Virtual Screening for Dipeptide Aggregation: Toward Predictive Tools for Peptide Self-Assembly

Pim W. J. M. Frederix,†,‡ Rein V. Ulijn,† Neil T. Hunt,‡ and Tell Tuttle*†

†WestCHEM, Department of Pure and Applied Chemistry, University of Strathclyde, 295 Cathedral Street, Glasgow, G1 1XL, United Kingdom
‡SUPA, Department of Physics, University of Strathclyde, 107 Rottenrow East, Glasgow, G4 0NG, United Kingdom

ABSTRACT: Several short peptide sequences are known to self-assemble into supramolecular nanostructures with interesting properties. In this study, coarse-grained molecular dynamics is employed to rapidly screen all 400 dipeptide combinations and predict their ability to aggregate as a potential precursor to their self-assembly. The simulation protocol and scoring method proposed allows a rapid determination of whether a given peptide sequence is likely to aggregate (an indicator for the ability to self-assemble) under aqueous conditions. Systems that show strong aggregation tendencies in the initial screening are selected for longer simulations, which result in good agreement with the known self-assembly or aggregation of dipeptides reported in the literature. Our extended simulations of the diphenylalanine system show that the coarse-grain model is able to reproduce salient features of nanoscale systems and provide insight into the self-assembly process for this system.

SECTION: Biophysical Chemistry

Self-assembling peptide materials are currently widely studied for potential applications in biomedicine and nanotechnology. They can form various nanostructures from 1D fibers and tubes to sheets and 3D networks with tunable supramolecular functionalities such as electronic conductivity and bioactivity. These features, combined with their benign toxicology and ease of preparation, position peptide-based nanostructures at the forefront of next-generation, soft nanomaterials design. 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.

Dipeptides, which consist of only two amino acids, have been chosen to apply our virtual screening procedure to, as they are the smallest of known peptide self-assemblers. They are also of interest in their own right and have shown interesting morphologies and functionalities. The most famous and well-studied example is diphenylalanine (FF), as first described by Reches and Gazit in 2003 and recently reviewed by Yan et al. FF is known to form tubular structures upon dilution from solvents into water or simply sonication in deionized water, but can be manipulated to follow different supramolecular order in other solvents. Other dipeptide examples include FW and IF, which form tubular structures and fibrous hydrogels, respectively. In a systematic survey, Görbitz showed that many more dipeptides can form crystalline structures with microporous unit cells. However, the formation of crystalline structures in saturated solutions does not necessarily imply self-assembly at low concentrations in solvent. Here we report a computational approach that allows us to determine, in an initial rapid screening, which peptides will exhibit aggregation in water. The aggregation of the peptides is a necessary precondition for their self-assembly, which can then be studied using longer, larger, and more detailed simulations.

For many amyloid peptide fragments, different simulation approaches have shown aggregation of individual oligopeptide fragments. Wu and Shea have recently reviewed a number of coarse-grain methods for protein aggregation in the light of amyloid-related diseases and observed several mechanisms for fibrillization, and progress was made toward understanding the general requirements for self-assembly. In an effort to screen for peptide fibrillogenesis, lattice models have also been employed to determine contributions from sequence factors. Of specific interest for this study, for the phenylalanine dipeptide and tripeptide, replica-exchange MD was performed, which showed the formation of open or ring-like peptide networks together with a significant β-sheet content. 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 computational cost issues, although aggregation of self-assembling monomers often occurs on a microsecond time scale.

Received: August 4, 2011
Accepted: September 1, 2011
Published: September 02, 2011
have showed that one can use implicit solvent and coarse-grained (CG) models to improve the accessible time frame without losing agreement with all-atom MD simulations.27,28

In the current study, we have employed the coarse grain MARTINI force field for biomolecular simulations29,30 to model the aggregation ability for all possible combinations of two amino acids. This force field utilizes a 4:1 atom/CG-bead mapping to represent protein backbone and side chains (and a (2 to 3): 1 atom/CG-bead mapping for ring systems). As an example, Figure 1 displays the diphenylalanine peptide structure and CG-bead representation of the same peptide. There is a significant loss of atomic detail in coarse-graining, and this removes the ability of the method to reproduce features such as explicit H-bonding interactions that fine-tune the relative positioning of one peptide to another, which is important, for example, in the formation of β-sheets. However, the polarity, shape, and non-bonded interaction potential of the amino acids are implicitly included through the parametrization of functional groups based on comparison with experimental results.29,30 Therefore, although the detailed supramolecular structure is not available, coarse grain methods do provide insight into the driving force for the aggregation and consequently the self-assembly of peptides.

Utilizing the scripting tools in VMD,31 we have analyzed the output configuration of every simulated system defining the aggregation propensity (AP) as the ratio of the solvent-accessible surface area (SASA) of the dipeptide molecules in the initial minimized box to the SASA of the final configuration of the simulation. A ratio greater than two is interpreted as a high degree of aggregation. From Figure 2, which shows the AP for all 400 dipeptide combinations, it is apparent that F and W residues promote aggregation, whereas small or charged residues decrease the score. In addition, we note that the graph is roughly symmetric. There are differences for some combinations, however, most notably SF and FS, where the AP of the former is 0.6 higher. (See the Supporting Information for details.) Most dipeptides have a score close to 1, indicating that they do not exhibit a propensity to aggregate, and thus their self-assembly in aqueous conditions can be excluded.

The AP scores were compared to known experimental work on nine different dipeptides in aqueous solution, as summarized in Table 1. Although preparation methods vary throughout the literature, our data accurately predicts the self-assembly/aggregation, or lack there of, for those dipeptides for which experimental data is available. From this table, the data in Figure 2 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 very interesting to test more complex examples and those on the borderline (i.e., 1.5 ≤ AP ≤ 2.5) in an effort to break this rule. For example, HS is an interesting candidate with an AP score of 1.66 (see Supporting Information 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 of the nature of the nanostructures. Globular, fibrous, and tubular structures have been observed among the results, but these features cannot be distinguished by this score. Work is continuing in our lab to identify a more suitable, generic descriptor for quantifying the supramolecular structure.

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, to test whether the CG model is able to converge (either qualitatively, or even semiquantitatively) toward the nanostructures observed experimentally, the simulation time and system size should be further increased to allow these structures to form.

Table 1. Comparison of Self-Assembly Score with Experiments

| dipeptide | reference | assessment method | observed structure | AP* |
|-----------|-----------|-------------------|--------------------|-----|
| FF        | 32        | rheology          | none               | 1.1 |
| FF        | 9−11,33   | TEM, AFM          | tubes, vesicles    | 3.2 |
| FK        | 32        | rheology          | none               | 1.2 |
| FW        | 9,34      | TEM               | tubes, aggregates  | 3.5 |
| IF        | 14        | TEM, SEM          | fibers             | 2.3 |
| VF        | 14        | TEM               | none               | 1.8 |
| WF        | 9c        | TEM               | aggregates        | 3.5 |
| WW        | 9c        | TEM               | aggregates        | 3.2 |
| WY        | 9c        | TEM               | aggregates        | 2.1 |

*Aggregation propensity score is the ratio of the solvent-accessible surface area (SASA) of the dipeptide molecules in the initial minimized box to the SASA of the final configuration of the simulation. Supporting Information. Amorphous.
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 3a) to see whether the structures observed experimentally in an aqueous environment are in agreement with the final structure from the simulation.

In the final snapshot of the 1.5 μs simulation (Figure 3d), the FF monomers have generated a tubular nanostructure in agreement with that observed experimentally (Table 1). The 1D nanotube has an outer diameter of 60 ± 7 Å and an inner diameter of 29 ± 5 Å (Figure 3d). 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. and Görbitz of crystallized FF nanotubes (van der Waals diameter of inner channel ~10 Å). Moreover, we also note that the dihedral angle between the side chains rotates from 180° in the starting monomer structures (Figure 1) to an average value of 0° in the final tube like nanostructure (Figure 4). The rotation of the side chains away from the ideal 180° value for the dipeptide is also in agreement with the X-ray structures observed by Görbitz. In the initial system (i.e., after minimization), the backbone dihedral angles are clustered around 120 and 180° (green columns in Figure 4). However, as the simulation continues, we observe that the side chains rotate during the aggregation process to adopt a dihedral angle between the side chains centered around 0°.

After 1 ns of simulation, the majority of the dipeptides have already reached this equilibrium value (blue columns in Figure 4), although the distribution is not as sharp as is achieved in the final snapshot at 1.5 μs (red columns in Figure 4). Although a direct comparison of the distance and dihedral parameters is not possible because of 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 3a), we observe an initial formation of sheet-like aggregates (Figure 3b).
0.2 \mu s. These sheet-like structures grow and accumulate to a point where they are then able fold up to form hollow vesicles after \(\sim 0.5 \mu s\) (Figure 3c). Extending the simulation time to up to a total of 1.5 \mu s results in the continuous accumulation of these vesicles, which fuse together into 1D nanotubes (Figure 3d).

Therefore, whereas the FF simulation shows that after 400 ns the final nanostructure that results from the self-assembly process may not be reached, the initial aggregation stages of self-assembly are already in process 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, whereas 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 Supporting Information (Figure S1). The AP for all experimentally studied dipeptides after extended, 4 \mu s simulations (Figure S2; see the Supporting Information) is displayed in Table S1 (see Supporting Information) and does not show significant changes with respect to the values in Table 1, supporting the validity of the 400 ns value of the AP.

In conclusion, the MARTINI coarse-grain force field has been applied to predict the aggregation ability of all 400 dipeptides of the 20 gene-encoded amino acids. Good agreement was found comparing the predicted aggregation propensity and supramolecular structure to experimental results from literature for various dipeptides including diphenylalanine. From the FF simulation, it is apparent that 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. In all MD simulations, care has to be taken when considering simulation time and system size because some effects may be time- or concentration-dependent. 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. 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.

The self-assembly of biomolecules often occurs on a time scale that exceeds the microsecond region, and as such atomistic methods are too computationally expensive to model multiple large systems. Moreover, as was observed in the FF simulation leading to the nanotube formation, there is a critical size requirement for the system to be modeled before self-assembly can be observed. That is, in the case of FF, we see initial sheet formation; the sheets are only able to form vesicles (the precursor stage to tube formation) once they have reached a sufficient size to fold on themselves. Atomistic simulations will not typically be of sufficient size to reach safely and surpass these critical size regimes because of the cost associated with such large systems at this level of theory. Thus, although detailed information may be lost, 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. The results of the initial screening and subsequent production simulations for systems of interest can then form the basis for more detailed atomistic studies.

### COMPUTATIONAL METHODS

A simulation for each dipeptide (in their zwitterionic form) was set up using the GROMACS molecular dynamics package. A cubic box with 300 dipeptides, placed randomly with a minimum distance of 3 Å between them, was solvated in standard MARTINI CG water (four water molecules per bead) to a final concentration of \(\sim 0.4\) M. A Berendsen thermostat and barostat\(^\text{36}\) were used to keep the temperature at 303 K and pressure at 1 bar, respectively. Bond lengths in aromatic side chains and the backbone—side-chain bonds in I, V, and Y were constrained using the LINCS algorithm.\(^\text{37}\) All boxes were energy minimized using the steepest descent integrator and then equilibrated for \(4 \times 10^6\) time steps of 25 fs.\(^\text{38-41}\) The extended FF simulation was performed in duplicate, and reported tube sizes were averaged over 10 different positions in either simulation. For all experimentally studied dipeptides, a similar simulation over \(40 \times 10^6\) steps was performed to study the effect of a longer simulation on the AP.

The total screening simulation time equates to 100 ns, but because of the smoothness of the CG potentials, this roughly equates to an effective 400 ns of atomistic simulation time.\(^\text{29,42}\) Throughout this Letter, all times refer to the simulation time speeded 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 screen excludes the possibility of performing these MD simulations using all-atom models on modest computing resources. The complete, 400 ns simulation for a single dipeptide system described above can be carried out in ca. 4 h on a four-core 2.6 GHz Opteron compute node. More details of the simulation can be found in the Supporting Information.

### ASSOCIATED CONTENT

Supporting Information

Computational details, a complete list of the AP scores, outputs for IF, WF, FW, and WW after 400 ns, and all extended simulations. This material is available free of charge via the Internet at [http://pubs.acs.org](http://pubs.acs.org).

### AUTHOR INFORMATION

**Corresponding Author**

*E-mail: tell.tuttle@strath.ac.uk.*

### ACKNOWLEDGMENT

R.V.U. would like to thank the Leverhulme Trust (Leadership Award) and EPSRC (Advanced Fellowship) for funding. N.T.H. acknowledges EPSRC and ERC for funding. T.T. thanks the EPSRC and the Glasgow Centre for Physical Organic Chemistry for funding.

### REFERENCES

1. Zelcer, M.; Ulijn, R. V. Next-Generation Peptide Nanomaterials: Molecular Networks, Interfaces and Supramolecular Functionality. *Chem. Soc. Rev.* 2010, 39, 3351–3357.

2. Hartgerink, J. D.; Benlash, E.; Stupp, S. L. Self-Assembly and Mineralization of Peptide-Amphiphile Nanofibers. *Science* 2001, 294, 1684–1688.

3. Ghadiri, M. R.; Granja, J. R.; Milligan, R. A.; McRee, D. E.; Khazanovich, N. Self-Assembling Organic Nanotubes Based on a Cyclic Peptide Architecture. *Nature* 1993, 366, 324–327.
(4) Reches, M.; Gazit, E. Designed Aromatic Homo-Dipeptides: Formation of Ordered Nanostructures and Potential Nanotechnological Applications. Phys. Biol. 2006, 3, S10–S19.

(5) Smith, A. M.; Williams, R. J.; Tang, C.; Coppo, P.; Collins, R. F.; Turner, M. L.; Saiani, A.; Ulijn, R. V. Fmoc-Diphenylalanine Self-Assemblies to a Hydrogel via a Novel Architecture Based on π → π Interlocked β-Sheets. Adv. Mater. 2008, 20, 37–41.

(6) Zhang, S.; Holmes, T.; Lockshin, C.; Rich, A. Spontaneous Assembly of a Self-Complementary Oligopeptide to Form a Stable Macromolecular Membrane. Proc. Natl. Acad. Sci. U.S.A. 1993, 90, 3334–3338.

(7) Xu, H. X.; Das, A. K.; Horie, M.; Shaik, M.; Smith, A. M.; Luo, Y.; Lu, X.; Collins, R.; Liem, S. Y.; Song, A.; et al. An Investigation of the Conductivity of Peptide Nanotube Networks Prepared by Enzyme-Trigged Self-Assembly. Nanoscale 2010, 2, 960–966.

(8) Silva, G. A.; Czeisler, C.; Niece, K. L.; Beniash, E.; Harrington, D. A.; Kessler, J. A.; Stupp, S. I. Selective Differentiation of Neural Progenitor Cells by High-Epitope Density Nanofibers. Science 2004, 303, 1532–1535.

(9) Reches, M.; Gazit, E. Casting Metal Nanowires within Discrete Self-Assembled Peptide Nanotubes. Science 2003, 300, 625–627.

(10) Yan, X.; Zhu, P.; Li, J. Self-Assembly and Application of Diphenylalanine-Based Nanostructures. Chem. Soc. Rev. 2010, 39, 1877–1890.

(11) Kim, J.; Han, T. H.; Kim, Y.-I.; Park, J. S.; Choi, J.; Churchill, D. G.; Kim, S. O.; Hee, H. Role of Water in Directing Diphenylalanine Assembly into Nanotubes and Nanowires. Adv. Mater. 2010, 22, 583–587.

(12) Han, T. H.; Oh, J. K.; Lee, G.-J.; Pyun, S.-L; Kim, S. O. Hierarchical Assembly of Diphenylalanine into Dendritic Nanoarchitectures. Colloids Surf., B 2010, 79, 440–445.

(13) Su, Y.; Yan, Wang, A.; Fei, J.; Cui, Y.; He, Q.; Li, J. A Peony-Flower-Like Hierarchical Mesocrystal Formed by Diphenylalanine. J. Mater. Chem. 2010, 20, 6734–6740.

(14) Sanchez-de Groot, N.; Parella, T.; Aviles, F.; Vendrell, J.; Ventura, S. Ile-Phe Dipeptide Self-Assembly: Clues to Amyloid Formation. Biophys. J. 2007, 92, 1732–1741.

(15) Görbiz, C. H. The Structure of Nanotubes Formed by Diphenylalanine, The Core Recognition Motif of Alzheimer’s β-Amyloid Polypeptide. Chem. Commun. 2006, 2332–2334.

(16) Görbiz, C. H. Microporous Organic Materials from Hydrophobic Dipeptides. Chem.—Eur. J. 2007, 13, 1022–1031.

(17) Hall, C. K.; Wagener, V. A. Computational Approaches to Fibril Structure and Formation. In Amyloid, Prions, and Other Protein Aggregates, Part B; Khéterpal, I., Wetzel, R., Eds.; Academic Press: Boston, 2007; pp 338–365.

(18) Meli, M.; Morra, G.; Colombo, G. Investigating the Mechanism of Peptide Aggregation: Insights from Mixed Monte Carlo-Molecular Dynamics Simulations. Biophys. J. 2008, 94, 4414–4426.

(19) Mousseau, N.; Derreumaux, P. Exploring the Early Steps of Amyloid Peptide Aggregation by Computers. Acc. Chem. Res. 2005, 38, 885–891.

(20) Sanay, D.; Hamley, I. W.; Aleman, C. Modeling the Tetraphenylalanine-PEG Hybrid Amphiphile: From DFT Calculations on the Peptide to Molecular Dynamics Simulations on the Conjugate. J. Phys. Chem. B 2011, 115, 8937–8946.

(21) Wu, C.; Shea, J. E. Coarse-Grained Models for Protein Aggregation. Curr. Opin. Struct. Biol. 2011, 21, 209–220.

(22) Hauser, C. A. E.; Deng, R. S.; Mishra, A.; Loo, Y. H.; Khoe, U.; Zhuang, F. R.; Cheong, D. W.; Accardo, A.; Sullivan, M. B.; Riekel, C.; et al. Natural Tri- To Hexapeptides Self-Assemble in Water to Amyloid β-Type Fiber Aggregates by Unexpected Alpha-Helical Intermediate Structures. Proc. Natl. Acad. Sci. U.S.A. 2011, 108, 1361–1366.

(23) Thirumalai, D.; Klimov, D. K.; Dima, R. I. Emerging Ideas on the Molecular Basis of Protein and Peptide Aggregation. Curr. Opin. Struct. Biol. 2003, 13, 146–159.

(24) Li, M. S.; Co, N. T.; Reddy, G.; Hu, C. K.; Straub, J. E.; Thirumalai, D., Factors Governing Fibrillogenesis of Polypeptide Chains Revealed by Lattice Models. Phys. Rev. Lett. 2010, 105, 218101.