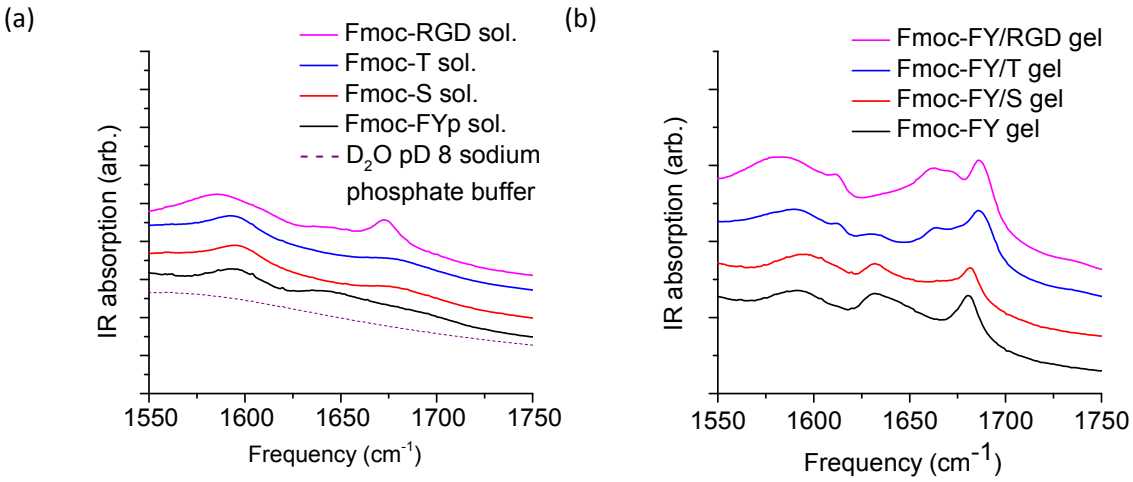


Fig. 4.7. Amide I FTIR spectra of Fmoc-amino acid and Fmoc-peptide solutions and gels. Spectra are vertically offset for clarity. (a) Solutions of surfactants and gelator precursor Fmoc-FY^P at a 10 mM concentration. A spectrum for a 0.1 M sodium phosphate buffer at pD = 8 without peptides is added to compare the background absorption. (b) Mixed Fmoc-FY / Fmoc-X hydrogels at a total monomer concentration of 20 mM. The path length is 25 μm for all samples.

For all samples, the detection of a broad vibrational absorption around 1595 cm^{-1} resulting from the asymmetric stretching of the C-terminal COO^- group, indicates that a relatively large proportion of C-terminal groups are deprotonated. In RGD-containing samples, there is a contribution to this peak around 1585 cm^{-1} from the aspartic acid side chain and a contribution at 1673 cm^{-1}, which can be attributed to the asymmetric vibration of the $CN_3H_5^+$ in arginine side chains,[52] although the latter is assigned by Cheng *et al*. to the presence of trifluoroacetic acid (TFA) in a pure Fmoc-RGD hydrogel. [21] The fact that no TFA is used and it is only present in the RGD-containing samples, however, precludes this assignment. Fmoc-S, -T and -RGD solutions show faint, broad absorptions around 1678 cm^{-1}, assigned to the absorption of the carbamate group. Finally, a weak, broad absorption centred around 1648 cm^{-1} can be assigned to unstructured, random coil-like amide bonds in the Fmoc-FY^P and -RGD peptides.

Upon addition of alkaline phosphatase, the charged phosphate group was removed from the tyrosine side chain. This led to the formation of hydrogels for all samples containing the Fmoc-FY component. Figure 4.7(b) shows the FTIR spectra of gels of Fmoc-FY by itself and mixed systems Fmoc-FY/X, (X=S,T,RGD)). Note that the amide I absorption has increased for all samples, caused by the ordered association of molecules (see Ch. 3). Extended amide I stacking interactions were observed in the gel state for all systems, but significant changes were observed upon the inclusion of different surfactants as discussed below.

The FTIR spectrum of the Fmoc-FY gel shows typical amide I peaks at 1630 cm^{-1} (with a tail towards 1650 cm^{-1}) and 1681 cm^{-1} which are associated with the presence of Fmoc-peptide β-sheet structures (see Chapter 3 for a discussion of IR spectral assignments of these systems). The fact that the gel amide I peaks remain largely unaffected upon the inclusion of Fmoc-S shows that the presence of Fmoc-S does not significantly disrupt the β-sheet structure, similar to the Fmoc-FF / Fmoc-S case discussed in the previous section.

Two new peaks are observed in the spectrum upon inclusion of the surfactants: the broad absorptions in the surfactant solutions shift to 1666 cm^{-1} and gain intensity upon gel formation, with RGD > T > S. Moreover, there is a peak present at 1611 cm^{-1} in the Fmoc-FY/Fmoc-X gels with similar behaviour (i.e. more intense from RGD > T > S). Based on their red-shifted frequencies compared to the solution phase and their fairly narrow width, these two peaks are tentatively assigned to modes arising from the inclusion of Fmoc-X in the fibrous core or on the surface of the nanostructure. No side chain absorptions are expected in this region and it should be noted that the IR peak pattern is significantly different from the results reported by Cheng *et al*. for a pure Fmoc-RGD hydrogel,[21] which indicates that the presence of Fmoc-FY significantly influences the self-assembly of the Fmoc-RGD peptide. The fact that the 1681 cm^{-1} peak, assigned in Chapter 3 to carbamate containing β-sheets, gains intensity relative to the 1630 peak suggests that a relatively higher amount of carbamate groups are

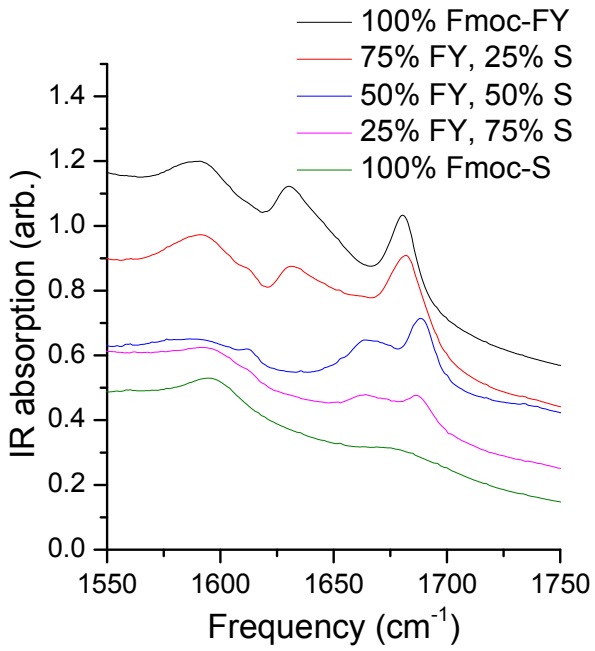


Fig. 4.8 Amide I FTIR spectra for Fmoc-FY/Fmoc-S samples at different ratios of monomers. Total monomer concentration 20 mM, path length 25 μm. Spectra are vertically offset for clarity.

incorporated in the hydrogen-bonding pattern. This could indicate that either the amide groups are losing their β-sheet-like arrangement, while the carbamate group remains H-bonded (possibly due to its proximity to the π-stacked fluorene rings), or Fmoc-T and Fmoc-RGD are intercalating the Fmoc-FY stack.

To study the effect of the enzymatic reaction and further test the core/shell hypothesis, peptide nanostructures were prepared by dissolving the Fmoc-FY gelator and surfactant directly in phosphate buffer at pH 7.3. FTIR spectra were taken for samples with 100:0, 75:25, 50:50, 25:75 and 0:100 ratios of Fmoc-FY to Fmoc-S (Fig. 4.8, total monomer concentration 20 mM). No significant differences were observed between the enzymatically or 'chemically' prepared Fmoc-FY gels. When the relative concentration of Fmoc-S was increased, peaks around 1611 and 1666 cm^{-1} became apparent, while the 1638 cm^{-1} peak loses intensity. For this non-enzymatically prepared hydrogel, this is consistent with a more disruptive type of packing (c.f. Fmoc-FY/T and Fmoc-FY/RGD gels),[60] where Fmoc-S intercalates the Fmoc-FY β-sheet structure, in contrast with the orthogonal (non-intrusive) packing observed in the phosphatase-prepared FY/S gels. When the concentration of surfactant was further increased to 75%, a viscous solution was obtained and the cooperative transition dipole coupling effect present in β-sheets starts to diminish as there are less extended structures, resulting in a loss of intensity for all peaks in the amide I region. Finally, the Fmoc-S solution again showed only the broad 1678 and 1595 cm^{-1} carbamate and COO^- peaks, as discussed above.

These results show that depending on the surfactant and the preparation method, different supramolecular structures can be accessed. The more orthogonal self-assembly observed for the alkaline phosphatase catalysed Fmoc-FY / Fmoc-S hydrogels is expected to be most suitable for cell culture, thanks to its hydrophilic surface. The results of Molecular Dynamics simulations of these systems are described in section 4.4.2.3, where further evidence for an orthogonal or disruptive mode of co-assembly is given.

4.4.2.3 MD simulations

MD simulations were carried out for the Fmoc-FF/Fmoc-S co-assembling discussed in the previous section. For every simulation, Fmoc-peptides were randomly placed in a cubic water box of 125 nm^3 (30 molecules) or 225 nm^3 for the mixed systems (60 molecules). Pure Fmoc-FF (Sim. **1**) and Fmoc-S (**2**) solutions were performed as a comparison with the mixed system of 30 Fmoc-FF and 30 Fmoc-S molecules (**3**). As discussed in section 4.2.2, gelation often takes place upon a change of the protonation state of the gelator molecules.[25,26] To test whether MD simulations would reproduce this behaviour, simulation **4** contained both protonated and deprotonated Fmoc-FF and Fmoc-S. Table 4.2 displays the contents of the 4 simulations carried out:

In all simulations containing Fmoc-FF, one-dimensional structures were observed, which formed across the periodic boundaries a few tens of nanoseconds into the simulation. For the pure Fmoc-S simulation, no significant aggregation was observed. Fig. 4.9. shows the final snapshots of all the simulations. Links to movies of the full simulations **3** and **4** can be found in Appendix 7.

Inspection of the observed fibrous nanostructure observed in simulations **1** and **3** revealed that the structure is mainly formed by aggregation of the phenylalanine side chains, rather than stacking of the fluorene rings, which were pointing into the solvent. Although for small Fmoc-peptides such as Fmoc-AA others have reported that not all Fmoc-groups may be buried in the hydrophobic fibre core,[9] aromatic stacking should

Table 4.2. Number of peptides and simulation length of Fmoc-FF / Fmoc-S simulations

#	**Fmoc-FF**	**Fmoc-FF-COO^-**	**Fmoc-S**	**Fmoc-S-COO^-**	**Sim. time (ns)**
1	30	-	-	-	60
2	-	-	30	-	60
3	30	-	30	-	100
4	15	15	15	15	200

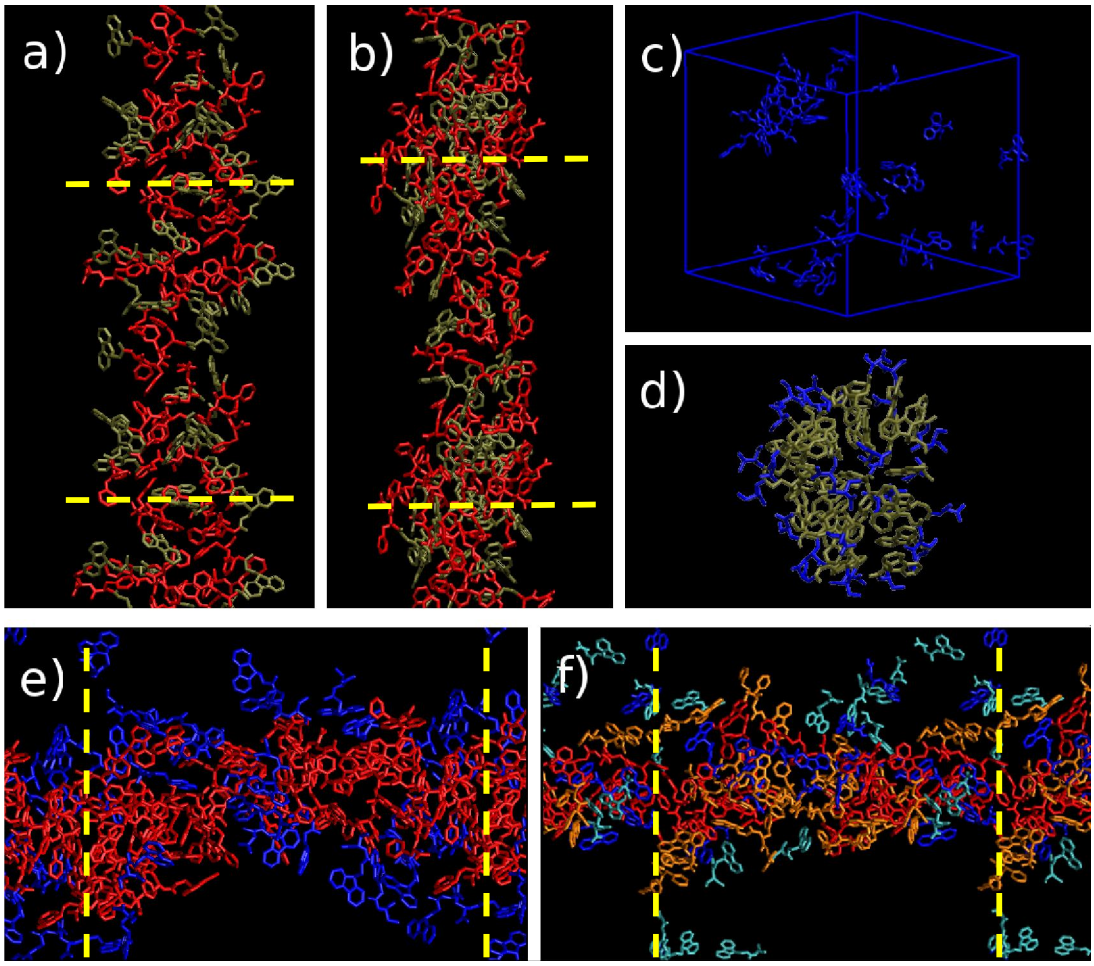


*Fig. 4.9. Final snapshots of simulations **1-4**. Hydrogen atoms and water molecules have been omitted for clarity. Red: Fmoc-FF, blue: Fmoc-S, orange: Fmoc-FF-COO$^-$, cyan: Fmoc-S-COO$^-$. Yellow dashed lines indicate one periodic unit cell. a,b) **1**, Fmoc-FF, with two different parameter sets (see text). Fmoc-groups are tan-coloured, c,d) **2**, Fmoc-S, with two different parameter sets, c) with cubic simulation box, e) **3**, Fmoc-FF/Fmoc-S, f) **4**, Fmoc-FF/Fmoc-FF-COO$^-$/Fmoc-S/Fmoc-S-COO$^-$.*

still be favourable compared to solvation of the Fmoc-group. This was also demonstrated using coarse-grain level simulations of Fmoc-*D*-Ala-*D*-Ala.[61] The pure Fmoc-S simulation showed little aggregation of monomers, in contrast with the experiments that indicate micellar or spherical aggregate formation. On the other hand, it was evident

from simulations **3** and **4** that the proposed core/shell structure is confirmed, as Fmoc-FF molecules form the core of the 1D structure, and Fmoc-S molecules were loosely bound to the outside of the core. Also, the self-assembly behaviour as a function of protonation state (sim. **4**) agreed with experimental evidence: in general, gelation was observed upon protonation as discussed above. During the simulation, the protonated Fmoc-FF molecules formed an elongated structure within the first 20-30 ns of the simulation, followed by binding of Fmoc-FF-COO^- to this structure. In other words, even though Fmoc-FF-COO^- was deprotonated, it still had a high tendency to aggregate and form nanostructures. This is confirmed by the work of Tang *et al*., showing that even at pH 9-9.5 self-supporting hydrogels could be formed, while still a majority (around 66%) of the Fmoc-FF molecules are in their deprotonated, charged state.[25] Again, Fmoc-S reversibly bound to the outside of this hydrophobic centre. Fmoc-S-COO^- did not bind and instead stayed in solution with occasional reversible binding events to the nanostructure, in line with our expectations.

It should be noted that the results from the MD simulations indicate that the parametrization of the Fmoc-group (see section 4.3.3) in the CHARMM force field may not accurately represent reality, as Fmoc-groups seem to be exposed to solution rather than interacting with each other, i.e. the group is too hydrophilic because of its relatively high point charges. A new set of parameters for the Fmoc-group was therefore developed in conjunction with Dr. Kenno van Ommeslaeghe[62] and the ParamChem automated parametrization program.[63–65] This set can be found in Appendix 5 and mainly involves lowered point charges on fluorene ring atoms to make them more comparable with simple aromatic systems such as phenylalanine in terms of polarity. Fig. 4.9(a,b) shows a comparison between 100 ns simulations of Fmoc-FF with both the sets of parameters. When the new parameter set was used (Fig 4.9(b,d)), it became clear that the Fmoc-groups were buried to a similar extent as the hydrophobic phenylalanine chains, which is more realistic considering the apolar nature of the fluorene ring. Moreover, the protonated Fmoc-S molecules now formed spherical aggregates, in agreement with the experimental work. This strongly indicates that the results in this section have to be

treated with caution, although some of the experimentally observed properties are accurately reproduced, and it is recommended to use the updated charges for future work. Note that this may also influence the results from section 4.4.1. Initial simulations indicate that the Fmoc-SF-OMe 64-mer is somewhat more stable when the new parameters are used. This is, however, not expected to influence the agreement with the WAXS data presented in section 4.4.1.

Finally, these new parameters were used to perform a 50 ns simulation of the mixed Fmoc-FY / Fmoc-S system discussed in section 4.4.2.2. We observed the spontaneous formation of a 1D nanostructure from an initial box with 30 well-solvated Fmoc-FY-OH and Fmoc-S-OH monomers each. Fig. 4.10 shows the simulation result and indicates that Fmoc-S can intercalate the hydrophobic Fmoc-FY stack, although the hydrophilic part of the surfactant (C-terminus and serine side chains) was pointing out into solution. Due to time constraints, no further simulations could be performed, but these results suggest a better agreement between experiment and simulation with the new Fmoc-group parameters.

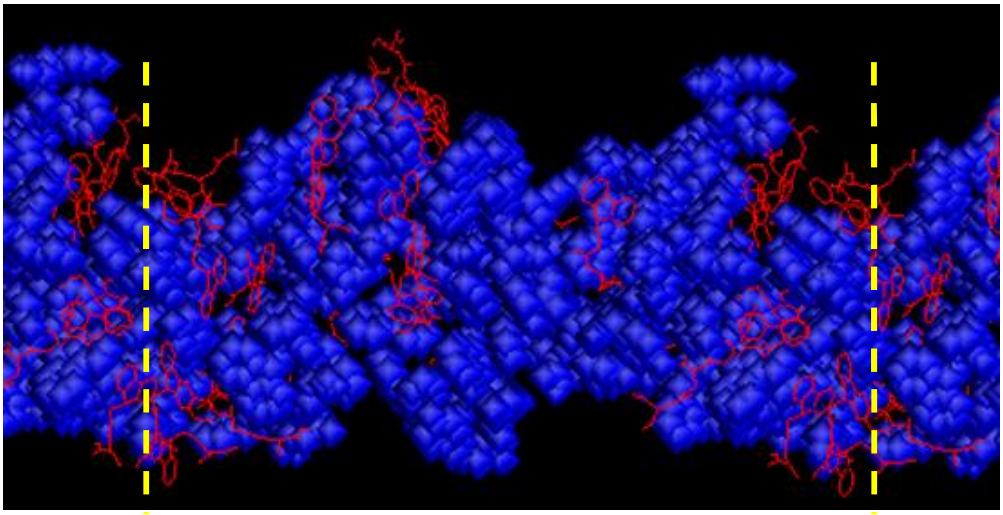


Fig. 4.10. MD simulation results from a 50 ns simulation of 30 Fmoc-FY and 30 Fmoc-S molecules. Blue, VdW representation: Fmoc-FY. Red, lines representation: Fmoc-S. Yellow lines indicate periodic boundaries.

4.4.3 Case study 3: Self-assembly of FF and IF

4.4.3.1 FTIR spectroscopy

The FTIR spectrum of dried FF dipeptide nanotubes was reported by Reches and Gazit and is displayed in Fig. 4.11(a). The spectrum shows a typical 1630 cm^{-1} absorption, assigned to an extend β-sheet conformation. No experimental work was performed on this peptide by our group.

The FTIR spectrum of IF in the gel state is displayed in Fig. 4.11(b) and shows strong peaks at 1654 and 1571 cm^{-1}, in agreement with the report from Sanchez-de Groot *et al*.[28] This spectrum indicates a totally different conformation of the peptides in the IF hydrogel compared to the FF nanotubes: the peak around 1571 cm^{-1} is assigned to carboxylate groups participating in strong salt bridges with the positive N-termini.[28] Moreover, the 1654 cm^{-1} band is too high in frequency for a typical β-sheet[52] and was reported to be independent of the presence of gel fibres by Sanchez-de Groot: in solutions at low peptide concentrations or high temperature, this peak showed only a shift of 1 cm^{-1} with respect to the gel phase. These observations indicate the environment of the amide group does not change upon gelation and the amide group of IF instead H-bonds with water molecules. This indicates head-to-tail (H-T) interactions are the dominant driving force in IF peptides in water. Interestingly, when the IF gels were dried, analogously to the case of the FF dipeptides, they showed 1618 cm^{-1} and 1678 cm^{-1} peaks,[28] which suggests the peptides undergo a morphological change going from the aqueous to the solid phase. This indicates the discrepancy between hydrogel and crystal phases, as is discussed further in the next section.

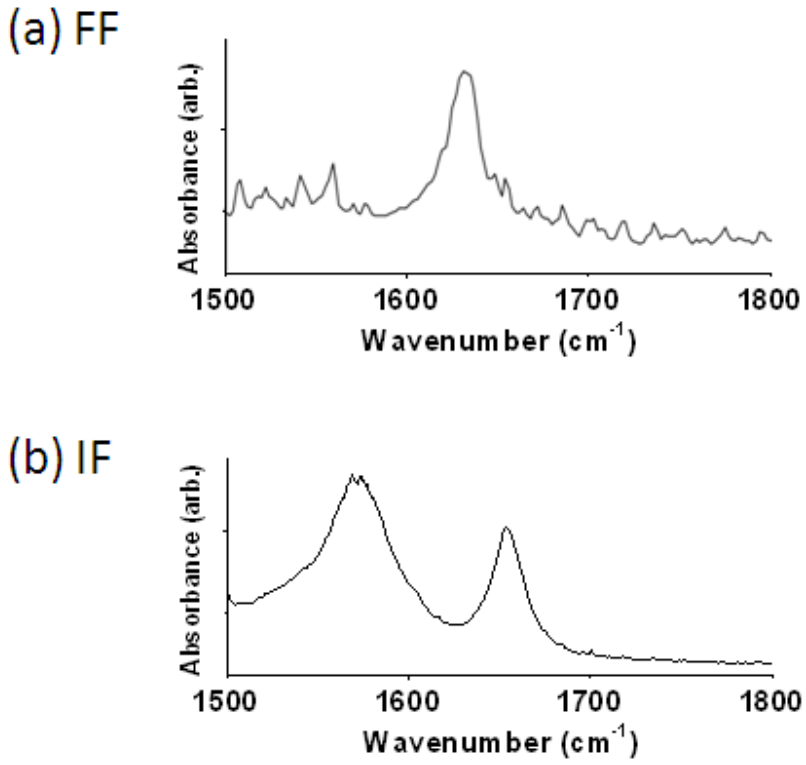


Fig. 4.11 Amide I FTIR spectra of dipeptides nanostructures formed by aqueous dilution of a dipeptide stock solution in hexafluoroisopropanol. (a) Spectrum of dried FF nanotubes, adapted from ref. 27. (b) IF hydrogel spectrum at a 1.5% w/v peptide concentration. Path length 25 µm.

4.4.3.2 MD simulations

MD simulations were performed for both dipeptides in an effort to access the 1D nanostructures described above. The results of 100 ns simulations of 40 randomly located dipeptides in a cubic water box in the NPT ensemble are given in Fig. 4.12. The link to a full simulation for FF is supplied in Appendix 7.

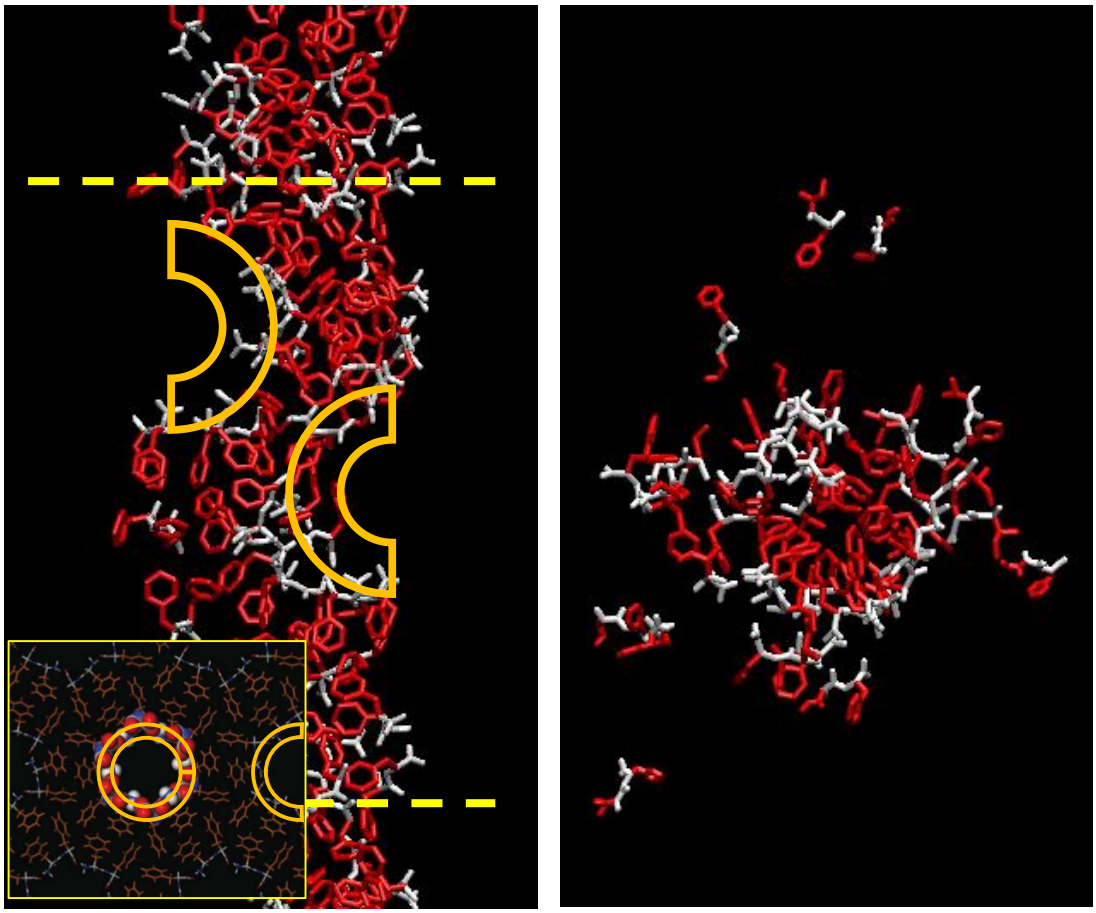


Fig. 4.12. Results from 100 ns atomistic MD simulations of 40 dipeptides. a) Diphenylalanine (FF) and b) isoleucinyl-phenylalanine (IF). Side chains are coloured in red, backbone in white. Yellow lines indicate periodic boundary conditions. Semicircles in a) show the possible indications for a microporous macrostructure like in the crystalline phase, which is displayed in the inset (obtained from ref. 29), as discussed in the text.

Analysis of the FF simulation revealed the dipeptides form a 1D nanostructure (Fig. 4.12(a)) with significant aromatic stacking of the phenylalanine side chains, although the charged termini of the peptides also interact (white in Fig. 4.12). The macroporous structure discovered by Reches and Gazit[27] and later clarified by X-ray Diffraction (XRD) on Chemical Vapour Deposition (CVD) samples by Görbitz[29] was not observed. This was not expected, as there are not enough molecules in the simulation box for this

structure to be feasible. However, Fig. 4.12(a) shows that our results agree with the conclusion of Tamamis *et al*. and Guo *et al*. that aromatic interactions are a main driving force in self-assembly of FF in solution.

The simulation on isoleucinyl-phenylalanine (IF) shows a less ordered result after 100 ns (Fig. 4.12(b)), with only 34 out of 40 molecules in the major cluster (monomers are defined to be in the same cluster if their nearest-neighbour distance is smaller than 3.5 Å). However, it is apparent that the simulation shows aggregation of the dipeptide monomers, in agreement with the experimental[28] report: a hydrophobic core is formed with both Ile and Phe side chains while most peptide backbones are interacting with each other on the interface with water. This packing is consistent with the reported IF crystal structure that shows a layered structure with hydrophobic/hydrophilic faces.[30] On the other hand, during the simulation it became apparent that lateral interactions between layers are also favourable. The attractive forces are mainly caused by salt bridges between C- and N-termini and disordered hydrophobic groups. This contributes to the fact that the simulation does not show an elongated structure across the periodic boundaries like for FF in Fig. 4.12(a). Moreover, the fact that the dipeptide is not symmetrical leads to less efficient packing in the size-limited solution phase simulation. It could not be conclusively determined from this simulation if the structure found is more in agreement with the hydrogel or the 'dried fibre' results of Sanchez-de Groot *et al*., but the prominent head-to-tail (H-T) interactions observed in the simulation point towards the hydrogel structure (see below).

For both simulations, the average number of hydrogen bonds was measured for the last 2.5 ns of the simulations (cutoff distance 3.4 Å, cutoff N-H---O angle ±25°). Coincidentally, for both IF and FF simulations an average of 32 H-bonds was observed (see Table 4.3). For FF, most of these were between C- and N-termini (25 H-bonds),

Table 4.3. Average number of hydrogen bonds between possible donors and acceptors in the last 2.5 ns of FF and IF simulations. Hydrogen bonds were counted with a donor-acceptor distance of 3.4 Å and a cutoff N-H---O angle ±25°).

H-bond donor / acceptor	**FF**	**IF**
N-terminus / C-terminus	25	20
Amide NH / C-terminus	7	10
Amide NH / Amide C=O	0	1
N-terminus / Amide C=O	0	1
TOTAL	**32**	**32**

while on average 7 H-bonds were present between C-terminus and amide NH. Interestingly, no amide-amide H-bonds were observed. Similar observations were made for the IF simulation: an average of 32 H-bonds were present per frame, mainly between C- and N-termini (20) and C-termini and the amide NH (10). Only very few H-bonds were measured between amide C=O and N-terminus (1), or amide C=O and amide NH of other monomers (1). This result is in agreement with the data from Rissanou *et al*. on FF who show H-T interactions are important,[33] but we propose they can also arise as a *consequence* of molecules being in close proximity due to aromatic interactions. Note that there is a strong variation around the average value (e.g. a standard deviation of 6 H-bonds was found for the last 2.5 ns in the simulations in the case of N-to-C-terminus interactions for FF, using a frame rate of 20 frames/ns), which shows the reversible nature of non-covalent bonds between the dipeptides.

Rissanou *et al*. observed similar hydrogen bonding between the termini and concluded that this is the main driving force in FF assembly from very similar atomistic simulations, as aggregation is less prominent in methanol solution.[33] However, they also reported that no self-assembly was observed in experiments containing the dialanine peptide in water, which has the same hydrogen bonding capabilities, but not the hydrophobic phenyalanine side chains. We therefore conclude from our results and those of Tamamis *et al*.[31] and Guo *et al*.[34] that aggregation is likely to be caused by a combination of hydrogen bonding and hydrophobic interactions, but the hydrophobic

interactions dominate the assembly of FF in aqueous solution. However, the crystalline phase results from Görbitz do also give important indications to explain various other observations: no straight, zipper-like packed hydrophobic core was observed (see Fig. 4.12), but rather smaller branched regions that seem to resemble the microporous crystal structure of FF nanotubes (displayed in the inset in Fig. 4.12(a)). This is attributed to the coulomb interactions between the head and tail groups, as found from the H-bond analysis. Regardless, it should be noted that it is dangerous to directly compare aqueous nanostructures with the crystal phase, as the absence of solvent has crucial influences on the molecular packing, especially for coulombic interaction such as head-to-tail (H-T) interactions between the C- and N-termini. As an example, Houton *et al*. used XRD to show a totally different arrangement between 2-naphtol dipeptides that formed gels and those that formed crystals.[66] In one example, they managed to grow crystals from the Nap-AA gel phase and observed that the packing is principally different between the two states of matter. This could be comparable to the observation of Sanchez-de Groot for the IF peptide, discussed above.

Although the size and timescale of the simulations reported here exceeds the atomistic simulations reported previously,[31,33] there is still a large discrepancy with the lab scale. The experimental fibre thickness is much larger than the computational thickness in this case, which can be explained by the small system size. Additionally, gelation generally takes place on a seconds' to weeks' timescale, while the simulation extends over only 100 ns. The use of periodic boundary conditions and the fact that the peptide concentration is approximately 10 times higher in the simulation compared to the experiment (e.g. 0.44 M vs. 0.048 M, respectively, for IF) speeds up the assembly significantly. Moreover, Ash and co-workers argued that the assembly of individual nanostructures can take place on a microsecond time-scale,[67] although macroscopic gelation requires entanglement and association of short fibres which are kinetically much more limited.

Finally, it should be noted that there may be an extra bias towards 1D structure formation in our simulation due to the usage of periodic boundary conditions. Therefore, interactions across the box boundary provide an extra bonus in free energy that would not be there in an infinitely large water box. However, this is considered a relatively small contribution to the overall morphology and the obvious advantage is that there are less 'edge effects' like in the Fmoc-X_1X_2-OMe simulations (section 4.4.1). For example, Mu *et al*. still observed 1D fibres in their MD simulations on Fmoc-AA, while the size of their water box prevents interactions between the periodic images.[9]

In conclusion, it can be said that atomistic MD simulations of dipeptides and derivatives are a valuable tool in predicting the propensity and structure of potential self-assembling building blocks. Both aromatic stacking and hydrogen bonding contributed to the formation of nanostructures for both FF and IF dipeptides to varying degrees. In any case, the computational cost of these simulations (100 ns ≈ 18,000 CPU hours for these relatively small systems) currently prevents large amounts of peptides to be studied this way. Therefore, different methods such as coarse-grain MD need to be employed for screening purposes. Chapter 6 extensively discusses these methods, and dipeptide self-assembly.

4.5 Discussion and future directions

In this chapter, three principally different systems where self-assembly plays a role are discussed. First of all, on the basis of the combination of computational geometry optimisations and IR results it was possible to obtain a comprehensive picture of the nanoscale arrangement of Fmoc-X_1-X_2-OMe structures that is consistent with all experimental data (section 4.4.1). However, it was also shown that when the system was put in a dynamic simulation, this structure proved less stable than expected on the basis of the minimization, which can be attributed to either methodological errors (e.g. the parameters of the force field, as discussed in 4.4.1 and 4.4.2.3) or the fact that other

structures may be more stable. Moreover, these systems exhibited a complex IR absorption pattern that can evidently not be related to traditional protein spectroscopy, as was also concluded from Chapter 3 and it is therefore hard to give conclusive band assignments.

In general, with the current simulation system sizes, it was evident that most of the molecules are still at the 'edge' of an extended nanostructure, which can introduce artefacts. It is expected that the growing availability of high performance computing will slowly diminish this problem. In any case, we are currently far from predicting nanoscale architectures on the basis of any single technique. However, the more nanostructures are unravelled by a combination of experimental and theoretical methods, the more informed our models become.

For the co-assembled systems of Fmoc-dipeptide gelators and Fmoc-amino acid surfactants fully atomistic simulations were performed without a biased starting structure, i.e. with randomly oriented well-solvated peptides as a starting point. Using this approach, compelling agreements between experimental observables and results from the simulations were observed on a 50–100 ns computational time scale: a core/shell nanostructure was found from the Fmoc-FF / Fmoc-S co-assembly, in agreement with experimental evidence from FTIR, AFM, fluorescence spectroscopy and SLS, and the assembly behaviour in different C-terminal protonation states predicted is consistent with the commonly accepted model of gelation upon protonation.[25,26] It was however noticed that the degree and mode of assembly are very sensitive to the parameterisation of the molecules in the system. A newly developed set of parameters of the Fmoc-moiety has recently been validated in our lab and it is expected that this will aid the quality of MD predictions of Fmoc-peptide nanostructures. It will be interesting to see if MD is sensitive to small changes in amino acid side chains by future work on the Fmoc-FY / Fmoc-X systems: IR spectra predict increasing deviations from the structure without surfactant when respectively S, T and RGD are introduced into the system.

For these mixed systems, it was very difficult to assign all the bands in the IR spectrum to specific vibrational modes in the nanostructure as many different interactions between amide groups are possible when more complexity is introduced. It is therefore important that spectral predictions by density functional theory are being developed, as described in Ch. 4; these give information on the particular local modes that are contributing to a particular absorption. The development of methods for calculating IR spectra from MD simulations (i.e. DFT-MD, see ref. 68) could accelerate the advancement of our understanding of these systems. For the future, perhaps larger atomistic simulations will be able to determine the H-bonding pattern with more certainty by looking at structural averaging.

Finally, simulations were performed for the simple dipeptide systems of FF and IF and analysed their IR spectra with respect to these simulations. Surprising effects were noticed when comparing the hydrogel phase to dried hydrogel samples for the IF dipeptide.[28] Moreover, a large difference was noted between IF and FF spectra: the former shows absorption frequencies indicative of head-to-tail bonding of the termini without secondary structure formation of the amide groups, while the latter shows a typical β-sheet peak without the spectral signature for these H-T interactions.[27] The level of hydrophobicity of the side chains is expected to play a role in this effect. Again, this unexpected, dramatic effect provided a good test case for the sensitivity of MD simulations. Although differences between the processes of self-assembly of the two dipeptides were noticed (e.g. more lateral association of small aggregates for this system through H-T interactions, rather than 1D self-assembly as observed for FF), it is not possible to currently draw such conclusions on MD results alone.

Another observation that was made from the simulations presented in this chapter is that although the observed morphology and shape were often predicted correctly (mainly 1D structures), the size of the construct (a few nm) often does not resemble the experimentally found values (up to 100s of nm). Again, this can be attributed to the

limited number of molecules in the box, but it can be envisioned that this can prevent optimal stacking of the structures, which in turn leads to 'defects' in, e.g., H-bonding patterns.

Altogether, the importance of using MD in studying the aggregation of peptides is certainly not disputed here. As one of few currently available techniques it provides a direct and intuitive insight with atomic detail into the structure of self-assembled nanostructures. Various groups have used all-atom MD to support their data with success (for examples see Ch. 2.3.6). However, care has to be taken with limited system sizes, starting from biased structures and kinetic traps. Importantly, analysis tools to extract information from an ensemble of molecules will have to be developed be able to draw the link between MD results and experimental observables such as H-bonding patterns (IR), aromatic stacking (fluorescence / UV-Vis), chirality (circular dichroism), etc.

4.6 Conclusion

Three different systems have been studied in this chapter: Firstly, Fmoc-X_1X_2-OMe nanostructures (X_1 = S, T, X_2 = F, L); secondly mixed-molecule system composed of hydrogelators Fmoc-FF or Fmoc-FY and surfactant Fmoc-X (X = S, T and RGD) and, finally, self-assembling unprotected dipeptides FF and IF. For each system, FTIR spectroscopic measurements and MD simulations were performed.

In the case of the Fmoc-dipeptide methyl esters discussed in section 4.4.1, IR absorption patterns indicated a β-sheet-like arrangement of the peptide backbone, with a relatively high component around 1650 cm^{-1} that indicated a frustrated sheet. Together with energy minimizations based on the atomistic CHARMM force fields it was possible to provide a π-β bilayer-type packing of the monomers, which could also explain the WAXS diffraction data and the very unusual two-dimensional mode of self-assembly found for

Fmoc-SF-OMe by TEM. However, in dynamic simulations the nanostructures did lose some of their order.

For IR spectra and simulations of co-assembled structures of Fmoc-FF with Fmoc-S, Fmoc-T, or Fmoc-RGD and Fmoc-FY with Fmoc-S, subtle differences in the monomers gave rise to distinct changes in the spectra. Peaks at 1611 and 1666 cm^{-1} are proposed to be a spectral signature for a core/shell structure that is in agreement with structural characterisation reports by light scattering, AFM and fluorescence spectroscopy. Moreover, the importance of the pH (and hence ionisation state) of these peptides was apparent from the models, consistent with experimental reports.[25,26] This proves atomistic MD can provide important information for the discovery of new biomaterials that need to be used at physiological pH. In this example it was correctly predicted that Fmoc-S does not form fibrous structures by itself, but can assemble on the surface of a spontaneously assembled Fmoc-FF fibre to form thicker, coated fibres.

Finally, MD was used to predict the self-assembly of small peptides starting from randomly oriented, well-solvated peptides on a ~100 ns time-scale. Diphenylalanine (FF) and isoleucinyl-phenylalanine (IF) were shown to form fibrous nanostructures in reasonable agreement with experimental results. The importance of hydrophobic and head-to-tail ionic interactions was discussed in the light of literature reports and IR spectra of these systems and aromatic stacking was found to be the most important driving force, especially in the case of FF. This computational approach opens the possibility to test new peptides *in silico* for self-assembly properties before experimental efforts are needed.

While the nanostructure morphology was generally correctly predicted for all three case studies discussed in this chapter, the exact conformation on a molecular level of the aromatic peptide amphiphiles could currently not be resolved from these simulations, especially when starting from a random orientation of molecules in solution. Moreover, the importance of hydrogen bonding, indicated by the information-rich amide I IR

spectra, was not always evident from the simulation results of these dynamic systems, possibly due to the limited simulation size. It is expected that improved force field parameters for the Fmoc-group and advances in computational power will lead to the development of a concise model for the self-assembled state of these molecules. Moreover, new methods such as described in Chapter 3 will be required to get an insight in the assignment of IR spectra, which are currently hard to interpret for complex systems.

4.7 References

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

(2) 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.

(3) Jayawarna, V.; Roy, S.; Holly, R. J.; Javid, N.; Zelzer, M.; Lamprou, D. A.; Urquhart, A. J.; Frederix, P. W. J. M.; Hunt, N. T.; Oreffo, R. O. C.; Richardson, S. M.; Merry, C. L. R.; Dalby, M. J.; Ulijn, R. V. *In preparation.*

(4) Abul-Haija, Y. M.; Roy, S.; Frederix, P. W. J. M.; Javid, N.; Jayawarna, V.; Ulijn, R. V. *Small, Under review.*

(5) Merrifield, R. B. *J. Am. Chem. Soc.* **1963**, *85*, 2149–2154.

(6) Merrifield, R. B. In *Hypotensive Peptides*; Erdös, E. G.; Back, N.; Sicuteri, F.; Wilde, A. F., Eds.; Springer Berlin Heidelberg, 1966; pp. 1–13.

(7) Das, A. K.; Hirst, A. R.; Ulijn, R. V. *Faraday Discuss.* **2009**, *143*, 293–303.

(8) 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*, 960–966.

(9) Mu, X.; Eckes, K. M.; Nguyen, M. M.; Suggs, L. J.; Ren, P. *Biomacromolecules* **2012**, *13*, 3562–3571.

(10) 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*, 37–41.

(11) MacKerell, A. D.; Banavali, N.; Foloppe, N. *Biopolymers* **2000**, *56*, 257–265.

(12) 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*, 671–690.

(13) Ulijn, R. V.; Bibi, N.; Jayawarna, V.; Thornton, P. D.; Todd, S. J.; Mart, R. J.; Smith, A. M.; Gough, J. E. *Mater. Today* **2007**, *10*, 40–48.

(14) Silva, G. A.; Czeisler, C.; Niece, K. L.; Beniash, E.; Harrington, D. A.; Kessler, J. A.; Stupp, S. I. *Science* **2004**, *303*, 1352–1355.

(15) Gelain, F.; Bottai, D.; Vescovi, A.; Zhang, S. *PLoS ONE* **2006**, *1*, e119.

(16) Higuchi, A.; Ling, Q.-D.; Chang, Y.; Hsu, S.-T.; Umezawa, A. *Chem. Rev.* **2013**.

(17) Jayawarna, V.; Richardson, S. M.; Hirst, A. R.; Hodson, N. W.; Saiani, A.; Gough, J. E.; Ulijn, R. V. *Acta Biomater.* **2009**, *5*, 934–943.

(18) Jayawarna, V.; Ali, M.; Jowitt, T. A.; Miller, A. F.; Saiani, A.; Gough, J. E.; Ulijn, R. V. *Adv. Mater.* **2006**, *18*, 611–614.

(19) Rimmer, S.; Johnson, C.; Zhao, B.; Collier, J.; Gilmore, L.; Sabnis, S.; Wyman, P.; Sammon, C.; Fullwood, N. J.; MacNeil, S. *Biomaterials* **2007**, *28*, 5319–5331.

(20) D'Souza, S. E.; Ginsberg, M. H.; Plow, E. F. *Trends Biochem. Sci.* **1991**, *16*, 246–250.

(21) Cheng, G.; Castelletto, V.; Jones, R. R.; Connon, C. J.; Hamley, I. W. *Soft Matter* **2011**, *7*, 1326.

(22) Yang, Z.; Gu, H.; Fu, D.; Gao, P.; Lam, J. K.; Xu, B. *Adv. Mater.* **2004**, *16*, 1440–1444.

(23) Sadownik, J. W.; Leckie, J.; Ulijn, R. V. *Chem. Commun.* **2011**.

(24) 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*, 1089–1094.

(25) Tang, C.; Smith, A. M.; Collins, R. F.; Ulijn, R. V.; Saiani, A. *Langmuir* **2009**, *25*, 9447–9453.

(26) Adams, D. J.; Butler, M. F.; Frith, W. J.; Kirkland, M.; Mullen, L.; Sanderson, P. *Soft Matter* **2009**, *5*, 1856–1862.

(27) Reches, M.; Gazit, E. *Science* **2003**, *300*, 625–627.

(28) Sanchez-de Groot, N.; Parella, T.; Aviles, F.; Vendrell, J.; Ventura, S. *Biophys. J.* **2007**, *92*, 1732–1741.

(29) Görbitz, C. H. *Chem. Commun.* **2006**, 2332–2334.

(30) Görbitz, C. H. *Acta Crystallogr. C* **2004**, *60*, o371–o373.

(31) Tamamis, P.; Adler-Abramovich, L.; Reches, M.; Marshall, K.; Sikorski, P.; Serpell, L.; Gazit, E.; Archontis, G. *Biophys. J.* **2009**, *96*, 5020–5029.

(32) Görbitz, C. H. *Acta Crystallogr. B* **2010**, *66*, 84–93.

(33) Rissanou, A. N.; Georgilis, E.; Kasotakis, E.; Mitraki, A.; Harmandaris, V. A. *J. Phys. Chem. B* **2013**.

(34) Guo, C.; Luo, Y.; Zhou, R.; Wei, G. *ACS Nano* **2012**, *6*, 3907–3918.

(35) 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*, 1781–1802.

(36) Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 5648–5652.

(37) Stephens, P. J.; Devlin, F. J.; Chabalowski, C. F.; Frisch, M. J. *J. Phys. Chem.* **1994**, *98*, 11623–11627.

(38) Schäfer, A.; Horn, H.; Ahlrichs, R. *J Chem Phys* **1992**, *97*, 2571–2577.

(39) *TURBOMOLE*; TURBOMOLE GmbH, 2012.

(40) Hanwell, M. D.; Curtis, D. E.; Lonie, D. C.; Vandermeersch, T.; Zurek, E.; Hutchison, G. R. *J. Cheminformatics* **2012**, *4*, 17.

(41) Humphrey, W.; Dalke, A.; Schulten, K. *J. Mol. Graph.* **1996**, *14*, 33–38.

(42) Cornelissen, J. J. L. M.; Rowan, A. E.; Nolte, R. J. M.; Sommerdijk, N. A. J. M. *Chem. Rev.* **2001**, *101*, 4039–4070.

(43) Cui, H.; Muraoka, T.; Cheetham, A. G.; Stupp, S. I. *Nano Lett.* **2009**, *9*, 945–951.

(44) Shao, H.; Parquette, J. R. *Chem. Commun.* **2010**, *46*, 4285–4287.

(45) Brack, A.; Orgel, L. E. *Nature* **1975**, *256*, 383–387.

(46) Nam, K. T.; Shelby, S. A.; Choi, P. H.; Marciel, A. B.; Chen, R.; Tan, L.; Chu, T. K.; Mesch, R. A.; Lee, B.-C.; Connolly, M. D.; Kisielowski, C.; Zuckermann, R. N. *Nat. Mater.* **2010**, *9*, 454–460.

(47) Naskar, J.; Palui, G.; Banerjee, A. *J. Phys. Chem. B* **2009**, *113*, 11787–11792.

(48) Debnath, S.; Shome, A.; Das, D.; Das, P. K. *J. Phys. Chem. B* **2010**, *114*, 4407–4415.

(49) Chen, L.; Morris, K.; Laybourn, A.; Elias, D.; Hicks, M. R.; Rodger, A.; Serpell, L.; Adams, D. J. *Langmuir* **2010**, *26*, 5232–5242.

(50) Marshall, K. E.; Morris, K. L.; Charlton, D.; O'Reilly, N.; Lewis, L.; Walden, H.; Serpell, L. C. *Biochemistry (Mosc.)* **2011**, *50*, 2061–2071.

(51) Fleming, S.; Frederix, P. W. J. M.; Ramos-Sasselli, I.; Hunt, N. T.; Ulijn, R. V.; Tuttle, T. *Langmuir* **2013**, *29*, 9510–9515.

(52) Barth, A.; Zscherp, C. *Q. Rev. Biophys.* **2002**, *35*, 369–430.

(53) Stein, S. E. In *NIST Chemistry WebBook, NIST Standard Reference Database Number 69, Eds. P.J. Linstrom and W.G. Mallard*; National Institute of Standards and Technology: Gaithersburg MD, 20899, http://webbook.nist.gov, 2013.

(54) Adams, D. J.; Morris, K.; Chen, L.; Serpell, L. C.; Bacsa, J.; Day, G. M. *Soft Matter* **2010**, *6*, 4144–4156.

(55) Braun, H.-G.; Cardoso, A. Z. *Colloids Surf. B Biointerfaces* **2012**, *97*, 43–50.

(56) Wang, Y.; Xu, J.; Wang, Y.; Chen, H. *Chem. Soc. Rev.* **2013**, *42*, 2930–2962.

(57) Smith, A. M.; Collins, R. F.; Ulijn, R. V.; Blanch, E. *J. Raman Spectrosc.* **2009**, *40*, 1093–1095.

(58) Hughes, M.; Birchall, L. S.; Zuberi, K.; Aitken, L. A.; Debnath, S.; Javid, N.; Ulijn, R. V. *Soft Matter* **2012**, *8*, 11565–11574.

(59) Venyaminov, S. Y.; Prendergast, F. G. *Anal. Biochem.* **1997**, *248*, 234–245.

(60) Fleming, S.; Debnath, S.; Frederix, P. W. J. M.; Ulijn, R. V. *Submitted.*

(61) Zhang, H.; Wang, H.; Xu, G.; Yuan, S. *Colloids Surfaces Physicochem. Eng. Asp.* **2013**, *417*, 217–223.

(62) Vanommeslaeghe, K. Aromatic stacking interactions. *CHARMM Dev. Forum*.

(63) ParamChem/CGenFF program; 2011.

(64) Vanommeslaeghe, K.; MacKerell, A. D. *J. Chem. Inf. Model.* **2012**, *52*, 3144–3154.

(65) Vanommeslaeghe, K.; Raman, E. P.; MacKerell, A. D. *J. Chem. Inf. Model.* **2012**, *52*, 3155–3168.

(66) 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*, 9797–9806.

(67) Ash, W. L.; Zlomislic, M. R.; Oloo, E. O.; Tieleman, D. P. *Biochim. Biophys. Acta BBA - Biomembr.* **2004**, *1666*, 158–189.

(68) Gaigeot, M.-P. *Phys. Chem. Chem. Phys.* **2010**, *12*, 3336–3359.

Chapter 5: Virtual screening for peptide aggregation: towards predictive tools for peptide self-assembly*

* Part of this chapter has been published as Frederix, P. W. J. M.; Ulijn, R. V.; Hunt, N. T.; Tuttle, T., *Virtual screening for peptide aggregation: towards predictive tools for peptide self-assembly,* J. Phys. Chem. Lett. 2011, 2, 2380–2384.[1]

5.1 Abstract

Several short peptide sequences are known to self-assemble into supramolecular nanostructures with interesting properties. Herein the use of a coarse-grained molecular dynamics approach is described to rapidly screen all 400 combinations of gene-encoded dipeptides and all 8000 tripeptides and predict their ability to aggregate as a potential precursor to their self-assembly. A simulation protocol and scoring method, based on relative solvent accessible surface areas, were proposed and evaluated, which allowed a rapid determination of whether a given peptide sequence was likely to aggregate under aqueous conditions. Furthermore, a method of selecting a subset of peptides based on the hydrophobicity of the amino acids was proposed for when large amounts of peptides have to be modelled or when modest computational resources are available. Systems that show strong aggregation tendencies in the initial screening protocol were selected for longer simulations, which resulted in good agreement with the known self-assembly or aggregation of peptides reported in the literature. Moreover, by carrying out extended simulations of experimentally studied systems (in particular diphenylalanine and triphenylalanine), it was showed that the coarse grain model is able to reproduce salient features of the nanoscale systems and provide insights into the self-assembly mechanism.

5.2 Introduction and objectives

5.2.1 Dipeptides

Self-assembling peptide materials are currently widely studied for potential applications in biomedicine and nanotechnology.[2] They can form various nanostructures from one-dimensional fibers[3] and tubes[4–6] to sheets[7] and 3D networks with tunable supramolecular functionalities such as electronic conductivity[8] and bioactivity.[9] These features, combined with their benign toxicology[10] and ease of preparation, position peptide based nanostructures at the forefront of next-generation, soft nanomaterials design, as discussed in Chapter 2. The high degree of flexibility in the types of nanostructures that can be produced is due to the vast chemical space that is available when considering the 20 gene-encoded amino acids and various functional groups that can be combined to create self-assembling peptide based units. As such, a method to rapidly screen these various combinations in the computer before investing time and resources in the laboratory is highly beneficial to the future development of novel materials.

Here a virtual screening procedure has been applied to both dipeptides and tripeptides. In the first half of this chapter dipeptides are discussed, followed by the tripeptides in section 5.4.

Dipeptides consist of only two amino acids and are among the smallest of known peptide self-assemblers and have been reported to form nanostructures with interesting morphologies and functionalities.[11,12] The most famous and well-studied example is diphenylalanine (FF) as first described by Reches and Gazit in 2003[11] and more recently reviewed by Yan *et al*.[12] FF is known to form tubular structures upon dilution from solvents into water or simply sonication in deionised water,[13] but can be manipulated to adopt different supramolecular order in other solvents.[14,15] Other dipeptide examples include FW[11] and IF,[16] which form tubular structures and fibrous hydrogels, respectively. For atomistic simulation results on FF and IF, see Ch. 4.5. In a systematic survey, Görbitz showed that many more dipeptides can form crystalline structures with

microporous unit cells.[17,18] However, the formation of crystalline structures in saturated solutions does not necessarily imply self-assembly at low concentrations in solvent, as discussed in Ch. 4.

Here a computational approach that allows us to determine, in an initial rapid screening, which peptides will exhibit aggregation in water is reported. The aggregation of the peptides is a necessary pre-condition for their self-assembly, which can then be studied using longer, larger and more detailed simulations. Note that although in proteins many residues may stack (e.g. F and Y side chains have been shown to stack in various solvents[19]), examples of small peptides are more rare as the charged C- and N-termini and polarity of the backbone play a relatively important role in preventing aggregation.

For many amyloid peptide fragments different simulation approaches have shown aggregation of individual oligopeptide fragments.[20–23] Wu and Shea have recently reviewed a number of coarse-grain molecular dynamics (MD) methods for protein aggregation in the light of amyloid-related diseases and observed several mechanisms for fibrillization.[24] They conclude coarse-grain methods are especially suitable for providing insight in the nucleation, growth mechanisms and macroscale structure of peptide assemblies thanks to the availability of large scale simulations. A different coarse-grain approach was undertaken by Thirumalai's group: in an effort to screen for peptide fibrillogenesis, they employed lattice models to determine contributions from the amino acid sequence to monomer aggregation.[25] Since the free energy landscape of partially unfolded, aggregation-prone conformers is considered to be the major determinant for the fibrillization time scale,[26] this method is also suitable for finding aggregating sequences by exhaustively sampling the conformational and sequence space.

Of specific interest for the results on di- and tripeptides reported here, Tamamis *et al*. have performed replica-exchange MD (see section 2.3.5) for the phenylalanine dipeptide and tripeptide.[27] They observed the formation of open or ring-like peptide networks through head-to-tail interaction, consistent with crystallographic evidence, together with

significant β-sheet content by amide-amide interactions. However, these simulations are often limited to small systems sizes (e.g., 12 dipeptides were used in the replica-exchange study) or short simulation times due to computational cost issues, although aggregation of self-assembling monomers often occurs on a μs timescale.[28] Villa *et al*. have showed that one can use implicit solvent and coarse-grained (CG) models for small peptides to improve the accessible time frame without losing agreement with all-atom MD simulations.[29,30]

When more computational resources become available, more detail could be introduced into the screening simulations. For example, Georgoulia and Glykos have recently employed atomistic simulations to study the foldability of 8640 tetra- and pentapeptides containing tryptophan in the search for quickly folding peptides. Applying a series of scoring functions based on characteristics of fast folding processes (e.g. ring-closure events or limited root mean square deviation for folded peptides) they continuously narrowed down the selection of interesting peptides. This allowed the simultaneous increase the simulation box size and run time, finally identifying the two best foldable peptides from their subset (RDKWP and RWPD). However, these simulations were performed on single peptides and in modelling self-assembly processes, significantly more monomers are required. Therefore, coarse-grain approaches can provide more information for the systems under investigation here.

In the current study, we have employed the coarse grain MARTINI force field for biomolecular simulations[31,32] to model the aggregation ability for all possible combinations of two and three amino acids. This force field utilizes a 4-to-1 atom:CG-bead mapping to represent protein backbone and side chains (and a 2-3:1 atom:CG-bead mapping for ring systems). As an example, Figure 5.1 displays the diphenylalanine peptide structure and CG-bead representation of the same peptide. The force field was originally modelled on small peptides and proteins and thus used without further parametrization and is discussed further in section 5.3, where representations for all

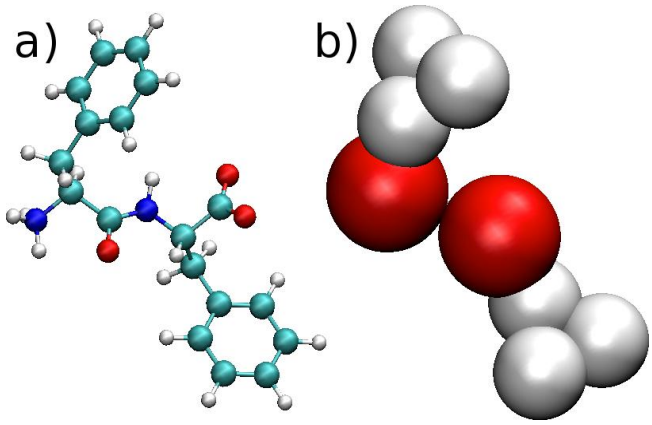


Fig. 5.1. a) Structure of the FF dipeptide starting structure. b) MARTINI coarse-grain representation. Red beads represent the peptide backbone, white beads the side chain.

amino acid residues are given. Obviously, there is a significant loss of atomic detail involved in coarse-graining, which removes the ability of the method to reproduce features such as H-bond directionality. This may cause overestimation of some intrapeptide or even inflexible interpeptide interactions: hydrogen bond donor and acceptor may not be able to form a H-bond because of steric effects in the atomistic picture, but interact within the CG force field, as this interaction is only represented by a coulombic point charges. However, the polarity, shape and non-bonded interaction potential of the amino acids are implicitly included in the force field through the parametrization of functional groups based on comparison with experimental results.[31,32] Thus, although the detailed supramolecular structure is not available, coarse grain methods do provide insights into the driving force for the aggregation and consequently the self-assembly of peptides, as the interaction strengths of peptides are accurately represented.

5.2.2 Tripeptides

For many naturally occurring self-assembling systems the peptide chain is longer than two amino acids, for example in the amyloid peptides related to, among others, Alzheimer's disease.[33] In the field of prebiotic chemistry and the formation of the first

proteins and enzymes, it is possible that complex proteins and enzymes have arisen from larger and larger self-replicating or catalytic peptides, although even dipeptide SH has been proposed to have catalytic activity towards peptide bond formation.[34] For (bio)materials scientists, longer self-assembling peptides can be useful to have more handles for introducing functionality and complexity into their nanostructures. We have therefore chosen to set the first step towards larger peptides by performing MD simulations on all tripeptides. As mentioned in the previous section, the chemical toolbox for peptides is large and the number of possible combinations with N amino acid residues rapidly increases as 20^N, creating 8000 different tripeptides.

A number of research groups have reported self-assembling tripeptides. Here, only tripeptides without, or with small protecting groups (t-butoxycarbonyl (Boc), acetyl (Ac) and methyl esters (OMe)) on the termini are discussed. Reches and Gazit reported the self-assembly of CFF into nanospheres rather than nanotubes for the FF dipeptide, which they attributed to the presence of a thiol.[35] Tamamis *et al*. reported on the formation of plate-like FFF structures using various types of spectroscopy, microscopy and MD simulations on a small amount of phenylalanine di- and tripeptides.[27] The group of Hartley has shown the self-assembly of LFF,[36] VFF and FFV[37] and their stereoisomers in which the first residue is a D-amino acid. They observed that changes in chirality on the molecular level also have dramatic consequences on the supramolecular morphology, attributed to steric effects when the tripeptides assemble into an antiparallel β-sheet conformation. Note that all these examples have at least two phenylalanine residues (termed a 'phenylalanine zipper' by Hartley). When the charges at the termini are removed by protecting groups, more sequences are prone to self-assemble. Hauser *et al*. reported acetylated Ac-IVD,[38] Ac-IVF, Ac-IVW and Ac-IVY[39] to form hydrogels (IVD and IVF) or aggregates (IVW and IVY). Subbalakshmi reported Boc-WII-OMe forms self-assembling fibrous hydrogels, in contrast with Boc-LII-OMe which formed ordered spheres.[40]

To test our method further, find new candidates for biomaterials and examine the self-assembling systems reported in literature, we have applied the MARTINI force field CG screening protocol discussed in the previous sections to all 8000 tripeptides. However, the large number of studied systems requires a more careful selection of simulation parameters due to limited availability of computational power. Therefore, a short initial simulation can be performed, in which initial aggregation into small assemblies can take place in the simulation box. We show that at this point, a selection of interesting cases can be made, depending on the peptide property of interest: this does not have to be just based on a peptide's propensity to aggregate, but can also take, e.g., hydrophobicity, aggregate shape or the presence of certain combinations of amino acids into account. For selected cases, an extended simulation can then be performed to let the (supramolecular) structure converge further.

The remainder of this chapter is divided as follows: after the details of the coarse-grain methods and simulation parameters, the dipeptide results are discussed in section 5.4. Afterwards, the protocol for obtaining valuable screening results is extended to the tripeptide systems (section 5.5). The chapter concludes with suggestions for further steps towards longer peptides or more efficient simulations (section 5.6) and the conclusions (section 5.7) that can be drawn from the work presented here.

5.3 Coarse-grain force field and simulation details

We have employed the coarse grain MARTINI force field for biomolecular simulations[31,32] to model peptide self-assembly. Applications and design principles of this force field have recently been discussed by Marrink and Tieleman.[41] Briefly, it utilizes a 4-to-1 atom:CG-bead mapping to represent protein backbone and side chains (and a 2-3:1 atom:CG-bead mapping for ring systems), which means that 4 heavy atoms (C, N, O, hydrogens are not counted) are represented as 1 single, spherical particle. Interaction energies between the various bead types are defined in the force field parameters. Fig. 5.2 shows a representation of all the amino acids.

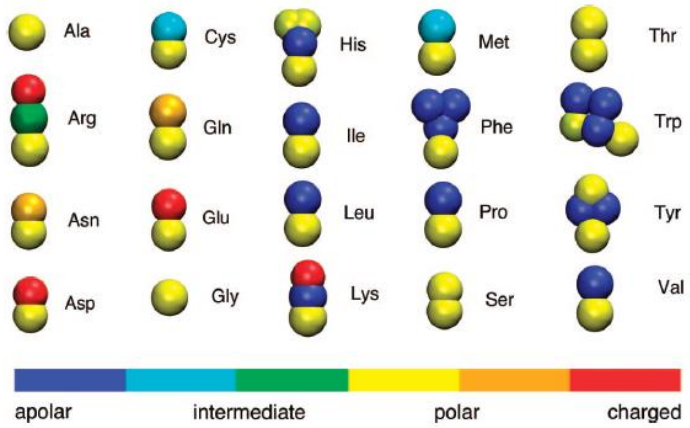


Fig. 5.2 Martini coarse-grain representation of all 20 gene-encoded amino acids. Different colours represent different types of beads, as indicated by the legend. Image obtained from ref. 22.

The force field was used without further parametrization. Obviously, there is a significant loss of atomic detail (e.g. chirality, H-Bond directionality, etc.) involved in coarse-graining. However, many interactions are implicitly included into the force field by extensive parametrization of functional groups based on comparison with experimental results.[31,32]

Please note that the original MARTINI force field parameters were recently improved by Singh and Tieleman and de Jong *et al*.,[42,43] who compared MARTINI simulation results with the Wimley-White whole-residue hydrophobicity scales.[44,45] This version of MARTINI (v. 2.2) has been used for the tripeptide simulations only, as the dipeptide simulations had already been performed prior to the force field update. Further improvements in the performance of polar residues were made when this version was used in conjunction with the polarised water model discussed above (v. 2.2P), by putting the charge off-centre in charged beads (D,E,K,R) and using polarised particles for uncharged, polar residues (N,Q,S,T). This change allows for stronger interactions between polar residues which were previously prohibited by the large VdW radius of the beads (i.e. the off-centre point charges can come closer to each the charges in other

beads; the attractive Coulomb force falls off with 1/r). The possible consequences of this change in method are discussed at the end of section 5.4.

In this study, a simulation for each peptide (in its zwitterionic form) was set up using the GROMACS molecular dynamics package.[46] A cubic box with 300 peptides, placed randomly with a minimum distance of 3 Å between the molecules in all directions, was solvated in standard MARTINI CG water (4 water molecules per bead) to a final concentration of ~ 0.4 M for dipeptides, and 0.26 for tripeptides (same number of amino acids/volume). A Berendsen[47] thermostat (coupling constant 1.0 ps) and barostat (coupling constant 3.0 ps) were used to keep the temperature at 303 K and pressure at 1 bar, respectively, as recommended in the MARTINI force field.[41] Bond lengths in aromatic side chains and the backbone-side chain bonds in I, V and Y were constrained using the LINCS algorithm.[48] All boxes were energy minimized using the steepest descent integrator and then equilibrated for $4 \cdot 10^6$ time steps of 25 fs* for dipeptides. For tripeptides, $5 \cdot 10^5$ steps were performed, followed by a $4 \cdot 10^6$ steps simulation in polarisable water[49] for selected cases (see below). For particularly interesting cases, larger or longer simulations were performed in a 21.6 × 21.6 × 21.6 nm box.

For the 8000 tripeptides, an initial screening phase of 50 ns was performed. Only for selected cases (see next section for selection criteria) a full 400 ns simulation was run to let the (supramolecular) structure converge further. For these extended MD runs, we have used the newly developed MARTINI polarisable water instead of the standard CG water. This water features 3 beads for 4 water molecules, unlike the standard 4:1 MARTINI water, and has 2 opposite point charges at constrained distances from the centre of the beads. This modification has been shown to account better for the orientational

* Recently there has been some discussion about the validity of taking large time steps in CG MD, see Winger, M.; Trzesniak, D.; Baron, R.; van Gunsteren, W. F. *Phys. Chem. Chem. Phys.* **2009**; Marrink, S. J.; Periole, X.; Tieleman, D. P.; de Vries, A. H. *Phys. Chem. Chem. Phys.* **2010**, *12*, 2254 and van Gunsteren, W. F.; Winger, M. *Phys. Chem. Chem. Phys.* **2010**, *12*, 2257. However within the current system setup we found the simulations to be well-behaved.

polarisability of real water.[49] Although this increases the computational cost of the simulation, it is considered to be of importance as the charge density in concentrated short peptides with charged termini is relatively high. Note that its use is likely to decrease the interactions between the charged termini of the peptides, which would possibly influence the results for the dipeptides. A simulation of the FF dipeptide using the polarisable water model, however, did not reveal any difference on AP or morphology of the nanotube formed (data not shown).

During the simulation with periodic boundary conditions, a neighbourlist was updated every 10 steps with a cut-off distance for electrostatic and Van der Waals interactions of 1.2 nm and a VDW switch distance of 0.9 nm using the respective shifted potentials. The standard MARTINI dielectric constant of 15 was used (2.5 for simulations using polarisable water). Typical files for minimization and production runs containing all simulation details are given in Appendix 1.

Due to the smoothness of the CG potentials, a simulation time of 100 ns roughly corresponds to an effective 400 ns of atomistic simulation time.[31,50] Throughout this chapter, all times refer to the simulation time sped up by this factor of 4. The size of our systems, the time scales required and the number of systems that are assessed in a virtual screening, excludes the possibility of performing these MD simulations using all-atom models on modest computing resources. On the other hand, as an example, the complete, 400 ns simulation for a single dipeptide system described above can be carried out in *ca.* four hours on a 4-core 2.6 GHz Opteron computing node.

At the end of the simulation the solvent-accessible solvent area (SASA), calculated with VMD[51] with a 1.4 Å rolling sphere radius, of the initial box was divided by the SASA of the final box to give the aggregation propensity (AP).

Dihedral angle profiles were extracted using the g_angle routine in the GROMACS packages using a 5° bin width. All dihedrals from beads SC1-BBe-BBe-SC1 were taken

into account. Values were extracted on the minimized structure and average from 0.75–1.25 ns ('after 1 ns') and from an average of the last 10 ns of the simulation ('end'). The width and standard deviation of the tube and tube core diameters for the diphenylalanine simulation were determined by measuring peptide and water atom centre distances, respectively, at 10 different planes throughout the tube, orthogonal to the longest tube dimension.

5.4 Dipeptide results and discussion

5.4.1 Dipeptide Aggregation Propensities

For all 400, initially solvated dipeptides, 400 ns simulations were performed as described above. The solvent-accessible surface area (SASA) of peptides, calculated using the scripting tools in VMD, is a valuable system output for the degree of aggregation of the system: when peptides aggregate less surface area is exposed to water. Therefore, the Aggregation Propensity (AP) was defined as the ratio of the SASA of the dipeptide molecules in the initial minimised box to the SASA of the final configuration of the simulation. A large ratio is interpreted as a high degree of aggregation. From Figure 5.3, which shows the AP for all the 400 dipeptide combinations, it is apparent that F and W residues promote aggregation, while small or charged residues decrease the score. In addition, we note that the graph is roughly symmetric. There are differences for some combinations though; most notably SF and FS where the AP of the former is 0.6 higher (see Appendix 6 for a list of all dipeptide AP scores). Most dipeptides have a score close to 1 indicating they do not exhibit a propensity to aggregate, and thus their self-assembly in aqueous conditions can be excluded.

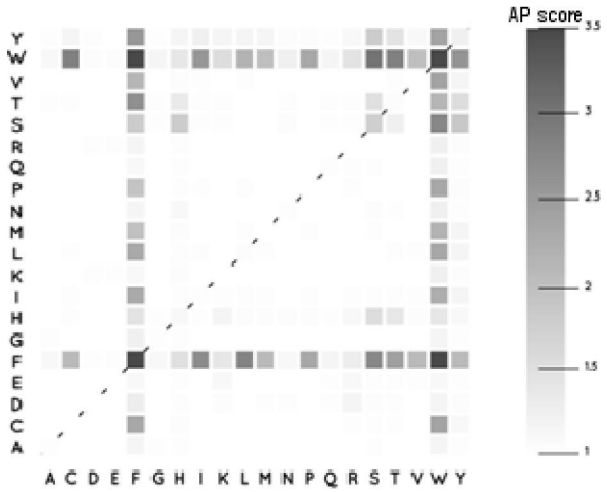


Fig. 5.3. Two-dimensional grids indicating the AP score for dipeptides. Horizontal and vertical axes show the amino acid one-letter codes for the first (N-terminus) and second (C-terminus) amino acid, respectively.

The AP scores were compared to all known experimental work on dipeptides in aqueous solution, as summarised in Table 1. Although preparation methods vary throughout the literature, our data accurately predicts the self-assembly/aggregation, or lack thereof, for those dipeptides for which experimental data are available. From this table, the data in Figure 5.3 and the final configurations of the MD simulations, we propose that an AP > 2 indicates a good candidate for further investigation into whether the system is likely to self-assemble. Clearly, this is an empirical, observation-based assignment and it will be interesting to test more complex examples and those on the borderline in an effort to break this rule. For example, HS is an interesting candidate with an AP score of 1.66 (see Appendix 6 for the full list of AP scores) and does not contain any hydrophobic amino acids. Nonetheless, within the systems studied thus far the AP score provides a useful and apparently robust indicator of the desired property.

Comparison of the SASA for the initial and final configurations of the simulations provides a clear indication of whether aggregation occurs and as such whether self-assembly may occur. However, the AP score provides no details as to the nature of the

Table 5.1. Comparison of self-assembly score with experiments

Dipeptide	Reference	Assessment method	Observed structure	AP[a]
FE	[52]	Rheology	None	1.1
FF	[11–13,53]	TEM, AFM	Tubes, vesicles	3.2
FK	[52]	Rheology	None	1.2
FW	[11,54]	TEM	Tubes, aggregates	3.5
IF	[16]	TEM, SEM	Fibers	2.3
VF	[16]	TEM	None	1.8
WF	[11*b*]	TEM	Aggregates[c]	3.5
WW	[11*b*]	TEM	Aggregates[c]	3.2
WY	[11*b*]	TEM	Aggregates[c]	2.1

[a] Aggregation Propensity score is the ratio of the solvent-accessible surface area (SASA) of the dipeptide molecules in the initial minimised box to the SASA of the final configuration of the simulation. [b] Supplementary information. [c] Amorphous

nanostructures. Globular, fibrous and tubular structures have been observed amongst the results, but these features cannot be distinguished by this score.

5.4.2 Dipeptide morphological features

We have shown that the tendency for dipeptides to aggregate can be determined within a 400 ns simulation window on the relatively small (300 dipeptide) systems used in the virtual screening procedure. However, in order to test whether the CG model is able to converge (either qualitatively or even semi-quantitatively) towards the nanostructures observed experimentally the simulation time and system size should be further increased to allow these structures to form. We have explored this possibility in the case of diphenylalanine by studying a larger system (1600 dipeptides in a box with double the dimensions of the original box, Figure 5.4a) in order to see whether the structures observed experimentally in an aqueous environment[11–13,53] are in agreement with the final structure from the simulation. A link to a movie of the full simulation can be found in Appendix 7.

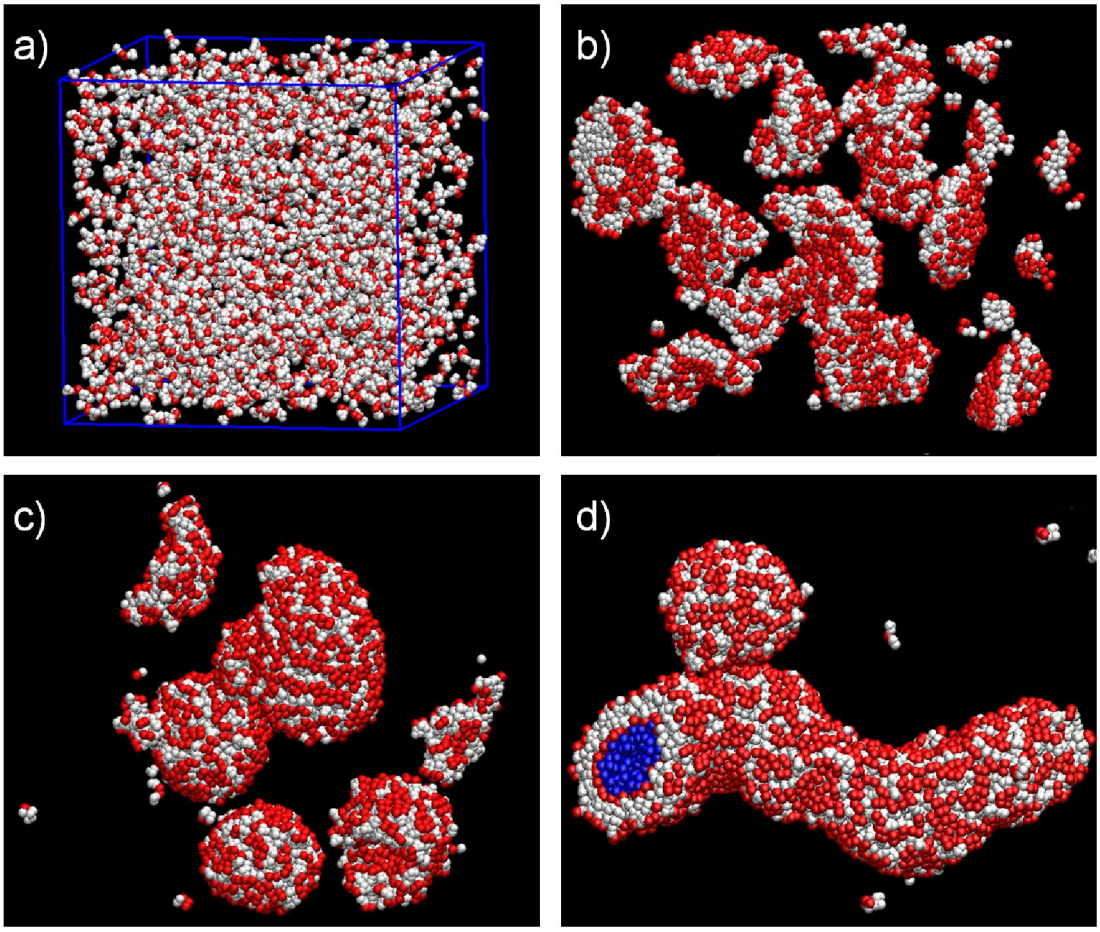


Fig. 5.4. Snapshots at different time points in the extended FF dipeptide MD simulations. Red: backbone beads, white: side chain beads. Simulation performed in explicit CG water, with beads omitted from the figure for clarity. a) 0 μs; randomly placed dipeptides. The periodic box (blue lines) is indicated in the figure. b) 0.2 μs; formation of sheet-like aggregates. c) 0.5 μs; vesicle formation by sheet folding. d) 1.5 μs; fused vesicles forming a hollow tube. The end of the tube is cut off to show water beads inside (blue).

In the final snapshot of the 1.5 μs simulation (Figure 5.4d) the FF monomers have generated a tubular nanostructure in agreement with that observed experimentally in terms of morphology (Table 1). The 1D nanotube in our simulation has an outer diameter

of 60 ± 7 Å and an inner diameter of 29 ± 5 Å (Figure 5.4d). Note that the experimentally observed nanostructures are much larger (approximately 100 nm in diameter), but are thought to consist of small hollow fibrils laterally aggregated as shown by Görbitz (see Fig. 4.10).[17] Although a direct comparison of aqueous nanostructures and X-ray crystallographic data on solid phase crystals is dangerous, as discussed in Ch. 4.5, the size of the water core (inner diameter) of our simulated tube agrees reasonably well with the X-ray diffraction analysis by Kim *et al*.[13] and Görbitz[17] of crystallised FF nanotubes (van der Waals diameter of inner channel ~ 10Å). Moreover, we also note that the dihedral angle between the sidechains rotates from 180º in the starting monomer structures (Figure 5.1) to an average value of 0º in the final tube-like nanostructure (Figure 5.5). The rotation of the sidechains away from the ideal 180º value for the dipeptide is also in agreement with the X-ray structures observed by Görbitz[18] and the atomistic simulations by Tamamis *et al*.[27] In the initial system (i.e., after minimization) the backbone dihedral angles are clustered around 120º and 180º (black line in Figure

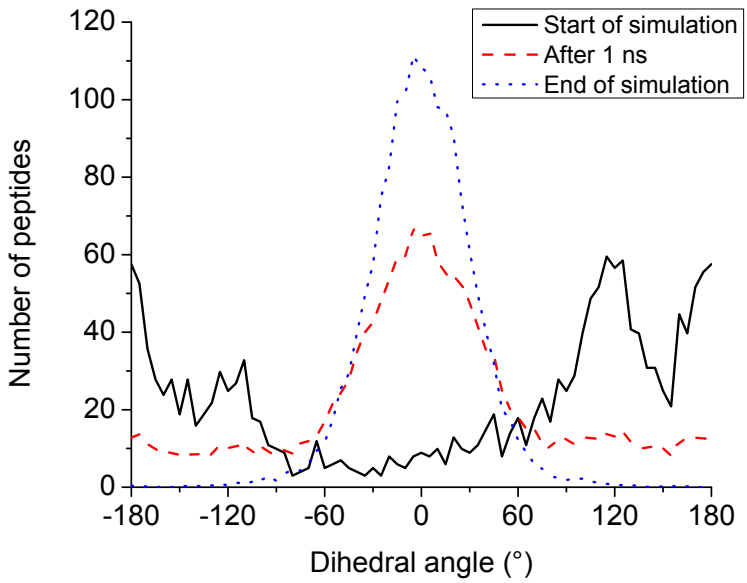


Fig. 5.5. Sidechain-backbone-backbone-sidechain dihedral angle distribution (binwidth of 5º) at the start of the simulation (0 μs, blue), after 1 ns (green) and at the end of the 1.5 μs simulation (red).

5.5). However, as the simulation continues we observe that the sidechains rotate during the aggregation process to adopt a dihedral angle between the sidechains centred around 0º. After 1 ns of simulation the majority of the dipeptides have already reached this average value (red line in Figure 5.5) although the distribution is not as sharp as is achieved in the final stages of the 1.5 μs (blue line in Figure 5.5). Although a direct comparison of the distance and dihedral parameters is not possible due to the lack of atomic detail in the CG model, the similar size and observed rotation in both the experimental and simulated systems is compelling.

In addition to the final tubular structure, the CG simulation also provides new information with respect to the mechanism by which the nanotubes are formed. From the initial random distribution of dipeptide monomers (Figure 5.4a) we observe an initial formation of sheet-like aggregates (Figure 5.4b) after only 0.2 μm. These sheet-like structures are formed by hydrophobic interactions of the side chains, while the polar backbone is exposed to solvent on the outside of the bilayer-like sheets. The sheets grow and accumulate to a point where they are then able fold up to form hollow vesicles after ~0.5 μs (Figure 5.4c). Extending the simulation time to up to a total of 1.5 μs results in the continuous accumulation of these vesicles, which fuse together into 1D nanotubes (Figure 5.4d). Thus, while the FF simulation shows that after 400 ns the final nanostructure that results from the self-assembly process may not be reached, the aggregation process is sufficiently advanced to give an accurate AP score (*vide infra*) and as such the virtual screening protocol outlined above is effectively able to screen for

the possibility of the system to self-assemble. Indeed, for dipeptides IF, WF and WW the final structures of their virtual screening simulations spanned one of the box dimensions to form fibrous assemblies, whilst FW also showed a tendency to form tubular structures, in agreement with the literature in Table 1. Results of these simulations can be found in the Fig. 5.6.

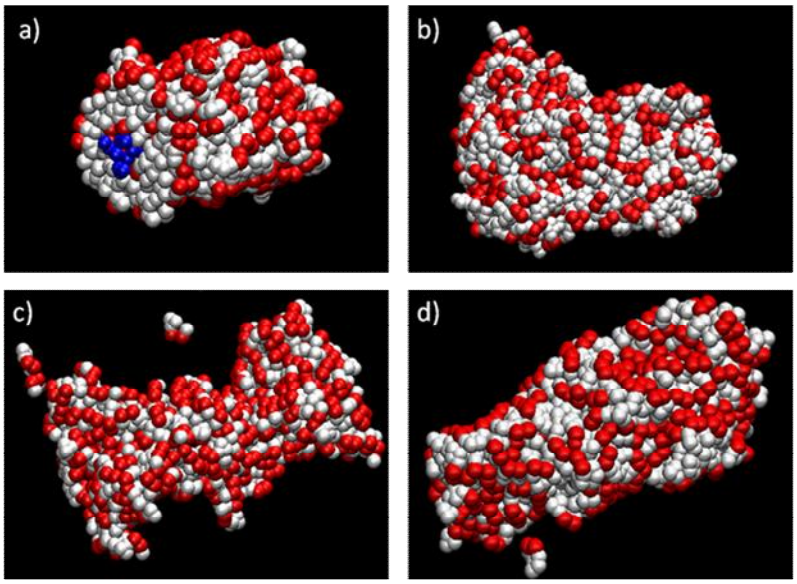


Fig. 5.6: Result of 400 ns simulations for aggregating dipeptides mentioned in the text. Every simulation box contains 300 dipeptides. Red: backbone beads, white: side chain beads. a) FW, open end of tube shows water core in blue, b) WW, no water core, c) IF, no water core and d) WF, some water beads reside in the middle of the aggregate, but no continuous core is observed.

5.4.3 Dipeptide extended simulations

To study the degree of convergence of both the structures and the AP score, extended 4 μs simulations were performed for all experimentally studied dipeptides (Table 5.1), with 1600 peptides similar to the FF case discussed above. The final structures of these simulations are displayed in Fig. 5.7. FE and FK show well-solvated peptides. VF, IF and WY now show similar amorphous aggregates, where still a relatively large amount of peptides are solvated. WW and WF show amorphous fibres that span the simulation box. FF and FW show hollow nanostructures in agreement with experiments, with FF showing an ordered nanotubular assembly as discussed before.

The AP for all experimentally studied dipeptides after 4 μs is displayed in Table 5.2 and does not show significant changes with respect to the values after 400 ns, supporting the

validity of the 400 ns value of the AP. However, VF crosses the empirical threshold of AP > 2 and shows clustered dipeptides, which are not observed experimentally by TEM or FTIR.[16] This indicates that extra care has to be taken considering borderline cases in which the AP ≈ 2. For the case of the dipeptides VF and IF, for example, small differences in side chain size could contribute to the presence or absence of aggregation. These small size differences are not included in the CG model.

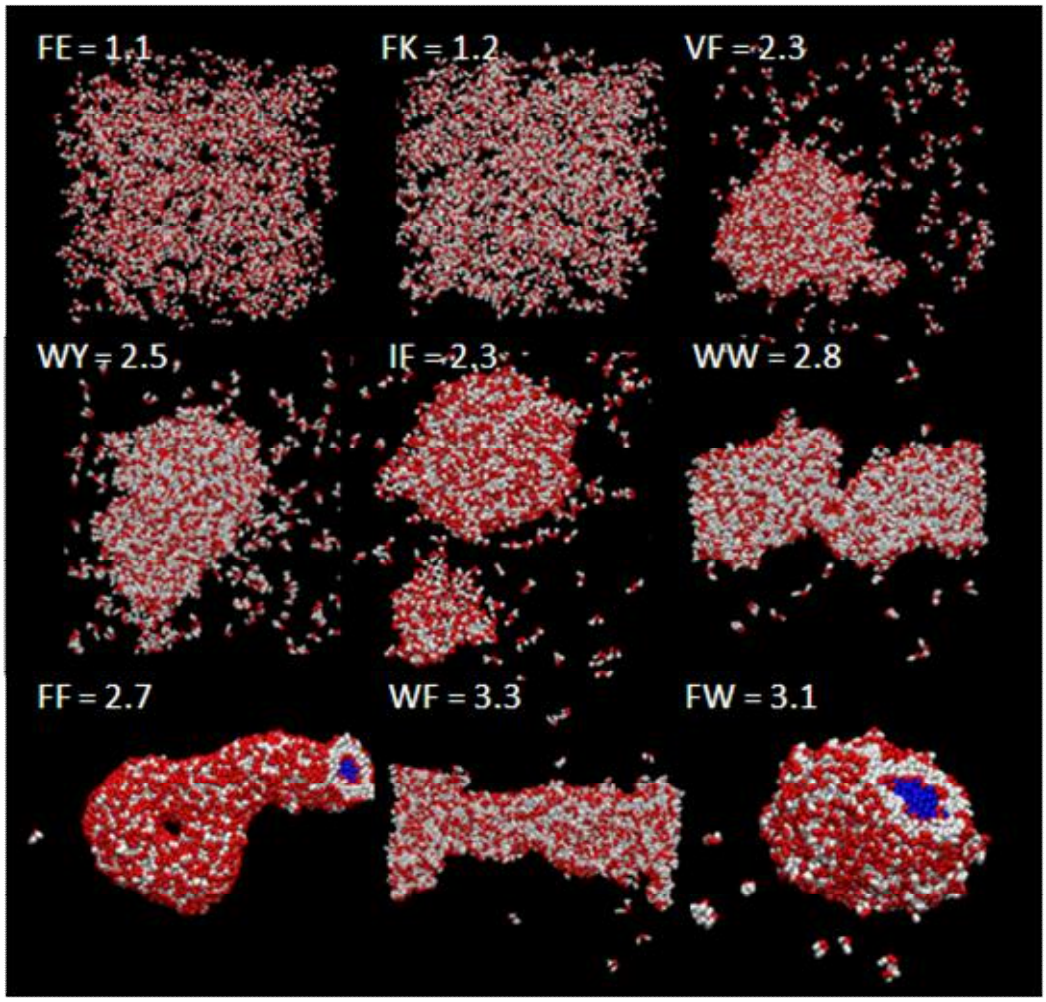


Fig. 5.7: Result of 4 μs simulations for experimentally studied dipeptides mentioned in the main text and Table 5.2. Every simulation box contains 1600 dipeptides. Red: backbone beads, white: side chain beads. For hollow nanostructures, the water core is shown in blue. The dipeptides and their respective AP scores are indicated in the top left corners of the panels.

Table 5.2. Change in Aggregation Propensity for simulations extended to 4 μs for the experimentally studied dipeptides.

Dipeptide	Observed structure	AP 0.4 μs	AP 4 μs
FE	None	1.1	**1.1**
FF	Tubes, vesicles	3.2	**2.7**
FK	None	1.2	**1.2**
FW	Tubes, aggregates	3.5	**3.1**
IF	Fibers	2.3	**2.3**
VF	None	1.8	**2.3**
WF	Aggregates	3.5	**3.3**
WW	Aggregates	3.2	**2.8**
WY	Aggregates	2.1	**2.5**

Finally, it should be noted that after these simulations were performed, Singh and Tieleman reported[42] that even though the MARTINI force field has been parametrised to reproduce the partitioning coefficients of biomolecules at the interface between water and cyclohexane,[32] with the standard MARTINI water model the charged amino acids D, E, K and R are significantly too favourable in hydrophobic environments (such as apolar solvents) compared to the Wimley-White whole-residue hydrophobicity scale.[44,45] This may have an effect on the aggregation of peptides containing these amino acids in water too: the energetic penalty for aggregation of charged residues together with hydrophobic amino acids is too low in a large nanostructure. However, dipeptides containing these amino acids did not show significant aggregation anyway and this effect would therefore not influence the conclusions. Furthermore, Singh and Tieleman found that F, W and P were slightly too hydrophobic, as their energy gain for dimerisation in water was too large compared to atomistic simulations and experimental values. This may have lead to an overestimation of the AP for peptides containing these residues in our study,

explaining why not all the peptides with an AP > 2 may form nanostructures experimentally. In future simulations, the update version of MARTINI (v2.2) is recommended, as employed for the case of tripeptides in the next section.

5.5 Tripeptides

5.5.1 Scoring method

The aqueous solubility of peptides is a matter that requires careful consideration in the field of peptide and protein biomaterials. When peptides are too hydrophilic, the propensity for self-assembly is generally low, but when they are too hydrophobic, insolubility prevents the controlled formation of extended networks in water. Our MD simulations start from a well-solvated system by definition and our previously defined AP score cannot distinguish between self-assembled and precipitated aggregates. Therefore, a hydrophobicity-corrected score AP_H was developed to rank peptides with both high AP and low hydrophobicity higher than peptides with equally high APs, but hydrophobic residues. A common measure of lipophilicity is the logarithm of the 1-octanol/water partition coefficient (logP).[55] This value is linearly proportional to the free energy change when molecules are transferred from the water to the octanol phase, $\Delta G_{water\text{-}oct}$ (for derivation, see Appendix A6.1). The $\Delta G_{water\text{-}oct}$ values for all amino acids have been reported by Wimley and White:[44,45] Table 5.3 gives the values for all 20 gene-encoded amino acids in the relevant side chain charge state at pH 7, which is used in the simulations. Other hydrophobicity scales were suggested as reviewed by Eisenberg *et al*.,[56] but the Wimley-White scale was chosen as it is determined from experiments on short peptides and independent of protein tertiary structure.[44] Moreover, this scale was used as the target data in the determination of the latest parameters of the MARTINI force field by De Jong *et al*., who reported a comparison of Wimley-White and MARTINI partitioning free energies.[43]

Table 5.3. Hydrophobicity scores ($\Delta G_{water\text{-}octanol}$, kcal/mol) for all 20 amino acids, sorted from hydrophobic to hydrophilic, taken from ref. 45.

Residue	**$\Delta G_{water\text{-}octanol}$**		**Residue**	**$\Delta G_{water\text{-}octanol}$**
Trp	−2.09		Thr	0.25
Phe	−1.71		Ser	0.46
Leu	−1.25		Ala	0.5
Ile	−1.12		Gln	0.77
Tyr	−0.71		Asn	0.85
Met	−0.67		Gly	1.15
Val	−0.46		Arg^+	1.81
Cys	−0.02		Lys^+	2.80
His	0.11		Glu^-	3.63
Pro	0.14		Asp^-	3.64

Aggregation Propensities (AP) were calculated using VMD, analogous to the previous section. Hydrophobicity-corrected scores AP_H were calculated from normalised AP and logP scores using the empirically founded equation 5.1:

$$AP_H = \left(\frac{AP(trip) - AP_{\min}}{AP_{\max} - AP_{\min}} \right)^{\alpha} \cdot \left(\frac{logP(trip) - logP_{\min}}{logP_{\max} - logP_{\min}} \right) \qquad (5.1)$$

where *AP(trip)* is the ratio of solvent accessible surface areas at the start and finish of an MD run for a given tripeptide, as for dipeptides. As *logP(trip)* is a unitless number linearly proportional to $\Delta G_{water\text{-}oct}$ (discussed above) and is normalised in Eq. 5.1, it was chosen to define it simply as the sum of the Wimley-White whole-residue hydrophobicities $\Delta G_{water\text{-}oct}$ (kcal/mol) for the tripeptide, given by:

$$logP(trip) = \sum_{i=1}^{3} (\Delta G_{\text{water-octanol}})_i \qquad (5.2)$$

In equation 5.1, α is an arbitrary coefficient that can be used to determine the weight of the normalised AP score to the AP_H value. For this study, α =2 was used to give a good compromise between AP and hydrophobicity (for validation, see Fig. 5.11). Depending on the desired properties, the exponent can be decreased (increased) to include more (less) hydrophilic peptides. Other algorithms can be used to select a subset of peptides with an optimal balance between hydrophobicity and AP, such as the Pareto optimality of a dataset (see, e.g., ref. 57), although this description provides a method with a possibility for intuitive weighting of AP and hydrophobicity. Due to computational cost restraints only the top 400 scoring peptides indicated after 50 ns by this protocol to be used in extended simulations.

5.5.2 Tripeptide aggregation propensities and evaluation of the AP_H score.

50 ns simulations were performed for all 8000 tripeptides studied. From the last frame of these simulations, the Aggregation Propensity was calculated and displayed in Fig. 5.8. It is immediately clear from the diagram that the most hydrophobic residues (W, F) promote aggregation, while especially the negatively charge residues in the first or second position have a prohibitive effect on aggregation, as expected. Y, H and to a lesser extent S and T, have a positive effect on the AP, attributed to their aromaticity and/or hydrogen-bonding capabilities.

When this is compared with the AP_H scores displayed in Fig. 5.9, a similar behaviour was noticed. However, it is apparent that some of the higher scoring peptides are not in the hydrophobic domains. Most notably, peptides with a positively charged amino acid (lysine, and to a lesser extent arginine) at the N-terminus and a complementary charge (aspartic or glutamic acid) at the C-terminus score relatively high. An example is KFD, which ranks highest in the list of AP_H scores (see Table 5.4). This can also be seen from Fig. 5.10, which shows the cumulative AP_H score for amino acids in the first, second and third position (sum of all AP_H's for the 400 peptides with that particular amino acid in that particular position): e.g. the AP scores with aspartic acid (D) in position 3 is on

average much higher than at position 1 and 2. Peptides with alternating charged and hydrophobic residues have been reported to readily self-assemble into β-sheets[58] and even gelate water in the case of FEFKFEFK.[52] Moreover, strong salt bridges can form between the 'doubly charged' termini of these peptides. This is consistent with the report on KFFE, a tetrapeptide related to Alzheimer's disease, by Bellesia and Shea, who used all-atom MD to show that hydrophobic interactions between the diphenylalanine cores only played a minor role in the aggregation of this particular fragment.[59] This observation could be important for designing new biomaterials.

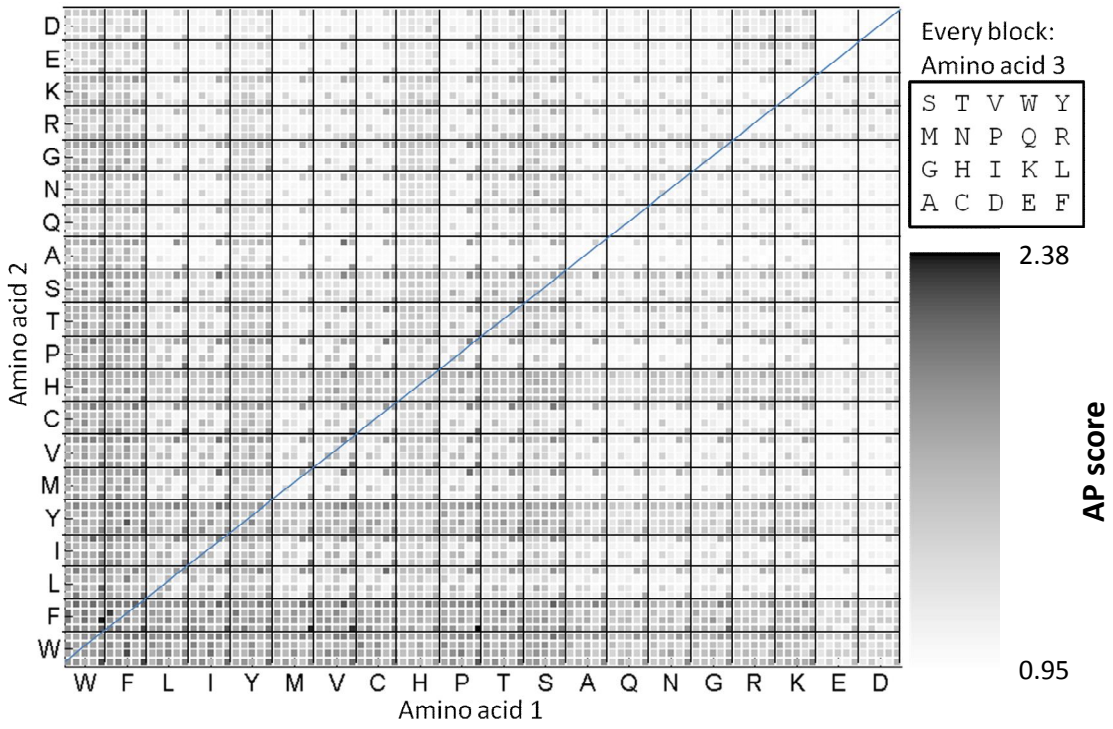


Fig. 5.8. AP score for all 8000 combinations of three amino acids after a 50 ns simulation. Within every rectangle, the third amino acid is represented by the position of the coloured square at the positions indicated in the legend on the right. The amino acids are ranked by decreasing Wimley-White hydrophobicity on the axes.

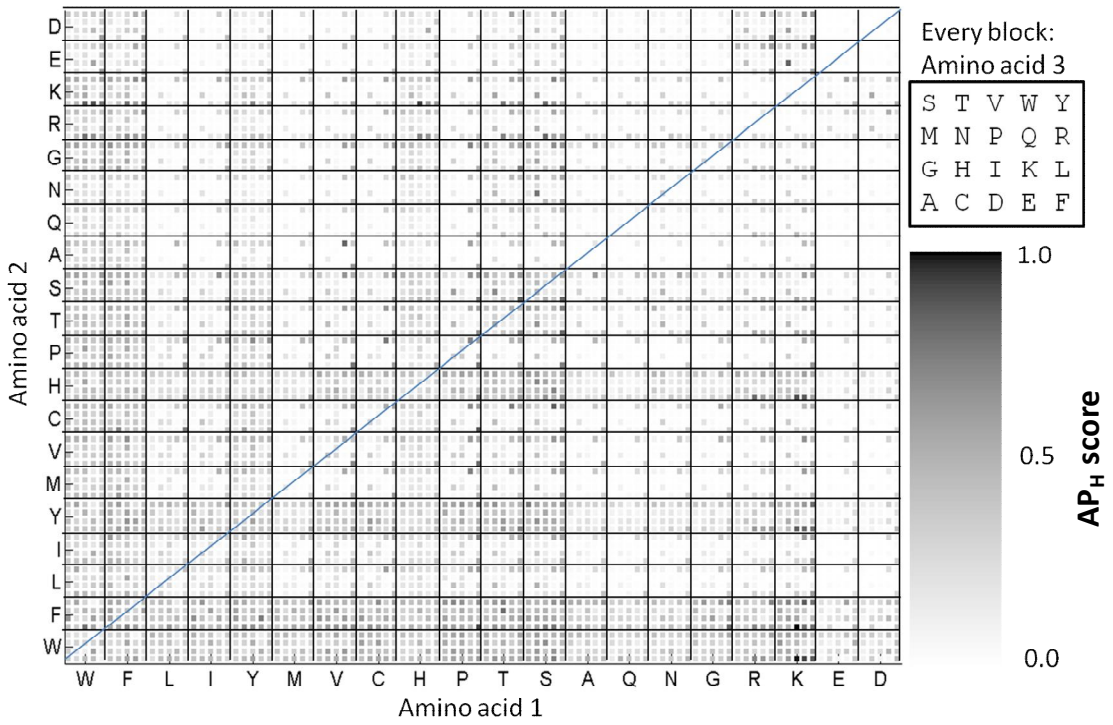


Fig. 5.9. Normalised AP_H score for all 8000 combinations of three amino acids after a 50 ns simulation. Within every rectangle, the third amino acid is represented by the position of the coloured square at the positions indicated in the legend on the right. The amino acids are ranked by decreasing Wimley-White hydrophobicity on the axes.

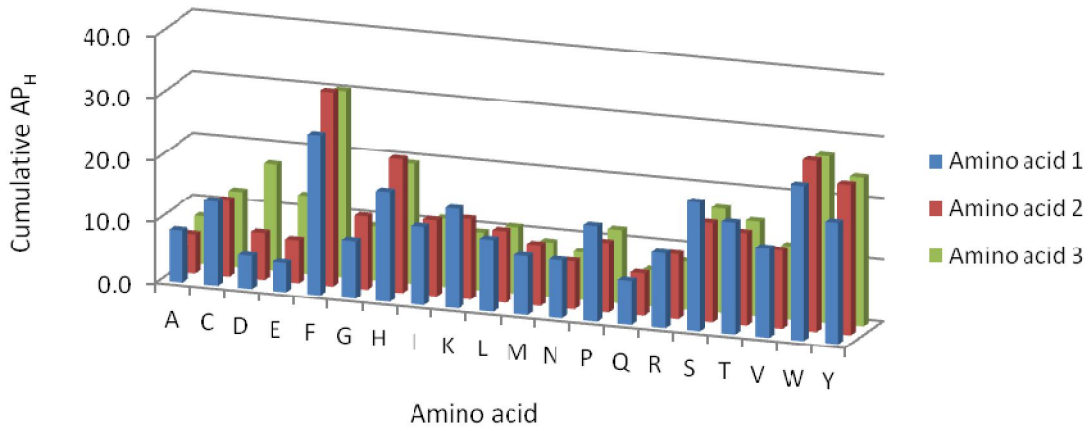


Fig. 5.10 Cumulative AP_H scores for tripeptides with the amino acid on the x-axis in the N-terminal (blue), middle (red) and C-terminal position (green).

Another observation from Figs. 5.9 and 5.10 is that peptides containing T or S stand out more as interesting candidates using the AP_H, attributed to the hydrogen bonding capabilities of their side chains. It was also noticed that hydrophobic residues in the 2nd and 3rd position have a particularly positive effect on the aggregation propensity with respect to those in the first position. The origin of this behaviour is unknown, but it should be noted that it was reported that tripeptide ^{D}VFF forms a stable hydrogel faster than its isomer ^{D}FFV experimentally.[37] Aliphatic residues (I, L, V) had an average effect on the AP_H, as could be expected, and showed no particular preferences for their position in the tripeptide. Note that the structures of the high-scoring peptides are discussed in section 5.5.3. These observations can be taken into account when designing new biomaterials

After the initial 50 ns screening phase a selection of tripeptides was made based on the resulting AP and AP_H. The empirical cutoff determined from the dipeptides was applied, selecting those peptides with an AP > 2 as a good candidate for self-assembly (see below for discussion). All 124 tripeptides satisfying this condition were selected for further simulation regardless of their hydrophobicity. Moreover, the top 400 peptides according to the AP_H score have been included in the full 400 ns simulation phase. 53 peptides fall into both categories. Fig. 5.11 shows an overview of the tripeptide APs as a function of their *logP(trip)* with the respective selections indicated. Overall, an inverse relation between hydrophilicity and AP was observed, although a wide range of AP scores was noticed, even for peptides with similar logP(trip) values. This was especially the case for peptides with intermediate hydrophilicity values (approximately -2 to 7 in Fig. 5.11). These observations validate the fact that the aggregation propensity cannot be determined from the logP value alone, but the MD simulation is necessary.

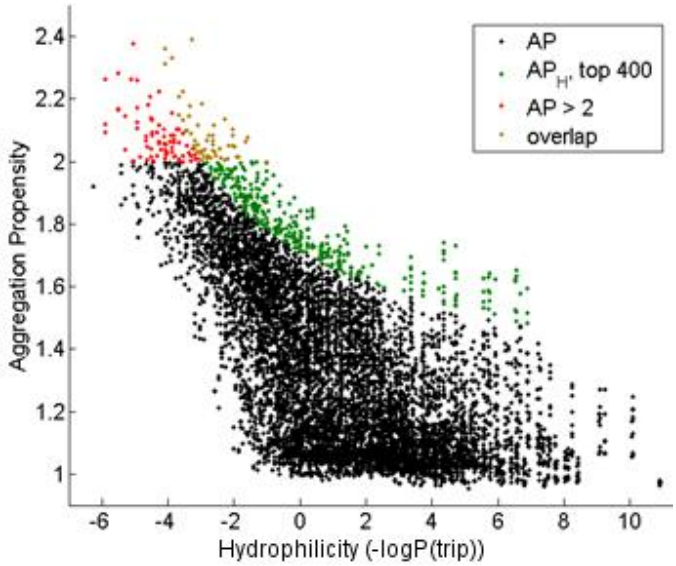


Fig. 5.11. Aggregation Propensity as function of hydrophilicity for all 8000 tripeptides. Red dots represent all tripeptides with AP > 2. Green dots represent the top 400 tripeptides from the AP_H score. Brown dots represent the overlap between these two categories.

It is not entirely straightforward to conclude that the cut-off for aggregating peptides lies at the same AP value as in the dipeptide simulations. Fig. 5.12 demonstrates that, as the surface area per tripeptide is larger than that for dipeptides, there is an increased scope for decreasing the solvent accessible surface area. Therefore, it is reasonable to assume that the AP for tripeptides needs to be somewhat higher than that for dipeptides (where the cut-off was set at AP = 2 for 400 ns) to indicate a good self-assembly candidate. On the other hand, the initial screening is only 50 ns in the case of tripeptides due to computational cost restraints. In addition, as the simulation box is 55% larger compared to dipeptides to keep the same concentration of amino acid residues, it is expected that reaching full aggregation takes a relatively longer time. Indeed, in many of the simulation boxes a number of smaller aggregates were noticed. Therefore, based on this and the results presented in Fig. 5.11, a score of 2 was still deemed to be suitable as an empirical cut-off at 50 ns. This would need to be re-examined for different systems.

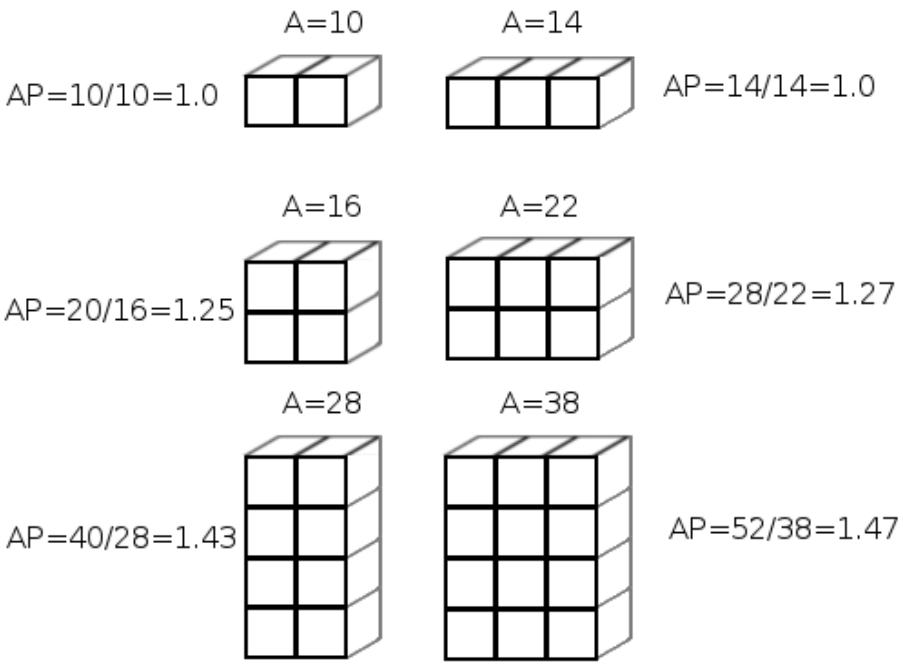


Fig. 5.12. Demonstration of principally different AP scores for dipeptides (left column) and tripeptides (right column). Each cube stands for 1 amino acid, A is the area. Even for 1D aggregation, the AP for tripeptides is higher when the same number of peptides aggregate.

Table 5.4 gives a list of selected tripeptides with their AP at 50 and 400 ns, and how their AP and AP_H scores ranked out of the 8000 systems. When compared with experimental results, all unprotected tripeptides that were reported to form self-assembled nanostructures experimentally (CFF, FFF, LFF, VFF and FFV), had an AP > 2 after 50 ns. After the 400 ns simulations, the AP scores increased for all these peptides, ranging from 2.42 for FFF to 3.04 for CFF. Due to the hydrophobicity of these peptides, they did not rank high in the AP_H list as listed in Table 5.4. The acetylated peptides reported by Lakshmanan and Hauser[39] score relatively poorly in both scoring systems, with only IVW having a high enough AP to fall within our candidate category. Ac-IVD, reported as the gelator with the most stable hydrogels, scores particularly low in our method. It is apparent that the fact that these peptides are acetylated has a major impact on their

Table 5.4. AP scores and ranking for self-assembling tripeptides from literature (top half) and top candidates from screening method (bottom half). Scores > 2 are highlighted in green, scores < 2 are red.

Trip.	**Ref.**	**# AP_H** [c]	**# AP** [c]	**AP 50 ns**	**AP 400 ns**
CFF	[35]	703	69	2.05	3.04
FFF	[27]	2422	7	2.26	2.42
FFV	[37]	1319	89	2.03	2.92
VFF	[37]	159	4	2.33	2.84
LFF	[36]	2390	58	2.07	2.70
IVD [a]	[38,39]	4084	4115	1.26	1.04
IVF [a]	[39]	1357	283	1.92	3.10
IVW [a]	[39]	490	35	2.10	2.48
IVY [a]	[39]	667	491	1.85	3.00
LII [b]	[40]	3030	1148	1.69	2.53
WII [b]	[40]	3339	752	1.77	2.72
KFD	This work	1	959	1.73	1.70
PFF	This work	4	1	2.39	3.51
KHD	This work	7	1480	1.63	1.92
TSF	This work	12	134	1.99	2.58
FYI	This work	107	10	2.22	2.24
FHF	This work	301	44	2.09	3.22
VHI	This work	461	613	1.81	2.59
AAA	This work	7815	7765	1.00	1.00

[a:] Ac-IV*X*-OH, [b:] Boc-*X*II-OMe, [c:] ranking out of 8000 at 50 ns.

assembly potential, attributed to the removal of the charge from the N-terminus, and to a smaller extent the introduction of another amide functionality that can participate in hydrogen bonding. The same was valid for the peptides studied by Subbalakshmi,[40] where Boc and methyl ester groups screen the charges at the N- and C-terminus, respectively. A small difference in AP was noted between WII and LII, supporting the

notion that the former may form more extended self-assembled nanostructures than the latter.

Furthermore, Table 5.4 lists a small number of tripeptides from across the range of candidates indicated by our protocol for more extensive screening. This lists contains a few 'surprising' tripeptides: e.g. KHD, which contains only hydrophilic residues and VHI which had a relatively high AP after 400 ns, but does not contain any of the three hydrophobic aromatic residues F, W and Y. Further, several tripeptides that ranked high at either the AP or AP_H scores, or both, were chosen for further inspection of their structures in the next section. AAA was simulated as a negative control. Experimental tests of the aggregation properties of the selected peptides are ongoing in our lab.

5.5.3 Tripeptide structural features

The CG structures of all peptides in Table 5.4 are displayed in Fig. 5.12. It is apparent that after 400 ns, most aggregating peptides formed a single cluster containing all 300 tripeptides, except from IVD, which showed no aggregation and LII, which showed several molecules that were still solvated. FFF and CFF formed the most ordered structures, where backbone beads lined up to allow for regular stacking of the phenylalanine side chains. FFV, VFF and LFF showed large aggregates, which is consistent with their assembly into fibres observed experimentally. Interestingly, IVF, IVW and IVY all formed extended structures, while IVD did not show any signs of aggregation as predicted at the 50 ns stage. This again indicates that acetylation of this group is important to obtain nanostructures experimentally for the case of IVD, while for the other IV*X* peptides, self-assembly could possibly be found without acetylation. The fact that they do not form transparent hydrogels, but rather an opaque hydrogel (IVF) and aggregates (IVW, IVY) may indicate the peptides are too hydrophobic when Ac-protected on the N-terminus.[39] LII and WII showed significant aggregation after 400 ns, with APs of 2.53 and 2.72, respectively. Although a direct comparison with experiments is ambiguous because of the capping groups in the report by Subbalakshmi,[40] it is

promising that these peptides are eventually predicted to aggregate even without the (small) capping groups.

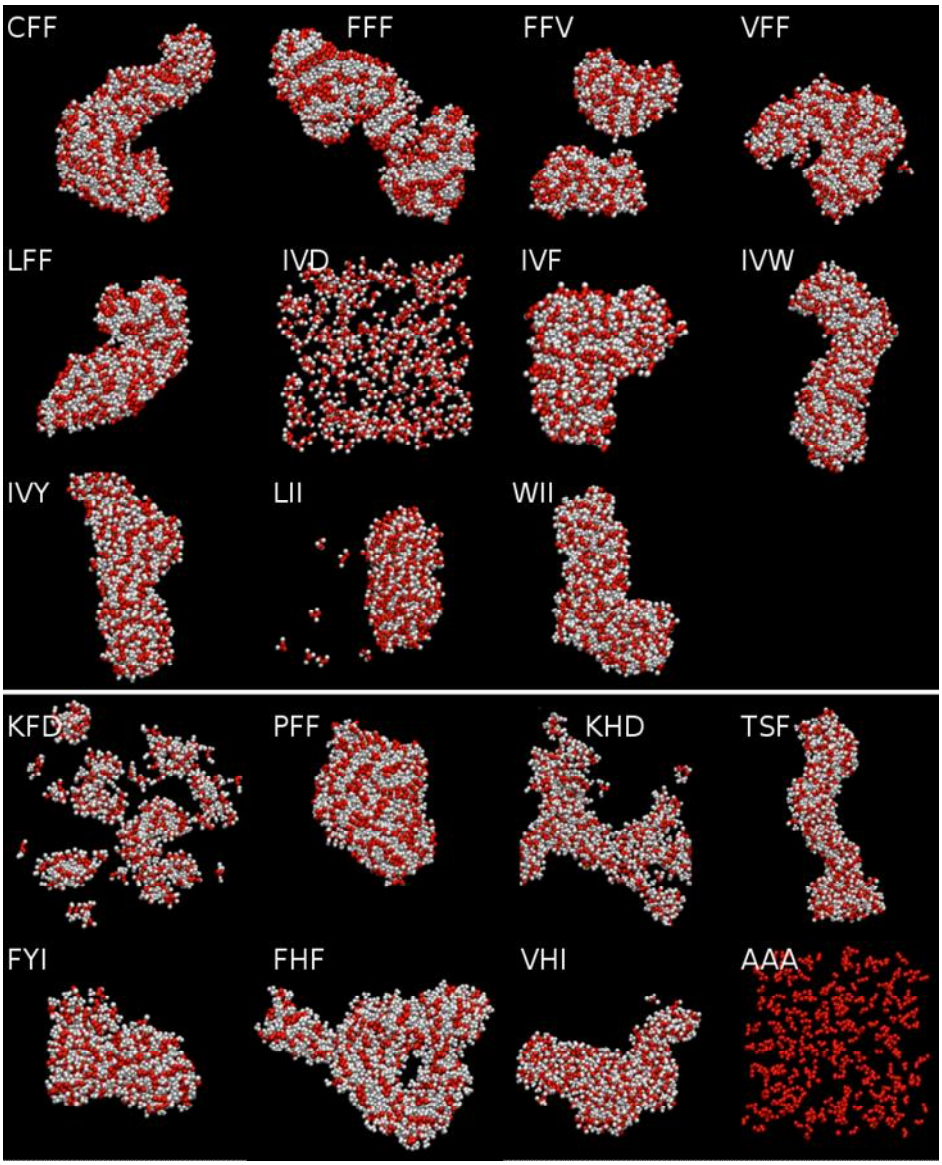


Fig. 5.12. Configurations of all peptides in Table 5.4 after a 400 ns simulation in polarized coarse-grain water. Top half: tripeptides experimentally studied in literature. Bottom half: interesting cases indicated in this work. Red, backbone beads. White, side chain beads.

Next, several interesting candidates for peptide nanostructures were predicted from our hydrophobicity corrected AP_H scores. The bottom half of Fig. 5.12 shows the structures for the tripeptides selected in Table 5.4. Tripeptides KFD and KHD showed aggregation of peptides, but seem to have a low propensity for forming stable extended aggregates, as indicated by their low AP scores, even after 400 ns (1.70 and 1.92, respectively). TSF and FHF form fibrous 1D nanostructures, which are predicted to develop into elongated fibres in experiments. PFF has an exceptionally high AP score (2.38 after 50 ns and 3.51 after 400 ns) and is therefore expected to be a strong candidate for a self-assembling system that has not been reported yet. FYI and VHI exhibit amorphous aggregates, and are envisioned to be useful in testing the limitations of our screening method when comparing with future experimental results, by providing examples of peptides with intermediate scoring results.

From the experimentally studied systems, FFF was taken as a subject of further investigation, analogous with our study of the FF dipeptide in section 5.4.3. Its morphology after 400 ns resembles the bilayer structure initially observed for FF (see Fig. 5.4), followed by formation of 1D structures and, to a lesser extent, spheres. However, no water core was observed in the self-assembled nanostructure (although <10 water beads reside within the structure), consistent with the report by Tamamis *et al*., who did not observe a hollow core for FFF assemblies.[27] The results of an extended simulation of 800 ns containing 1200 peptides in a 24 x 24 x 24 nm box (see Fig. 5.13) shows similar results to the 400 ns simulation.

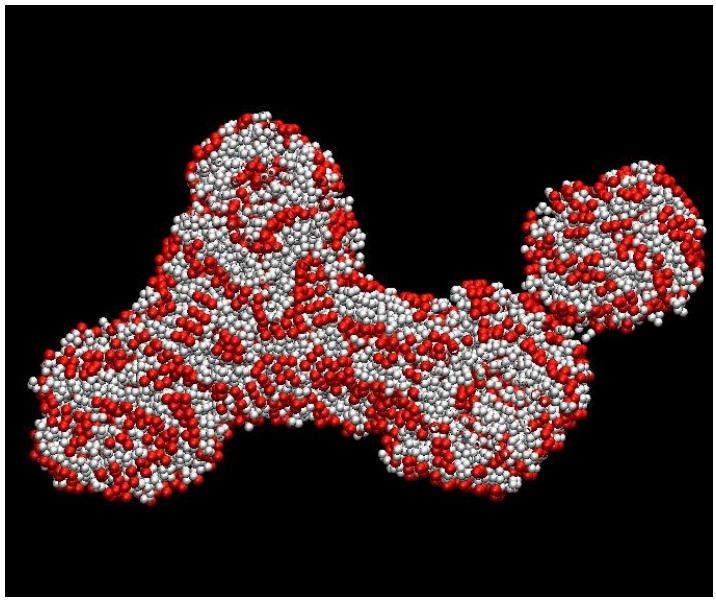


Fig. 5.13. Final shapshot of a 0.8 μs simulation of 1200 FFF peptides. Red, backbone beads. White, side chain beads.

Equivalent to the discussion on dipeptides with Fig. 5.5, Fig. 5.14 shows the side chain-backbone-backbone-side dihedral angles. The figure shows a similar behaviour to that for the FF dipeptide, where even after 1 ns most of the groups have rotated to accommodate intramolecular π-π stacking of the phenylalanine rings. This deviation from the typical β-sheet dihedral angles that may be expected for longer, hydrophobic peptides at minimization is consistent with the results from Tamamis *et al*.[27] They suggest phenylalanine side chains are pointing to the same side of their connecting amide bond in FF, although they also find that in the FFF tripeptide the third phenylalanine can point to the other side of the connecting amide bond. Their peptides, however, were not confined in an ordered nanostructure, due to the small size of the simulation box (12 peptides). Moreover, in the structural analysis performed by Han *et al*. on Boc-protected FFF, all phenylalanine side chains point to the same side of the peptide.[60] They used quantum mechanical calculations to show that this peptide adopts a turn-like conformation, leading to the formation of spherical nanostructures. Although these structures were obtained by evaporation of the ethanol solvent, the agreement is compelling.

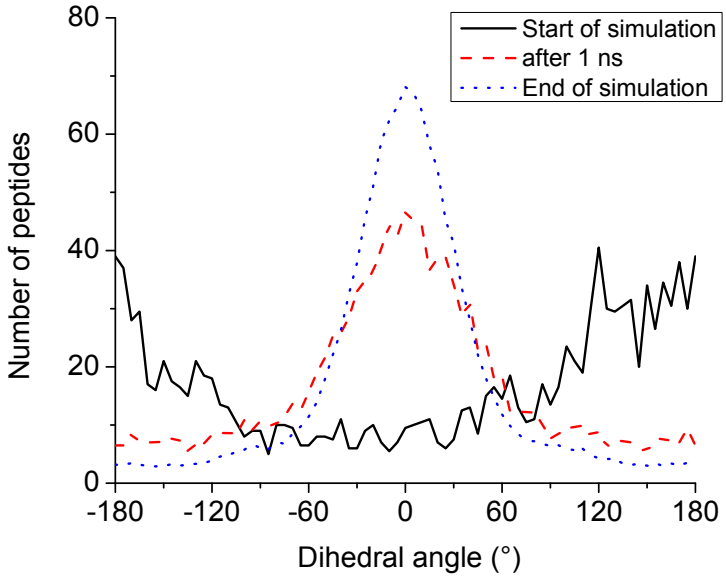


Fig. 5.14. Sidechain-backbone-backbone-sidechain dihedral angle distribution (binwidth of 5°) at the start of the simulation (0 μs, blue), after 1 ns (green) and at the end of a 0.8 μs simulation (red) for the FFF tripeptide.

5.6 Directions for future research

The results on di- and tripeptides show that the CG methodology presented here is suitable for screening for new biomaterials. For dipeptides, good agreement between experimental and theoretical propensity to aggregate was found, while extended simulations can even reproduce salient morphological features of the nanostructures formed. Naturally, the first step is to test if the tripeptides that are predicted to aggregate do actually form nanostructures when tested in the laboratory. Currently, synthesis of KHD, PFF, KFD, TSF and FYI is ongoing to examine this. A combination of IR spectroscopy and AFM or TEM will be employed to test for aggregation of the selected peptides.

It can be desirable to extend this approach to even larger peptides, or combinations of short peptides (c.f. core/shell nanostructures discussed in Ch. 4), to introduce more

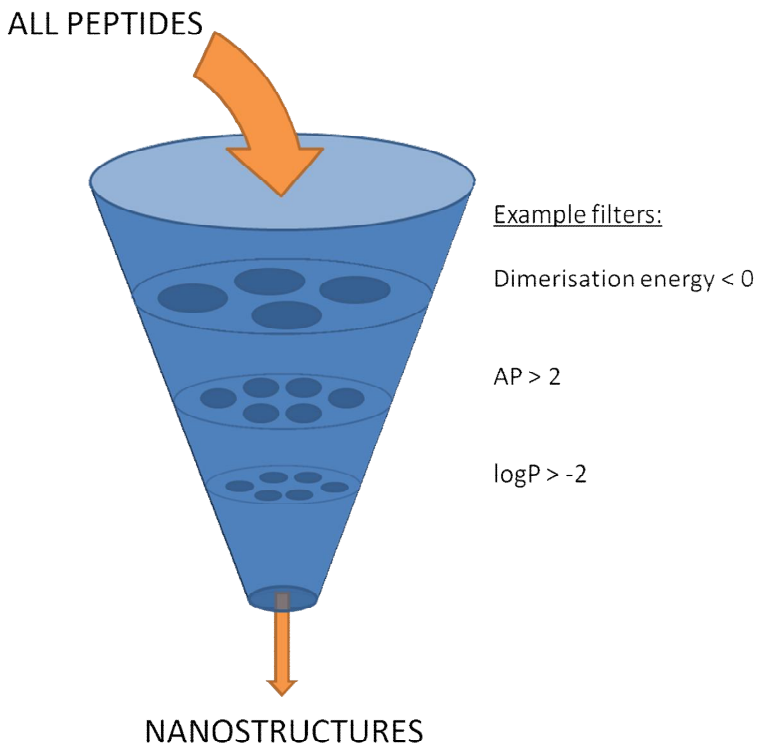


Fig. 5.15 Schematic representation of screening for nanostructures.

complexity, functional handles, particular structural and morphological motifs or to include specific epitopes for biology-related research. It will become increasingly difficult to test all possible combinations of amino acids using just the AP as described here. New 'filters' will have to be introduced to limit the amount of computations necessary to obtain the most suitable candidates for self-assembling nanostructures (see Fig. 5.15). It was shown in this chapter that the AP_H score can provide such a filter for selecting the peptides with the right balance between hydrophobicity and hydrophilicity. As another example, MD simulations can be performed to calculate the binding energy of a dimer with negligible computational cost. If this number is too positive (larger than a certain cutoff, e.g. > 0), the peptide could be disregarded as a candidate for self-assembly, as aggregation of is energetically not favourable enough. Another potential screening filter can be provided by looking at the structure of the produced nanostructures. Fuhrmans and Marrink have recently used Minkowski functionals to

automatically extract nanostructure morphology from a set of atomic coordinates and applied this to structures produced in the MARTINI force field.[61] These functionals are sensitive to the local curvature of nanostructures and could therefore be particularly useful when the shape of the nanostructure (sphere, fibre, tube, etc.) is of importance. For example, the FF and FFF results show that leading to the nanotube formation, there is not only a critical aggregate size before self-assembly can be observed, but these peptides also undergo specific morphological transitions. That is, initially bilayer sheets are formed, which only form vesicles (the precursor stage to tube formation) when the sheets have reached a sufficient size to fold on themselves. Atomistic simulations will not always be of sufficient size to reach and surpass these critical size regimes, due to the cost associated with such large systems at this level of theory. Thus, although detailed information may be lost, coarse-grained models are very useful tools in the discovery of new self-assembling structures, since they can easily be employed to screen a large amount of different molecules for their aggregation properties.

5.7 Conclusion

The self-assembly of biomolecules often occurs on a timescale that exceeds the microsecond region, and as such atomistic methods as discussed in Ch. 4 are too computationally expensive to model multiple large systems. In this chapter, the usage of the MARTINI coarse-grain MD force field has been investigated in predicting the aggregation ability of all 400 dipeptides and 8000 tripeptides of the 20 gene-encoded amino acids. A scoring method based on solvent-accessible surface area has been proposed and good agreement was found comparing the predicted aggregation propensity to experimental results from literature for various di- and tripeptides, including diphenylalanine and triphenylalanine.

For the case of dipeptides, studied initially, it was found that the AP score after 400 ns provides a convenient, affordable and apparently robust screening model for the ability of dipeptides to aggregate and hence their potential for self-assembly. It is apparent from

extended simulations on diphenylalanine that a structural equilibrium has not been reached after the 400 ns used in calculating the AP score. However, this is not the goal of the initial screening protocol, which is intended to identify systems that can potentially self-assemble and as such warrant further investigation. Additionally, the produced score did not significantly change after 400 ns and is therefore a good measure for a peptide's aggregation propensity even when structural features need more time to converge.

In all MD simulations care has to be taken when considering simulation time and system size, since some effects may be time or concentration-dependent. Nonetheless, when extended simulations were employed, good agreement was found between the morphological architecture predicted and that in experimental reports. This strongly suggests that for more accurate structural information and insights into the mechanism of formation of those systems identified as having a potential for self-assembly, longer and larger simulations can be performed to allow the system to converge.

Diphenylalanine (FF) was studied in more detail as experimental structural information is available in literature. The simulation results accurately show the formation of hollow nanotube, in agreement with experiment. Considering the coarse-grain nature of the simulation, pseudo-dihedral angles and tube dimensions compared to data from X-ray diffraction and atomistic simulations.

Secondly, the first step towards larger peptides was set by studying all 8000 tripeptides in a similar fashion. Due to the rapid increase in the number of possible combinations of amino acids a slightly different approach was adopted. An initial screening phase of 50 ns was performed, followed by a full 400 ns simulation using the polarisable MARTINI parameters for selected cases. All peptides with a high AP were selected, similar to the dipeptide approach, but a new scoring system was also introduced to identify high-scoring tripeptides with hydrophilic residues, important for the solubility of the assembling molecules. This new score, termed AP_H, contains a weighted contribution from the Wimley-White hydrophobicity value (logP) of the tripeptides. This method

allows for the determination of peptides with a desired balance between hydrophilicity and aggregation potential, as only a very weak correlation is observed between logP and AP. The top 400 peptides identified by this method were also selected for extended simulation.

For the tripeptides that have been reported to form self-assembling nanostructures in literature, aggregation was correctly observed in our simulations. The exception to this is IVD, which did not show any aggregation on a 400 ns time scale. This is attributed to the necessary acetyl-capping in experiments. FFF was studied more extensively and showed structural features resembling data from more accurate methods (atomistic MD and semi-empirical QM simulations). In contrast with FF, no hollow water core was observed, which is in agreement with literature. Furthermore, several effects of amino acid order in a short tripeptide sequence were observed: aggregation was preferentially observed for tripeptides with hydrophobic amino acids (especially F, W and Y) in both position 2 and 3, while positively and negatively charged amino acids were most favourable in positions 1 and 3, respectively.

Future work will focus on validation of the CG approach for tripeptides by comparing with experimental results on the peptides that have a high AP_H score. In conclusion, CG models are a very useful tool in the discovery of new biomaterials because they can easily be employed to screen a large amount of different molecules for their self-assembly properties. This can be aided further by the development of additional filters or protocols to screen for the self-assembly of larger and larger biomolecules. The results of the initial screening and subsequent production simulations for systems of interest can then form the basis for more detailed atomistic and experimental studies.

5.6 References

(1) Frederix, P. W. J. M.; Ulijn, R. V.; Hunt, N. T.; Tuttle, T. *J. Phys. Chem. Lett.* **2011**, *2*, 2380–2384.

(2) Zelzer, M.; Ulijn, R. V. *Chem. Soc. Rev.* **2010**, *39*, 3351–3357.

(3) Hartgerink, J. D.; Benlash, E.; Stupp, S. L. *Science* **2001**, *294*, 1684–1688.

(4) Ghadiri, M. R.; Granja, J. R.; Milligan, R. A.; McRee, D. E.; Khazanovich, N. *Nature* **1993**, *366*, 324–327.

(5) Reches, M.; Gazit, E. *Phys. Biol.* **2006**, *3*, S10–S19.

(6) 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*, 37–41.

(7) Zhang, S.; Holmes, T.; Lockshin, C.; Rich, A. *Proc. Natl. Acad. Sci. U. S. A.* **1993**, *90*, 3334 –3338.

(8) 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*, 960–966.

(9) Silva, G. A.; Czeisler, C.; Niece, K. L.; Beniash, E.; Harrington, D. A.; Kessler, J. A.; Stupp, S. I. *Science* **2004**, *303*, 1352–1355.

(10) Kumar, P.; Pillay, V.; Modi, G.; Choonara, Y. E.; du Toit, L. C.; Naidoo, D. *Recent Patents Drug Deliv. Formul.* **2011**, *5*, 24–51.

(11) Reches, M.; Gazit, E. *Science* **2003**, *300*, 625–627.

(12) Yan, X.; Zhu, P.; Li, J. *Chem. Soc. Rev.* **2010**, *39*, 1877.

(13) Kim, J.; Han, T. H.; Kim, Y.-I.; Park, J. S.; Choi, J.; Churchill, D. G.; Kim, S. O.; Ihee, H. *Adv. Mater.* **2010**, *22*, 583–587.

(14) Han, T. H.; Oh, J. K.; Lee, G.-J.; Pyun, S.-I.; Kim, S. O. *Colloids Surf. B Biointerfaces* **2010**, *79*, 440–445.

(15) Su, Y.; Yan, X.; Wang, A.; Fei, J.; Cui, Y.; He, Q.; Li, J. *J. Mater. Chem.* **2010**, *20*, 6734.

(16) Sanchez-de Groot, N.; Parella, T.; Aviles, F.; Vendrell, J.; Ventura, S. *Biophys. J.* **2007**, *92*, 1732–1741.

(17) Görbitz, C. H. *Chem. Commun.* **2006**, 2332–2334.

(18) Görbitz, C. H. *Chem. Eur. J.* **2007**, *13*, 1022–1031.

(19) Chelli, R.; Gervasio, F. L.; Procacci, P.; Schettino, V. *J. Am. Chem. Soc.* **2002**, *124*, 6133–6143.

(20) Hall, C. K.; Wagoner, V. A. In *Amyloid, Prions, and Other Protein Aggregates, Part B*; Academic Press, 2006; Vol. 412, pp. 338–365.

(21) Meli, M.; Morra, G.; Colombo, G. *Biophys. J.* **2008**, *94*, 4414–4426.

(22) Mousseau, N.; Derreumaux, P. *Acc. Chem. Res.* **2005**, *38*, 885–891.

(23) Zanuy, D.; Hamley, I. W.; Alemán, C. *J. Phys. Chem. B* **2011**, *115*, 8937–8946.

(24) Wu, C.; Shea, J.-E. *Curr. Opin. Struct. Biol.* **2011**, *21*, 209–220.

(25) Li, M. S.; Co, N. T.; Reddy, G.; Hu, C.-K.; Straub, J. E.; Thirumalai, D. *Phys. Rev. Lett.* **2010**, *105*, 218101.

(26) Thirumalai, D.; Klimov, D.; Dima, R. *Curr. Opin. Struct. Biol.* **2003**, *13*, 146–159.

(27) Tamamis, P.; Adler-Abramovich, L.; Reches, M.; Marshall, K.; Sikorski, P.; Serpell, L.; Gazit, E.; Archontis, G. *Biophys. J.* **2009**, *96*, 5020–5029.

(28) Ash, W. L.; Zlomislic, M. R.; Oloo, E. O.; Tieleman, D. P. *Biochim. Biophys. Acta BBA - Biomembr.* **2004**, *1666*, 158–189.

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

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

(31) Marrink, S. J.; Risselada, H. J.; Yefimov, S.; Tieleman, D. P.; de Vries, A. H. *J. Phys. Chem. B* **2007**, *111*, 7812–7824.

(32) Monticelli, L.; Kandasamy, S. K.; Periole, X.; Larson, R. G.; Tieleman, D. P.; Marrink, S.-J. *J. Chem. Theory Comput.* **2008**, *4*, 819–834.

(33) Barrow, C. J.; Small, D. H. *Abeta Peptide and Alzheimer's Disease: Celebrating a Century of Research*; Springer, 2007.

(34) Gorlero, M.; Wieczorek, R.; Adamala, K.; Giorgi, A.; Schininà, M. E.; Stano, P.; Luisi, P. L. *FEBS Lett.* **2009**, *583*, 153–156.

(35) Reches, M.; Gazit, E. *Nano Lett.* **2004**, *4*, 581–585.

(36) Marchesan, S.; Waddington, L.; Easton, C. D.; Winkler, D. A.; Goodall, L.; Forsythe, J.; Hartley, P. G. *Nanoscale* **2012**, *4*, 6752–6760.

(37) Marchesan, S.; Easton, C. D.; Kushkaki, F.; Waddington, L.; Hartley, P. G. *Chem. Commun.* **2012**, *48*, 2195–2197.

(38) 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*, 1361–1366.

(39) Lakshmanan, A.; Hauser, C. A. E. *Int. J. Mol. Sci.* **2011**, *12*, 5736–5746.

(40) Subbalakshmi, C.; Manorama, S. V.; Nagaraj, R. *J. Pept. Sci.* **2012**, *18*, 283–292.

(41) Marrink, S. J.; Tieleman, D. P. *Chem. Soc. Rev.* **2013**, *42*, 6801–6822.

(42) Singh, G.; Tieleman, D. P. *J. Chem. Theory Comput.* **2011**, *7*, 2316–2324.

(43) De Jong, D. H.; Singh, G.; Bennett, W. F. D.; Arnarez, C.; Wassenaar, T. A.; Schäfer, L. V.; Periole, X.; Tieleman, D. P.; Marrink, S. J. *J. Chem. Theory Comput.* **2013**, *9*, 687–697.

(44) Wimley, W. C.; Creamer, T. P.; White, S. H. *Biochemistry (Mosc.)* **1996**, *35*, 5109–5124.

(45) White, S. H.; Wimley, W. C. *Biochim. Biophys. Acta BBA - Rev. Biomembr.* **1998**, *1376*, 339–352.

(46) Hess, B.; Kutzner, C.; van der Spoel, D.; Lindahl, E. *J. Chem. Theory Comput.* **2008**, *4*, 435–447.

(47) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; DiNola, A.; Haak, J. R. *J. Chem. Phys.* **1984**, *81*, 3684.

(48) Hess, B. *J. Chem. Theory Comput.* **2008**, *4*, 116–122.

(49) Yesylevskyy, S. O.; Schäfer, L. V.; Sengupta, D.; Marrink, S. J. *PLoS Comput Biol* **2010**, *6*, e1000810.

(50) Marrink, S. J.; de Vries, A. H.; Mark, A. E. *J. Phys. Chem. B* **2004**, *108*, 750–760.

(51) Humphrey, W.; Dalke, A.; Schulten, K. *J. Mol. Graph.* **1996**, *14*, 33–38.

(52) Guilbaud, J.-B.; Vey, E.; Boothroyd, S.; Smith, A. M.; Ulijn, R. V.; Saiani, A.; Miller, A. F. *Langmuir* **2010**, *26*, 11297–11303.

(53) Yan, X.; Cui, Y.; He, Q.; Wang, K.; Li, J.; Mu, W.; Wang, B.; Ou-yang, Z. *Chem. Eur. J.* **2008**, *14*, 5974–5980.

(54) Amdursky, N.; Molotskii, M.; Gazit, E.; Rosenman, G. *J. Am. Chem. Soc.* **2010**, *132*, 15632–15636.

(55) Leeson, P. D.; Springthorpe, B. *Nat. Rev. Drug Discov.* **2007**, *6*, 881–890.

(56) Eisenberg, D.; Weiss, R. M.; Terwilliger, T. C.; Wilcox, W. *Faraday Symp. Chem. Soc.* **1982**, *17*, 109–120.

(57) Godfrey, P.; Shipley, R.; Gryz, J. *VLDB J.* **2007**, *16*, 5–28.

(58) Brack, A.; Orgel, L. E. *Nature* **1975**, *256*, 383–387.

(59) Bellesia, G.; Shea, J.-E. *Biophys. J.* **2009**, *96*, 875–886.

(60) Han, T. H.; Ok, T.; Kim, J.; Shin, D. O.; Ihee, H.; Lee, H.-S.; Kim, S. O. *Small* **2010**, *6*, 945–951.

(61) Fuhrmans, M.; Marrink, S.-J. *J. Mol. Model.* **2011**, *17*, 1755–1766.

Chapter 6: Applications of biomaterials in energy research: Encapsulation of [FeFe]-Hydrogenase Model Compounds in Peptide Hydrogels*

* Parts of this chapter have been published as Frederix, P. W. J. M.; Kania, R.; Wright, J. A.; Lamprou, D. A.; Ulijn, R. V.; Pickett, C. J.; Hunt, N. T. *Encapsulating [FeFe]-hydrogenase model compounds in peptide hydrogels dramatically modifies stability and* photochemistry, Dalton Trans. **2012**, *41*, 13112–13119.[1]

Declaration of contributions to published articles: The author was responsible for sample preparation, FTIR (incl. gel melting), TRIR, data analysis and writing the manuscript. Solvent samples were provided by Dr. Rafal Kania. AFM was performed by Dr. Dimitrios Lamprou and the hydrogenase compound was provided by Dr. Joseph Wright.

6.1 Abstract

Hydrogenase active site mimics are a promising route towards the environmentally friendly production of hydrogen. Here, the encapsulation of existing mimics in Fmoc-dipeptide hydrogels has been studied, as well as the effects of the gel environment on the photochemistry of the mimic using Fourier Transform Infrared spectroscopy (FTIR) and ultrafast time-resolved infrared (TRIR) spectroscopy.

[FeFe]-hydrogenase model compound (μ-$S(CH_2)_3S)Fe_2(CO)_4(PMe_3)_2$ [**1**] has been encapsulated in a Low Molecular Weight (LMW) hydrogelator (Fmoc-Leu-Leu). FTIR, gel melting and TRIR experiments reveal significant contrasts in chemical environment and photochemistry between the encapsulated molecules and solution phase systems. Specifically, the gel provides a more rigid hydrogen bonding environment, which restricts isomerisation following photolysis while imparting significant increases in stability relative to a similarly aqueous solution.

Furthermore, we have studied the generality of this approach by performing similar experiments on **1** in a more hydrophilic hydrogel (Fmoc-Tyr-Ser). The stability of the mimics in an aqueous environment again benefitted from encapsulation, with some surprising effects on the photochemistry of the compound: FTIR and TRIR experiments indicated a more flexible H-bonding environment, which stabilized the polar isomer of **1**, in contrast with the more hydrophobic Fmoc-LL gel. Since understanding and ultimately controlling the mechanistic role of ligands near Fe centres is likely to be crucial in exploiting artificial hydrogenases, these gels may offer a new option for future materials design involving catalysts.

Finally, the photochemistry of a bridging hydride [FeFe]-hydrogenase mimic in the presence of photosensitizer $Ru(bpy)_3^{2+}$ was studied. It was shown that $Ru(bpy)_3Cl_2$ acts as a base on the mimic in acetonitrile and irradiation at 532 nm further promotes

deprotonation via an intermediate. These results are expected to help towards introducing the full photocatalytic cycle of hydrogen evolution within the gel phase.

6.2 Introduction and objectives

6.2.1 Hydrogen production by hydrogenase enzyme mimics

The hydrogenase (H_2-ase) enzymes have attracted a significant amount of attention as a result of their ability to catalyze the reversible conversion of protons and electrons to H_2.[2] The potential use of active site mimics as new catalysts in hydrogen fuel cells is particularly attractive as it offers the possibility of replacing the platinum-based systems currently in use with more economical materials. For the development of a fuel cell based on hydrogenase mimics, a number of requirements have to be met. Obviously, the anode and cathode components containing the active catalyst are crucial. These materials have to convert the inputs to H_2 gas using low overpotentials (difference between thermodynamic and experimental voltages required, which results in energy loss in the form of heat[3]). Ideally, H_2 will be produced by water electrolysis and therefore only needs clean water (source of protons) and electricity as continuous inputs.

On the other hand, semi-permeable membranes and suitable electrolytes for diffusing protons from anode to cathode are important in a typical fuel cell design. Multiple types of fuel cells (i.e. with various active compounds, membranes and electrolytes) have been identified as key areas of research and development by the UK and USA governments.[4,5] However, even when taking into account that water-splitting systems will be further commercialized, it is envisioned that the required production of H_2 cannot be sustained with current methods and by 2030 about 50% of hydrogen gas will still be obtained from steam reformation of natural gas,[6] which would still put a stress on the supply of fossil fuels. This stresses the scope for the development of new production methods.

Inspired by nature, we have looked into the active site of hydrogenase enzymes, which are proteins that are naturally occurring in certain anaerobic bacteria.[7] They contain a diiron or nickel-iron active site (Fig. 6.1(a,b)) that plays the dominant role in the reversible reduction of protons to form H_2.[7] However, since synthesizing the entire 122

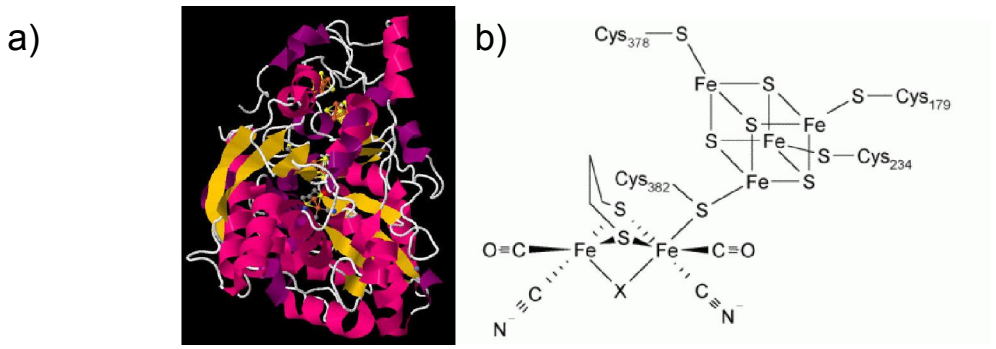


Fig. 6.1 a) Structure of the [FeFe]-hydrogenase enzyme from desulfovibrio desulfuricans, PDB 1HFE[13]. b) Active site of the enzyme in a) showing the [FeFe] center connected to a Fe_4S_4-cluster.

kDa enzyme complex (for the species with the highest turnover) in large quantities is not viable for commercial purposes, the active site is modelled by relatively simple metal carbonyl complexes. An active area of research is changing the ligands of the iron centre to control the stability of the reaction intermediates, tuning the catalytic pathway; and large libraries of active site mimics have been reported.[8–12]

It has been shown that several such mimics possess catalytic activity for hydrogen production,[12,14] however, their efficiency is significantly less than that of the natural enzyme and they retain the undesirable oxygen sensitivity observed in the natural system. An approach to solving these issues may be the incorporation of model compounds into an environment that, like the enzymatic active site pocket, can influence their catalytic effectiveness and stability, but which can also be synthetically tailored for the optimization of future applications, such as inclusion in the electrode of a fuel cell. Examples of such chemical and physical immobilization of [FeFe]-H_2-ase model systems have included covalent attachment to polyethylene glycol (PEG) beads[15] and oligopeptide chains,[16] and incorporation into the backbone[17,18] and side chains[19,20] of polymers, often with retention of catalytic activity. Of particular interest here, it has been shown that the accurate positioning of ligands around Fe-derived active site mimics proved to be crucial for improving the catalytic rate.[21] Similar observations have been

made in studies of closely related systems based upon Ni coordination chemistry. Recently, Dubois and co-workers surpassed the rate of catalysis of native hydrogenase enzymes (10^5 versus 10^4 molecules/s, respectively) by tuning the spatial arrangement of the pendant amines around a nickel centre.[22] It has also been shown that including peptide moieties into the outer coordination sphere of a Ni-core hydrogen formation catalyst improved the activity of the complex dramatically.[23]

A route to immobilization yet to be explored is the use of peptide-based, low-molecular weight hydrogelators.[24] As discussed in the previous chapters, these form 3D networks by non-covalent interactions and produce self-supporting gels at low concentrations (~10 mmol L^{-1}) in water.[25–28] They have shown biocompatibility and conductivity[29] as well as being generally low-cost and highly tuneable in terms of structure, chemical properties and formation routes. From a hydrogenase perspective, the key advantage of these gels lies in the combination of a peptide-based scaffold with a 'rigid' aqueous environment. As significant potential exists to modify the peptide environment created near the encapsulated compound, they may be well suited for artificial H_2-ase exploitation. Indeed, several studies have shown that incorporation of enzymes[30–33] and enzyme mimics[34,35] in peptide-based hydrogels increases their stability or activity. Moreover, providing an aqueous environment for generally water-insoluble H_2-ase mimics is advantageous because only relatively low overpotentials at mM acid concentrations were needed for the reduction of protons in water[36,37] and combining water-splitting systems with H_2 production is one of the future goals in energy research, as discussed.

A number of studies have addressed the reactivity and dynamics of the artificial H_2-ases in organic solutions, including spectroelectrochemistry,[38] reaction kinetics studies[39] and photolytic methods.[40–42] Wright et al. showed that soaking a cationic electropolymerized poly(pyrrole) film in a solution containing $(\mu\text{-pdt})Fe_2(CO)_4(CN)_2^{2-}$ (pdt = propanedithiolate) affords the immobilization of the active site mimic into the polymer matrix.[43] This approach improved the stability of protonated products upon reaction with HCl vapour dramatically. However, the role of an encapsulating material on the

photochemistry or ultrafast dynamics of these species at room temperature has not been studied, although the effects of UV irradiation of (μ-S_2)-$Fe_2(CO)_6$ in low temperature Nujol matrices have been observed.[44] Additionally, we can study the effect of encapsulating small molecules on the gel matrix, to provide us with further information on the IR spectroscopy of peptide nanostructures as discussed in the previous chapters. The background on ultrafast infrared spectroscopy on hydrogenase enzyme active site mimics is provided in the next section.

The first part of this chapter is focussed on the H_2-ase active site mimic (μ-pdt)$Fe_2(CO)_4(PMe_3)_2$ (**1**, Fig. 6.2(a)) and the study of it using infrared absorption and ultrafast time-resolved infrared (TRIR) spectroscopy. These techniques were employed to observe the effects of non-covalently trapping **1** in an Fmoc-LL (**2,** Fig. 6.2(b)) dipeptide hydrogel (Fmoc = 9-fluorenylmethoxycarbonyl) on the chemical environment and photochemistry of the mimic. Compound **1** has been shown be a good candidate for study in the H_2 production cycle by several research groups, due to an electron-rich metal centre that promotes facile protonation.[36,39,45–47] As well as being of interest in terms of possible materials development, such a study also has relevance to the chemistry of model systems and to the active site of the enzyme itself, which features significant interaction of the diiron cluster with peptide side chains in the hydrophobic active site pocket. Additionally, to investigate the possibilities of our method to optimise the environment of the mimic further, we have performed test experiments regarding

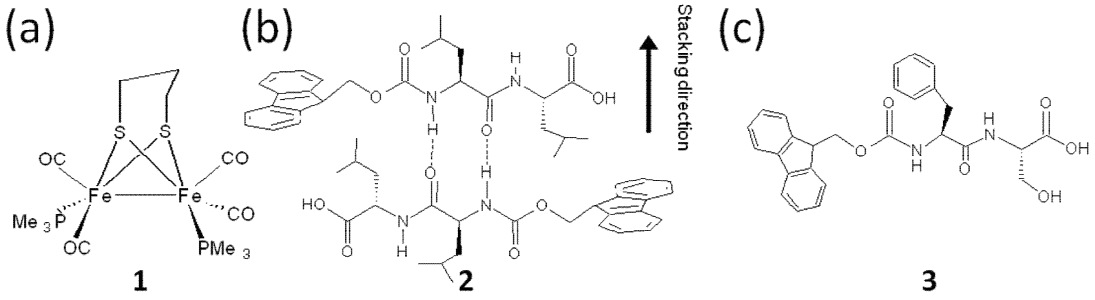


*Fig. 6.2. (a) . (a) Schematic diagram of the structure of (μ-pdt)$Fe_2(CO)_4(PMe_3)_2$ **1** in the dibasal configuration. (b) Chemical structure of two stacked gelator molecules in an anti-parallel configuration, Fmoc-LL, **2**. (c) Chemical structure of Fmoc-YS, **3**.*

encapsulation of **1** in an Fmoc-YS (**3**, Fig. 6.2(c)) hydrogel, which features more hydrophilic side chains. This is discussed in section 6.4.4.

Eventually, the development of photocatalytic H_2 production by incorporating these water splitting systems with light-harvesting complexes can diminish the requirement for indirect electricity supply to hydrogenase fuel cells. This would allow for more small-scale, mobile energy generation, crucial for transport applications. To achieve this, Na *et al*. showed that [FeFe]-hydrogenase mimics can be combined with photosensitizers such as ruthenium tris-2,2'-bipyridine ($Ru(bpy)_3^{2+}$) and a proton/electron donor such as ascorbic acid to evolve H_2 in solution.[48] It has been well documented that $Ru(bpy)_3$-compounds can transfer electrons to a variety of systems upon excitation with visible light, including hydrogenases (see ref. 48 and references therein). In the presence of $Ru(bpy)_3^{2+}$ and $\lambda \approx 400$ nm light the following reactions can occur:[49]

$$Ru(bpy)_3^{2+} \xrightarrow{h\nu} Ru(bpy)_3^{2+*} \quad (6.1)$$

$$Ru(bpy)_3^{2+*} + FeFe^{n+} \longrightarrow Ru(bpy)_3^{3+} + FeFe^{(n-1)+} \quad (6.2)$$

Moreover, this photosensitizer has been incorporated in gel-phase systems in a number of studies. For example, Kurimura *et al*. showed retention of the $Ru(bpy)_3^{2+}$ redox properties in a gelatine hydrogel,[50] while Perine showed excellent diffusion properties of electrolytes for its redox chemistry in a polysaccharide hydrogel.[51] It was even shown that $Ru(bpy)_3$-N,N-dimethylanilline can act as an initiator of a poly(N-isopropylacrylamide) (PNIPAM) hydrogel, after which the $Ru(bpy)_3^{2+}$ could be used as a fluorescent probe for the microscopic hydrogel environment.[52] Therefore, a fully gel-encapsulated system involving a photosensitizer and a hydrogenase-like catalyst in a proton-rich environment is a promising candidate for H_2 evolution. E.g. ascorbic acid can act as both a sacrificial proton and electron donor to recycle the protonated catalyst and the photosensitizer, respectively. Fig. 6.3 summarises the general idea for the photocatalytic production of H_2.

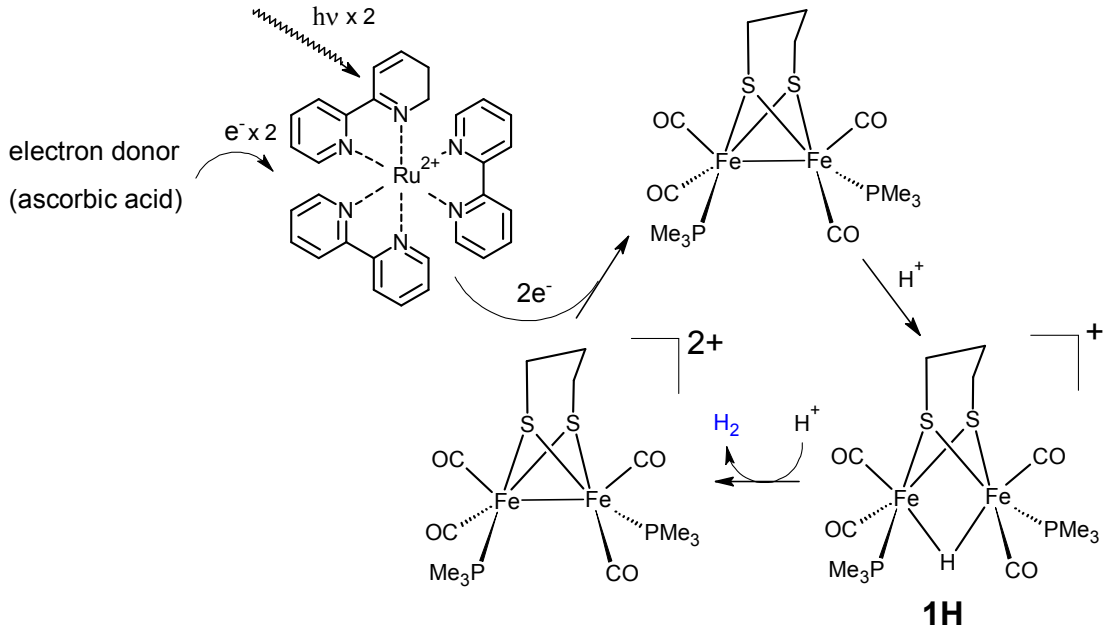


*Fig. 6.3. Proposed mechanism for photocatalytic evolution of hydrogen using Ru(bpy)$_3^{2+}$ and the bridging hydride [FeFe] hydrogenase active site mimic **1H**.*

For these experiments the bridging-hydride compound studied by Jablonskytė *et al*., (μ-pdt,H)$Fe_2(CO)_4(PMe_3)_2^+$ or **1H** (see bottom right in Fig. 6.3) was used, because of its favourable redox potential towards reduction by the excited ruthenium compound.[53] Moreover, this type of paramagnetic compound is a likely intermediate in the evolution of H_2 as shown in the proposed catalytic mechanism in Fig. 6.3, based on the results of Na and Wang *et al*.[48,54] Therefore, it is interesting to study its photochemistry and stability. Jablonskytė and co-workers used spectroelectrochemistry to show catalytic activity of the system in the presence of a proton source (50 mM HBF_4 • 3 Et2O, 5 equivalents) in MeCN electrolyte, indicated by the disappearance of the FTIR bands of the parent complex **1H**, fully consistent with electrocatalysis involving the generation of a 37-electron radical and fast protic attack to yield dihydrogen and restore the parent hydride cation **1H**.

We have performed initial experiments testing the stability of **1H** under 532 nm and 355 nm radiation and studied the light-induced changes in its FTIR spectrum in the presence of $Ru(bpy)_3Cl_2$ in acetonitrile. Although the H_2 evolution reaction in the gel phase has not yet been performed here, these results provide crucial information for future work on the immobilisation and application of these systems for H_2 production.

6.2.2 Time-Resolved Infrared Spectroscopy

Time resolved infrared (TRIR) spectroscopy employs two (or more) ultrafast laser pulses in a pump-probe scheme. The development of ultrafast lasers (pulse duration picoseconds of shorter) over the last few decades has allowed the study of dynamic processes on the molecular scale (see, e.g., refs. 55,56). When both pump and probe beams have an IR wavelength, the vibrational dynamics of the ground and first vibrational excited state (v=1) can be studied (see Fig. 6.4). In UV_{pump}-IR_{probe} TRIR spectroscopy, employed here, the UV pump beam photons have enough energy to excite the molecule to an electronically excited state, which can lead to subsequent dissociation of a ligand or chemical group. The 2nd pulse, which is in the IR range and thus has a much lower photon energy, is fed through an optical delay line that uses the finite speed of light ($2.998 \cdot 10^8$ m/s ≈ 0.3 mm/ps) to arrive at the sample after a short time delay (typically 0 to hundreds of picoseconds). This pulse is able to detect a depletion of the population of the ground state molecules ('bleaches'), when the spectrum is subtracted by a "UV pump off" spectrum: fewer molecules are in the ground state than without UV excitation (see Fig. 6.4(b). Moreover, the presence of any transient species and excited state parent molecules with an IR signature can be probed.

In his chapter, the pump-probe spectroscopy of iron carbonyl complexes will be discussed. Therefore, a brief overview of their spectroscopy will be given here. Hypothetical potential energy surfaces for the ground and excited state of an iron

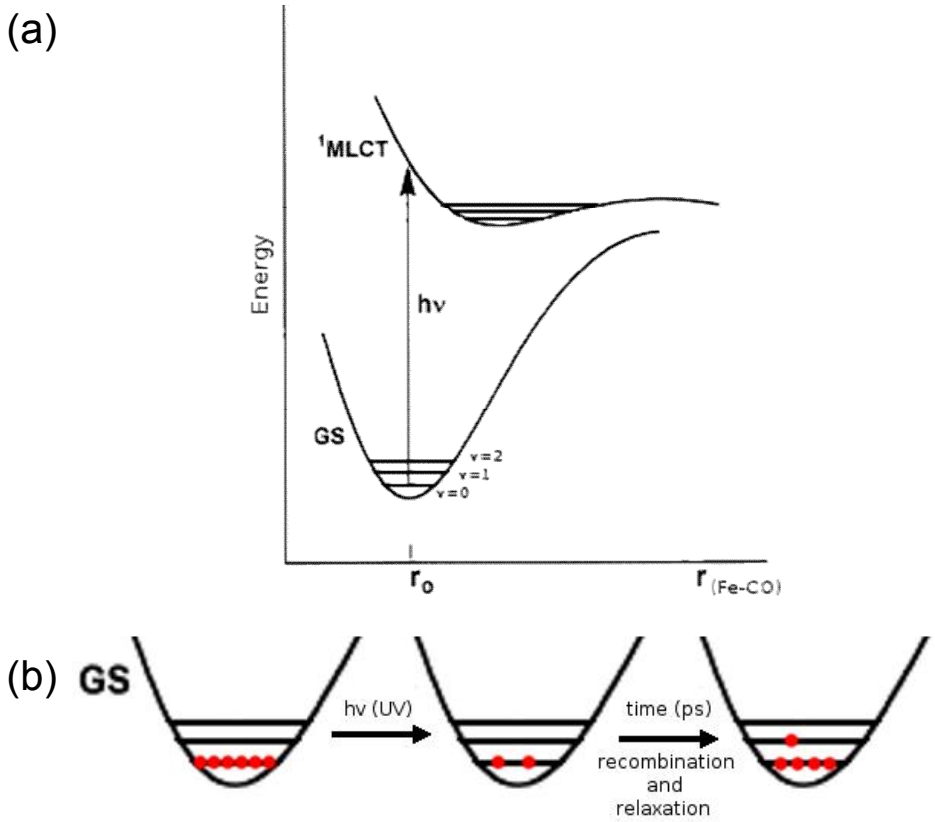


Fig. 6.4. (a) Hypothetical potential energy surfaces of a ground state iron carbonyl complex and its metal-to-ligand charge transfer (MLCT) state. Only three vibrational levels of the ground state and bound vibrational levels of the excited state are shown.(b) Population of the ground state before UV excitation (t < 0), directly after excitation (t=0) where the ground state is depopulated, and after a short time delay Δt, in which partial geminate recombination has created a vibrational hot parent molecule.

carbonyl complex are displayed in Fig. 6.4(a). The metal-to-ligand charge transfer (MLCT) state is an electronically excited state where the electrons from the metal-ligand bond have been localised on either the metal or the ligand (i.e. $M^{\cdot}\cdots L^{+}$ or $M^{+}\cdots L^{\cdot}$), giving the bond an ionic rather than a covalent character. This state is often dissociative or weakly bound as shown in Fig. 6.4(a), leading to quick dissociation of the ligand upon

UV excitation of the complex. The unsaturated metal centre left behind usually has a different IR signature than the ground state of the parent ion.

However, the dissociating ligand cannot always diffuse away from the parent ion in the condensed phase due to collisions with solvent. The ligand's 'solvent cage' therefore induces recombination of the two separated fragments. This process is called geminate recombination and generally occurs on a time scale of a few 100 ps for metal-ligand complexes. Geminate recombination or relaxation of the MLCT often leaves the molecule in a vibrationally 'hot' state, i.e. not the v=0 ground state. These molecules show up in TRIR spectra slightly redshifted with respect to the ground state absorption, due to the anharmonicity of the potential energy well, discussed in the section 4.2.1. Vibrational relaxation can then take place (generally on a time scale of tens of picoseconds) to reform the ground state parent complex. These two processes can generally be distinguished from the intensity decay of transient absorptions and bleaches as the time delay between pump and probe pulses increases. They are sensitive to the environment of the molecules and as such hold important clues to the structure and dynamics of the molecule under study.

The principle of data collection is schematically represented in Fig. 6.5. As discussed above, bleaches arise from depopulation of the ground state (middle of Fig. 6.4(b)), while transient absorptions can be a sign of new molecular species or vibrationally hot parent molecules.

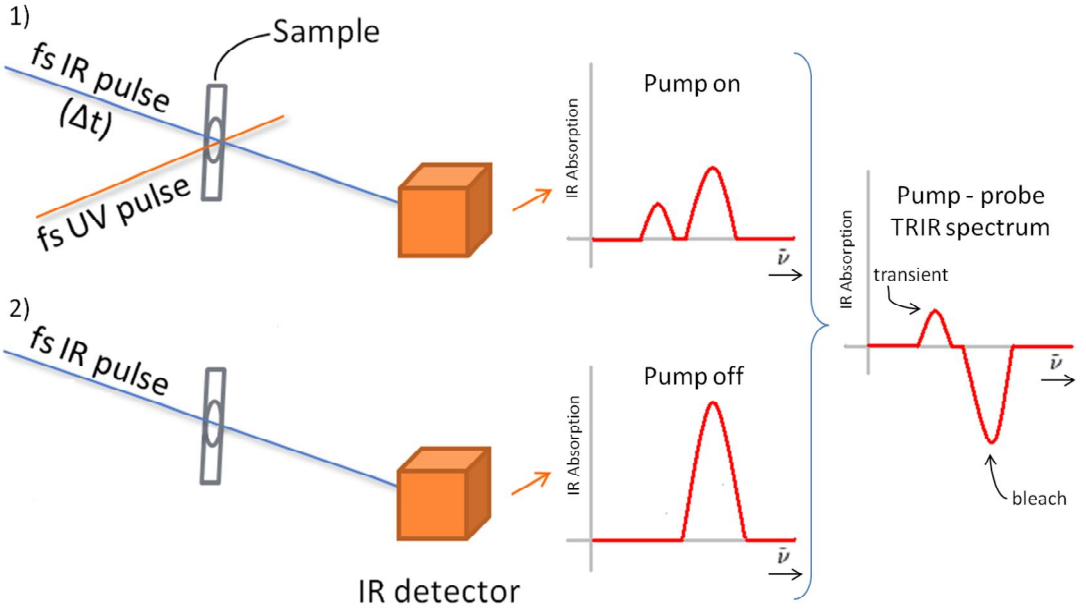


Fig. 6.5. Schematic of UV_{pump}-IR_{probe} TRIR experiments. An optical chopper in the UV path provides 'pump on' (1) and 'pump off' (2) spectra. The subtraction of the two provides the pump-probe spectrum.

6.3 Experimental details

6.3.1 Materials

Fmoc-LL was synthesized using standard peptide synthesis by Dr. S. Roy. Fmoc-YS was synthesized by Dr. L. S. Birchall. The hydrogenase mimic was supplied by Dr. J. A. Wright at the University of East Anglia. (μ-pdt)-$Fe_2(CO)_4(PMe_3)_2$ (pdt = propanedithiolate) was prepared according to established methods described in ref. 36 and 57, respectively. D_2O (99.9 atom% D), n-heptane, acetonitrile (MeCN), methanol (MeOH) were purchased from Sigma-Aldrich. Because of the oxygen sensitivity of hydrogenase compounds, all organic solvents were deoxygenated by bubbling nitrogen through them for 10 minutes prior to use and dried over CaH_2 pellets where appropriate.

6.3.2 Preparation of gels encapsulating hydrogenase mimics

Fmoc-LL (or Fmoc-YS) hydrogels were prepared for encapsulation of hydrogenase mimics as follows. All sample manipulations were performed under nitrogen atmosphere to avoid oxidation of the hydrogenase mimic. For a 1 mL hydrogel, 20 μmol of gelator molecule and 18 ± 2 mg (roughly 40 μmol, variable due to absence of a balance in the glove box) of the hydrogenase mimic were dissolved in 100 μL of deaerated methanol by sonication, as the mimics are generally water-insoluble. After addition of 900 μL 0.1 M sodium phosphate buffer in D_2O at pH 8 and vigorous shaking for 5 minutes the viscosity of the solution started to increase quickly. The sample was then left for 1 hour to rest during which time a self-supporting hydrogel formed. For comparison experiments, Fmoc-LL gels without mimic also contained 10% MeOH.

6.3.3 Time-resolved infrared spectroscopy

Ultrashort laser pulses (800 nm, 35 fs, 1 kHz) were produced via a regeneratively amplified Ti:Sapphire laser system. These pulses were then used to pump two white light seeded optical parametric amplifiers. 4^{th} harmonic generation of the output of one of these was used to produce UV pump pulses with a wavelength of 355 nm (resonant with the metal-ligand charge transfer band of iron carbonyls) with a power of roughly 3 mW at the sample position. The second optical parametric amplifier (OPA) was used to produce the mid IR probe pulses via Difference Frequency Generation (DFG) of the "signal" and "idler" beams. The time delay between the pump and probe pulses was controlled using an optical delay line while the respective polarizations of the pump and probe beams were set to the magic angle throughout. After overlapping the beams in the sample, the probe was coupled into a spectrometer equipped with a 64-channel HgCdTe array detector. Chopping the pump pulse train at half the laser repetition rate facilitated collection of pump on – pump off difference TRIR spectra. A schematic of our experimental setup is displayed in Figure 6.6, where the "UV path" and only the probe

arm of the 'black IR path' going into detector 1 were used for the experiments discussed in this chapter.

Samples were housed between two 2-mm CaF_2 windows separated by a 50 or 200 μm spacer for gel and liquid samples, respectively. For liquid samples, a flow cell arrangement was used to refresh the sample continuously between laser shots, but the hydrogels are too viscous to apply this technique. The cell holder was continuously rastered horizontally and vertically during the measurements to avoid sample degradation through prolonged UV irradiation. Monitoring of the sample quality via FTIR was performed periodically throughout the experiment.

Resulting pump-probe spectra were frequency-calibrated using the peaks in the FTIR spectrum as positions for the bleaches. Resulting spectra were analysed in IGOR Pro[58] using a global fitting routine. For all bleaches and transient peaks, the peak areas were simultaneously extracted for all time delays by fitting the spectra to Gaussian functions. Initially, all absorption frequencies and linewidths were kept constant, while peak amplitudes were being fitted. Subsequently, frequencies and linewidths were slowly released to allow for the slight frequency shifts or broadening that can occur with increasing pump-probe time delays, until a satisfactory fit was obtained. Peak areas were fitted to mono- or biexponential functions where possible to extract rate constants.

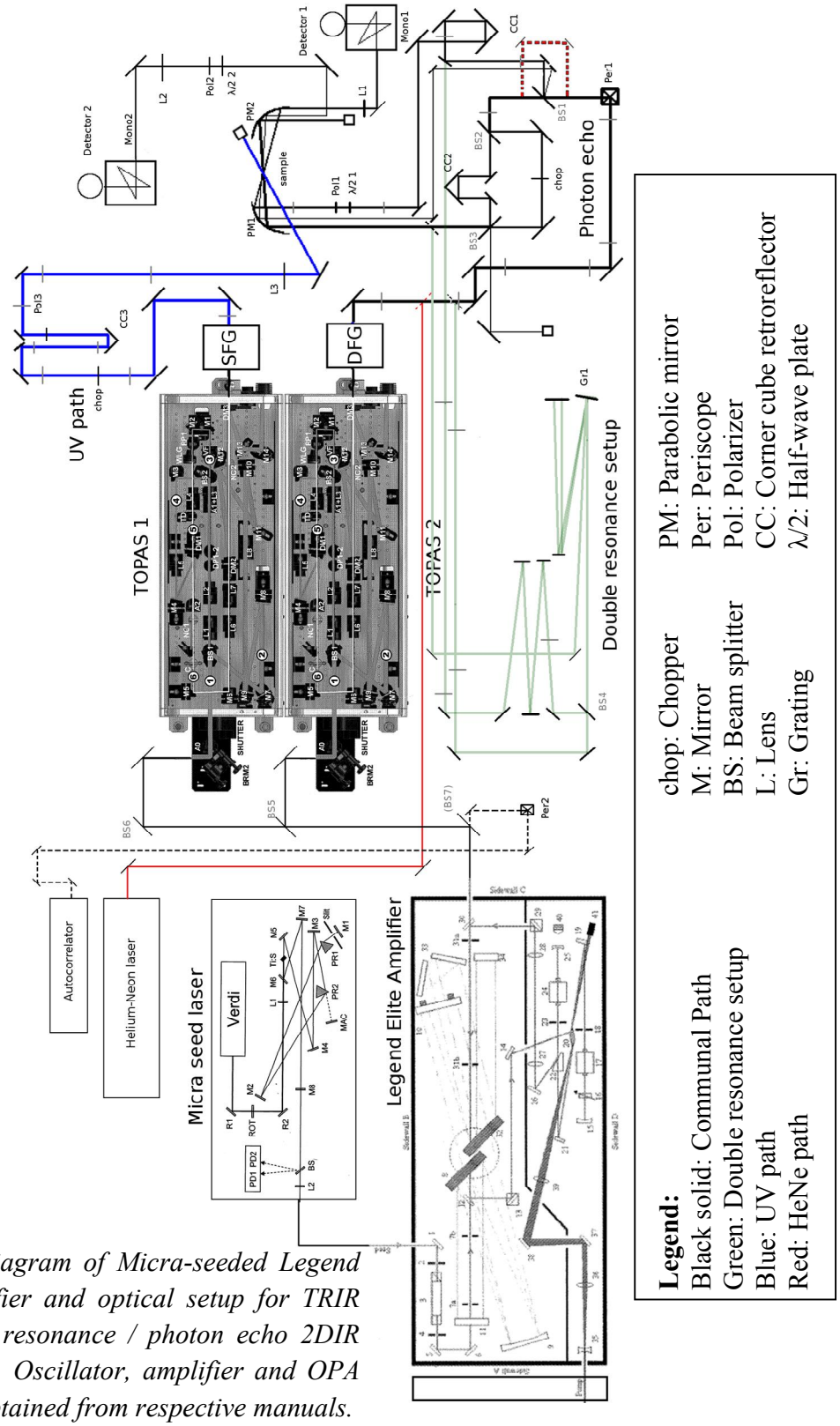


Fig. 6.6. Diagram of Micra-seeded Legend Elite Amplifier and optical setup for TRIR and double resonance / photon echo 2DIR experiments. Oscillator, amplifier and OPA diagrams obtained from respective manuals.

6.3.4 Fourier transform infrared spectroscopy

FTIR spectra from the gel samples were obtained using the Bruker Vertex 70 spectrometer in the range of 800–4000 cm^{-1}. Samples were housed between two CaF_2 plates (thickness, 2 mm) using a 50 μm spacer. Spectra were obtained by averaging over 25 scans with a resolution of 1 cm^{-1}, and corrected for atmospheric background absorptions from gaseous H_2O and CO_2 by subtracting a scan from an empty sample compartment. Because the broad background absorption from liquid D_2O and HOD in the sample may vary from sample to sample, no attempt was made to subtract this.

6.3.5 Steady-state UV excitation with IR probe

The stability of (μ-((H)($S_2C_3H_6$))-$Fe_2(CO)_4(PMe_3)_2$ under UV irradiation was tested using the 3rd and 2nd harmonic of a Nd:YAG laser (Advanced Optical Technology Ltd.) at 355 and 532 nm, respectively. The UV beam was attenuated to provide a power around 4 mW at 1.25 kHz (3.5 μJ/pulse) at the sample position (unfocussed, beam diameter around 1 cm). The stability of the compound was probed with FTIR spectroscopy using the same procedure as described in 6.3.4.

As the photochemical processes took place on multiple time-scales, varying laser repetition rates were used for the experiments that studied the photochemistry of 20 mM (μ-((H)($S_2C_3H_6$))-$Fe_2(CO)_4(PMe_3)_2$ in acetonitrile in the presence of 1 mM $Ru(bpy)_3Cl_2$ (see section 6.4.5). Details are given in Table 6.1.

Table 6.1. Irradiation scheme for 1 mM $Ru(bpy)_3^{2+}$ + 20 mM (μ-((H)($S_2C_3H_6$))-$Fe_2(CO)_4(PMe_3)_2$ in MeCN, totalling 195 M shots, λ=532 nm.

Laser Repetition Rate (Hz)	Total time (s)	Number of shots (first:step:last)
100	50	1,000 : 1,000 : 5,000
1250	36	10,000 : 10,000 : 50,000
10000	1945	100,000 : 100,000 : 1,000,000 1,000,000 : 1,000,000 : 15,000,000 22,000,000 : 5,000,000 : 182,000,000 195,000,000

6.3.6 Hydrogenase mimic vibrational frequency calculations

DFT calculations were performed using Gaussian09.[59] The structures of two isomers were optimized *in vacuo* and harmonic IR frequencies calculated using the LANL2DZ basis set with effective core potential for iron atoms and the 6-311+G(d,p) basis set for all others. The M06L functional, designed for transition metal complexes was used.[60] An empirical scaling factor of 0.97 (estimated here) was used to overlay experimental and calculated results. As these calculations are performed at DFT level on a single molecule with a triple zeta basis set, the intramolecular basis set superposition error can be considered negligible.[74]

6.3.7 Fluorescence spectroscopy

Fluorescence emission spectra were measured on a Jasco FP-6500 spectrofluorometer with light measured orthogonally to the excitation light with a scanning speed of 100 nm min^{-1}. The sample was excited at 280 nm (bandwidth 3 nm, within Fmoc-absorption band) and emission was recorded between 300 and 600 nm. The emitted light was measured with a 3 nm bandwidth with a medium response and a 1 nm data pitch.

6.3.8 Atomic Force Microscopy (AFM, Dr. D.A. Lamprou)

For the AFM experiments on hydrogenase mimic-containing gels, 5 μL of the diluted (with Millipore water, x 10) sample solution was deposited onto a freshly cleaved mica surface (G250-2 Mica sheets 1" × 1" × 0.006"; Agar Scientific Ltd, Essex, UK), and was left to air dry for 1 h, then rinsed with 200 μL of water (Millipore, 18.2 MΩ cm), and air dried for a further 1 h before AFM imaging. The images were obtained by scanning the mica surface in air under ambient conditions using a BrukerMultiMode with NanoScope IIID Controller Scanning Probe Microscope (Digital Instruments, Santa Barbara, CA, USA) operated in tapping mode. Silicon probes were used (FESP; nominal length 225 μm, width = 28 μm, tip radius = 8 nm, resonant frequency = 75 kHz, spring constant = 2.8 N m^{-1}; Bruker Instruments SAS, Dourdan, France). AFM scans were obtained at 512 × 512 pixels resolution. Typical scanning parameters were as follows: tapping frequency 69 kHz, integral and proportional gains 0.3 and 0.5, respectively, set point 0.5 – 0.7 V and scanning speed 1.0 Hz. The images were analyzed by using Bruker Image Analysis Nanoscope software Version 6.14r1.

6.4 Results and discussion

Fmoc-LL was chosen for its capability to form a homogeneous, translucent hydrogel[61] and was successfully used to encapsulate compound **1**. The gelator compound and hydrogenase mimic were suspended in 100 μL of deaerated methanol by sonication, as the mimics are generally water-insoluble. After addition of 900 μL 0.1 M sodium phosphate buffer in D_2O at pH 8 and vigorous shaking for 5 minutes the viscosity of the solution started to increase quickly. The sample was then left for 1 hour to rest during which a self supporting gel (Fig. 6.7(a)) was observed to form with the same uniform red colouration as solutions of compound **1**. Note that for comparison experiments the gel without mimic also contained 10% methanol.

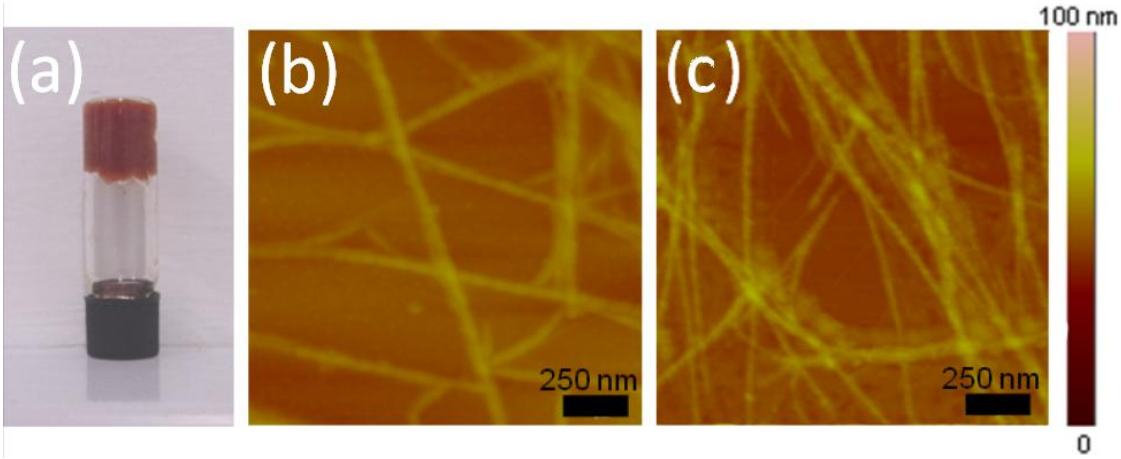


*Fig. 6.7. (a) Photograph showing red self-supporting Fmoc-LL hydrogel formed with inclusion of **1**. (b,c) AFM images of the dried Fmoc-LL hydrogel showing fibril formation with and without **1**, respectively. Scale bars: 250 nm, colour scale 0–100 nm height.*

AFM images (Fig. 6.7(b,c)) of Fmoc-LL hydrogels (dried on mica) with and without compound **1** show the presence of a fibrous network for both samples as has been previously demonstrated for Fmoc-LL alone.[61] No significant change was observed in either fibre thickness or morphology upon inclusion of the mimic.

The remainder of this section is split into five parts. Initially the results of FTIR studies are shown on the gel-encapsulated hydrogenase mimic, followed by a description of gel melting experiments and time resolved studies of the photochemistry. Subsequently, results from tests of this hydrogenase in a different hydrogel are reported. Finally, our efforts towards a complete light-driven catalytic system for H_2 formation are discussed.

6.4.1 FTIR experiments

The FTIR spectrum of the gel in the CO ligand stretching region contains peaks at 1887, 1896 and 1931 cm^{-1} alongside a weaker peak at 1964 cm^{-1}, as displayed in Fig. 6.8(a). In order to gain insight into the chemical environment of **1** in the gel, comparisons of the

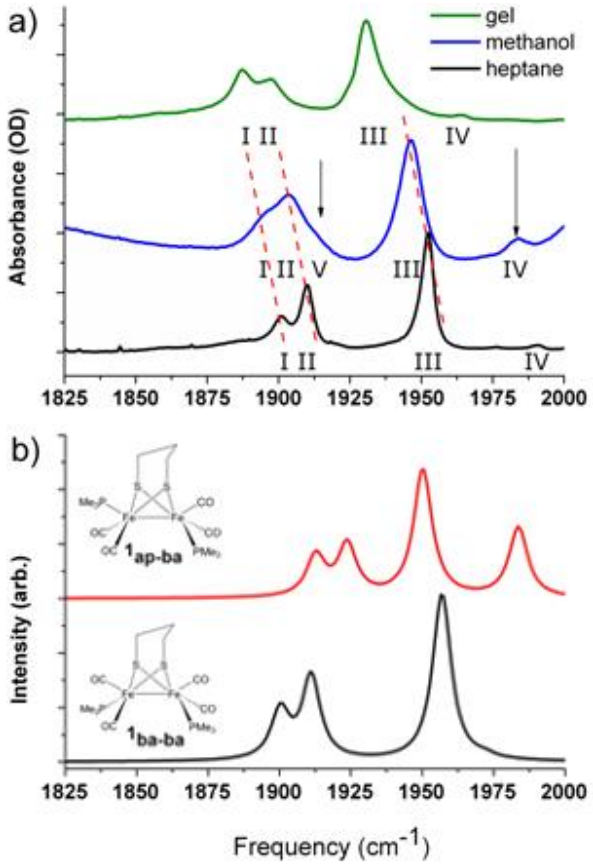


*Fig. 6.8 (a) FTIR spectrum in the CO stretching region of the mid IR of **1** in Fmoc-LL hydrogel (~40 mM of **1**, 20mM of Fmoc-LL, path length 50 μm) and in solutions of heptane and methanol (~1.5 mM, path length 200 μm). (b) Comparison of DFT-predicted IR spectra for **1** in the dibasal ($\mathbf{1}_{ba\text{-}ba}$) and apical-basal ($\mathbf{1}_{ap\text{-}ba}$) isomeric forms with structures shown. M06L/6-311+G(d,p) DFT frequencies have been scaled by 0.97 to overlay calculated and experimental spectra.*

infrared spectrum of the gel sample with those of **1** in solution were carried out. As the gel is 90% aqueous in composition, the logical comparison would be with a solution of **1** in water. However, this was not feasible due to the inherent instability and insolubility of the solute. Given that the two most likely analogues of the environment of **1** in the gel involve the aliphatic side chains of leucine or the presence of a hydrogen bonding matrix, the two solvents chosen for comparison were the hydrophobic n-heptane and protic,

hydrophilic methanol. Fig. 6.8(a) shows the FTIR spectra in the CO stretching region of **1** in methanol and heptane alongside that of the gel sample. The peak patterns in the CO stretching region observed for compound **1** in the gel are sufficiently similar to those in solution in terms of relative peak positions and intensities to preclude breakdown of **1** upon gel formation. The three main CO stretching bands (labelled I-III in Fig 6.8(a)) and the weak high frequency peak (IV) described above are clearly observed. A similar pattern of lines are observed in heptane (also labelled I-IV), though the gel phase sample shows a red-shift of some 13-21 cm^{-1} and the full width half maximum (FWHM) in the gel is 9 cm^{-1} vs. 6 cm^{-1} in heptane, suggesting that there is a significant influence of the gel state on the interactions of **1** with its environment. The methanol sample also exhibits lines that are attributable to the same transitions I-IV. These are also red shifted by 5 cm^{-1} and broadened (FWHM = 10 cm^{-1}) with respect to the heptane sample as would be expected for a polar, H-bonding solvent. It is however interesting that the red-shift is significantly less than that exhibited by the gel-phase sample suggesting that the aqueous nature of the gel may be influencing the larger red shift in the latter environment.

In the case of the methanol solution, it is noticeable that the highest frequency peak (IV) has a larger amplitude compared to the equivalent peak in heptane. Additionally, when the red-shift of peaks I and II is taken into account (see red dashed lines), an additional peak near 1912 cm^{-1} is revealed, this is labelled V and these differences are marked with arrows in Fig. 6.8(a). Such solvent-dependent changes in the spectrum of **1** have been observed previously and can be attributed to the presence of different isomers of **1**.[46] The reported solid-state structure[36] of **1** features the two trimethylphosphine ligands in a dibasal arrangement ($\mathbf{1}_{\mathbf{ba\text{-}ba}}$, Fig. 6.8(b)) while in polar d_6-acetone[36] and MeCN solutions a second isomer was observed in which the ligands occupy an apical-basal arrangement ($\mathbf{1}_{\mathbf{ap\text{-}ba}}$, Fig. 6.8(b)).[46] Density Functional Theory (DFT) calculations have been carried out and show that these represent the two lowest energy isomers of **1** and that they are separated in energy by only 4.8 kJ mol^{-1} with the $\mathbf{1}_{\mathbf{ap\text{-}ba}}$ form having the higher energy. Furthermore the $\mathbf{1}_{\mathbf{ap\text{-}ba}}$ isomeric form is predicted to possess a larger dipole moment (4.65 vs. 2.09 D) consistent with this becoming more apparent in polar media. DFT-predictions

of the IR spectra of the two isomeric forms in the CO stretching region are shown in Fig. 6.8(b). Four transitions are observed representing the various symmetric and antisymmetric combinations of the four carbonyl stretch vibrations. Although these calculations overestimate the transition frequencies, the relative peak positions have been shown to be reliable for this combination of functional and basis set.[60] As vibrational frequencies are commonly overestimated by DFT theory, the results have been scaled by an empirical scaling factor of 0.97 to correct for anharmonic effects and overlay the data with the experimental spectra (see also Ch. 3). It is noticeable from Fig. 6.8(b) that the simulated spectrum for $\mathbf{1}_{\mathbf{ap\text{-}ba}}$ displays a prominent high frequency absorption similar to that labelled IV in the experimental data. This is also consistent with work by Kania *et al*.[62] which used temperature-dependent FTIR and TRIR data to show that peak IV, and the correlated peak V, represent a reliable spectral signature for the presence of $\mathbf{1}_{\mathbf{ap\text{-}ba}}$. From the intensity of peak IV in methanol, we estimate roughly 20% of compound **1** is in the apical-basal isomeric form in this solvent.

Despite the polarity of the gel medium, it is noteworthy that little evidence for $\mathbf{1}_{\mathbf{ap\text{-}ba}}$ is observed (Fig. 6.8(a)): the high frequency peak (IV) remains low in intensity, while the more intense absorptions show band patterns more consistent with the heptane solution. In apparent contrast to this, the broadened linewidths and the large red shift of the gel spectrum indicate strong H-bonding to its environment. As mentioned above, the latter is unsurprising in a largely aqueous medium, but the line-broadening effects in water would be expected to result in a greater FWHM than for methanol solutions, which is not observed in the gel sample. Indeed, adding 10% D_2O to a pure methanol solution of **1,** (the highest water concentration possible while keeping **1** stable) dramatically broadens all transitions in the carbonyl region (see Fig. 6.9(a)). A possible explanation for these observations could be that the mobility giving rise to inhomogeneous line broadening associated with aqueous solvation is reduced in the gel and the environment does not facilitate stabilization of a mixture of isomers as would be expected for a polar solution.[63,64] The latter effect could be due to involvement of the leucine side chains or

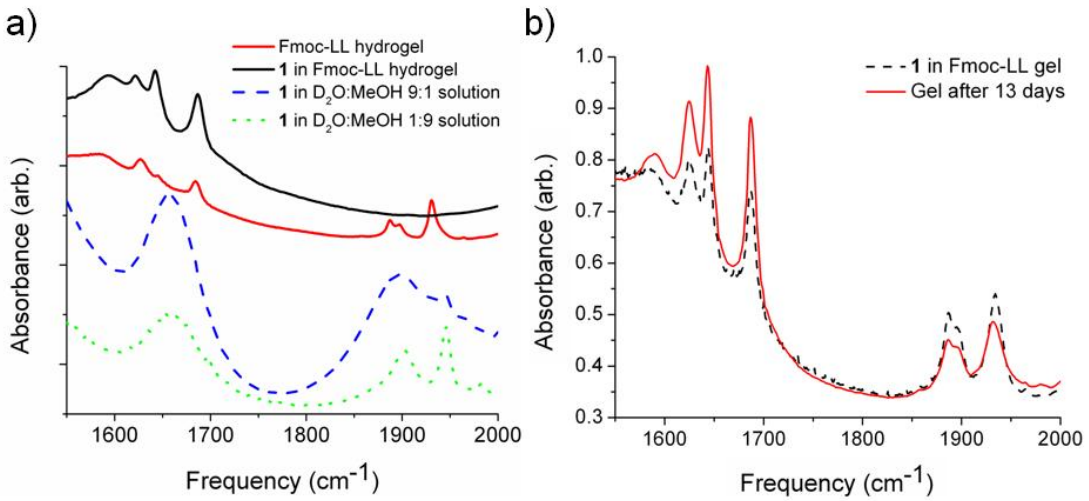


Fig. 6.9 (a) Comparison of FTIR spectra of an Fmoc-LL hydrogel, 1 in an Fmoc-LL hydrogel and solutions of 1 in D_2O:MeOH 9:1 and 1:9 mixtures, representing the solvent in the gel phase and the highest stable water concentration for 1, respectively. Spectra are vertically offset for clarity. (b) Comparison of fresh (1 hour) gel FTIR spectrum and the same sample after 13 days. In both spectra, the sloped background between 1600 and 1800 cm^{-1} can be attributed to water absorption.

the aromatic Fmoc group. Finally, it is noticeable that the intensity of the lowest frequency CO stretching mode (peak I) is larger in the gel than in both solutions. The reason for this is unclear but may result from the effects of the environment on the transition dipole moments of **1**.

FTIR absorption measurements of the gel with and without inclusion of **1** in both the amide I and the CO stretching region of the mid-IR are displayed in Fig. 6.9(a). The amide I region contains peaks at 1625, 1643 and 1686 cm^{-1} that indicate the formation of a β-sheet-based supramolecular structure in agreement with previous data for Fmoc-LL[61] and typical spectra as found in Chapter 3 and 4. Upon inclusion of **1**, these peaks are unaffected in terms of position, although a loss of intensity was noted indicating the gel is somewhat weakened due to the presence of the mimic. Four additional peaks are

observed located at 1887, 1896, 1931 and 1964 cm^{-1}, attributable to CO ligand stretching vibrational modes of **1** as discussed above.

Compound **1** is extremely unstable in aqueous solution: precipitation and/or degradation of the mimic takes place within 30 min for all solution with >10% v/v D_2O, leaving no IR signature of the mimic using the method for FTIR outlined above. However, when a solution phase spectrum was taken within 5 minutes of sample preparation a spectral signature of an unstable mixture of **1** in D_2O:MeOH 9:1 without gelator can be obtained as shown in Fig. 6.9(a). This spectrum exhibits a dramatic line broadening in comparison with the gel phase, in agreement with the lack of the rigid environment. Despite the instability of the mimic in aqueous solvent mixtures, the mimic was shown by FTIR spectroscopy to be stable in the gel environment for up to two weeks as shown in Fig. 6.9(b). This is an important result, as it shows the gel phase can provide an aqueous environment without degradation of the mimic.

To investigate the environment of **1** in the gel further and understand the nature of the stabilising effect observed, FTIR experiments as a function of temperature were carried out.

6.4.2 Gel melting experiments

Fig. 6.10 shows the spectra of gels without (a) and with (c) compound **1** incorporated as the temperature was increased above the macroscopic melting point of the gel. For the Fmoc-LL hydrogel without the active site mimic, the amide I region peaks due to the β-sheet structure disappear gradually, reaching half the initial intensity at $T_{1/2}$ = 70 °C (see Fig. 6.10(b)). This is consistent with breakdown of the supramolecular hydrogen bonding structures that form the fibrils in the gel matrix.[65] Simultaneously, the broad peak at 1590 cm^{-1} associated with the deprotonated C-terminus of **2** gains in intensity, confirming our assignment in Chapters 3 and 4 and previous work on the apparent pK_a shifts (the molecule is expected to be in its fully ionized state at pH 8) observed in Fmoc-dipeptides

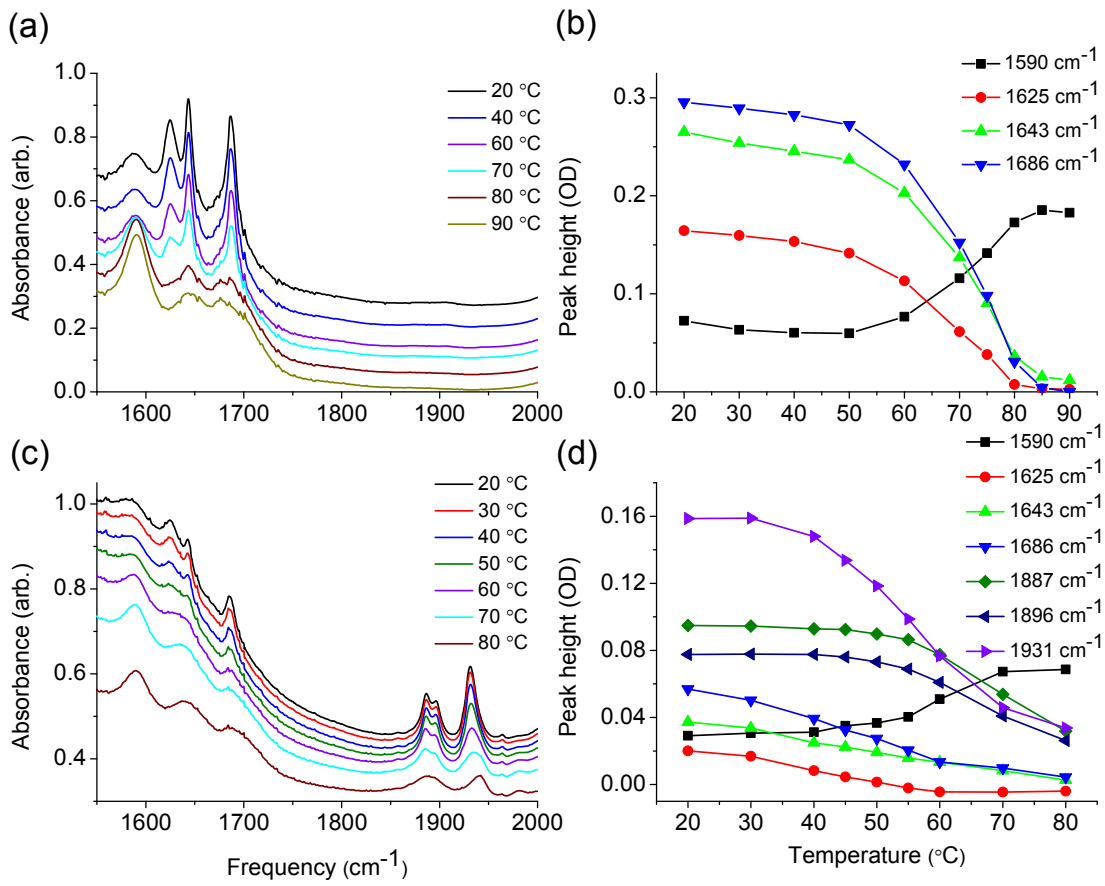


*Fig. 6.10. Gel melting experimental data. (a) Amide I and CO stretching region of the mid IR spectrum of an Fmoc-LL hydrogel at various temperatures. Spectra are vertically offset for clarity. (b) Peak heights of features in (a) as a function of temperature. (c) Amide I and CO stretching region for an Fmoc-LL gel with **1** incorporated, at various temperatures. Spectra are vertically offset for clarity. (d) Peak heights of features in (c) as a function of temperature.*

upon gelation.[66] A self-supporting gel is no longer observed at temperatures above the melting temperature, but upon cooling, all peaks recover fully.

Gels prepared with **1** included show a similar behaviour of the amide I peaks albeit with a lower temperature melting point ($T_{1/2}$ = 50 °C, see Fig. 6.10(c,d)), indicating some

effects on gel properties due to the presence of **1**, although AFM (Fig. 6.7(b,c)) shows fibrous networks for both samples, which indicates no major structural changes have occurred. Interestingly, while the peaks due to the CO stretching vibrations of **1** also show changes with temperature, these do not occur simultaneously with the changes in the amide I region. Significant broadening and blue-shifting of the CO peaks, which is consistent with a decrease in the rigidity and strength of the H-bonding due to exposure to non-gel phase water, occurs at $T_{1/2}$ = 65–70 °C (see Fig. 6.10(c,d)). This indicates that **1** is not yet exposed to a truly aqueous environment when the temperature required to disrupt the β-sheet structures is reached. Upon further heating, a non-reversible decrease in amplitude and further broadening of the CO peaks suggested decomposition or precipitation of the H_2-ase mimic. These observations were confirmed visually: upon heating of the gel to 50 °C a homogeneous red solution was obtained, which turned back into a red hydrogel when cooled. However, further heating to 80 °C lead to irreversible precipitation of **1,** though a translucent hydrogel was reformed on top of the precipitate at room temperature.

The fact that hydrogenase mimic degradation occurs at higher temperatures than the disappearance of the characteristic β-sheet peaks suggests compound **1** is not directly associated with the β-sheet and indicates that a certain degree of residual structuring is present even after the β-sheet is lost. This structure is likely to be due to maintained π-stacking interactions between Fmoc units that persist when the β-sheet interactions are broken and confirms previous observations that, for Fmoc-dipeptide gelation, molecular order is driven by π-stacking. These are induced before gelation, and the molecules are only locked into a β-sheet structure upon cooling.[65,67] To probe this effect, fluorescence spectroscopy was performed during the Fmoc-LL gel melting process, as shown in the Fig. 6.11. The intensity of the characteristic excimer peak at 386 nm, which is attributable to Fmoc π-stacking interactions upon fibrillization,[68] was monitored during the melting of the gel. This showed a strong decrease above 75 °C consistent with dissociation of Fmoc stacks. It is however interesting to note that the monomer peak around 328 nm, which shows quenching upon initial aggregation[69] did not recover in a

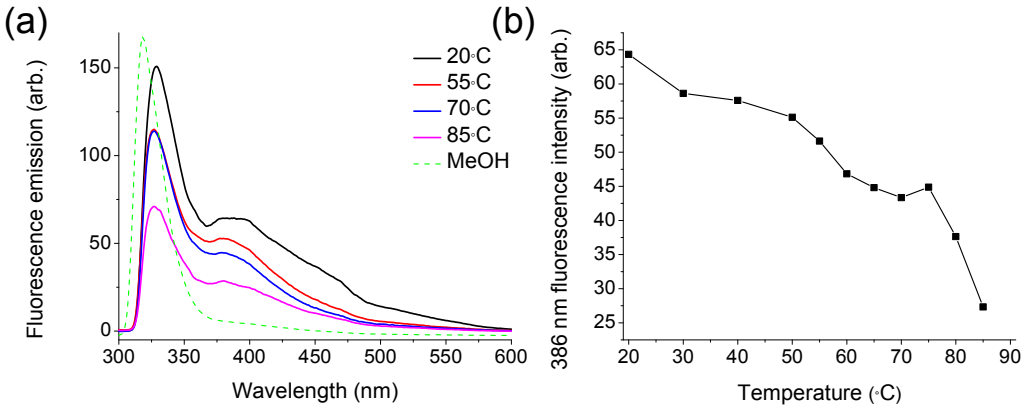


Fig. 6.11. (a) Fluorescence emission spectra of an Fmoc-LL hydrogel at various temperatures. The dashed line represents the emission in methanol at 20 °C. Measurements were performed in a 1 cm cuvette with an excitation wavelength 280 nm and light collected orthogonally to the excitation light. (b) Intensity at 386 nm as function of temperature.

correlated fashion suggesting a disruption into smaller units rather than complete disintegration of the structures. Unfortunately, the strong UV absorption of **1** prevents performing fluorescence spectroscopy on gels encapsulating the hydrogenase mimic as it absorbs both the excitation light and the fluorescing light from the Fmoc moieties.[46] However, the observation that J-aggregate fluorescence emission and CO stretch absorption both undergo a sharp loss of intensity at a temperature significantly higher than the β-sheet disappearance (c.f. Figs. 6.10 and 6.11) suggests a close relationship between the presence of the aromatic stacking and stabilization of **1** in an aqueous environment.

6.4.3 TRIR experiments

The dynamics, environment, and photochemistry of **1** in the hydrogel were studied using ultrafast TRIR measurements. In these experiments, a UV pump pulse (355 nm) tuned to resonance with the metal-to-ligand charge transfer band (MLCT, see section 6.1.2) of the

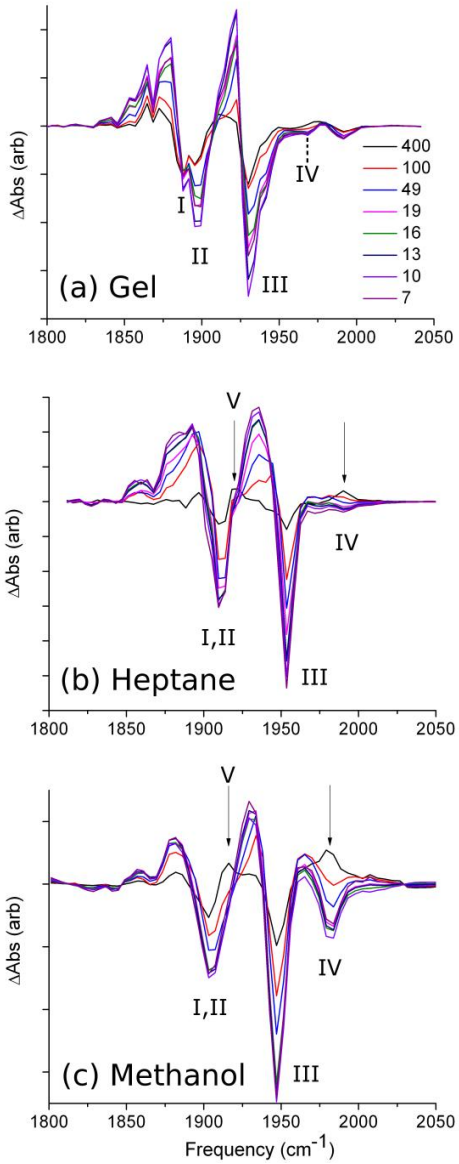


*Fig 6.12. TRIR difference spectra of **1** in (a) 40mM in a 20 mM Fmoc-LL hydrogel, path length 50 μm, (b,c) 1.5 mM in heptane and methanol, path length 200 μM. The data range displayed corresponds to UV_{pump}-IR_{probe} time delays from 7-400 ps. Arrows indicate growing photoproduct peaks.*

Fe-CO bond was used to cause photolysis of a CO ligand while an IR probe pulse was used to monitor the state of the sample. Such controlled creation of a vacant coordination site is an important step in understanding mechanistic pathways of the model compound.

The time-resolved difference spectra collected for **1** in the hydrogel are shown in Fig. 6.12 where they are compared with data from heptane and methanol solutions for a range of pump-probe time delays. In the gel sample (Fig. 6.12(a)), negative peaks are observed from the earliest pump-probe time delays at frequencies that exactly match the absorptions of the CO stretching modes of **1** observed via FTIR spectroscopy**.** These bleaches are labelled I-IV analogously with the peaks in Fig. 6.8 and indicate immediate loss of the parent molecule following photolysis.

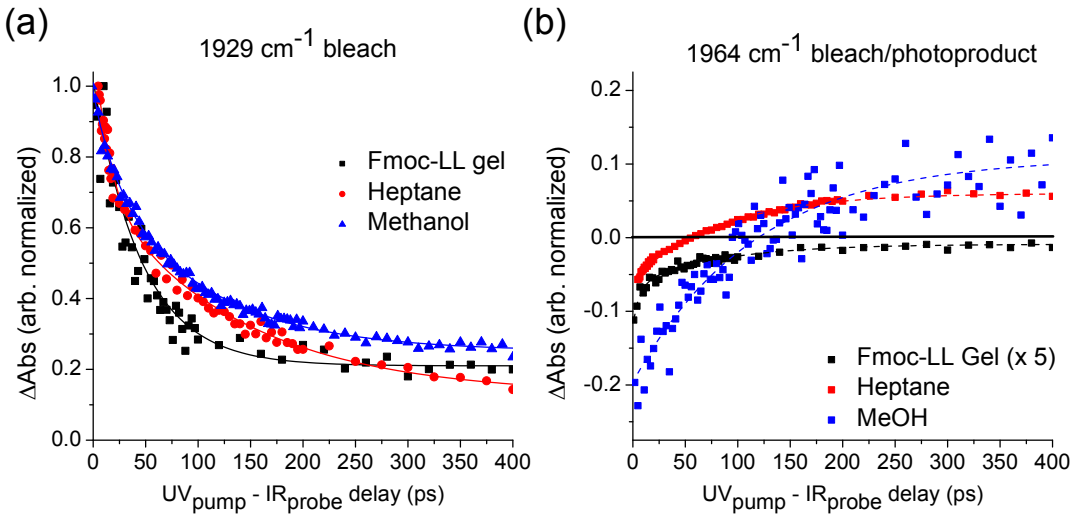


Fig. 6.13 Time-dependent peak area of the main bleach (peak III, 1929 cm^{-1}, normalized and absolute value) for 1 in the hydrogel, heptane and methanol. (b) Peak area for the high frequency peak (IV, 1964 cm^{-1}) that shows growth of a photoproduct peak for only heptane and methanol. Solid lines indicate fits to exponential functions (see text).

*Table 6.2. Selected exponential decay TRIR lifetimes of **1** in an Fmoc-LL hydrogel, extracted from the data in Fig. 6.12. Standard deviations to the fit are indicated.*

Frequency (cm^{-1})	**τ_1 (ps)**	**τ_2 (ps)**
1853 (transient)	8 ± 1	51 ± 5
1897 (bleach)	-	44 ± 2
1922 (transient)	-	74 ± 5
1929 (bleach)	-	58 ± 3
1964 (bleach)	6 ± 1	81 ± 8

Positive, transient absorption peaks shifted several wavenumbers to lower frequency in comparison to the bleaches were also observed; these are assignable to transient tricarbonyl intermediates and vibrationally-hot parent molecules formed following excitation. Fitting revealed that the peaks recover in a similar exponential manner to the parent bleaches and therefore this behaviour is entirely consistent with a simple geminate recombination process of the CO and tricarbonyl intermediate following photolysis.[41,42] It is noted that some evidence for faster (<10 ps) dynamics was observed in the peaks at 1853 and 1964 cm^{-1} and the 1853 and 1922 cm^{-1} peaks also exhibited some narrowing and blue-shifting on a similar timescale, which are suggestive of vibrational cooling, but this was not clearly resolved in all peaks.

The data for **1** in heptane and methanol solutions show similar spectral features (Fig. 6.12 (b) and (c)). In contrast to the gel sample, fitting (Fig. 6.13) has revealed clear biexponential dynamics with cooling rates of 35 ± 15 ps and a longer geminate recombination timescale of 100-200 ps (see ref. 62). It is interesting to note that, in the hydrogel, both vibrational cooling and geminate recombination appear to occur significantly faster than is observed in solution. Taken together, these results further suggest the presence of a somewhat immobile solvent pocket environment for the gel-phase H_2-ase compound, increasing the geminate recombination rate as well as apparently accelerating vibrational cooling via interaction with the H-bonding network. It should be noted that these effects could also be explained by binding of the mimic to the

Fmoc-LL fibres. However, this would not be consistent with the large redshift of CO modes in the gel, which indicates full exposure to a H-bonding solvent, and the lack of change in fibre characteristics in the AFM upon inclusion of **1** (N.B. the mimic and Fmoc-LL are present in roughly a 2:1 stoichiometric ratio making full incorporation into the fibres unlikely). On the other hand, Chapter 4 showed that binding to the surface of a hydrophobic core does not necessarily change the amide I modes of the fibre (e.g. the case for the Fmoc-FF / Fmoc-S system), which implies that interactions with the fibres cannot be excluded on the basis of the IR spectra.

It is also instructive to consider the TRIR difference spectra at the longer accessible pump-probe delay times (~400 ps, Fig. 6.12). In the case of **1,** both in the solution phase and in the gel, the bleaches do not fully recover (see Fig. 6.13) indicating that the parent molecule population is still depleted on these timescales. The positive peaks formed at early times also do not recover fully in the gel material, which shows small positive features even beyond 400 ps, suggesting a long lived tricarbonyl species. In previous studies of the hexacarbonyl analogue of **1** in solution, long-lived positive features due to generation of a significant quantity of solvent adduct species have been observed.[42] Similar long-lived features are seen in the 400 ps spectrum of **1** in methanol in the 1850–1900 cm^{-1} region but they are not present for **1** in heptane solution.[62] This indicates a certain amount of stabilization of the tricarbonyl photoproduct in the gel, possibly by an interaction with a molecule from its first solvation shell. The identity of this species is not clear but the data suggest it is unlikely to be due to a leucine side chain as no heptane coordination is observed to the unsaturated iron centre after photolysis. The most likely candidates are thus the more strongly coordinating methanol or water with the latter being statistically most likely.

In the case of both solution samples, a new peak is observed near 1912 cm^{-1} (peak V) and the bleach due to the highest frequency CO stretching vibration (peak IV) inverts to become a positive peak on timescales that match the geminate recombination timescale

for each solvent (see arrows in Figs 6.12(b,c) and dynamics in Fig. 6.13). Comparisons with the DFT simulations and FTIR spectra discussed above indicate that the new features in the solution-phase spectra are due to an increased quantity of $\mathbf{1}_{\mathbf{ap\text{-}ba}}$ formed following photolysis suggesting that recombination occurs with a bias towards the higher energy isomer. Peaks IV and V do not recover on timescales of 400 ps though DFT calculations showed that the energy barrier to isomerisation is ~38 kJ/mol suggesting that this process requires photolysis of a carbonyl ligand to occur on these timescales rather than being thermally activated.[62] Interestingly, this inversion is not seen in the gel sample spectrum in which peak IV simply recovers toward the baseline (Fig. 6.13(b)). This is an important result indicating that recombination occurs in the same isomeric ratio to that of the parent molecule. This may arise as a result of the faster recombination time or the fact that the more rigid local environment (either solvent or fibres) precludes motion of the large trimethylphosphine groups and limits rotation of the whole molecule on the timescales allowed for recombination. Note that the position of peak IV (dashed line in Fig. 6.12(a)) is determined from the FTIR spectrum and the features observed at higher frequency are most probably due to a UV-induced breakdown product of **1** accumulating during the measurement.

An open question remains whether structural isomerisation of the carbonyl ligands is possible in the gel material; these groups are significantly less sterically-hindered than the trimethylphosphine ligands and may be free to change coordination geometry as is suspected to happen in the enzyme reaction mechanism. Indeed, the fact that the enzyme features relatively small CO and CN ligands may be important in this; coordination from protein side chains in the hydrophobic enzyme pocket may have implications for ease of movement of more bulky ligands during the catalytic cycle in addition to the roles played by CN in dissipating vibrational energy as recently discussed using 2D-IR spectroscopy.[70] Finally, the fact that a proportion of the CO ligands removed from the H_2-ase active site mimic do not return on long timescales, as evidenced by the incomplete bleach recovery, strongly suggests that the gels are gas-permeable, which may well be crucial to future applications.

6.4.4 (μ-pdt)$Fe_2(CO)_4(PMe_3)_2$ in a hydrogel with hydrophilic side chains

Tunability of the environment is one of the great advantages of peptide hydrogels as a material. Usually, the propensity towards self-assembly of a gelator depends strongly on their hydrophobicity, but Hughes *et al*. showed that also Fmoc-dipeptides with relatively hydrophilic side chains such as tyrosine (Y) and serine (S) can form hydrogels.[71] Fmoc-YS (**3**) was used to encapsulate hydrogenase mimic **1** to study the tunability of the environment and to see if the method described in the previous section can be extrapolated to other hydrogels.

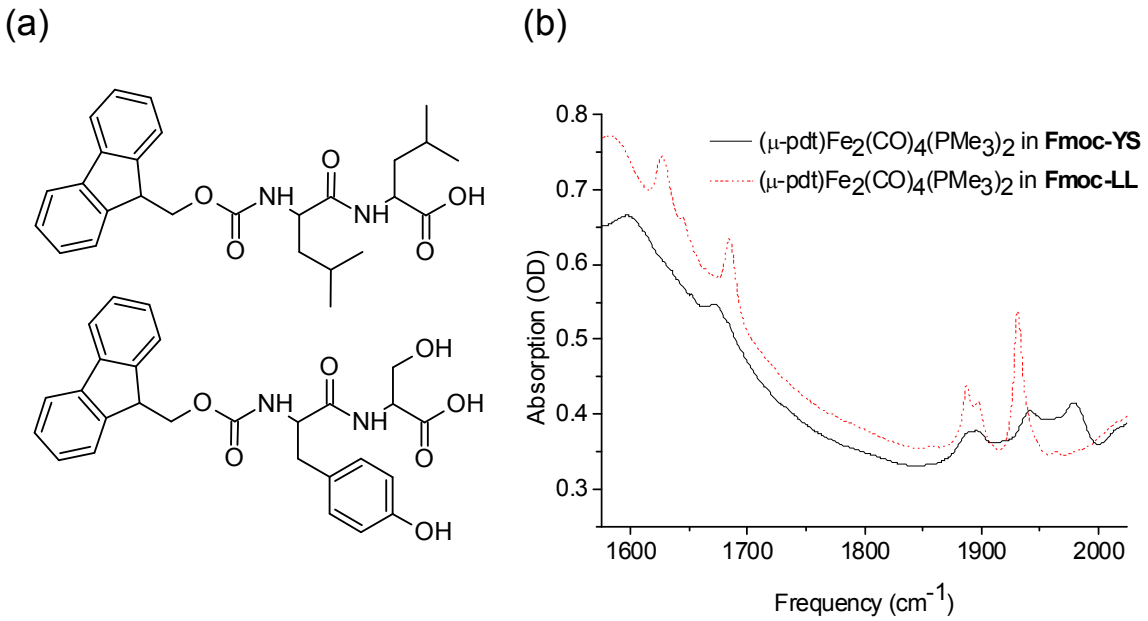


*Fig. 6.14. (a) Fmoc-LL (**2**) and Fmoc-YS (**3**). (b) FTIR comparison between **1** in a 'hydrophilic' (Fmoc-YS) and a 'hydrophobic' (Fmoc-LL) hydrogel.*

Fmoc-YS has been shown to form a self-supporting fibrous hydrogel with an elastic modulus of G' = 3.4 kPa when formed by the subtilisin-catalyzed ester hydrolysis from its methylated precursor compound Fmoc-YS-OMe.[71] Here, we use the same preparation method as for Fmoc-LL hydrogels discussed in this chapter. As such, Fmoc-YS was directly added to a solution of **1** in MeOH and diluted with pH 8 phosphate buffer, upon which a weak gel was formed.

Fig. 6.14 shows a comparison between the FTIR spectra of **1** encapsulated in Fmoc-LL and Fmoc-YS. Notably, the spectrum of the Fmoc-YS gel does not show the usual amide I peaks clearly. This is quite unusual for Fmoc-dipeptides, but in agreement with data from Hughes *et al*.[71] It is expected that the hydrophilicity (and therefore flexibility in aqueous media) of the terminal Ser amino acid prevents the formation of extended β-sheets, and fibres are mainly held together by aromatic π-π interactions of the Fmoc-groups. The peak around 1675 cm^{-1} indicates the Fmoc-carbamate may still be involved in a hydrogen bonding network, which is consistent with the results from Yang *et al*., who showed the Tyr side chain itself is hydrophobic (or aromatic) enough to form nanostructures even without a second amino acid.[72]

Furthermore, it is apparent that the metal carbonyl region shows distinctly different peaks in the Fmoc-YS gel compared to the Fmoc-LL environment. In the former, linewidths are much larger and the two higher frequency peaks are blue-shifted, indicating a more flexible, weakly H-bonding environment for the mimic. The intensity of the high frequency peak associated with $\mathbf{1}_{\mathbf{ap\text{-}ba}}$ at 1976 cm^{-1} (*cf.* 1964 cm^{-1} in the Fmoc-LL gel) is much higher and a contribution at 1917 cm^{-1} provides further evidence of the presence of $\mathbf{1}_{\mathbf{ap\text{-}ba}}$. These observations all resemble a more polar, less H-bonding environment, not unlike the spectrum of **1** in acetonitrile reported by Kania *et al*.[62] Tentatively, a small contribution around 1964 cm^{-1} can be assigned to $\mathbf{1}_{\mathbf{ap\text{-}ba}}$ that is more tightly bound to the fibre core.

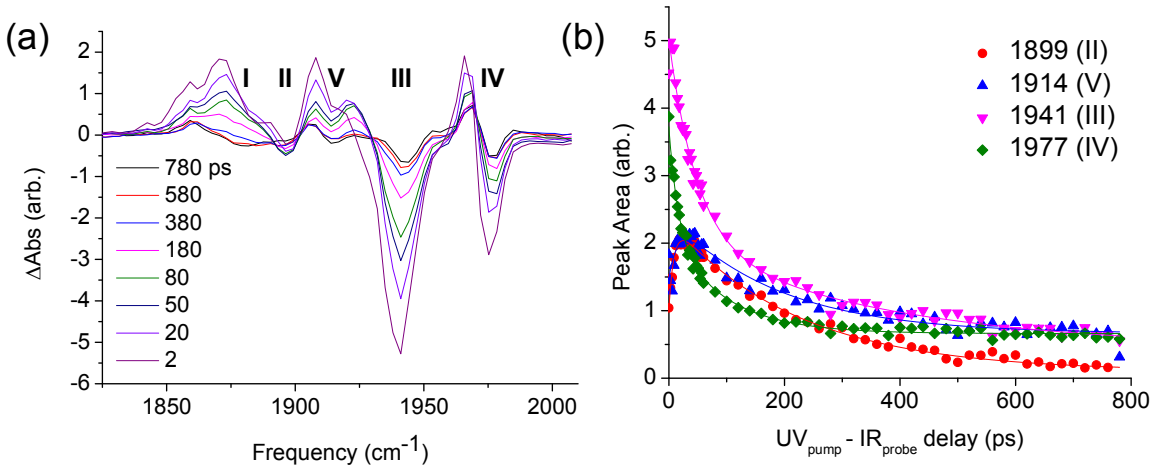


*Fig. 6.15. (a) TRIR difference spectra of **1** in an Fmoc-YS hydrogel. The data range displayed corresponds to UV_{pump}-IR_{probe} time delays from 2-780 ps. (b) Time-dependent peak area of the bleaches at the indicated frequencies. Solid lines represent the biexponential fit to the data.*

To probe the differences in environment in this gel further, UV_{pump}-IR_{probe} TRIR spectroscopy was performed consistent with the method described in the previous section. The averaged pump-probe spectra (Fig. 6.15(a)) show bleaches at the same frequencies as found in the FTIR spectrum: 1889, 1899, 1914, 1941 and 1977 cm^{-1}, labelled I through V analogous to Figs. 6.8 and 6.12. Transient absorption were noted shifted several wavenumbers to lower frequencies, and in some cases overlap with other bleaches (see Table 6.3 for full assignment). Note that the features between 1900 and 1925 cm^{-1} are assumed to be a small bleach (peak V, 1914 cm^{-1}) on top of the broader transient absorption (1910 cm^{-1}) associated with the 1941 cm^{-1} bleach. The bleaches recover as the time delay between pump and probe pulses is extended from 2 to 780 ps as displayed in Fig. 6.15(b).

*Table 6.3. Selected exponential decay TRIR lifetimes of **1** in an Fmoc-YS hydrogel, extracted from the data in Fig. 6.15. Standard deviations to the fit are indicated.*

Frequency (cm^{-1})	**τ_r (ps)** [a]	**τ_1 (ps)**	**τ_2 (ps)**
1859 (transient I, photoproduct)	10 ± 1	-	226 ± 70
1872 (transient I,II) [b]	-	32 ± 4	725 ± 8.7x10^2
1880 (bleach I / transient II) [b]	[c]	[c]	190 ± 22
1899 (bleach II / transient V)	9 ± 1	-	217 ± 11
1910 (transient III)	[c]	[c]	224 ± 25
1914 (bleach V / isomer)	15 ± 5	-	184 ± 27
1941 (bleach III)	-	49 ± 7	368 ± 1.8x10^2
1967 (transient IV) [b]	-	42 ± 8	266 ± 1.9x10^2
1977 (bleach IV / isomer)	-	16 ± 2	96 ± 11

[a]*: This represents a rise time rather than a decay time*

[b]*: Peak overlap or low resolution contributes to relatively large error*

[c]*: Data could only be fitted to a single exponential*

Averaged TRIR spectra were analysed by fitting the traces to a series of nine Gaussian lineshapes using a global fitting routine, all peak areas were extracted and fitted to a biexponential function where possible; the extracted time decay constants are given in Table 6.3. The analysis is somewhat obscured by the broader linewidth of absorptions in the Fmoc-YS gel and the weak absorption observed in the FTIR spectrum at 1918 cm^{-1} (attributed to $\mathbf{1}_{\mathbf{ap\text{-}ba}}$). The latter completely overlaps with the transient absorption of the main bleach, while its associated transient peak is expected at the position of the 1899 cm^{-1} bleach. Together with the relatively low resolution of our experimental setup this is responsible for the high uncertainty in the extracted decay rates of peaks at 1872, 1880 and 1966 cm^{-1}.

Several peaks exhibit a 30–50 ps component (τ_1) which is assigned to vibrational cooling and decay of the metal-to-ligand charge transfer (MLCT) excitation. Moreover, most

peaks exhibit a 180–270 ps decay rate (τ_2), consistent with geminate recombination of dissociated CO ligands and parent ions. The exception is the 1976 cm^{-1} bleach, which seems to recover much faster. However, this is likely to be caused by the isomerisation observed for (polar) solution phase samples, which contributes to a more positive signal at the bleach position, therefore artificially accelerating its recovery. The fact that the associated transient absorption does not recover as quickly as the bleach provides strong evidence for this effect being caused by the isomerisation rather than a different molecular species (e.g. solvent adducts). Moreover, the relatively strong presence of peak V and its fast initial growth and slightly lower τ_2 are consistent with the occurrence of photoinduced isomerisation in this particular gel matrix. However, the gel environment still limits this process to an extent in comparison with solution, indicated by the fact that the 1976 cm^{-1} peak remains negative even at long pump-probe delays, in contrast with solution phase samples.

In contrast with the strong, hydrophobic Fmoc-LL gel, no fast (<10 ps) decay dynamics were detected, although for peaks at 1859, 1899 and 1914 cm^{-1} growth of the respective band was observed on that time scale. This is likely a combined effect of overlapping decaying transient and growing photoproduct absorptions (either $\mathbf{1}_{\mathbf{ap\text{-}ba}}$ or a tricarbonyl/solvent adduct species). It is apparent that the bleach recovery occurs significantly slower than in the Fmoc-LL gel (Fig. 6.12 / Table 6.2) and the essentially static component, which indicates permanent loss of the dissociated CO ligand, is much smaller. The comparison between the main bleach recovery of **1** in both Fmoc-YS (1941 cm^{-1}) and Fmoc-LL given in Fig. 6.16 shows this clearly. This discrepancy is attributed to the more solvent-like behaviour in the weak, hydrophilic-natured hydrogel and is consistent with data obtained in polar solvents such as methanol.[62]

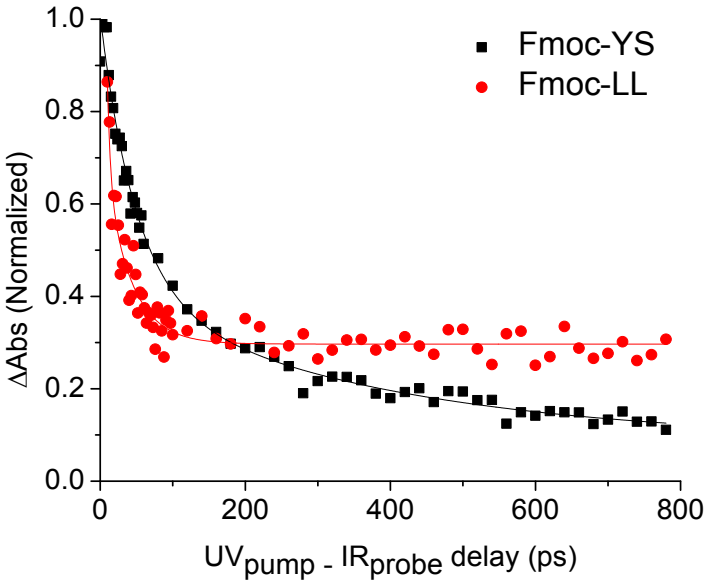


*Fig. 6.16. Comparison between main bleach decay of **1** in Fmoc-YS and Fmoc-LL hydrogels*

In conclusion, the FTIR spectrum of **1** in the Fmoc-YS hydrogel displays similar features to solution phase samples in terms of line broadening and shifting compared, more particularly methanol. Additionally, its ultrafast dynamics have a distinctly different behaviour from in the strong, rigid Fmoc-LL gel and resemble the dynamics observed in a polar solution. However, the gel still stabilises the mimic to an extent, as samples are stable over a prolonged period of time compared to aqueous solution. These results show that the environment of the mimic can be tuned with significant effects on the photochemistry.

6.4.5 Towards light-driven formation of hydrogen

It has been shown in section 6.4.3 and 6.4.4 that hydrogels can provide an aqueous environment for the catalyst, enabling water to be used as a proton source. For transport or other mobile applications, electrons are ideally supplied by a solar energy source, or

alternatively by batteries, but these sources need to be 'wired' to the catalyst. A well known family of molecules that act as an electron transfer agent upon irradiation with visible light are the ruthenium tris-(2,2'-bipyridine) compounds ($Ru(bpy)_3^{2+}$)[49] as discussed in the introduction. Initial efforts to combine this sensitizer with a hydrogenase mimic are described here. As stated in the introduction, (μ-pdt,H)$Fe_2(CO)_4(PMe_3)_2^+$ or **1H** (see bottom right in Fig. 6.3) was used, because of its favourable redox potential towards reduction by the excited ruthenium compound.[53]

It is important to assess the stability of **1H** before studying the dynamics of the H_2 evolution using wavelengths in the UV/Vis region of the spectrum. Therefore, experiments have been performed in order to study the photochemistry of **1H** under light produced by the 2nd or 3rd harmonic from an Nd:YAG laser at 532 or 355 nm, respectively. A 20 mM solution of 1H in MeCN was irradiated for up to 400 seconds with a 4 mW laser beam (for experimental details, see section 6.3.5). FTIR spectra were recorded at various points of time (Figs. 6.17 and 6.18). Peak heights were analysed as a function of time and fitted with (bi)exponential decay functions.

It is apparent from Fig. 6.17 that significant spectral changes occur upon 355 nm irradiation: starting material absorption (peaks IV and V) quickly decrease in favour of a photoproduct (peaks I, II and III).* Fitting reveals exponential decay/growth lifetimes of 26 ± 2 s for peaks II-V, while peak I has a distinctly slower growth lifetime of 101 s (see Table 6.4 for full details). Peak II and III were also observed in the FTIR spectrum of a sample that was kept in the dark over the course of 24 hours (data not shown). However, peak I at 1941 cm^{-1} is not observed in the 'dark' sample without laser light, indicating this absorption arises from a different molecular species than the photodissociation product of **1H**.

* Note the numbering of the peaks is independent from the peaks discussed in section 6.4.3 and 6.4.4.

Irradiation at 532 nm causes similar changes in the spectrum. Prolonged data collection over 10^4 seconds* allowed the fitting of the peak heights to a biexponential decay curve: the exponential decay lifetime of the starting material was somewhat slower at 42 s at this wavelength, with a slow component of $1.0–2.0 \times 10^3$ sec. Peaks II and III grow with a similar time constant (57 and 66 s), but again decay slowly at long time scales. Peak I shows a constant biexponential increase with exponential growth constants of 74 and 2.4×10^3 s.

These results indicate **1H** has limited stability under UV/Vis radiation, which has to be taken into account when performing pump-probe experiments to determine the ultrafast dynamics of the hydrogen evolution, for example by using a flow cell geometry for solution phase samples. On the other hand, it does signify the approach of the hydrogel encapsulation, which was shown to have a limiting effect on structural changes (such as isomerisation) in the mimic in the previous sections. The first step in the future work will be to perform the photochemical study of these compounds in the gel phase.

* These experiments were performed at a laser repetition rate of 1.25 kHz. Therefore, 10,000 seconds is equivalent to 12.5 million laser shots.

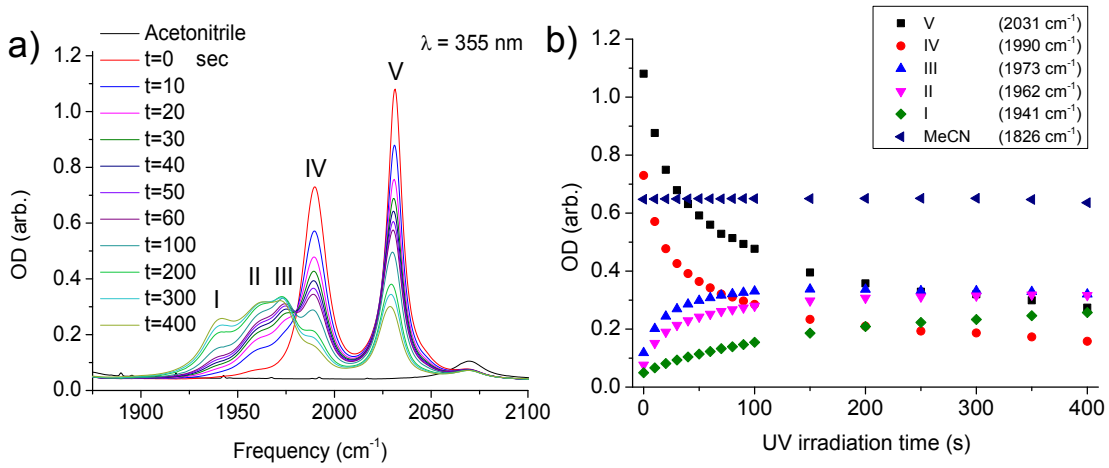


*Fig. 6.17. a) FTIR spectra of 20 mM of **1H** in MeCN solution under UV irradiation with λ=355 nm. b) Peak heights as a function of irradiation time of peaks I–V indicated in a). The acetonitrile peak at 1826 cm^{-1} was used to check sample quality.*

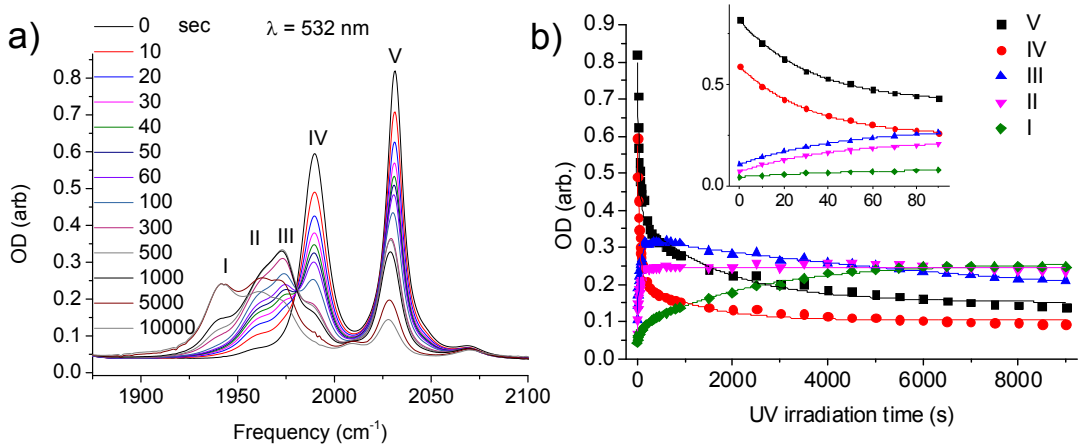


*Fig. 6.18. a) FTIR spectra of 20 mM of **1H** in MeCN solution under UV irradiation with λ=532 nm. b) Peak heights as a function of irradiation time of peaks I–V indicated in a). Solid lines are biexponential fits to the data. The inset shows a magnification of the 0–90 s range.*

*Table 6.4. Exponential growth/decay constants for FTIR peaks of irradiation of **1H** in MeCN solution, extracted from the data in Figs. 6.17 and 6.18. Standard deviations to the fit are indicated.*

Peak #, λ = 355 nm	**τ_1 (s)**	**τ_2 (s)**
I (photoproduct)	101 ± 7	[a]
II (photoproduct)	27 ± 2	[a]
III (photoproduct)	23 ± 1	[a]
IV (decay of starting material)	26 ± 2	[a]
V (decay of starting material)	28 ± 2	[a]
Peak #, λ = 532 nm	**τ_1 (s)**	**τ_2 (s)**
I (photoproduct)	$74 \pm 1x10^1$	$2.4x10^3 \pm 0.1x10^3$
II (photoproduct)	57 ± 3	[a]
III (photoproduct)	66 ± 3	$9.4x10^3 \pm 2.3x10^3$ [b]
IV (decay of starting material)	42 ± 2	$1.2x10^3 \pm 0.2x10^3$
V (decay of starting material)	43 ± 3	$1.7x10^3 \pm 0.2x10^3$

[a] *Data could only be fitted to a monoexponential curve.*

[b] *This is likely a mix of both a decaying and a growing component with similar time constants.*

Additionally, the stability of the ruthenium photosensitizer under UV/Vis irradiation was assessed in deuterated acetonitrile (MeCN-d_3), as $Ru(bpy)_3^{2+}$ absorption bands around 1425, 1447, 1466 and 1606 cm^{-1} are obscured by non-deuterated MeCN. This experiment was also performed in the presence of sacrificial proton and electron donor, ascorbic acid ($C_6H_8O_6$, H_2A). Fig. 6.19 shows the FTIR spectra of 10 mM $Ru(bpy)_3Cl_2$ in both neat MeCN-d_3 and a saturated H_2A solution in MeCN-d_3. No significant loss of intensity of was recorded over the course of 10 million laser shots with no discernible

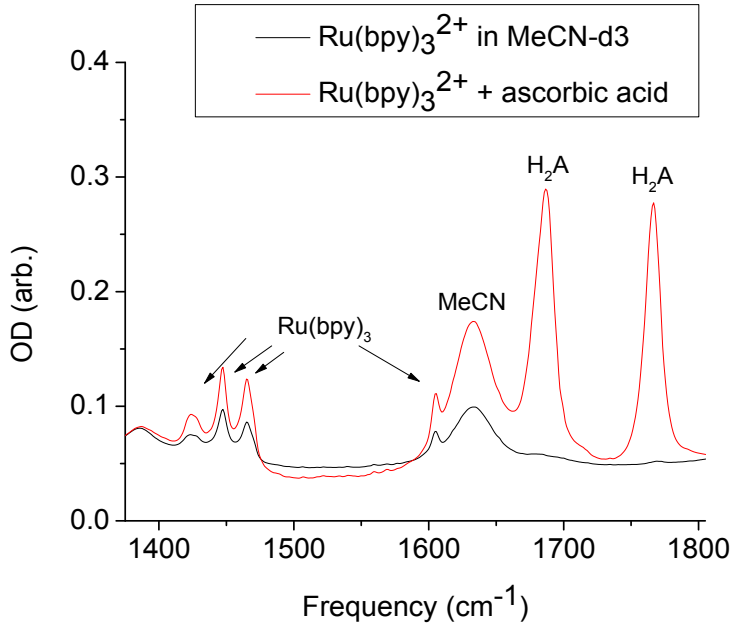


Fig. 6.19. FTIR spectra of $Ru(bpy)_3^{2+}$ in deuterated acetonitrile (CD_3CN) before and after addition of 16 mg of ascorbic acid (H_2A).

change in $Ru(bpy)_3^{2+}$ absorption and a ~12% decrease (by peak height) in H_2A absorption. The stability of the photosensitizer and proton source upon irradiation are important for future experiments on this system.

Finally, the photochemical behaviour of **1H** was studied in the presence of $Ru(bpy)_3^{2+}$ ions over the course of 2×10^8 laser shots at 532 nm. Although this is outside the main MLCT absorption band of the ruthenium compound,[49] distinct changes in the metal carbonyl region of the FTIR spectrum were observed (Fig. 6.20). Upon mixing of 20 mM of hydrogenase mimic and 1 mM of $Ru(bpy)_3Cl_2$ two new absorption bands were observed around 1900 and 1941 cm^{-1} (Fig. 6.20(a)), even without irradiation. As the new peaks are in a roughly 1:20 stoichiometric ratio to the initial peaks, these bands are assigned to deprotonation of **1H** to form compound **1**, in agreement with the line positions and shapes of **1** in MeCN from Kania *et al.*[62] This was further confirmed by experiments with 0.1 and 5 mM of ruthenium compound, which showed a linear change

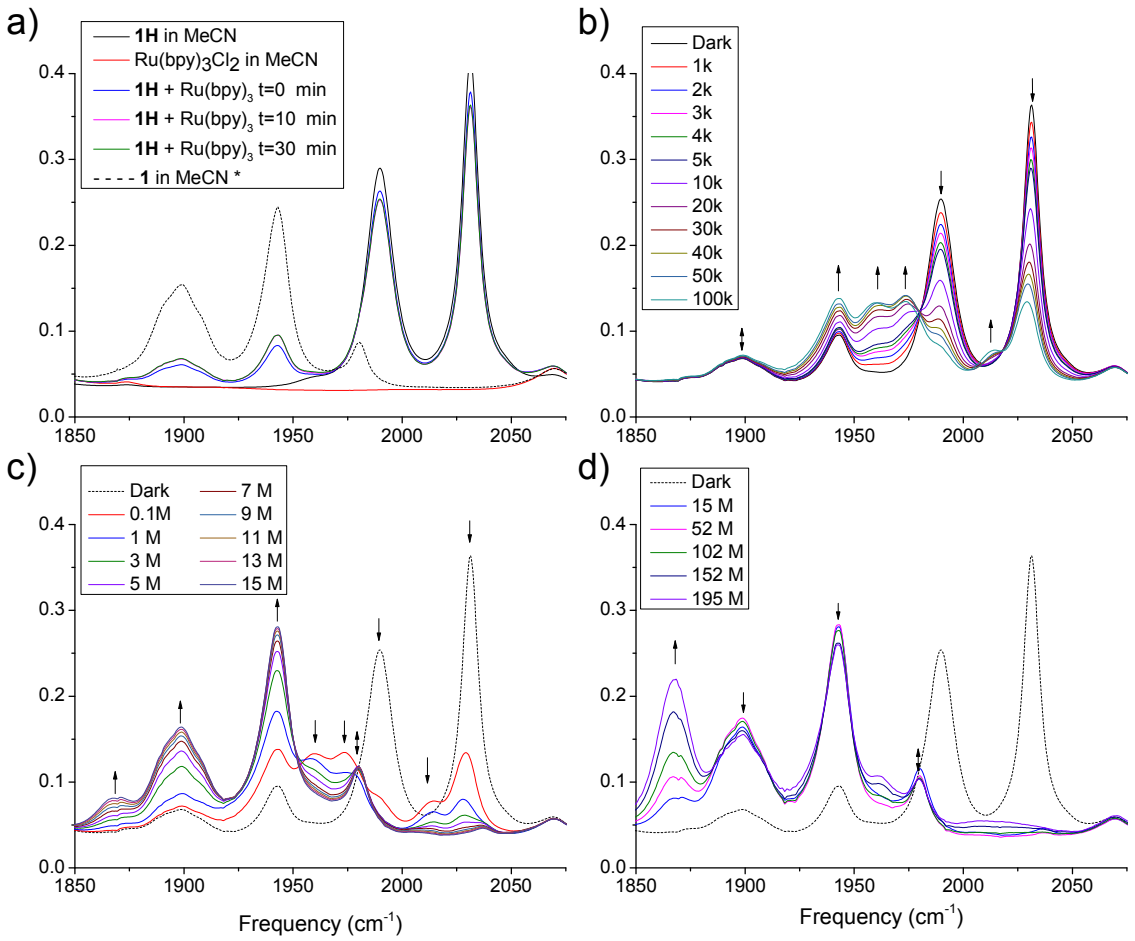


*Fig. 6.20. FTIR spectra of 532 nm irradiation of **1H** (20mM) in the presence of 1 mM of $Ru(bpy)_3^{2+}$ in MeCN. Arrows indicate direction of peak growth/decrease. a) Separate and mixed starting materials in the dark. Graph shows a comparison with the data on deprotonated compound **1** from ref. 62. b) 0–1.00x10^5 laser shots, c) 0–1.50x10^7 laser shots, d) 0–1.95x10^8 laser shots.*

in peak intensity for the new bands. After this instantaneous process (on a seconds time scale), no further changes were observed over 30 min. in the dark. Upon low dosages of radiation (0–10^5 laser shots), it quickly becomes apparent from Fig. 6.20(b) that **1H** initially converts to its photoproduct, also observed in the experiments without $Ru(bpy)_3^{2+}$ (Fig. 6.17 and 6.18), with gain of peaks at 1961 and 1976 cm^{-1} and slow

increase in the 1941 cm^{-1} absorption. However, prolonged irradiation revealed further conversion of this intermediate photoproduct to a different molecular species: from 1–15 million laser shots (Fig. 6.20(c)) the deprotonated product indicated by the peaks at 1900 and 1941 cm^{-1} grows in at the cost of the intermediate found in Fig. 6.20(b). Notice a stable peak at 1980 cm^{-1} (likely because of cancelling growth of **1** and decrease of **1H** bands) that corresponds to the presence of $\mathbf{1}_{\mathbf{ap\text{-}ba}}$ in the polar MeCN solvent.[62] A slowly increasing absorption at 1867 cm^{-1} in Fig. 6.20 (c,d) indicates the formation of a breakdown product and was paired with the appearance of black particulate in the sample cell.

Although it is not clear what the intermediate structure represents, it is clear that 1) $Ru(bpy)_3Cl_2$ acts as a base to **1H** in acetonitrile and 2) irradiation at 532 nm further promotes deprotonation via an intermediate. FTIR spectra of preliminary experiments where ascorbic acid was added to a deprotonated MeCN solution of **1H** and $Ru(bpy)_3Cl_2$ are consistent with a reversible protonation, crucial for the catalytic mechanism discussed in above. Irradiation at the correct wavelength for excitation of the MLCT state of the photosensitizer (around 450 nm[49]) in the presence of ascorbic acid and the mimic therefore has good promise for catalytic H_2 generation if the results from Na *et al*. can be extrapolated to our system.[48] In this thesis only the groundwork for this process has been provided due to time constraints, but future work will focus on reaching this target.

Encapsulation of these reactants in an Fmoc-dipeptide hydrogel will benefit the stability of the compounds and give the advantage of an aqueous environment. Again, the latter is important, as water is an ideal source of protons, considering its abundance and proton relay capabilities. Fmoc-peptides do not have strong UV/Vis absorption bands with $\lambda >$ 300 nm, so should not interfere with $Ru(bpy)_3^{2+}$ excitation. Moreover, the development of conducting peptide hydrogels[29,73] could open the usage of the hydrogel scaffold to transport electrons between photosensitizer and catalytic centre. Altogether, although

technical development is still required, this system has great potential for commercial hydrogen production.

6.5 Conclusion

In the search for environmentally friendly ways of hydrogen gas production, we have studied the possibilities that the quickly developing hydrogenase enzyme active site mimics can offer. Specifically, previously studied mimic (μ-pdt)$Fe_2(CO)_4(PMe_3)_2$ (**1**) was studied in a peptide hydrogel, formed by the self-assembly of small protected dipeptide monomers. The contrasting behaviour of the diiron carbonyl in the gel phase material to that in the studied solvents is striking. Encapsulating compound **1** in a fibrous Fmoc-LL hydrogel results in FTIR spectra that are not fully consistent with either polar (methanol) or apolar (heptane) environments with indications of a rigid H-bonding environment. Dynamical studies point to encapsulation of the solute in a rigid 'solvent cage' that accelerates cooling and geminate recombination while preventing isomerisation processes found in solutions. Melting studies suggest some relation with the π-stacking structural aspects of the fibrils though the environment is clearly not fully hydrophobic and is complex. The implications for the use of gel phase materials in catalytic hydrogen production are interesting, where the lack of 'solvent mobility' and isomerisation potential may prove to be beneficial to its rate or prohibitive in preventing chemical interconversions. Moreover, the stability of the enzyme mimic in an aqueous environment where the presence of water can improve catalytic rates is enhanced dramatically.

The initial results concerning different gels suggest this approach is not limited to compound **1** or Fmoc-LL hydrogels, opening up the possibility for, for example, a more acidic or hydrophilic environment. **1** in a much more hydrophilic and weaker Fmoc-YS gel shows much more solution-like behaviour, indicated by broader linewidths, blueshifted absorption and the strong presence of the $\mathbf{1}_{\mathbf{ap\text{-}ba}}$ isomer in the FTIR spectrum.

Nonetheless, the mimic is still stabilized in the gel compared to solution, especially considering the aqueous gel medium. TRIR data show much slower vibrational dynamics compared to the mimic in a rigid Fmoc-LL gel, but isomerisation is still somewhat limited with respect to solution phase results.

A new catalyst, bridging hydride [FeFe] mimic **1H** shows good promise for catalysing hydrogen evolution. It has been previously shown that the redox potentials of this compound are favourable[53] and here we have studied its rich photochemistry in the presence of photosensitizer $Ru(bpy)_3^{2+}$ and electron/proton donor ascorbic acid. Future work will focus on developing this system towards a fully functional catalytic H_2-producing system for applications in energy management. The inherent stability and accessibility to water that a gel environment offers for the catalyst and photosensitizer could be crucial in achieving cost-effective H_2 generation using bio-inspired catalysts.

6.6 References

(1) Frederix, P. W. J. M.; Kania, R.; Wright, J. A.; Lamprou, D. A.; Ulijn, R. V.; Pickett, C. J.; Hunt, N. T. *Dalton Trans.* **2012**, *41*, 13112–13119.

(2) Mulder, D. W.; Shepard, E. M.; Meuser, J. E.; Joshi, N.; King, P. W.; Posewitz, M. C.; Broderick, J. B.; Peters, J. W. *Structure* **2011**, *19*, 1038–1052.

(3) Dahiya, R. P. *Progress in Hydrogen Energy*; Springer, 1987.

(4) Hydrogen and Fuel Cell Association, U. K. http://www.ukhfca.co.uk/the-industry/fuel-cells/ (accessed Mar 18, 2013).

(5) Department of Energy, U. S. A. http://www1.eere.energy.gov/hydrogenandfuelcells/fuelcells/pdfs/fc_comparison_chart.pdf (accessed Mar 18, 2013).

(6) UK H2 Mobility - Synopsis of Phase 1 Results; 2013.

(7) Frey, M. *ChemBioChem* **2002**, *3*, 153–160.

(8) Rakowski DuBois, M.; DuBois, D. L. *Chem. Soc. Rev.* **2009**, *38*, 62.

(9) Ezzaher, S.; Capon, J.-F.; Gloaguen, F.; Pétillon, F. Y.; Schollhammer, P.; Talarmin, J.; Kervarec, N. *Inorg. Chem.* **2009**, *48*, 2–4.

(10) Wilson, A. D.; Shoemaker, R. K.; Miedaner, A.; Muckerman, J. T.; DuBois, D. L.; Rabowski-DuBois, M. *Proc. Natl. Acad. Sci.* **2007**, *104*, 6951 –6956.

(11) Thomas, C. M.; Liu, T.; Hall, M. B.; Darensbourg, M. Y. *Inorg. Chem.* **2008**, *47*, 7009–7024.

(12) Tschierlei, S.; Ott, S.; Lomoth, R. *Energy Environ. Sci.* **2011**, *4*, 2340.

(13) Nicolet, Y.; Piras, C.; Legrand, P.; Hatchikian, C. E.; Fontecilla-Camps, J. C. *Struct. Fold* **1999**, *7*, 13–23.

(14) Tard, C.; Pickett, C. J. *Chem. Rev.* **2009**, *109*, 2245–2274.

(15) Green, K. N.; Hess, J. L.; Thomas, C. M.; Darensbourg, M. Y. *Dalton Trans.* **2009**, 4344–4350.

(16) Roy, S.; Shinde, S.; Hamilton, G. A.; Hartnett, H. E.; Jones, A. K. *Eur. J. Inorg. Chem.* **2011**, *2011*, 1050–1055.

(17) Ru, X.; Zeng, X.; Li, Z.; J. Evans, D.; Zhan, C.; Tang, Y.; Wang, L.; Liu, X. *J Polym. Sci. Polym. Chem* **2010**, *48*, 2410–2417.

(18) Zhan, C.; Wang, X.; Wei, Z.; Evans, D. J.; Ru, X.; Zeng, X.; Liu, X. *Dalton Trans.* **2010**, *39*, 11255–11262.

(19) Ibrahim, S.; Woi, P. M.; Alias, Y.; Pickett, C. J. *Chem. Comm.* **2010**, *46*, 8189–8191.

(20) Wang, M.; Chen, L.; Li, X.; Sun, L. *Dalton Trans*. **2011**, *40*, 12793–12800.

(21) Liu, Y.-C.; Tu, L.-K.; Yen, T.-H.; Lee, G.-H.; Chiang, M.-H. *Dalton Trans.* **2011**, *40*, 2528.

(22) Helm, M. L.; Stewart, M. P.; Bullock, R. M.; DuBois, M. R.; DuBois, D. L. *Science* **2011**, *333*, 863–866.

(23) Jain, A.; Lense, S.; Linehan, J. C.; Raugei, S.; Cho, H.; DuBois, D. L.; Shaw, W. J. *Inorg. Chem.* **2011**, *50*, 4073–4085.

(24) Estroff, L. A.; Hamilton, A. D. *Chem. Rev.* **2004**, *104*, 1201–1218.

(25) Ghadiri, M. R.; Granja, J. R.; Milligan, R. A.; McRee, D. E.; Khazanovich, N. *Nature* **1993**, *366*, 324–327.

(26) Reches, M.; Gazit, E. *Science* **2003**, *300*, 625–627.

(27) Zhang, Y.; Gu, H.; Yang, Z.; Xu, B. *J. Am. Chem. Soc.* **2003**, *125*, 13680–13681.

(28) Zelzer, M.; Ulijn, R. V. *Chem. Soc. Rev.* **2010**, *39*, 3351–3357.

(29) 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*, 960–966.

(30) Wang, Q.; Yang, Z.; Wang, L.; Ma, M.; Xu, B. *Chem. Comm.* **2007**, 1032–1034.

(31) Wang, Q.; Yang, Z.; Gao, Y.; Ge, W.; Wang, L.; Xu, B. *Soft Matter* **2008**, *4*, 550–553.

(32) Das, D.; Roy, S.; Debnath, S.; Das, P. K. *Chem. Eur. J.* **2010**, *16*, 4911–4922.

(33) Gao, Y.; Zhao, F.; Wang, Q.; Zhang, Y.; Xu, B. *Chem. Soc. Rev.* **2010**, *39*, 3425–3433.

(34) Wang, Q.; Yang, Z.; Zhang, X.; Xiao, X.; Chang, C. K.; Xu, B. *Angew. Chem. Int. Ed.* **2007**, *46*, 4285–4289.

(35) Wang, Q.; Yang, Z.; Ma, M.; Chang, C. K.; Xu, B. *Chem. Eur. J.* **2008**, *14*, 5073–5078.

(36) Zhao, X.; Georgakaki, I. P.; Miller, M. L.; Yarbrough, J. C.; Darensbourg, M. Y. *J. Am. Chem. Soc.* **2001**, *123*, 9710–9711.

(37) Singleton, M. L.; Crouthers, D. J.; Duttweiler, R. P.; Reibenspies, J. H.; Darensbourg, M. Y. *Inorg. Chem.* **2011**, *50*, 5015–5026.

(38) Best, S. P.; Borg, S. J.; Vincent, K. A. In *Spectroelectrochemistry*; Kaim, W.; Klein, A., Eds.; Royal Society of Chemistry: Cambridge, 16:32:00; pp. 1–30.

(39) Jablonskytė, A.; Wright, J. A.; Pickett, C. J. *Dalton Trans.* **2010**, *39*, 3026–3034.

(40) Ridley, A. R.; Stewart, A. I.; Adamczyk, K.; Ghosh, H. N.; Kerkeni, B.; Guo, Z. X.; Nibbering, E. T. J.; Pickett, C. J.; Hunt, N. T. *Inorg. Chem.* **2008**, *47*, 7453–7455.

(41) Kaziannis, S.; Santabarbara, S.; Wright, J. A.; Greetham, G. M.; Towrie, M.; Parker, A. W.; Pickett, C. J.; Hunt, N. T. *J. Phys. Chem. B* **2010**, *114*, 15370–15379.

(42) Stewart, A. I.; Wright, J. A.; Greetham, G. M.; Kaziannis, S.; Santabarbara, S.; Towrie, M.; Parker, A. W.; Pickett, C. J.; Hunt, N. T. *Inorg. Chem.* **2010**, *49*, 9563–9573.

(43) Wright, J. A.; Webster, L.; Jablonskytė, A.; Woi, P. M.; Ibrahim, S. K.; Pickett, C. J. *Faraday Discuss.* **2010**, *148*, 359–371.

(44) Silaghi-Dumitrescu, I.; Bitterwolf, T. E.; King, R. B. *J. Am. Chem. Soc.* **2011**, *128*, 5342–5343.

(45) Gloaguen, F.; Lawrence, J. D.; Schmidt, M.; Wilson, S. R.; Rauchfuss, T. B. *J. Am. Chem. Soc.* **2001**, *123*, 12518–12527.

(46) Wright, J. A.; Pickett, C. J. *Chem. Comm.* **2009**, 5719–5721.

(47) Jablonskytė, A.; Wright, J. A.; Pickett, C. J. *Eur. J. Inorg. Chem.* **2011**, *2011*, 1033–1037.

(48) Na, Y.; Wang, M.; Pan, J.; Zhang, P.; Åkermark, B.; Sun, L. *Inorg. Chem.* **2008**, *47*, 2805–2810.

(49) Campagna, S.; Puntoriero, F.; Nastasi, F.; Bergamini, G.; Balzani, V. In *Photochemistry and Photophysics of Coordination Compounds I*; Balzani, V.; Campagna, S., Eds.; Topics in Current Chemistry; Springer Berlin Heidelberg, 2007; pp. 117–214.

(50) Kurimura, Y.; Hiraizumi, K.; Harakawa, T.; Yamashita, M.; Osada, Y.; Shigehara, K.; Yamada, A. *J. Chem. Soc. Faraday Trans.* **1990**, *86*, 609–613.

(51) Perine, S. C. *J. Electrochem. Soc.* **1989**, *136*, 220C–222C.

(52) Rivarola, C. R.; Biasutti, M. A.; Barbero, C. A. *Polymer* **2009**, *50*, 3145–3152.

(53) Jablonskytė, A.; Wright, J. A.; Fairhurst, S. A.; Peck, J. N. T.; Ibrahim, S. K.; Oganesyan, V. S.; Pickett, C. J. *J. Am. Chem. Soc.* **2011**, *133*, 18606–18609.

(54) Wang, W.; Rauchfuss, T. B.; Bertini, L.; Zampella, G. *J. Am. Chem. Soc.* **2012**, *134*, 4525–4528.

(55) Fermann, M. E.; Galvanauskas, A; Sucha, G. *Ultrafast Lasers: Technology and Applications*; CRC Press, 2002.

(56) Telle, H. H.; Ureña, A. G.; Donovan, R. J. *Laser Chemistry: Spectroscopy, Dynamics and Applications*; John Wiley & Sons, 2007.

(57) Liu, X.; Ibrahim, S. K.; Tard, C.; Pickett, C. J. *Coord. Chem. Rev.* **2005**, *249*, 1641–1652.

(58) *IGOR Pro*; Wavemetrics, Inc.

(59) M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G. A. Petersson, H. Nakatsuji, M. Caricato, X. Li, H. P. Hratchian, A. F. Izmaylov, J. Bloino, G. Zheng, J. L. Sonnenberg, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, J. A. Montgomery, Jr., J. E. Peralta, F. Ogliaro, M. Bearpark, J. J. Heyd, E. Brothers, K. N. Kudin, V. N. Staroverov, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, N. Rega, J. M. Millam, M. Klene, J. E. Knox, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, R. L. Martin, K. Morokuma, V. G. Zakrzewski, G. A. Voth, P. Salvador, J. J. Dannenberg, S.

Dapprich, A. D. Daniels, Ö. Farkas, J. B. Foresman, J. V. Ortiz, J. Cioslowski, *Gaussian 09 Revision A.1*.

(60) Zhao, Y.; Truhlar, D. G. *J. Chem. Phys.* **2006**, *125*, 194101.

(61) Das, A. K.; Collins, R.; Ulijn, R. V. *Small* **2008**, *4*, 279–287.

(62) Kania, R.; Frederix, P. W. J. M.; Wright, J. A.; Ulijn, R. V.; Pickett, C. J.; Hunt, N. T. *J. Chem. Phys.* **2012**, *136*, 044521–044521–9.

(63) Tamai, Y.; Tanaka, H.; Nakanishi, K. *Macromolecules* **1996**, *29*, 6750–6760.

(64) Tamai, Y.; Tanaka, H.; Nakanishi, K. *Macromolecules* **1996**, *29*, 6761–6769.

(65) 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*, 1089–1094.

(66) Tang, C.; Smith, A. M.; Collins, R. F.; Ulijn, R. V.; Saiani, A. *Langmuir* **2009**, *25*, 9447–9453.

(67) Chen, L.; Morris, K.; Laybourn, A.; Elias, D.; Hicks, M. R.; Rodger, A.; Serpell, L.; Adams, D. J. *Langmuir* **2010**, *26*, 5232–5242.

(68) Channon, K. J.; Devlin, G. L.; Magennis, S. W.; Finlayson, C. E.; Tickler, A. K.; Silva, C.; MacPhee, C. E. *J. Am. Chem. Soc.* **2008**, *130*, 5487–5491.

(69) Kang, H. K.; Kang, D. E.; Boo, B. H.; Yoo, S. J.; Lee, J. K.; Lim, E. C. *J. Phys. Chem.* **2005**, *109*, 6799–6804.

(70) Kaziannis, S.; Wright, J. A.; Candelaresi, M.; Kania, R.; Greetham, G. M.; Parker, A. W.; Pickett, C. J.; Hunt, N. T. *Phys. Chem. Chem. Phys.* **2011**, *13*, 10295.

(71) Hughes, M.; Birchall, L. S.; Zuberi, K.; Aitken, L. A.; Debnath, S.; Javid, N.; Ulijn, R. V. *Soft Matter* **2012**, *8*, 11565–11574.

(72) Yang, Z.; Gu, H.; Fu, D.; Gao, P.; Lam, J. K.; Xu, B. *Adv. Mater.* **2004**, *16*, 1440–1444.

(73) Chen, L.; Revel, S.; Morris, K.; Adams, D. J. *Chem. Commun.* **2010**, *46*, 4267–4269.

(74) Asturiol, D.; Duran, M.; Salvador, P. *J. Chem. Phys.* **2009**, *128*, 144108–144108–5

Chapter 7: Conclusions and outlook

A report on the study of the structure, dynamics and applications of peptide-based nanomaterials using a range of techniques is provided in this thesis. In particular, inputs from the field of infrared (IR) spectroscopy and computational chemistry have led to insights in the architecture of self-assembled peptides and peptide derivatives on the nanoscale.

Initially, the use of computational methods in the field of biomolecular self-assembly was evaluated by looking at a varied collection of systems in the literature. The modelling of self-assembly is an especially difficult challenge in this field due to the large amount of monomers that are needed to form self-assembled structures. Moreover, they are held together by weak, non-covalent bonds that are relatively hard to model. It was found that although the field of molecular dynamics (MD) simulations and quantum mechanical (QM) calculations in the context of low molecular weight biomolecules is still very much in development, promising examples have been reported with regards to reproducing experimental results. It is envisioned that especially MD simulations will be crucial in elucidating the design rules of peptide-based nanostructures in the next decade.

In Chapter 3 the use of IR spectroscopy in determining the secondary structure of these peptide materials was assessed. For two analogous hydrogel systems (Fmoc-AA, and Fmc-AA), an analysis of the IR absorption spectrum was performed, backed up by density functional theory frequency calculations. IR absorptions are very sensitive to H-bonding patterns in peptides and therefore can be a good probe for supramolecular structure, in this case β-sheets. On the other hand, the assignment of IR absorption bands can be difficult for these systems, as the well-developed theories for proteins do not apply to peptides. This has led to the ambiguous assignment of IR bands of β-sheets in the field of short peptide hydrogelators, for which a more accurate model has been presented here. Unfortunately, the DFT methods were not yet able to categorically distinguish between particular self-assembling modes.

The performance of both IR spectroscopy and all-atom MD simulations in three case studies of peptide nanostructures has been evaluated in Chapter 4. Both techniques were shown to be sensitive to small changes in the supramolecular structure as a consequence of variations in the amino acid side chains of the peptides under study, showing they can provide valuable information on nanostructure architecture. Although all-atom MD is still somewhat limited by computational costs, especially in mimicking spontaneous self-assembly from an unbiased, randomly distributed configuration of building blocks, it provides an intuitive picture of the molecular arrangement at atomic resolution. IR spectroscopy and MD simulations were employed to study three complex nanostructures involving peptides with atypical backbone dihedral angles, co-assembled states and dipeptides that change their arrangement upon drying of their hydrogels. The results from these systems indicated that although IR is a very sensitive technique, the origins of the bands are not always understood. When both IR and MD methods can be informed with more experimental and theoretical data for known structures, they are however envisioned to be able to give an accurate description of the self-assembled state for a general case of a peptide-based nanostructure.

In Chapter 5 the issue of limited simulation sizes that is always present in high-resolution computational studies was addressed by employing the MARTINI coarse-grain (CG) force field to peptide self-assembly. Although certain details on specific interactions are lost, using this method allowed for the development of a high-throughput screening method that has been applied to predict the aggregation propensity of all combinations of di- and tripeptides. It was shown that both the propensity of the modelled peptides to aggregate and the structural features of that aggregate are in good agreement with experimental studies. The chapter discusses further ways to screen for aggregating peptides with particular properties. This method can then be applied to larger peptides (or other biomolecules) with limited additional computational cost.

In Chapter 6 of this thesis, the obtained knowledge of the peptide nanostructures has been applied in the encapsulation of a catalyst with the goal of producing hydrogen as a

fuel. Encapsulation of a hydrogenase enzyme active site mimic into an Fmoc-dipeptide hydrogel dramatically increased the stability of the mimic compared to the solution phase, which is important as water is the most widely accessible source of protons. The particular location of the mimic in the gel was not fully consistent with either polar (water-like) or apolar (associated with fibres) situations, but indicates a rigid H-bonding environment. Time-resolved infrared spectroscopy suggests the hydrogel provides a rigid 'solvent cage' that accelerates cooling and recombination after UV-induced photolysis of its ligands, but prevents isomerisation processes found in solutions. The first steps towards tailoring the gel environment of the mimic were set with interesting changes in the spectroscopy between a gel with hydrophobic and a hydrophilic amino acids. Furthermore, an effort has been made towards photosensitized production of H_2 were set by incorporating photoactive electron transfer agents and stability studies of a hydrogenase mimic with a bridging hydride ligand.

Overall, the selection of methods developed in this thesis can be used towards various applications using peptide-based nanomaterials. As a particular example, the CG screening method presented in Ch. 5 can be used to identify suitable biomaterials for encapsulating hydrogenase mimics, e.g. by looking for aggregating amine-containing peptides; it has recently been shown that hydrogenase mimics often perform better in the presence of pendant amine groups near the catalytic site.[1] Moreover, the knowledge of peptide assembly allows the design of new structures that present certain functional moieties on their interface with the environment, effectively designing an artificial enzyme active site. Subsequently, FTIR and TRIR spectroscopy can be used to probe the interactions between the identified gel matrix and the catalyst. When redox-active sites are incorporated in the gel structure in an electrochemical setup, this system has great promise in the reversible reduction of protons in hydrogen fuel applications.

Reference

(1) Ezzaher, S.; Capon, J.-F.; Gloaguen, F.; Pétillon, F. Y.; Schollhammer, P.; Talarmin, J.; Kervarec, N. *Inorg. Chem.* **2009**, *48*, 2–4.

Appendix 1: Typical input parameters for NAMD, GROMACS, Gaussian and Turbomole calculations

NAMD EXAMPLE MINIMIZATION (5000 steps) + SHORT EQUILIBRATION (50 ps)

```
#############################################################
## ADJUSTABLE PARAMETERS                                   ##
#############################################################

structure          structure.psf
coordinates        coordinates.pdb
outputName         output

set temperature    300
margin             2

# Continuing a job from the restart files
if {0} {
set inputname      previousname
binCoordinates     $inputname.restart.coor
binVelocities      $inputname.restart.vel  ;# remove the
"temperature" entry if you use this!
extendedSystem     $inputname.xsc
}

firsttimestep      0


#############################################################
## SIMULATION PARAMETERS                                   ##
#############################################################

# Input
paraTypeCharmm      on
parameters          par_all22_36.prm

# NOTE: Do not set the initial velocity temperature if you
# have also specified a .vel restart file!
temperature         $temperature


# Periodic Boundary conditions
# NOTE: Do not set the periodic cell basis if you have also
# specified an .xsc restart file!
```

```
if {1} {
cellBasisVector1    64.1   0.0   0.0
cellBasisVector2     0.0  67.3   0.0
cellBasisVector3     0.0   0.0  69.7
cellOrigin          31.34 33.96 33.04
}
wrapWater           on
wrapAll             on


# Force-Field Parameters
exclude             scaled1-4
1-4scaling          1.0
cutoff              12
switching           on
switchdist          10
pairlistdist        13.5


# Integrator Parameters
timestep            1.0    ;# fs
rigidBonds          water  ;# larger steps require 'all'
nonbondedFreq       1
fullElectFrequency  2
stepspercycle       10


#PME (for full-system periodic electrostatics)
if {1} {
PME                 yes
PMEGridSpacing      1.0
}

# Constant Temperature Control
langevin            on    ;# do langevin dynamics
langevinDamping     5     ;# damping coefficient (gamma) of 5/ps
langevinTemp        $temperature
langevinHydrogen    off    ;# don't couple langevin bath to
hydrogens


# Constant Pressure Control (variable volume for equilibration)
if {1} {
useGroupPressure      yes ;# needed for 2fs steps
useFlexibleCell       no  ;# no for water box, yes for membrane
useConstantArea       no  ;# no for water box, yes for membrane

langevinPiston        on
langevinPistonTarget  1.01325 ;#  in bar -> 1 atm
langevinPistonPeriod  100
langevinPistonDecay   50
langevinPistonTemp    $temperature
```

```
}

# Slow heating (when starting from predefined structure)
If {0} {
reassignTemp          25
reassignIncr          25
reassignFreq          1000
reassignHold          300
}

restartfreq         1000     ;# 500steps = every 0.5ps
dcdfreq             1000
xstFreq             1000
outputEnergies      500
outputPressure      500


#############################################################
## EXECUTION SCRIPT                                        ##
#############################################################

# Minimization
if {1} {
minimize            5000
}
run 50000
```

NAMD EXAMPLE PRODUCTION RUN (here 50 ns)

```
structure          structure.psf
coordinates        minimized+equilibrated_structure.pdb
outputName         outputname

set temperature    300

# Continuing a job from the restart files
if {1} {
set inputname      outputname_min+eq
binCoordinates     $inputname.restart.coor
binVelocities      $inputname.restart.vel  ;# remove the
"temperature" entry if you use this!
extendedSystem     $inputname.xsc
}

firsttimestep      0

# Input
paraTypeCharmm      on
parameters          par_all22_36.prm
```

```
# Force-Field Parameters
exclude             scaled1-4
1-4scaling          1.0
cutoff              12
switching           on
switchdist          10
pairlistdist        13.5


# Integrator Parameters
timestep            2.0  ;# 2 fs for production runs
rigidBonds          all  ;# steps > 1 fs require 'all'
nonbondedFreq       1
fullElectFrequency  2
stepspercycle       20
wrapWater           on
wrapAll             on

#PME (for full-system periodic electrostatics)
if {1} {
PME                 yes
PMEGridSpacing      1.0
}

# Constant Temperature Control
langevin            on    ;# do langevin dynamics
langevinDamping     5     ;# damping coefficient (gamma) of 5/ps
langevinTemp        $temperature
langevinHydrogen    off   ;# don't couple langevin bath to hydrogens


# Constant Pressure Control (variable volume)
if {1} {
useGroupPressure      yes ;# needed for 2fs steps
useFlexibleCell       no  ;# no for water box, yes for membrane
useConstantArea       no  ;# no for water box, yes for membrane

langevinPiston        on
langevinPistonTarget  1.01325 ;#  in bar -> 1 atm
langevinPistonPeriod  100
langevinPistonDecay   50
langevinPistonTemp    $temperature
}

restartfreq         25000     ;# 25000steps = every 50ps
dcdfreq             25000
xstFreq             25000
outputEnergies      25000
outputPressure      25000

run 25000000 ;# 50ns NPT
```

GROMACS EXAMPLE MINIMIZATION (normal water)

```
integrator               = steep
; start time and timestep in ps
tinit                    = 0.0
dt                       = 0.025
nsteps                   = 5000
; number of steps for center of mass motion removal =
nstcomm                  = 1
comm-grps                =
emtol                    = 200

; OUTPUT CONTROL OPTIONS =
; Output frequency for coords (x), velocities (v) and forces (f) =
nstxout                  = 0
nstvout                  = 0
nstfout                  = 0
; Output frequency for energies to log file and energy file =
nstlog                   = 100
nstenergy                = 100
; Output frequency and precision for xtc file =
nstxtcout                = 0
xtc_precision            = 0
; This selects the subset of atoms for the xtc file. You can =
; select multiple groups. By default all atoms will be written. =
xtc-grps                 =
; Selection of energy groups =
energygrps               =

; NEIGHBORSEARCHING PARAMETERS =
; MARTINI - no need for more frequent updates
; or larger neighborlist cut-off due
; to the use of shifted potential energy functions.

; nblist update frequency =
nstlist                  = 10
; ns algorithm (simple or grid) =
ns_type                  = grid
; Periodic boundary conditions: xyz or no =
pbc                      = xyz
; nblist cut-off         =
rlist                    = 1.2

; OPTIONS FOR ELECTROSTATICS AND VDW =
; MARTINI - vdw and electrostatic interactions are used
; in their shifted forms. Changing to other types of
; electrostatics will affect the general performance of
; the model.
```

```
; Method for doing electrostatics =
coulombtype              = Shift
rcoulomb_switch          = 0.0
rcoulomb                 = 1.2
; Dielectric constant (DC) for cut-off or DC of reaction field =
epsilon_r                = 15
; Method for doing Van der Waals =
vdw_type                 = Shift
; cut-off lengths        =
rvdw_switch              = 0.9
rvdw                     = 1.2
; Apply long range dispersion corrections for Energy and Pressure =
DispCorr                 = No

; OPTIONS FOR WEAK COUPLING ALGORITHMS =
; MARTINI - normal temperature and pressure coupling schemes
; can be used. It is recommended to couple individual groups
; in your system seperately.

; Temperature coupling   =
tcoupl                   = no
pcoupl                   = no

; GENERATE VELOCITIES FOR STARTUP RUN =
gen_vel                  = no
gen_temp                 = 303
gen_seed                 = 474529

; OPTIONS FOR BONDS     =
; MARTINI - for ring systems constraints are defined
; which are best handled using Lincs.

constraints              = none
; Type of constraint algorithm =
constraint_algorithm     = Lincs
; Do not constrain the start configuration =
unconstrained_start      = no
; Highest order in the expansion of the constraint coupling matrix =
lincs_order              = 4
; Lincs will write a warning to the stderr if in one step a bond =
; rotates over more degrees than =
lincs_warnangle          = 30
```

GROMACS PRODUCTION RUN (normal water)

```
integrator              = md
tinit                   = 0.0
dt                      = 0.025
nsteps                  = 500000
nstcomm                 = 100

nstxout                 = 0
nstvout                 = 0
nstfout                 = 0
nstlog                  = 10000
nstenergy               = 10000
nstxtcout               = 0
xtc_precision           = 0

nstlist                 = 10
ns_type                 = grid
pbc                     = xyz
rlist                   = 1.2

coulombtype             = Shift
rcoulomb_switch         = 0.0
rcoulomb                = 1.2
epsilon_r               = 15
vdw_type                = Shift
rvdw_switch             = 0.9
rvdw                    = 1.2
DispCorr                = No

tcoupl                  = Berendsen
tau_t                   = 1.0 1.0
tc-grps                 = protein non-protein
ref_t                   = 303 303
Pcoupl                  = Berendsen
Pcoupltype              = isotropic
tau_p                   = 3.0 3.0
compressibility         = 3e-4 3e-4
ref_p                   = 1.0  1.0

gen_vel                 = yes
gen_temp                = 303
gen_seed                = 474529

constraints             = none
constraint_algorithm    = Lincs
unconstrained_start     = no
lincs_order             = 4
lincs_warnangle         = 30
```

EXAMPLE GAUSSIAN FREQUENCY CALCULATION INPUT

```
%nproc=8
%chk=fmoc-aa_opth.chk
%mem=16GB
#p B97D/SVP freq Denfit

Fmoc-AA monomer freq calc

0 1
 N    -0.637654    -0.882167     1.963521
 C     0.504337    -0.049468     1.647094
 C     1.587557    -0.265105     2.723000
 O     1.344508    -0.719057     3.839804
 H    -0.724796    -1.224613     2.919263
 H     0.901415    -0.356938     0.659281
 C     0.131011     1.450840     1.571296
 H    -0.679043     1.579582     0.832633
 H    -0.217864     1.802400     2.560460
 H     1.002428     2.059134     1.263259
 C    -1.610829    -1.121395     1.038362
 N     2.826021     0.155071     2.336268
 C     3.949394     0.228787     3.249494
 C     4.994356     1.195675     2.649950
 O     4.899466     1.694624     1.553321
 H     2.967056     0.571504     1.415778
 H     3.593169     0.656332     4.211681
 C     4.554436    -1.162725     3.549366
 H     3.756723    -1.803150     3.963254
 H     4.952867    -1.618073     2.624140
 H     5.371048    -1.099399     4.295242
 O     6.066696     1.446950     3.436748
 O    -1.588816    -0.732802    -0.116854
 O    -2.615159    -1.866892     1.609103
 C    -3.837927    -1.999540     0.871190
 C    -4.955255    -1.142723     1.511268
 H    -4.120618    -3.067891     0.907454
 H    -3.683074    -1.696976    -0.179770
 H    -8.728914    -3.083069    -0.732495
 C    -8.118680    -2.208752    -0.472079
 C    -8.471828    -0.938215    -0.968508
 H    -9.354123    -0.834370    -1.613357
 C    -7.707043     0.196814    -0.645863
 H    -7.984564     1.185092    -1.034107
 C    -6.581402     0.042612     0.181547
 C    -6.227332    -1.237832     0.682327
 C    -6.989415    -2.364820     0.357261
 H    -6.719455    -3.356531     0.743567
 C    -5.614987     1.026411     0.688256
 C    -5.536512     2.417931     0.504698
 H    -6.271561     2.942836    -0.118723
 C    -4.500936     3.127046     1.137900
 H    -4.424490     4.213931     1.004885
 C    -3.563302     2.456310     1.947811
 H    -2.763736     3.025115     2.439215
 C    -3.640964     1.062031     2.133036
 H    -2.907655     0.545594     2.761665
 C    -4.662748     0.350170     1.495321
 H    -5.112485    -1.504616     2.548363
 H     5.985012     0.972607     4.281580
```

EXAMPLE TURBOMOLE GEOMETRY OPTIMISATION INPUT

```
$title
fmoc-aa-h
$operating system unix
$symmetry c1
$redundant    file=coord
$maxcor 1600
$coord    file=coord
$user-defined bonds    file=coord
$atoms
n  1,12                                                                    \
   basis =n def2-SVP                                                       \
   jbas  =n def2-SVP
c  2-3,7,11,13-14,18,25-26,30-31,33,35-37,39-40,42,44,46,48                \
   basis =c def2-SVP                                                       \
   jbas  =c def2-SVP
o  4,15,22-24                                                              \
   basis =o def2-SVP                                                       \
   jbas  =o def2-SVP
h  5-6,8-10,16-17,19-21,27-29,32,34,38,41,43,45,47,49-50                   \
   basis =h def2-SVP                                                       \
   jbas  =h def2-SVP
$basis    file=basis
$rundimensions
   dim(fock,dens)=141369
   natoms=50
   nshell=234
   nbf(CAO)=530
   nbf(AO)=502
   dim(trafo[SAO<-->AO/CAO])=586
   rhfshells=1
$scfmo   file=mos
$closed shells
 a       1-101                                  ( 2 )
$scfiterlimit       300
$thize     0.10000000E-04
$thime        5
$scfdump
$scfintunit
 unit=30       size=0        file=twoint
$scfdiis
$scforbitalshift  automatic=.1
$drvopt
   cartesian  on
   basis      off
   global     off
   hessian    on
   dipole     on
   nuclear polarizability
$interconversion  off
   qconv=1.d-7
   maxiter=25
$optimize
   internal   off
   redundant  off
   cartesian  on
   global     off
   basis      off   logarithm
$coordinateupdate
   dqmax=0.3
   interpolate  on
   statistics    5
$forceupdate
   ahlrichs numgeo=0  mingeo=3 maxgeo=4 modus=<g|dq> dynamic fail=0.3
   threig=0.005  reseig=0.005  thrbig=3.0  scale=1.00  damping=0.0
$forceinit on
   diag=default
$energy    file=energy
```

```
$grad    file=gradient
$forceapprox    file=forceapprox
$lock off
$dft
   functional b-p
   gridsize   m3
$disp
$scfconv   6
$scfdamp   start=0.700  step=0.050  min=0.050
$tmole
$ricore     2000
$rij
$jbas    file=auxbasis
$marij
$last step     tm2molden
$end
```

Appendix 2: DFT frequency calculation results for additional Fmoc-AA and Fmc-AA conformations

Additional results discussed in the Chapter 4.

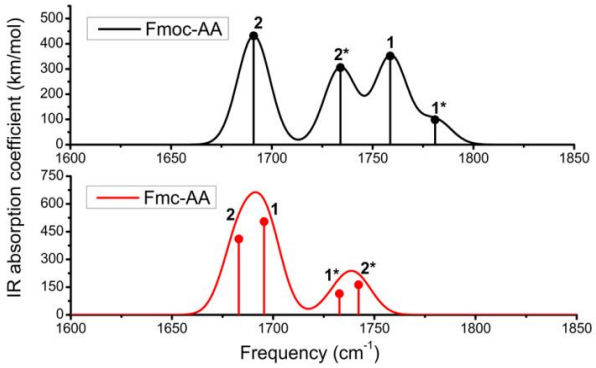


Fig. A2.1 Simulated Amide I spectra for Fmoc-AA and Fmc-AA parallel dimers. Spectra are generated applying an 8 cm^{-1} linewidth Gaussian to the calculated normal modes (vertical droplines). Amide I modes arising from amide groups with their carbonyl group pointing out of the stack are indicated with an asterisk (*).

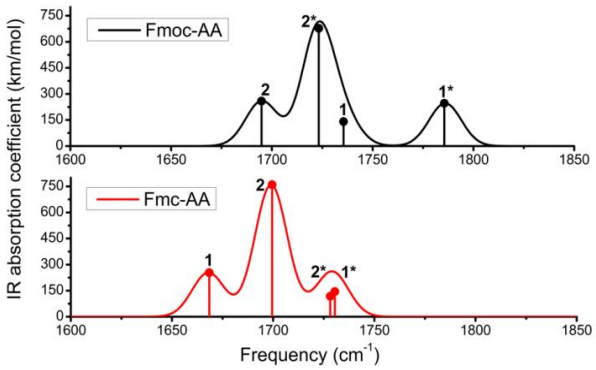


Fig. A2.2 Simulated Amide I spectra for Fmoc-AA and Fmc-AA antiparallel (ap1) dimers. Spectra are generated applying an 8 cm^{-1} linewidth Gaussian to the calculated normal modes (vertical droplines). Amide I modes arising from amide groups with their carbonyl group pointing out of the stack are indicated with an asterisk (*).

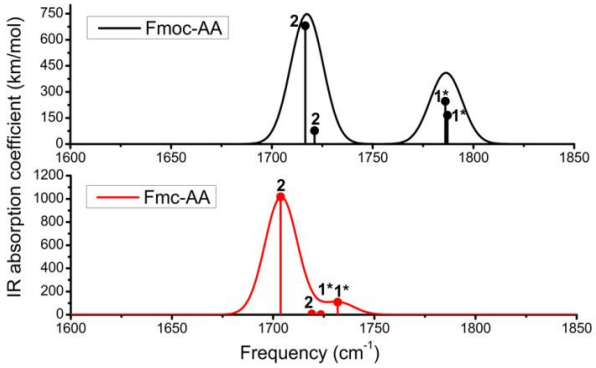


Fig. A2.3. Simulated Amide I spectra for Fmoc-AA and Fmc-AA antiparallel (ap2a) dimers. Spectra are generated applying an 8 cm^{-1} linewidth Gaussian to the calculated normal modes (vertical droplines). Amide I modes arising from amide groups with their carbonyl group pointing out of the stack are indicated with an asterisk (*).

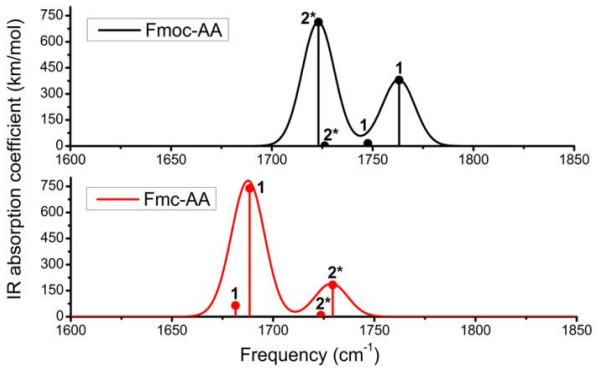


Fig. A2.4. Simulated Amide I spectra for Fmoc-AA and Fmc-AA antiparallel (ap2b) dimers. Spectra are generated applying an 8 cm^{-1} linewidth Gaussian to the calculated normal modes (vertical droplines). Amide I modes arising from amide groups with their carbonyl group pointing out of the stack are indicated with an asterisk (*).

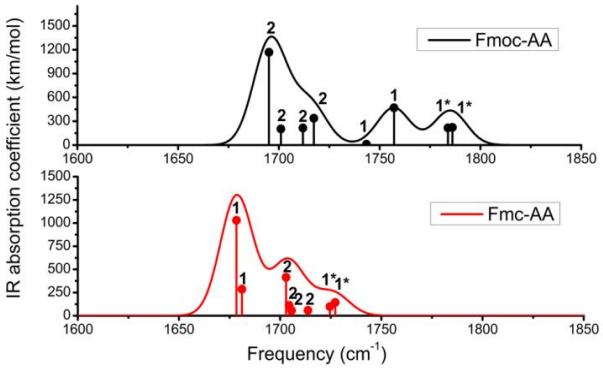


Fig. A2.5. Simulated Amide I spectra for Fmoc-AA and Fmc-AA antiparallel (ap2a) tetramers. Spectra are generated applying an 8 cm^{-1} linewidth Gaussian to the calculated normal modes (vertical droplines). Amide I modes arising from amide groups with their carbonyl group pointing out of the stack are indicated with an asterisk (*).

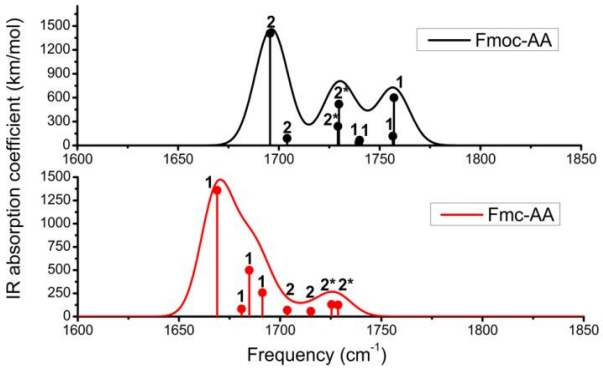
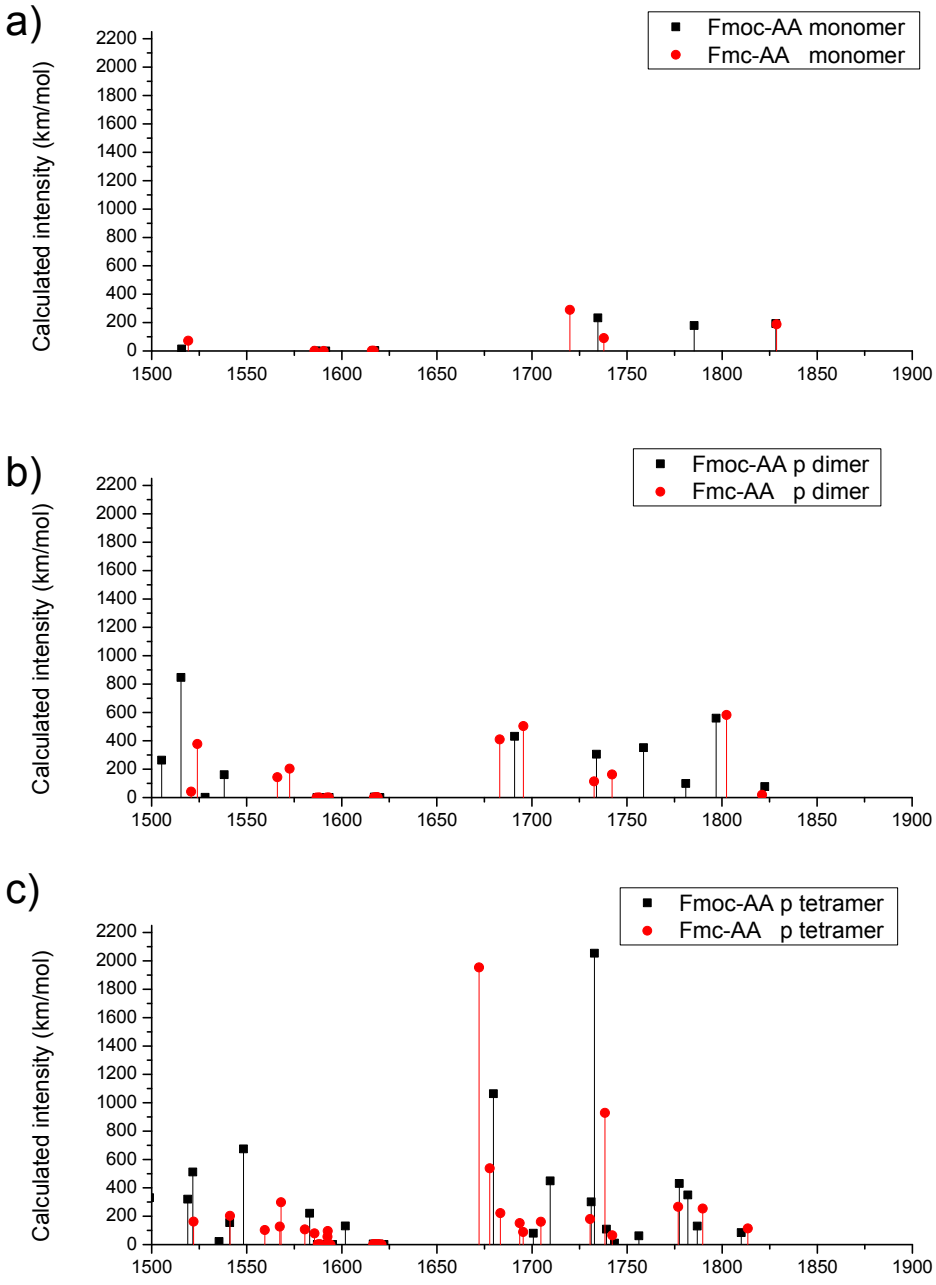
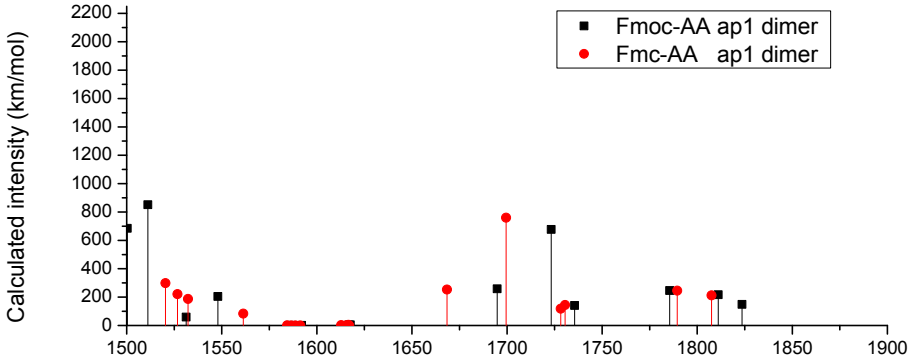


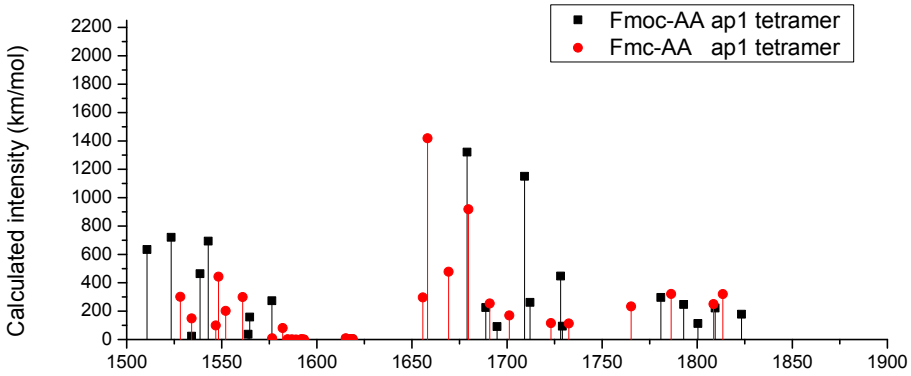
Fig. A2.6 Simulated Amide I spectra for Fmoc-AA and Fmc-AA antiparallel (ap2b) tetramers. Spectra are generated applying an 8 cm^{-1} linewidth Gaussian to the calculated normal modes (vertical droplines). Amide I modes arising from amide groups with their carbonyl group pointing out of the stack are indicated with an asterisk (*).


a)
Fmoc-AA monomer
Fmc-AA monomer
Calculated intensity (km/mol)
b)
Fmoc-AA p dimer
Fmc-AA p dimer
Calculated intensity (km/mol)
c)
Fmoc-AA p tetramer
Fmc-AA p tetramer
Calculated intensity (km/mol)

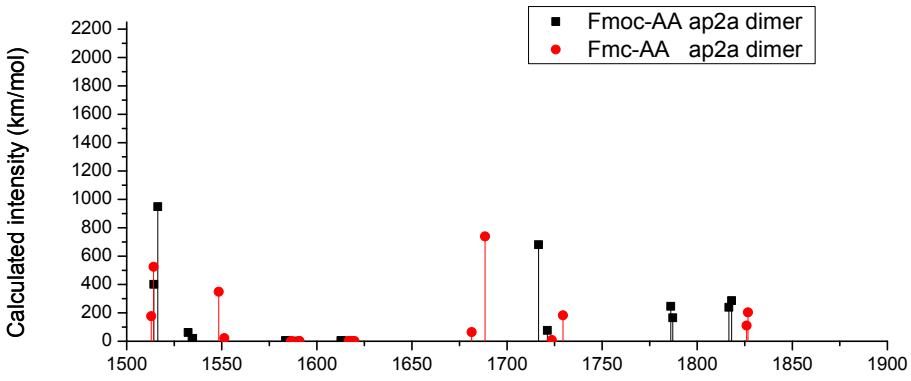
d)


Fmoc-AA ap1 dimer
Fmc-AA ap1 dimer
Calculated intensity (km/mol)
2200
2000
1800
1600
1400
1200
1000
800
600
400
200
0
1500
1550
1600
1650
1700
1750
1800
1850
1900

e)


Fmoc-AA ap1 tetramer
Fmc-AA ap1 tetramer
Calculated intensity (km/mol)
2200
2000
1800
1600
1400
1200
1000
800
600
400
200
0
1500
1550
1600
1650
1700
1750
1800
1850
1900

f)


Fmoc-AA ap2a dimer
Fmc-AA ap2a dimer
Calculated intensity (km/mol)
2200
2000
1800
1600
1400
1200
1000
800
600
400
200
0
1500
1550
1600
1650
1700
1750
1800
1850
1900

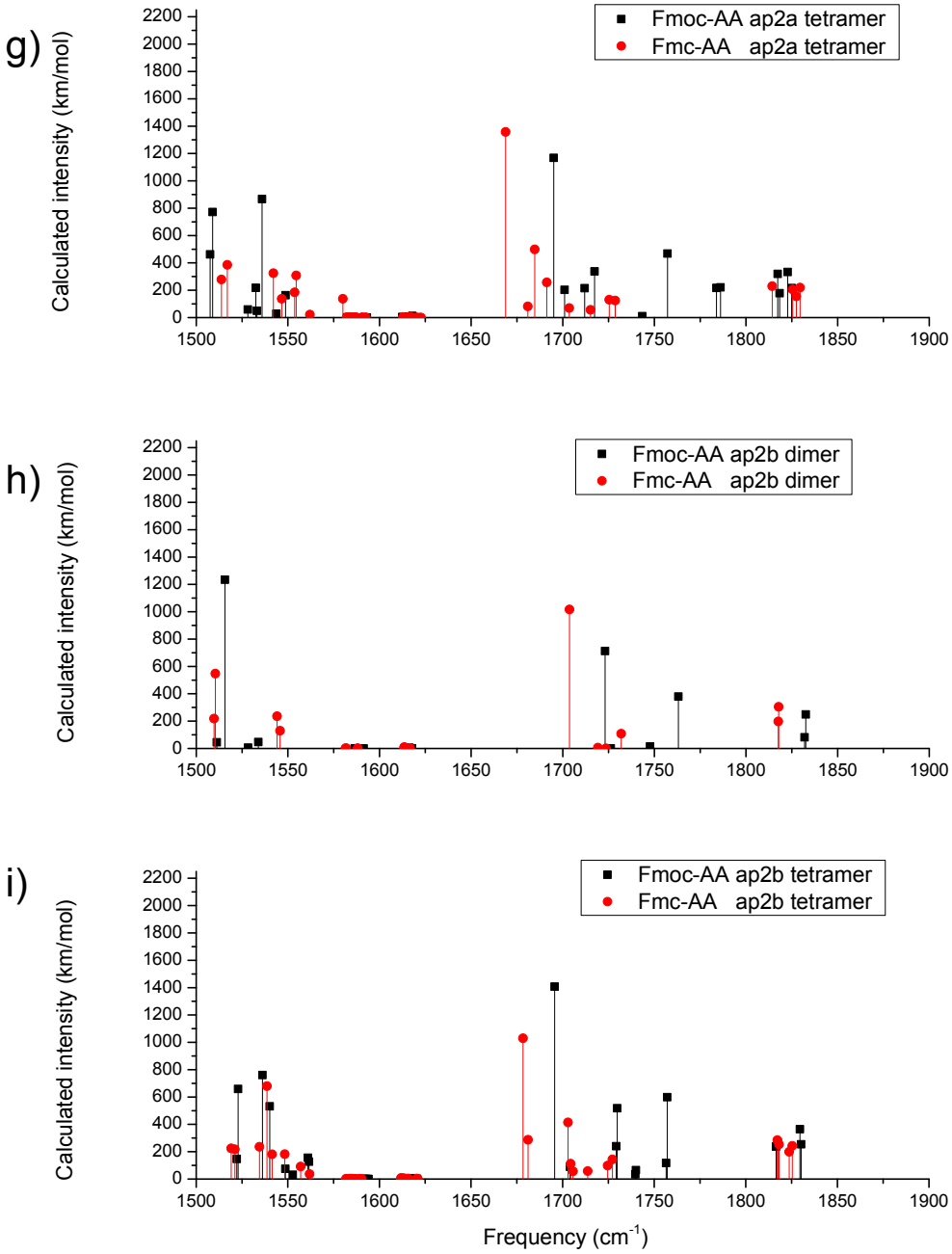


Fig. A2.7. Simulated stick spectra for Fmoc-AA and Fmc-AA monomer, dimer and tetramer models. No modes have been omitted from these spectra. a) monomers, b) parallel dimers, c) parallel tetramers, d) antiparallel 1 dimers, e) antiparallel 1 tetramers, f) antiparallel 2a dimers, g) antiparallel 2a tetramers, h) antiparallel 2b dimers, i) antiparallel 2b tetramers.

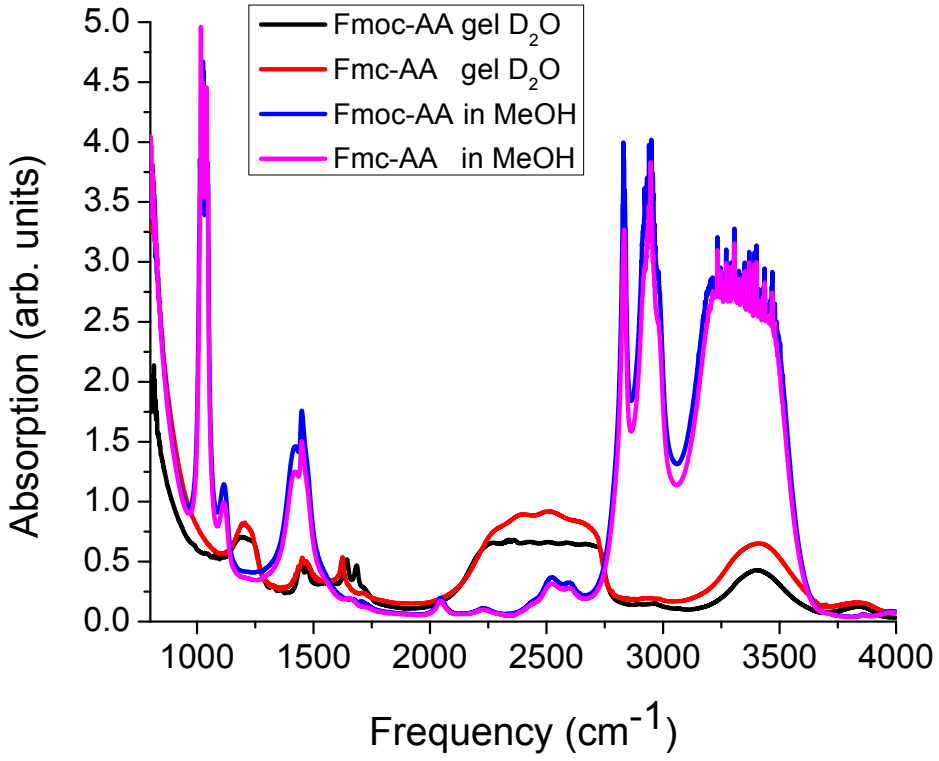


Fig. A2.8. FTIR spectra of 20 mM Fmoc-AA and Fmc-AA deuterated hydrogels and methanol solution in the range 800-4000 cm^{-1}.

Fig. A2.9. B97D/def2-SVP optimised structures of Fmoc-AA (left) and Fmc-AA monomers.

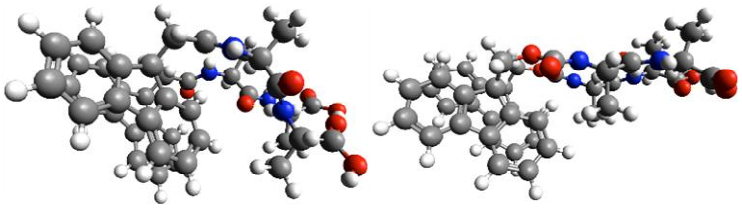


Fig. A2.10. B97D/def2-SVP optimised structures of Fmoc-AA (left) and Fmc-AA parallel dimers.

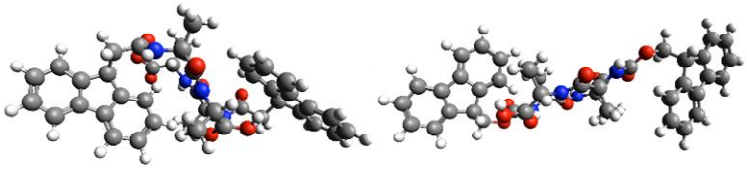


Fig. A2.11. B97D/def2-SVP optimised structures of Fmoc-AA (left) and Fmc-AA antiparallel 1 dimers.

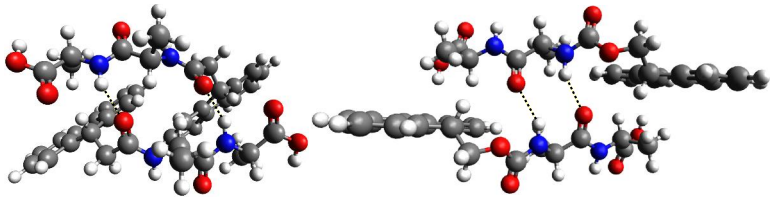


Fig. A2.12. B97D/def2-SVP optimised structures of Fmoc-AA (left) and Fmc-AA antiparallel 2a dimers.

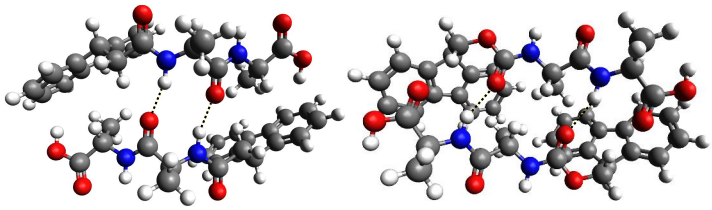


Fig. A2.13. B97D/def2-SVP optimised structures of Fmoc-AA (left) and Fmc-AA antiparallel 2b dimers.

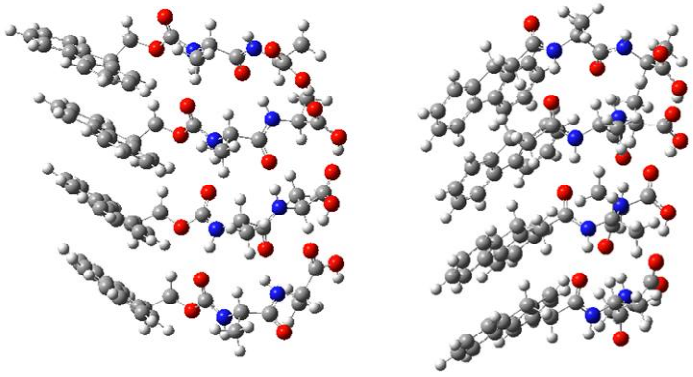


Fig. A2.14. B97D/def2-SVP optimised structures of Fmoc-AA (left) and Fmc-AA parallel tetramers.

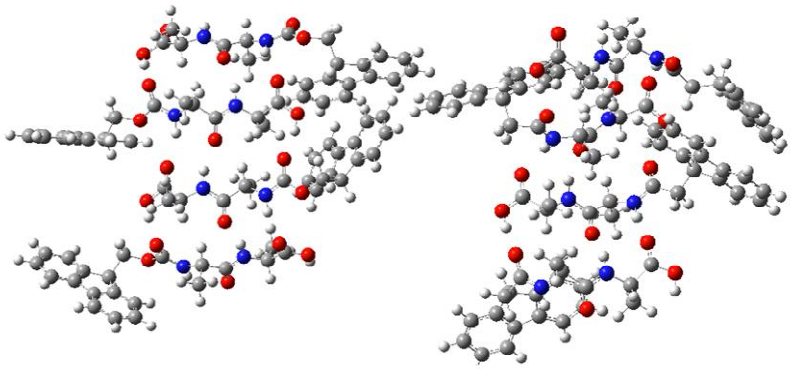


Fig. A2.15. B97D/def2-SVP optimised structures of Fmoc-AA (left) and Fmc-AA antiparallel 1 tetramers.

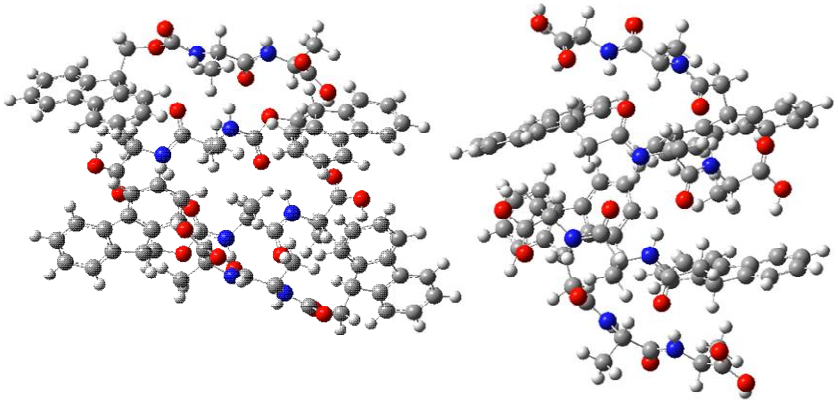


Fig. A2.16. B97D/def2-SVP optimised structures of Fmoc-AA (left) and Fmc-AA antiparallel 2a tetramers.

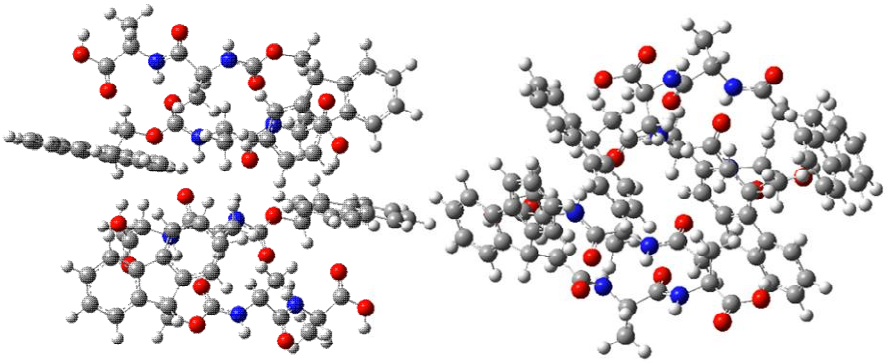


Fig. A2.17. B97D/def2-SVP optimised structures of Fmoc-AA (left) and Fmc-AA antiparallel 2b tetramers.

Appendix 3: Additional TEM images of Fmoc-SF-OMe nanostructures

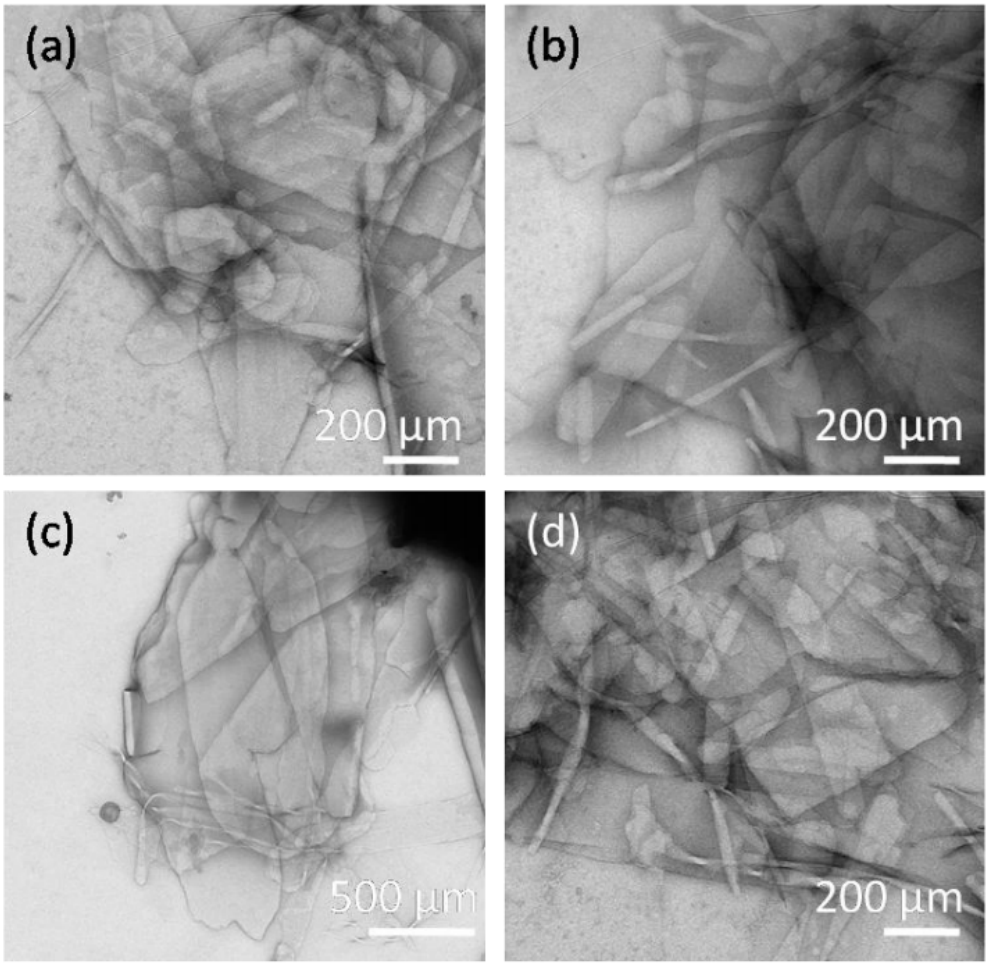


Fig. A3.1. TEM images of Fmoc-SF-OMe samples at t = 72 h. Extended (>200 µm) flat, sheet-like layers can be observed. Layers can stack to form spherulitic structures.

Appendix 4: Simulations results SF, SL, TL and TF

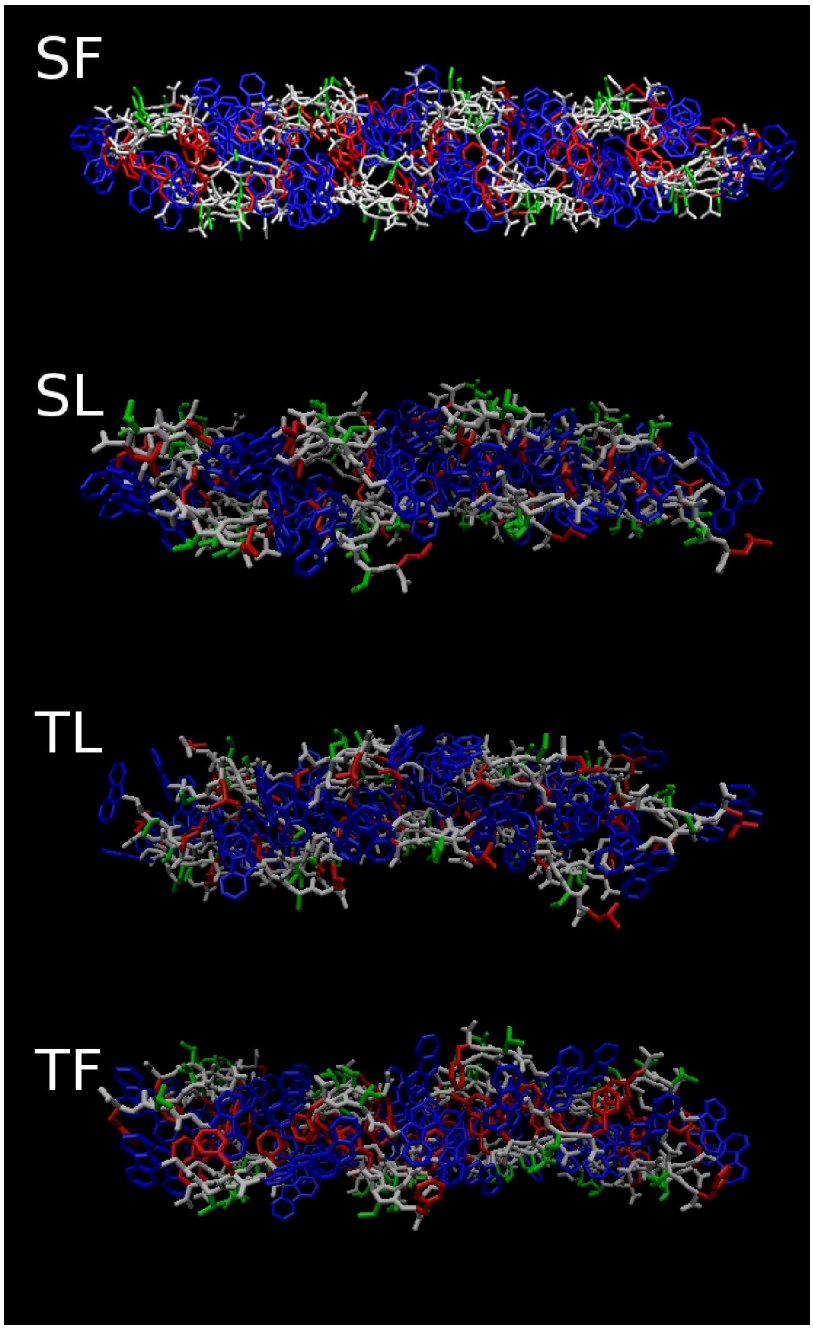


Fig. A4.1 Energy minimization results for SF, SL, TL and TF, top view.

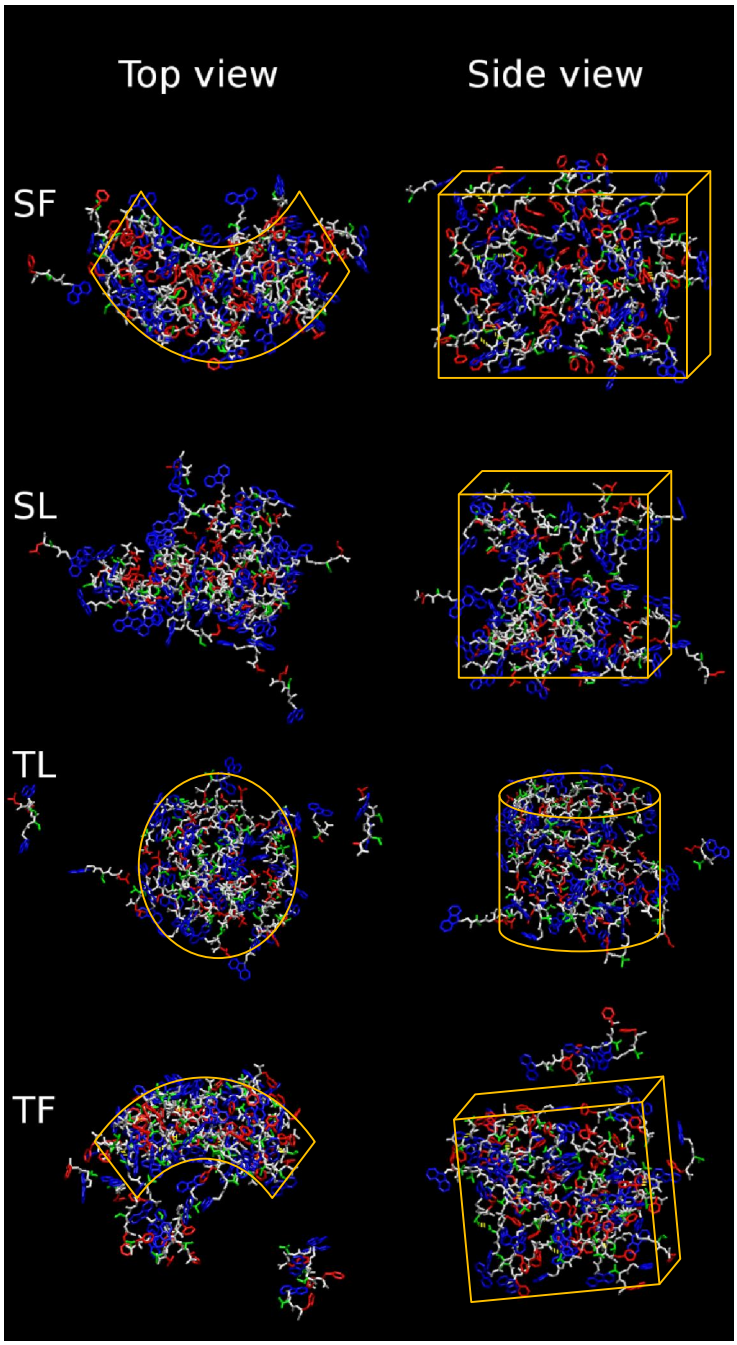


Fig. A4.2 Results of 100 ns simulations of SF, SL, TL and TF. Top view: from original sheet top view (see A3.1). Side view at 90° angle. As perspective is hard to gauge in the snapshots, various shapes aiding the eye have been added.

Appendix 5: MD parameters for Fmoc-group

A5.1 Topology

```
RESI FMO         0.00 !
!KEVO: charges adjusted to reflect updated acetone L-J.
!Resulting charges are unlikely to be good ==> re-optimize?
!Created by Pim Frederix May 24, 2010 PARTLY FROM FLUORENE
!Adapted after suggestions by Kenno (CHARMM forum)

GROUP
ATOM C1   C2R61  -0.115
ATOM H1   HGR61   0.115
ATOM C2   C2R61  -0.115 ! CHARGES SUGGESTED BY KENNO
ATOM H2   HGR61   0.115
ATOM C3   C2R61  -0.115
ATOM H3   HGR61   0.115
ATOM C4   C2R61  -0.115 !       H4          H7
ATOM H4   HGR61   0.115 !       |           |
ATOM C5   C2R67   0.000 !       C4          C7
ATOM C6   C2R67   0.000 !      //  \        /  \\
ATOM C7   C2R61  -0.115 ! H3--C3    C5----C6    C8--H8
ATOM H7   HGR61   0.115 !     |     ||    ||     |
ATOM C8   C2R61  -0.115 ! H2--C2    C13   C11   C9--H9
ATOM H8   HGR61   0.115 !      \\  /  \  /  \  //
ATOM C9   C2R61  -0.115 !       C1    C12   C10
ATOM H9   HGR61   0.115 !       |    /  \    |
ATOM C10  C2R61  -0.115 !      H1  H21   |  H10
ATOM H10  HGR61   0.115 !               |
ATOM C11  C2RC0   0.000 !          HF1--CF1--HF2
ATOM C12  C3C52  -0.090 !               |
ATOM H21  HGA2    0.090 !              OF1
ATOM C13  C2RC0   0.000 !               |
                                    !           C==OF2
                                       !        |
                                       !        TO AMINO ACID RESIDUE NH
GROUP
ATOM CF1  CG321    0.100              !By Pim
ATOM HF1  HGA2     0.090
ATOM HF2  HGA2     0.090
ATOM OF1  OG302   -0.490
ATOM C    CG2O1    0.840
ATOM OF2  OG2D1   -0.630

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 ! Deliberate distorsion!
IC C5   C11 *C6   C7   0.0000   0.00 180.00   0.00  0.0000 ! Deviating def to make above work
IC C7   C6   C11  C10  0.0000   0.00   0.00   0.00  0.0000 ! Deviating def to make above work
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
```

A5.2 Bond, angle and dihedral parameters not available in CHARMM27/36

```
BONDS
CG2O1   OG302    340.00     1.4300
CG2O1   NH1      370.00     1.3450
CG321   C3C52    195.00     1.5180

ANGLES
C2R67   C2R61   CG311     45.80     122.30
C2R61   CG311   C2R61     76.00     107.60
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
CG2O1   NH1     CT1       50.00     120.00
H       NH1     H         23.00     120.00

DIHEDRALS
C2R67   C2R61   CG311   C2R61       3.1000  2   180.00
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   OG302       0.1950  3     0.00
C3C52   CG321   OG302   CG2O1       0.0000  3     0.00
```

```
HGA2   C3C52  CG321  HGA2       0.1400  3     0.00
HGA2   C3C52  CG321  OG302      0.1400  3     0.00
CG321  OG302  CG2O1  OG2D1      0.9000  1     0.00
CG321  OG302  CG2O1  OG2D1      2.8500  2   180.00
CG321  OG302  CG2O1  NH1        0.6000  1   180.00
CG321  OG302  CG2O1  NH1        2.0000  2   180.00
HGA2   CG321  OG302  CG2O1      2.0500  2   180.00
OG302  CG2O1  NH1    H          2.5000  2   180.00
OG302  CG2O1  NH1    CT1        1.6000  1   180.00
CG2O1  NH1    CT1    CT3        0.0000  1   180.00
CG2O1  NH1    CT1    CT2        0.0000  1   180.00
CG2O1  NH1    CT1    C          0.0000  1     0.00
CG2O1  NH1    CT1    HB         0.0000  1   180.00
CG2O1  NH1    CT1    CD         0.0000  1   180.00
CG2O1  NH1    CT1    CC         0.0000  1   180.00
CG2O1  NH1    CT1    CT1        0.0000  1     0.00
OG2D1  CG2O1  NH1    H          2.5000  2   180.00
```

Appendix 6: AP and AP_H scores

A6.1 Validation of using the sum of the Wimley-White hydrophobicity components as a measure for the hydrophobicity of a peptide.

Considering the equilibrium

Amino acid (water) ⇆ *Amino acid (octanol)*

An equilibrium constant K_{eq}, in this case defined as the partition coefficient P can be written up as

$$K_{eq} = \frac{[Amino-acid]_{oct}}{[Amino-acid]_{wat}} \equiv P ,$$

In chemical equilibrium, the standard Gibbs free energy change $\Delta G°$ can be written as

$$\Delta G° = -RT \ln K_{eq}$$

Showing that the standard free energy change is linearly related to the logarithm of P

$$\Delta G° = \frac{-RT}{\log e} \log P$$

A6.2 AP scores for all dipeptides

1st / 2nd	A	C	D	E	F	G	H	I	K	L	M	N	P	Q	R	S	T	V	W	Y
A	1.1	1.0	1.0	1.0	1.2	1.0	1.1	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.2	1.1
C	1.0	1.0	1.0	1.0	2.0	1.0	1.1	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.1	1.0	2.1	1.1
D	1.0	1.0	1.0	1.0	1.2	1.0	1.1	1.0	1.1	1.0	1.0	1.0	1.0	1.1	1.2	1.1	1.1	1.0	1.2	1.1
E	1.0	1.0	1.0	1.0	1.1	1.0	1.1	1.0	1.1	1.0	1.0	1.0	1.0	1.1	1.1	1.1	1.0	1.0	1.1	1.1
F	1.1	1.9	1.1	1.1	3.2	1.2	1.4	2.3	1.4	2.4	1.8	1.2	2.1	1.2	1.3	2.3	2.1	1.8	3.5	1.8
G	1.0	1.0	1.0	1.0	1.2	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.2	1.1
H	1.0	1.1	1.0	1.0	1.4	1.1	1.1	1.1	1.2	1.1	1.1	1.1	1.1	1.0	1.1	1.5	1.3	1.1	1.3	1.1
I	1.0	1.1	1.0	1.0	2.0	1.0	1.1	1.1	1.0	1.1	1.1	1.0	1.0	1.0	1.0	1.0	1.0	1.0	2.0	1.2
K	1.0	1.0	1.1	1.0	1.2	1.0	1.1	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.2	1.1
L	1.0	1.0	1.0	1.0	2.0	1.0	1.1	1.1	1.0	1.1	1.1	1.0	1.0	1.0	1.0	1.0	1.0	1.0	2.0	1.2
M	1.0	1.0	1.0	1.0	1.7	1.0	1.1	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.9	1.2
N	1.0	1.0	1.0	1.0	1.2	1.0	1.1	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.1	1.1	1.0	1.2	1.1
P	1.0	1.0	1.0	1.0	1.7	1.0	1.1	1.0	1.0	1.1	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	2.0	1.1
Q	1.0	1.0	1.0	1.0	1.1	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.2	1.0
R	1.0	1.0	1.0	1.0	1.2	1.0	1.1	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.2	1.1
S	1.0	1.0	1.0	1.0	1.6	1.0	1.7	1.0	1.1	1.0	1.0	1.1	1.0	1.1	1.1	1.6	1.2	1.0	2.4	1.7
T	1.0	1.0	1.0	1.0	2.3	1.0	1.3	1.0	1.1	1.0	1.0	1.1	1.0	1.1	1.1	1.4	1.1	1.0	1.9	1.5
V	1.0	1.0	1.0	1.0	1.9	1.0	1.1	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	2.1	1.2
W	1.1	2.4	1.1	1.1	3.5	1.2	1.3	2.2	1.4	1.9	1.8	1.2	2.0	1.2	1.4	2.6	2.4	1.8	3.2	2.2
Y	1.1	1.2	1.1	1.0	2.2	1.1	1.2	1.2	1.2	1.2	1.2	1.1	1.1	1.1	1.1	1.7	1.4	1.2	2.1	1.2

A6.3 Tripeptides scoring AP > 2

pep	AP	logP	AP_H
PFF	2.39	-3.28	0.17
WFL	2.38	-5.05	0.07
MFF	2.36	-4.09	0.12
VFF	2.33	-3.88	0.13
FFM	2.31	-4.09	0.11
FWF	2.28	-5.51	0.04
FFF	2.26	-5.13	0.06
WWF	2.26	-5.89	0.02
FWI	2.26	-4.92	0.07
FYI	2.22	-3.54	0.12
VFW	2.22	-4.26	0.09
PWF	2.21	-3.66	0.12
IFF	2.21	-4.54	0.08
LCF	2.18	-2.98	0.14
WLL	2.18	-4.59	0.07
SFW	2.18	-3.34	0.12
IFW	2.17	-4.92	0.06
WFF	2.17	-5.51	0.03
FFW	2.17	-5.51	0.03
IMW	2.15	-3.88	0.10
FWP	2.15	-3.66	0.11
WIW	2.15	-5.30	0.04
CLW	2.15	-3.36	0.12
YWF	2.14	-4.51	0.07
GFF	2.14	-2.27	0.16
WFT	2.13	-3.55	0.11
WFM	2.13	-4.47	0.07
IWF	2.13	-4.92	0.05
WFW	2.12	-5.89	0.01
MFY	2.12	-3.09	0.12
VWF	2.12	-4.26	0.08
PMW	2.11	-2.62	0.14
WFI	2.11	-4.92	0.05
VAW	2.10	-2.05	0.16
IVW	2.10	-3.67	0.10
FFY	2.10	-4.13	0.08
SWW	2.10	-3.72	0.09
FFC	2.10	-3.44	0.10
MIF	2.10	-3.50	0.10
FWW	2.09	-5.89	0.01
MFW	2.09	-4.47	0.07
IWI	2.09	-4.33	0.07
TWF	2.09	-3.55	0.10
FHF	2.09	-3.31	0.11
ISW	2.08	-2.75	0.13
WMF	2.08	-4.47	0.06
IYW	2.08	-3.92	0.08
VVF	2.08	-2.63	0.13
PVF	2.08	-2.03	0.15
CWF	2.08	-3.82	0.09
YFF	2.08	-4.13	0.08
FLF	2.08	-4.67	0.06
PCF	2.08	-1.59	0.17
WCF	2.07	-3.82	0.09
LPF	2.07	-2.82	0.12
FVV	2.07	-2.63	0.13
TFF	2.07	-3.17	0.11
LFF	2.07	-4.67	0.06
FWV	2.06	-4.26	0.07
SCW	2.06	-1.65	0.16
IIF	2.06	-3.95	0.08
FCW	2.06	-3.82	0.08
WFP	2.06	-3.66	0.09
FYF	2.06	-4.13	0.07
WVF	2.06	-4.26	0.07
YFM	2.06	-3.09	0.11
YLW	2.06	-4.05	0.08
WFG	2.05	-2.65	0.12
CFF	2.05	-3.44	0.10
VPF	2.05	-2.03	0.14
VMW	2.05	-3.22	0.10
TYF	2.05	-2.17	0.14
VWW	2.05	-4.64	0.06
CPW	2.05	-1.97	0.15
FVF	2.04	-3.88	0.08
YFW	2.04	-4.51	0.06
FFT	2.04	-3.17	0.10
VYW	2.04	-3.26	0.10
YFC	2.04	-2.44	0.13

IWW	2.04	-5.30	0.03
WCW	2.04	-4.20	0.07
VYF	2.04	-2.88	0.11
ILF	2.04	-4.08	0.07
GFW	2.04	-2.65	0.12
SFF	2.04	-2.96	0.11
WPW	2.03	-4.04	0.07
IFT	2.03	-2.58	0.12
WLT	2.03	-3.09	0.10
FFV	2.03	-3.88	0.08
PWI	2.03	-3.07	0.10
WLC	2.03	-3.36	0.09
VWV	2.03	-3.01	0.11
LWY	2.03	-4.05	0.07
PLW	2.02	-3.20	0.10
LYW	2.02	-4.05	0.07
IIW	2.02	-4.33	0.06
HFF	2.02	-3.31	0.09
CWM	2.02	-2.78	0.11
LWV	2.02	-3.80	0.08
TFW	2.02	-3.55	0.09
IYY	2.02	-2.54	0.12
FFA	2.02	-2.92	0.11
FWY	2.02	-4.51	0.06
TFV	2.02	-1.92	0.14
FWC	2.02	-3.82	0.08
MWW	2.01	-4.85	0.05
TCW	2.01	-1.86	0.14
CMF	2.01	-2.40	0.12
IFP	2.01	-2.69	0.11
LIF	2.01	-4.08	0.07
MYF	2.01	-3.09	0.10
PWL	2.01	-3.20	0.10
FLP	2.01	-2.82	0.11
CFV	2.01	-2.19	0.13
FYL	2.01	-3.67	0.08
WVW	2.00	-4.64	0.05
FFS	2.00	-2.96	0.10
PFY	2.00	-2.28	0.12
AFW	2.00	-3.30	0.09
LWI	2.00	-4.46	0.06
LYF	2.00	-3.67	0.08
LWF	2.00	-5.05	0.04
IFY	2.00	-3.54	0.08
KWF	2.00	-1.00	0.16

Top 400 AP_H

pep	AP	logP	AP_H
KFD	1.73	4.73	0.19
KWD	1.74	4.35	0.19
HKD	1.65	6.55	0.18
PFF	2.39	-3.28	0.17
KWE	1.72	4.34	0.17
WKE	1.71	4.34	0.17
KHD	1.63	6.55	0.17
PCF	2.08	-1.59	0.17
KWF	2.00	-1.00	0.16
KFW	2.00	-1.00	0.16
KHE	1.62	6.54	0.16
TSF	1.99	-1.00	0.16
SCW	2.06	-1.65	0.16
WKD	1.69	4.35	0.16
KYD	1.64	5.73	0.16
KEH	1.62	6.54	0.16
GFF	2.14	-2.27	0.16
VAW	2.10	-2.05	0.16
SSF	1.96	-0.79	0.15
KYE	1.63	5.72	0.15
STF	1.97	-1.00	0.15
SRD	1.62	5.91	0.15
KYY	1.79	1.38	0.15
PVF	2.08	-2.03	0.15
SKD	1.59	6.90	0.15
TFP	2.00	-1.32	0.15
RHD	1.63	5.56	0.15
KYF	1.85	0.38	0.15
FKF	1.92	-0.62	0.15
KFF	1.92	-0.62	0.15
WFD	1.88	-0.16	0.15
WRD	1.69	3.36	0.15
FKD	1.65	4.73	0.15
PHF	2.00	-1.46	0.15
KFY	1.84	0.38	0.15
KFE	1.64	4.72	0.15
PPF	1.99	-1.43	0.15
SFE	1.73	2.38	0.15
SNH	1.78	1.42	0.15
WKF	1.94	-1.00	0.15
CPW	2.05	-1.97	0.15
SYS	1.84	0.21	0.14
VPF	2.05	-2.03	0.14
PYY	1.96	-1.28	0.14
YFK	1.83	0.38	0.14
TKD	1.58	6.69	0.14
FKY	1.82	0.38	0.14
KDF	1.63	4.73	0.14
LCF	2.18	-2.98	0.14
TCW	2.01	-1.86	0.14
GFY	1.95	-1.27	0.14
TGF	1.87	-0.31	0.14
SWE	1.73	2.00	0.14
STY	1.84	0.00	0.14
TYF	2.05	-2.17	0.14
PMW	2.11	-2.62	0.14
PSF	1.93	-1.11	0.14
TFV	2.02	-1.92	0.14
SKW	1.77	1.17	0.14
SFD	1.71	2.39	0.14
SPF	1.93	-1.11	0.14
PGW	1.90	-0.80	0.14
FFD	1.82	0.22	0.14
KFT	1.75	1.34	0.14
PPW	1.99	-1.81	0.14
RWD	1.66	3.36	0.14
TRD	1.59	5.70	0.14
PFT	1.94	-1.32	0.14
FCC	1.98	-1.75	0.14
RFY	1.87	-0.61	0.13
PIY	1.97	-1.69	0.13
RFD	1.64	3.74	0.13
SHK	1.66	3.37	0.13
SHN	1.74	1.42	0.13
YKD	1.58	5.73	0.13
SGF	1.83	-0.10	0.13
HRD	1.58	5.56	0.13
VVF	2.08	-2.63	0.13
TSH	1.76	0.82	0.13

FRY	1.86	-0.61	0.13	VFE	1.70	1.46	0.12
TTF	1.91	-1.21	0.13	KDY	1.55	5.73	0.12
FKW	1.89	-1.00	0.13	CMF	2.01	-2.40	0.12
YPV	1.89	-1.03	0.13	PSY	1.79	-0.11	0.12
PTF	1.91	-1.32	0.13	YGF	1.88	-1.27	0.12
VFF	2.33	-3.88	0.13	MFY	2.12	-3.09	0.12
CFV	2.01	-2.19	0.13	MFF	2.36	-4.09	0.12
SFY	1.98	-1.96	0.13	TFD	1.67	2.18	0.12
PCW	1.98	-1.97	0.13	SHD	1.59	4.21	0.12
CSY	1.82	-0.27	0.13	SWK	1.71	1.17	0.12
YFC	2.04	-2.44	0.13	RDF	1.61	3.74	0.12
SWP	1.93	-1.49	0.13	IFT	2.03	-2.58	0.12
FVV	2.07	-2.63	0.13	LPF	2.07	-2.82	0.12
SGW	1.84	-0.48	0.13	VCF	1.98	-2.19	0.12
ISW	2.08	-2.75	0.13	LFD	1.74	0.68	0.12
CFC	1.95	-1.75	0.13	KPF	1.71	1.23	0.12
IFG	1.94	-1.68	0.13	TYD	1.63	3.18	0.12
KVW	1.78	0.25	0.13	RYD	1.58	4.74	0.12
YFD	1.73	1.22	0.13	TKF	1.70	1.34	0.12
FDF	1.78	0.22	0.13	SVW	1.96	-2.09	0.12
SYH	1.81	-0.14	0.13	TWC	1.94	-1.86	0.12
SHF	1.88	-1.14	0.13	CTW	1.94	-1.86	0.12
VFG	1.87	-1.02	0.13	KWM	1.77	0.04	0.12
WRE	1.63	3.35	0.12	GFW	2.04	-2.65	0.12
KDW	1.60	4.35	0.12	GFT	1.80	-0.31	0.12
PYV	1.87	-1.03	0.12	KYW	1.78	0.00	0.12
TYT	1.81	-0.21	0.12	VAF	1.91	-1.67	0.12
FYI	2.22	-3.54	0.12	HRE	1.55	5.55	0.12
PFY	2.00	-2.28	0.12	IYY	2.02	-2.54	0.12
YPY	1.89	-1.28	0.12	PSW	1.89	-1.49	0.12
VFD	1.71	1.47	0.12	VWK	1.76	0.25	0.12
PFV	1.97	-2.03	0.12	SGY	1.72	0.90	0.12
SFW	2.18	-3.34	0.12	HKE	1.53	6.54	0.12
WFG	2.05	-2.65	0.12	AFY	1.94	-1.92	0.12
TFC	1.91	-1.48	0.12	SYY	1.84	-0.96	0.12
PFS	1.87	-1.11	0.12	RWE	1.61	3.35	0.12
STH	1.74	0.82	0.12	CYH	1.82	-0.62	0.12
FGS	1.79	-0.10	0.12	SCY	1.79	-0.27	0.12
VYY	1.95	-1.88	0.12	CFG	1.81	-0.58	0.12
FRD	1.61	3.74	0.12	KEW	1.58	4.34	0.12
VSW	1.97	-2.09	0.12	KWC	1.73	0.69	0.12

VFP	1.95	-2.03	0.12
SHS	1.71	1.03	0.12
SWH	1.89	-1.52	0.12
PLY	1.92	-1.82	0.12
SKE	1.52	6.89	0.12
TFT	1.86	-1.21	0.12
HKF	1.70	1.20	0.12
FTC	1.88	-1.48	0.12
TYI	1.89	-1.58	0.12
CLW	2.15	-3.36	0.12
FIS	1.98	-2.37	0.12
PWH	1.92	-1.84	0.12
VFK	1.73	0.63	0.12
KLF	1.77	-0.16	0.12
CFD	1.66	1.91	0.12
PWF	2.21	-3.66	0.12
PVW	1.99	-2.41	0.12
CYF	1.99	-2.44	0.12
PCY	1.80	-0.59	0.12
VFC	1.96	-2.19	0.12
TSY	1.76	0.00	0.12
WDF	1.77	-0.16	0.12
FMC	1.98	-2.40	0.12
LPY	1.91	-1.82	0.12
FKE	1.56	4.72	0.12
YFG	1.86	-1.27	0.12
THS	1.71	0.82	0.12
WEF	1.77	-0.17	0.12
YCC	1.81	-0.75	0.12
WSY	1.97	-2.34	0.12
LSF	1.99	-2.50	0.12
KDH	1.52	6.55	0.12
IYC	1.91	-1.85	0.12
FSM	1.92	-1.92	0.12
CPF	1.89	-1.59	0.11
SWT	1.87	-1.38	0.11
KMW	1.76	0.04	0.11
ASW	1.84	-1.13	0.11
FYT	1.95	-2.17	0.11
KIW	1.79	-0.41	0.11
THF	1.86	-1.35	0.11
PKW	1.71	0.85	0.11
MHF	1.96	-2.27	0.11
VWH	1.98	-2.44	0.11
VGW	1.86	-1.40	0.11
FYK	1.73	0.38	0.11
CWP	1.92	-1.97	0.11
KTF	1.68	1.34	0.11
EFW	1.77	-0.17	0.11
FFM	2.31	-4.09	0.11
RYF	1.80	-0.61	0.11
GFI	1.89	-1.68	0.11
FRE	1.59	3.73	0.11
PWV	1.97	-2.41	0.11
CGY	1.73	0.42	0.11
PFA	1.83	-1.07	0.11
SWC	1.89	-1.65	0.11
PFC	1.88	-1.59	0.11
KFH	1.69	1.20	0.11
VFS	1.89	-1.71	0.11
CFY	1.98	-2.44	0.11
VYV	1.88	-1.63	0.11
PTY	1.77	-0.32	0.11
IFP	2.01	-2.69	0.11
SHT	1.71	0.82	0.11
PSH	1.71	0.71	0.11
SCF	1.85	-1.27	0.11
KFV	1.71	0.63	0.11
REY	1.56	4.73	0.11
PHS	1.71	0.71	0.11
FFR	1.88	-1.61	0.11
VYF	2.04	-2.88	0.11
FFE	1.74	0.21	0.11
FSP	1.83	-1.11	0.11
KWT	1.69	0.96	0.11
CWM	2.02	-2.78	0.11
WPP	1.90	-1.81	0.11
YKF	1.73	0.38	0.11
TSW	1.85	-1.38	0.11
RHE	1.53	5.55	0.11
GFL	1.89	-1.81	0.11
KEY	1.53	5.72	0.11

FGF	1.95	-2.27	0.11	YYI	1.97	-2.54	0.11
FVS	1.88	-1.71	0.11	PWT	1.87	-1.70	0.11
CTF	1.86	-1.48	0.11	WHP	1.89	-1.84	0.11
WWK	1.85	-1.38	0.11	VKF	1.70	0.63	0.11
KIF	1.75	-0.03	0.11	PFG	1.76	-0.42	0.11
WFK	1.82	-1.00	0.11	SVY	1.78	-0.71	0.11
KWL	1.78	-0.54	0.11	TFF	2.07	-3.17	0.11
FWE	1.76	-0.17	0.11	FLP	2.01	-2.82	0.11
RFF	1.87	-1.61	0.11	CYC	1.79	-0.75	0.11
SFG	1.75	-0.10	0.11	CGF	1.77	-0.58	0.11
SFC	1.84	-1.27	0.11	RWY	1.81	-0.99	0.11
PHV	1.76	-0.21	0.11	PWC	1.90	-1.97	0.11
STS	1.68	1.17	0.11	PPY	1.76	-0.43	0.11
KHW	1.70	0.82	0.11	PYS	1.74	-0.11	0.11
IGY	1.79	-0.68	0.11	SWN	1.79	-0.78	0.11
VFH	1.92	-2.06	0.11	WEK	1.55	4.34	0.11
KSF	1.66	1.55	0.11	FHF	2.09	-3.31	0.11
CYS	1.76	-0.27	0.11	KEF	1.54	4.72	0.11
SKH	1.59	3.37	0.11	CKW	1.69	0.69	0.11
FYP	1.94	-2.28	0.11	RFE	1.57	3.73	0.11
YWD	1.70	0.84	0.11	KHF	1.67	1.20	0.11
TGS	1.65	1.86	0.11	SYT	1.73	0.00	0.11
VKW	1.73	0.25	0.11	TSS	1.67	1.17	0.11
TWD	1.65	1.80	0.11	FWD	1.74	-0.16	0.11
FYD	1.68	1.22	0.11	TKH	1.59	3.16	0.11
KFS	1.66	1.55	0.11	FRF	1.86	-1.61	0.11
RDW	1.59	3.36	0.11	SRE	1.51	5.90	0.11
SSY	1.73	0.21	0.11	SHP	1.69	0.71	0.11
PYT	1.76	-0.32	0.11	SGS	1.63	2.07	0.11
WYT	1.98	-2.55	0.11	SFH	1.81	-1.14	0.11
PAW	1.85	-1.45	0.11	WKY	1.73	0.00	0.11
FYC	1.96	-2.44	0.11	FFA	2.02	-2.92	0.11
WGT	1.79	-0.69	0.11	WFE	1.74	-0.17	0.11
VTW	1.94	-2.30	0.11	WFT	2.13	-3.55	0.11
FGC	1.78	-0.58	0.11	YFE	1.66	1.21	0.11
SFF	2.04	-2.96	0.11	GVF	1.80	-1.02	0.11
KSW	1.67	1.17	0.11	CCY	1.78	-0.75	0.11
YFM	2.06	-3.09	0.11	TFL	1.98	-2.71	0.11
WST	1.84	-1.38	0.11	SGH	1.64	1.72	0.11
IPY	1.87	-1.69	0.11	RYE	1.54	4.73	0.11
TWY	1.97	-2.55	0.11	IFA	1.93	-2.33	0.11

FGV	1.80	-1.02	0.11
FKS	1.65	1.55	0.11
VWV	2.03	-3.01	0.11
CYW	2.00	-2.82	0.11
YFS	1.89	-1.96	0.11
FSV	1.86	-1.71	0.11
TDW	1.63	1.80	0.11
SIY	1.83	-1.37	0.11
KWW	1.83	-1.38	0.11
FFG	1.92	-2.27	0.11
KIY	1.67	0.97	0.11
FWP	2.15	-3.66	0.11
CSF	1.82	-1.27	0.11
EKW	1.55	4.34	0.11
CWD	1.64	1.53	0.10
FSL	1.95	-2.50	0.10
FTP	1.82	-1.32	0.10
FFC	2.10	-3.44	0.10
TWG	1.77	-0.69	0.10
FTL	1.97	-2.71	0.10
TKY	1.61	2.34	0.10
PWI	2.03	-3.07	0.10
SRW	1.71	0.18	0.10
KFI	1.72	-0.03	0.10
WLT	2.03	-3.09	0.10
SMY	1.78	-0.92	0.10
HDK	1.49	6.55	0.10
THT	1.69	0.61	0.10
KSD	1.48	6.90	0.10
WKT	1.67	0.96	0.10
FYG	1.81	-1.27	0.10
GCW	1.79	-0.96	0.10
SRF	1.69	0.56	0.10
MFD	1.65	1.26	0.10
IKW	1.75	-0.41	0.10
WSS	1.80	-1.17	0.10
VHS	1.71	0.11	0.10
TTH	1.68	0.61	0.10
SYP	1.73	-0.11	0.10
WKS	1.66	1.17	0.10
FGI	1.85	-1.68	0.10
VMW	2.05	-3.22	0.10
TKW	1.67	0.96	0.10
SWA	1.80	-1.13	0.10
SYK	1.60	2.55	0.10
FFT	2.04	-3.17	0.10
SNY	1.68	0.60	0.10
SNF	1.74	-0.40	0.10
PAF	1.79	-1.07	0.10
GWY	1.84	-1.65	0.10
DKH	1.49	6.55	0.10
YCT	1.75	-0.48	0.10
CYT	1.75	-0.48	0.10
FFS	2.00	-2.96	0.10
SYF	1.87	-1.96	0.10
SVF	1.85	-1.71	0.10
WKI	1.74	-0.41	0.10
PIF	1.96	-2.69	0.10
AFC	1.80	-1.23	0.10
TYS	1.72	0.00	0.10
RHS	1.60	2.38	0.10
SYW	1.92	-2.34	0.10
KWH	1.67	0.82	0.10
PYM	1.80	-1.24	0.10
FPG	1.74	-0.42	0.10
STT	1.66	0.96	0.10
TFK	1.64	1.34	0.10
VYT	1.78	-0.92	0.10
FTI	1.95	-2.58	0.10
NSW	1.77	-0.78	0.10
TIY	1.83	-1.58	0.10
SFK	1.63	1.55	0.10
MIF	2.10	-3.50	0.10
CWC	1.89	-2.13	0.10

Appendix 7: Links to movies of full simulation results

1) 100 ns simulation of Fmoc-FF-OH and Fmoc-S-OH at atomistic level:

http://youtu.be/3iEFgHBwwIA

2) 200 ns simulation of Fmoc-FF-OH, Fmoc-FF-O^-, Fmoc-S-OH and Fmoc-S-O^- at atomistic level:

http://youtu.be/n29AR6eXK3s

3) 100 ns simulation of diphenylalanine at atomistic level:

http://youtu.be/cn-UaoNdgfc

4) 1500 ns simulation of diphenylalanine at coarse-grain level:

http://youtu.be/eQe5HrH-G5c