DEVELOPING COMPUTATIONAL TOOLS FOR THE STUDY AND DESIGN OF AMYLOID MATERIALS

A Dissertation

by

SAI VAMSHI REDDY JONNALAGADDA

Submitted to the Office of Graduate and Professional Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Chair of Committee,	Phanourios Tamamis
Committee Members,	Hae-Kwon Jeong
	Hung-Jen Wu
	Raymundo Arroyave
Head of Department,	Arul Jayaraman

December 2019

Major Subject: Chemical Engineering

ABSTRACT

The self-assembly of short peptides into amyloid structures is linked to several diseases but has also been exploited for the design of novel functional amyloid-based materials. Such materials are potentially biocompatible and biodegradable, while their unique molecular organization provides them with remarkable mechanical properties. Amyloid fibrils are among the stiffest biological materials and exhibit a high resistance to breakage. Apart from the aforementioned properties, they are particularly attractive due to their easy synthesis and the ability to be re-designed through mutations at sequence level, which can result in potential functionality.

Previous studies have reported the rational based design of functional amyloid materials, designed through primarily scientists' intuition, and their applications in several fields as agents for tissue-engineering, antimicrobial and antibacterial agents, drug carriers, materials for separation applications, etc.

The current work starts from the use of previously reported protocols for the computational elucidation of the structure of amyloids, leading to the formation of amyloid materials, and the investigation of the functional properties of rationally designed self-assembling peptides, and introduces a new approach for the computational design of functional amyloid materials, based on engineering and biophysical principles. In summary, we developed a computational protocol according to which an optimization-based design model is used to introduce mutations at non-β-sheet residue positions of an amyloid designable scaffold (amyloid with non-β-sheet forming residues at its termini). The designed amino acids are introduced to the scaffold in such a way so that they mimic how amino acids bind to particular ions/compounds of interest according to experimentally resolved structures (defined by us as materialphore models) and also aim at energetically stabilizing the bound conformation of the pockets. The optimum designs are

computationally validated using a series of simulations and structural analysis techniques to select the top designed peptides, which are predicted to form fibrils with specific ion/compound binding properties for experimental testing.

The computational protocol has been implemented first for the design of amyloid materials (i) binding to cesium ions, and in additional cases, for the design of amyloid materials (ii) serving as potential AD drug carriers, (iii) which could promote cell-penetration and possess DNA binding properties, and (iv) incorporating potential cell-adhesion, calcium and strontium binding properties. The computational protocol is also presented here as a step toward a generalized computational approach to design functional amyloid materials binding to an ion/compound of interest. This work can constitute a stepping stone for the functionalization of peptide/protein-based materials for several applications in the future.

ACKNOWLEDGEMENTS

I would like to thank my advisor Dr. Phanourios Tamamis who guided me through every phase in my PhD. He provided me tremendous support both in professional and personal life. I would like to thank my committee members, Dr. Hae-Kwon Jeong for guiding through experiments and allowing me to use his lab. I would also like thank Dr. Arul Jayaraman, Dr. Nazmul Karim, Dr. Katy Kao, Dr. Pushkar Lele, and Dr. Jodie Lutkenhaus for letting me use their lab and equipment to perform experiments.

Next, I thank Asuka Orr who worked with me during my 4 years of PhD. She provided me tremendous support in writing manuscripts and performing explicit solvent MD simulations. I would like to thank Joseph Jakubowski, Shujun He, Chang-Hyun Choi, Kendal Henderson, and Jacob Spies for developing or co-developing parts of the computational protocol and/or assisting in the analysis of the computationally designed peptides. I would also like to thank all the lab members of Dr. Tamamis' lab who have been a part of my journey.

I would like to thank Dr. Anna Mitraki and her lab members who collaborated with us on various studies. I also would like to thank Dr. Manos Gkikas and his lab members for collaborating with us. I also thank Dr. Bryan Tomlin for performing elemental analysis experiments, which were conducted at Center for Chemical Characterization and Analysis at Texas A&M University. I thank the Texas A&M High Performance Research Computing Center for making available state-of-the-art supercomputing facilities for my research. I would also like to thank all the staff in the chemical engineering department for helping promptly whenever required and providing a hassle-free work environment.

Finally, I am grateful to my family and friends for supporting me throughout my life.

CONTRIBUTORS AND FUNDING SOURCES

This work was supervised by a dissertation committee consisting of Dr. Phanourios Tamamis (chair), Dr. Hae-Kwon Jeong and Dr. Hung-Jen Wu of the Artie McFerrin Department of Chemical Engineering., and Dr. Raymundo Arroyave of the Department of Materials Science and Engineering.

All work for the dissertation was completed by the student with certain computational contributions made by various members of Dr. Tamamis' lab; the students are either explicitly mentioned in the thesis and/or in papers that were published or that will be published in the future. Experiment contributions have been made by several experimental groups, primarily involving Dr. Anna Mitraki's lab at the University of Crete, Greece, and Dr. Manos Gkikas' lab at the University of Massachusetts Lowell; their names are mentioned in the thesis or in papers that were published or will be published in the future. The amyloid fibrils prepared for elemental analysis for cesium binding were prepared by the student and Asuka Orr.

This work was made possible in part by the startup fund awarded to Dr. Phanourios Tamamis by the Artie McFerrin Department of Chemical Engineering at Texas A&M University. This work was also made possible by the Water Seed Grant under Grant Number 163029-00000 awarded to Dr. Phanourios Tamamis by Texas A&M Engineering Experiment Station (TEES).

TABLE OF CONTENTS

NOMENCLATURE

a	Born radius
AAindex	Amino acid index database
ACh	Acetylcholine
AChE	Acetylcholinesterase
AD	Alzheimer's disease
AFM	Atomic force microscopy
ASA	Accessible Surface Area
ATCC	American Type Culture Collection
Aβ	Amyloid-β
b	Bond
C3	Complement protein 3
CD	Circular dichroism
CMAP,φ,ψ	Backbone torsional correction
CO_2	Carbon dioxide
COM	Center of mass
CPPsite2.0	Database of Cell-Penetrating Peptides
D	Distance
DEE	Dead End Elimination
DMD	Discontinuous molecular dynamics
DMEM	Dulbecco's Modified Eagle's Medium
DTNB	5,5'-dithiobis-2-nitrobenzoic acid

ECM	Extracellular matrix
$E_{i,a,k,b}^{j,l}$	Pairwise interaction energy
ELR	Elastin-like recombinamer
F	Empirical factor
FACTS	Fast analytical continuum treatment of solvation
FBS	Fetal bovine serum
FESEM	Field Emission Scanning Electron Microscopy
GB	Generalized born
IAPP	Islet amyloid polypeptide
IFP	Inverse folding problem
IGF	Insulin like growth factors
k_B	Boltzman's constant
LJ	Lennard-Jones
MD	Molecular dynamics
NMDA	N-methyl-D-aspartate
P_1	Nematic order parameter
P_2	Polar order parameter
Pafig	Prediction of amyloid fibril-forming segments
PDB	Protein Data Bank
PN	Proteostatsis network
PSSM	Position specific scoring matrix
q	Point charge
r	Distance between atoms

REMD	Replica exchange molecular dynamics
rMSCs	Rat mesenchymal stem cells
S-layer	Surface layers
SALSA	Simple algorithm for sliding averages
SASA	solvent accessible surface area
SecStr	Possible conformational switches
SSNMR	Solid state nuclear magnetic resonance
SVM	Support vector machines
TEM	Transmission electron microscopy
ThT	Thioflavin-T
UB, S	Urey-Bradley
αTAT1	α-tubulin acetyl transferase protein
ε	Dielectric constant
θ	Valence angle
φ	Dihedral angle
ω	Improper angle

LIST OF FIGURES

LIST OF TABLES

1. INTRODUCTION

1.1 Molecular Self-assembly

Molecular self-assembly is the process in which simple molecules, usually in a less ordered state (random coil, disordered aggregate, or a solution), spontaneously form complex multicomponent three-dimensional ordered structures (a folded macromolecule or a crystal)[1]. The interactions occurring during the self-assembly process between the molecules are generally non-covalent (Coulomb interactions, hydrogen bonds, hydrophobic interactions, and Van der Waals interactions) in nature[1]. Molecular self-assembly is ubiquitous in various fields like biology, material science, and chemistry. Some of the examples in the field of chemistry and material science are the formation of crystals[2], formation of colloids[3], and phase separated polymers[4].

Molecular self-assembly is particularly important in the field of biology as it is centrally important to life. Self-assembly occurs in all organisms from very simple bacteria to mammals. For example, phospholipids, the most abundant cell membrane lipid, self-assemble naturally to form lipid bilayers which provide structural integrity for a cell[5,6,7]. Phospholipids are amphiphilic molecules that contain hydrophilic phosphate head group and hydrophobic fatty acid tail. Upon coming into contact with aqueous environment, phospholipids self-assemble spontaneously to avoid the energetic penalty of hydrophobic tail interacting with water. Thus, phospholipids organize themselves into lipid bilayers where hydrophobic tails interact with each other and hydrophilic heads interact with inside part of the cell or with the outside bulk water.

An additional paradigm of biomolecular self-assembly involves protein self-assembly, the self-organization of individual proteins into highly ordered protein architectures ranging from nm to μm in size. For example, two-dimensional arrays of proteinaceous subunits form bacterial cell surface layers (S-layers) through the process of self-assembly in hundreds of species of walled

bacteria and in almost all archaea[5,8,9]. Silk is another example of protein self-assembly, which is a remarkable biomaterial comprising protein fibers spun by insects and spiders, which is produced through self-assembly of predominantly two protein molecules[5,10,11,12]. Silk is produced for various functions like web construction, cocoon formation, and prey catching. Silk has outstanding mechanical properties including high strength and flexibility even though the interactions between the molecules making up silk are non-covalent in nature. Spider silk is considered to be remarkably stronger than a steel filament of same diameter but also highly flexible[10].

1.2 Amyloids

The aggregation of proteins into amyloid fibrils is an additional key example of molecular self-assembly[5]. Unlike the self-assembled structures described above which could be functional in nature, self-assembled amyloid structures may be relevant to diseases. The amyloid state of proteins was initially discovered in the context of diseases where the proteins self-assemble into amyloid fibrils, a key structural characteristic of amyloid diseases, that are insoluble and heterogenous extracellular protein deposits[13,14]. Amyloid fibrils share a common underlying structural architecture where the proteins self-assemble into β-sheets and the core structure of the β-sheets in the amyloid fibrils align perpendicular to the fibril axis. This alignment is termed as cross-β organization[14,15]. Currently, 50 different peptides or proteins are known to self-assemble into amyloid fibrils which are associated with human disease[14]. Some of the diseases include among others Alzheimer's disease (AD) where amyloid-β (40-42 amino acids long) forms senile plaques and tau protein forms neurofibrillary tangles as pathological features[14,16]; Parkinson's disease where α-synuclein forms Lewy bodies as pathological features[14,16]; type 2 diabetes where islet amyloid polypeptide (IAPP) forms pancreatic islet amyloid as pathological feature[14,16]. It has

been shown that the strength of specifically studied amyloid fibrils is comparable to steel and mechanical stiffness comparable to silk[17].

Despite the fact that amyloids were first identified and were linked to human diseases in which peptides and proteins misfold into amyloid fibrils, subsequent studies have indicated several functional amyloid fibrils that perform physiological roles in humans.[18] In addition, interestingly, naturally occurring sequences from amyloidogenic proteins or natural fibrous proteins can be extracted to form amyloids outside the native context of the entire protein. For example, while amyloid fibrils formed by the peptides such as amyloid-β (40-42 amino acids long) and islet amyloid polypeptide (37 amino acids long) are the pathological features of Alzheimer's disease and type 2 diabetes as mentioned above, respectively, short fragments such as KLVFFAE[19] extracted from Aβ as well as NFGAIL[20] extracted from IAPP form amyloid fibrils too and possess similar ultrastructural, molecular conformational and cytotoxic properties to that the ones formed by the full length corresponding polypeptide (Aβ and IAPP respectively). Also, following a minimalistic approach to find the smallest peptide that can self-assemble, Görbitz[21] in 2001 reported the self-assembly of a very short peptide, the Alzheimer's β-amyloid diphenylalanine structural motif. This change from larger proteins or peptide to short peptide fragments enabled the large-scale synthesis of amyloid fibrils and their use in various nanotechnological applications. Thus, the roles of naturally occurring sequences from disease related amyloid peptides/proteins or β-sheet rich peptide/protein regions has been expanded to include their use as elementary blocks for the formation of amyloid materials[10,22,23,24,25,26,27,28,29,30,31,32,33,34,35].

1.3 Amyloid Materials

Amyloids as materials (amyloid materials) show many favorable properties such as they (i) can be easily fabricated through self- assembly, (ii) have properties that can be tailored through mutations at the sequence level, (iii) can be produced via chemical synthesis or recombinant technologies in sufficient quantities, (iv) have biocompatible nature, (v) have ability to be biodegradable, and (vi) have high Young's modulus due to the dense hydrogen bonding network that is the characteristic of core of amyloid fibrils[36]. Thus, these advantages of amyloid materials make them attractive as artificial functional materials and they are involved in applications including but not limited to cell culture scaffolds, biosensors, photoluminescent materials, amyloid nanocomposites, adhesive amyloids, organic quantum dots, and vehicles for drug delivery[36].

Several examples of exploiting amyloid or amyloid-like self-assembly for the fabrication of materials have been reported in the literature (reviewed in refs[10,36]). For example, Reches and Gazit[31] in 2003 reported that they observed the "amyloid-like" self-assembly of the diphenylalanine peptide mentioned above into discrete, and stiff nanotubes which upon reduction with ionic silver followed by enzymatic degradation, resulted in the production of discrete nanowires with a long persistence length[31]. An analogue of the diphenylalanine peptide, diphenylglycine peptide also self-assembles and forms nanospherical structures efficiently and have remarkable stability[37]. Furthermore, Kasai and colleagues[38] combined the IKVAV containing amyloidogenic peptide A208 (2097-2108 peptide fragment from mouse laminin α1 chain) with RGD tripeptide to promote cell-adhesion on amyloid gel scaffolds. The authors inspired by the sequences of cell-adhesion active sites in the extracellular fibrils of fibronectin demonstrated that the RGD conjugated A208 peptide has self-assembling properties combined with cell-attachment properties, thus inferring the possible use of these systems in tissue engineering as cell culture

scaffolds[38]. Moreover, Mezzenga and colleagues[39] combined graphene and amyloid fibrils from the milk protein β-lactoglobulin and synthesized biodegradable nanocomposites in which amyloid fibrils provided the enzyme sensing substrate and the graphene layers contributed with electron conductivity and mechanical strength. The nanocomposites were shown to possess shape memory features which could be used in water sensing.

The functionality of such materials is an additional important aspect with regard to their potential application. Functionality can be related to the ability of the amyloid materials to bind ions, compounds or surface (cell attachment); such properties can enable amyloid materials' applicability in several areas including as tissue engineering, separations or drug delivery. Functional amyloid materials have been produced to bind ions which resulted in tissue engineering applications. For example, Mitraki and colleagues[25,40] rationally designed aspartate-rich self-assembling peptides binding calcium ion using insights of MD simulation provided by Tamamis et al[25]. They further investigated the cell adhesion and proliferation of the calcium bound mineralized aspartate-rich self-assembling peptides and observed that the designed peptide strongly supports cell adhesion and proliferation with significant increase of biomineralization.[25] The specific designed peptide can be used as a secondary scaffold on top of a primary laser-fabricated strong inert scaffold. Thus, the combinations of these scaffolds can be used as a support for the directed growth of cell into biomineralization units for bone tissue engineering applications.[25]

Functional amyloid materials have also been exploited to bind ions (e.g., arsenic[41], gold[42], mercury[42], lead[42], and palladium[42]) which could result in applications in the field of separations. For example, Mezzenga and colleagues[41,42] synthesized hybrid composite membranes consisting of β-lactoglobulin amyloid fibrils and activated carbon which can be used as an efficient tool to

remove heavy metal ion pollutants like gold[42], mercury[42], lead[42] and palladium[42] form wastewater. Additionally, the hybrid membranes were also able to capture radioactive compounds like uranyl acetate and phosphorus-32 from the wastewater[42]. The specific hybrid membranes were also used to remove arsenic (both the arsenate and arsenite oxidation forms) from contaminated ground water with efficiency as high as 99.6%[41]. Eisenberg and colleagues[32,33] designed amyloid fibers capable of capturing carbon dioxide from flue gas. They showed that the ε-amino group of an exposed lysine at high pH in an amyloid fibril can capture carbon dioxide via carbamate formation.

Apart from amyloid materials' applications in tissue engineering or separations, amyloid materials have also been suggested for drug delivery applications (i.e., amyloid materials potentially serving as carriers). For example, Silva and colleagues[43] investigated the use of amyloid material as potential vehicles for drug delivery. Specifically, they showed that the microtubes formed from the self-assembly of the peptide L-diphenylalanine can be conjugated to a hydrophilic model drug rhodamine. They observed that the model drug is embedded homogeneously into the microtubes and also the microtubes modulate the release of the drug. In addition, they have observed that cell viability in presence of the microtubes is high indicating that they have low toxicity. Marchesan and colleagues[44] reported the self-assembly of the tripeptide Leu-Phe-Phe with one of the most effective antibiotics for skin and eye infections, ciprofloxacin. They reported that the antibiotic appears bound within the hydrogel formed by the peptide and antibiotic via non-covalent interactions. They observed that the ciprofloxacin-peptide hydrogel showed anti-bacterial activity against specific bacteria, but the hydrogel showed no cytotoxicity in haemolysis assays of human red blood cells. Mezzenga and colleagues[45] investigated the ability of β-lactoglobulin amyloid fibrils to bind iron where the specific amyloid fibrils act as carriers for iron fortification. Traditional fortificants generally cause strong organoleptic changes in foods whereas iron

generally accumulates in solutions due to its strong oxidative tendency[45]. The specific amyloid fibrils which are biodegradable can act as an iron fortificant with minimal organoleptic changes in food and also stabilizes iron due to the fibrils' natural reducing properties[45]. The resulting material containing iron releases the ion during acidic and enzymatic *in vitro* digestion[45]. The specific material was also tested *in vivo* in rats and the high iron bioavailability was observed without accumulation in organs suggesting their safety and usefulness as efficient iron fortificant[45]. All in all, functional amyloid materials can be successfully designed to bind ions, compounds or surfaces which can have applications in field of tissue engineering, separations and drug carriers. Based on most of the cases outlined in the introduction, the design of such functional materials can be rationally achieved through intuition and can also be complemented through computational-theoretical methods. An important aspect related to functionalizing amyloid materials is associated with the ability to study their structure and their self-assembly process, which can be performed using a combination of experimental and computational methods, as outlined in chapter 2. An additional important aspect related to functionalizing amyloid materials, which is also outlined in chapter 2, is associated with the ability of computational methods to – not only predict their structure – but also enable the design of mutations to peptides and proteins, which could yield functionality.

2. STRUCTURAL DETERMINATION OF AMYLOID FIBRILS - IMPORTANCE OF COMPUTATIONAL METHODS IN STUDYING THE AMYLOID PROPENSITY AND FORMATION, AND IN PROTEIN DESIGN

2.1 Experimental Methods for the Characterization of the Structure of Amyloid Fibrils

Amyloid fibrils are identified by using florescent dye Thioflavin T and Congo red, a deep crimson dye adapted from fabric industry. Benzathiole dye Thioflavin-T (ThT), a potent florescent dye has large enhancement of florescent emission upon binding amyloid fibrils. Amyloid fibrils also bind to the dye Congo red, and emit a green birefringence in polarized light after the fibrils are stained with Congo red[15,46,47,48].

Electron and atomic microscopy can be used to characterize the typical macromolecular structure of amyloid fibrils as long, straight and unbranched protein fibrils with diameters up to 100 nm and length reaching several micrometers[13,15]. The X-ray diffraction pattern given by amyloid fibrils is cross-β, meaning that the core structure of the fibrils consists of a β-sheet conformation in which the β-sheet strands are perpendicular to the fibril axis and the hydrogen bonding directions are parallel to the fibril axis[13,15]. The cross-β X-ray diffraction pattern typically consist of two major reflections, a ~ 4.7 Å reflection along the fibril axis corresponding to the hydrogen bonding distance and ~ 8-10 Å reflection perpendicular to the fibril axis corresponding to the sidechain packing between the sheets (Figure 2.1). Circular dichroism (CD) spectroscopy is also a commonly used technique to characterize and study the secondary structure of amyloid fibrils[49,50] and the specific spectroscopy is predominantly based on excitation of electronic amide groups[50]. CD spectroscopy identifies the amyloid fibrils' characteristic β-sheet secondary structure based on the specific Φ, Ψ dihedral angles and H-bond patterns affecting the spectrum[50].

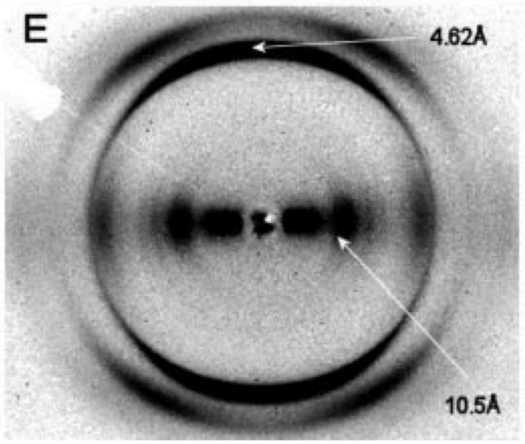


Figure 2.1 Example X-ray diffraction pattern of a 5-residue amyloid forming peptide with sequence GAITIG. The diffraction pattern depicted here is cross-β with ~ 4.7 Å reflection along the fibril axis and ~ 8-10 Å reflection perpendicular to the fibril axis. Figure taken from ref.[51]

More recently, X-ray crystallography, solid state nuclear magnetic resonance (SSNMR) spectroscopy, cryo-electron microscopy methods have been used to provide additional details into the structure of amyloid fibrils[49]. These recent methods have produced highly detailed composition of amyloid fibrils and their molecular architecture[49]. Specifically, X-ray crystallography was used to yield atomic-resolution structures of short amyloid peptides or short peptide segments from amyloidogenic proteins[15,52,53,54]. This was enabled by the fact that the short fibril-forming peptides form needle-shape microcrystals in which the fibrils span the entire length of the crystals[52]. However, longer amyloid fibrils are non-crystalline, insoluble and heterogeneous, so it is more challenging to decipher the atomic-resolution structure by using X-ray crystallography[52,55]. Thus, SSNMR method was used to produce the atomic-resolution structure of various amyloidogenic proteins[55]. The intrinsically disordered nature of a lot of amyloidogenic proteins didn't the allow the SSNMR to produce full-scale atomic model[56]. Thus, cryo-EM technique was successfully used to produce atomic structures of fibrils formed by the amyloidogenic proteins Aβ[57], Tau[58], and α-synuclein[56,59].

2.2 Computational Methods on the Study of Amyloid Propensity and Amyloid Formation

In addition to experimental methods providing invaluable insights into the identification and structural characterization of amyloid fibrils, computational methods have been developed as well to study the amyloid propensity of peptide or protein sequences, as well as the structures formed by amyloid-forming peptides and proteins. Computational methods have primarily focused on the use of (i) sequence-based bioinformatics methods to investigate aggregation propensity of the peptides or proteins and thus identify the peptides or proteins most likely to form amyloid fibrils, and (ii) structural modelling-based methods including simulations, where the energy of amino acids is represented by statistical potentials or all atom energy potentials, to gain insights into the amyloid formation properties for peptides or proteins. The development of sequence-based bioinformatics methods has been primarily facilitated by the significant improvements in computing and the use of big-data analysis and/or machine learning methods. The development of structural modelling-based has been primarily facilitated by the ability of computers to simulate amyloid formation which in turn has been enabled by improvements in computational power, advanced force fields, and simulation algorithms.

2.2.1 Sequence-Based Bioinformatics Methods

Several sequence-based algorithms in bioinformatics like TANGO[60,61,62], AGGRESCAN[63], Waltz[64], Amyloidogenic Pattern[65], Average Packing Density[66], FoldAmyloid[67], Hexapeptide Conformational Energy[68], Beta-strand contiguity[69], Pafig[70], SecStr (Possible Conformational Switches)[71], Amyloid Mutants[72], NetCSSP[73,74,75,76], AMYLPRED[77], PASTA[78], BETASCAN[79] have been developed to predict the amyloid propensity of a peptide or protein based on the sequence. In what follows, an outline of the key principles associated with the function of a portion of the aforementioned tools is provided.

TANGO[60,61,62], a statistical mechanics-based algorithm, is one of such methods that identifies β-aggregating region of peptide or protein sequence. TANGO[60,61,62] considers different competing conformations such as α-helix, β-aggregation, β-turn, α-helical aggregation, and random coil and assumes that every segment of a peptide or protein can be populated by the aforementioned according to the Boltzmann distribution. In addition, the algorithm also assumes that the core segments in β-aggregates are fully buried and they generally show a tendency to satisfy their hydrogen bonding potential. In order to predict the β-aggregating regions of a peptide or protein, the algorithm calculates the partition function of the conformational phase-space. It should be noted that the algorithm predicts the β-aggregation propensity of a peptide or protein.

AGGRESCAN[63], a web-based tool for the aggregation-prone regions in a peptide or protein sequences is based on an aggregation propensity scale developed based on *in vivo* experiments. The main assumption of the specific algorithm is that short and specific segments in a peptide or protein modulate the aggregation of the protein. The specific algorithm has successfully predicted the aggregation regions of disease-related proteins.[63] In addition, the algorithm can also help in identifying the aggregation prone regions of bioactive polypeptides thereby anticipating the problems during the storage or recombinant production of the specific peptides or proteins.[63]

Waltz[64], a web-based bioinformatics tool, uses a position specific scoring matrix (PSSM) to identify or predict amyloid forming peptide or protein sequences. The specific algorithm was trained based on sequence diversity of more than 200 amyloid hexapeptides investigated using various biophysical and structural methods. Interestingly, it was observed that different amino acids had strong position specific tendencies to form amyloid structures in disease related and functional amyloids. The specific computational algorithm assigns a score S_{total} which is calculated

as follows: ($S_{total} = \alpha_{profile} S_{profile} + \alpha_{physprop} S_{physprop} + \alpha_{struct} S_{struct}$). $S_{profile}$ is the score calculated from the log-based position specific scoring matrix (PSSM) for the amyloid propensity. The specific matrix was calculated based on the position specific tendencies for different amino acids as described above. $S_{physprop}$ is the score calculated based on nineteen physical properties that based describe the amyloid propensity and the specific score is the sum of the products of the amino acid frequency with the normalized property value of the respective amino acid for each position. S_{struct} is the score calculated from a position specific pseudo-energy matrix which was evaluated from structural modeling using amyloid backbone structures. It is worth noting that the specific algorithm can help user better distinguish between amorphous β-aggregates and amyloid peptide sequences.[64]

Additional sequence-based amyloid prediction tools in bioinformatics are summarized below along with their key properties.

- Amyloidogenic Pattern[65] uses a saturation scanning mutagenesis method on the *de novo* designed amyloidogenic peptide STVIIE to identify sequence pattern that could be involved in amyloid-like fibril formation.
- The computational algorithm, Average Packing Density[66], relates the average packing density of amino acid segments to the formation of amyloid fibrils. Later, the specific algorithm was extended to FoldAmyloid[67] by including hydrogen-bonding interactions between amino acids in the peptide or protein segments of interest.
- The computational algorithm, Hexapeptide Conformational Energy[68] identifies hexapeptide amino acid regions in a peptide or protein that fit as β-strands in a stacked β-sheet structure based on the crystal structure of the peptide NNQQNY, an amyloidogenic hexapeptide segment from the yeast prion protein Sup35.

- Beta-strand contiguity[69] is a computational algorithm that identifies and locates regions in a peptide or protein with high propensity for β-strand structure. The specific algorithm, also known as SALSA (Simple ALgorithm for Sliding Averages) calculates and assigns a β-strand propensity score by generating varying peptide segments (from 4 to 20 amino acids) within a peptide or protein sequence.
- Prediction of amyloid fibril-forming segments (Pafig)[70], a method based on Support Vector Machines (SVM), identifies hexapeptide sequences which can be associated with amyloid fibril formation. The specific method is trained based on 41 physicochemical methods selected from the Amino Acid index database (AAindex).
- SecStr (Possible Conformational Switches)[71] is a computational algorithm that uses the consensus secondary structure prediction program of same name to identify segments that have high tendency to form β-strands or α-helices.
- AmyloidMutants[72], a web-based tool, uses energy calculation to predict mutational and structural landscapes of amyloid fibrils.
- NetCSSP[73,74,75,76] is a computational method that calculates the hidden β-propensity for segments that appears to be α-helical in native state but have the ability to form β-strands.
- AMYLPRED[77] is a consensus amyloid prediction computational algorithm that employed five of the above described algorithms (Amyloidogenic Pattern[65], Average Packing Density[66], Hexapeptide Conformational Energy[68], Beta-strand contiguity[69], Pafig[70], SecStr[71]) to predict the aggregation prone segments in a peptide or protein. The method was later extended to include all the eleven aforementioned algorithms to predict the aggregation prone segments in a peptide or protein.

Such sequence-based bioinformatics methods have been utilized to predict the aggregation propensity of various peptides or proteins. For example, TANGO[60,61,62] was used to test the amyloidogenic propensity of 71 peptides that were not experimentally tested before[61]. The specific set of peptides were derived from the prion protein, lysozyme, and β2-microglobulin. TANGO correctly predicted the amyloidogenic propensity of 65 peptides indicating high accuracy[61].

2.2.2 Structural Modeling-Based Methods

Several types of structural modelling-based computational methods have been developed to study peptide or protein aggregation and amyloid formation, as the self-assembly process involves timescales that can exceed hours and also length scales varying from one to several hundred nanometers[80]. This resulted in a hierarchy of structural modeling-based computational methods, outlined below, from "extremely" coarse-grained computational models where the length scales are in the order of millimeters and time scales are in the order of seconds to the atomic-detail computational models where the length scales are primarily in the order of nanometers and time scales are in the order of nanoseconds up to microseconds.

Coarse-grained Computational Models

Coarse-grained computational methods can investigate systems with length scales in the order of μm and time scales in the order of μs, and they are frequently used to investigate amyloid aggregation. Molecular dynamics (MD) simulations is the standard method for simulating proteins where the velocities and positions are calculated by employing molecular mechanics[81]. Typically, in coarse-grained computational models, an ensemble of atoms is grouped into single bead, thus allowing longer simulation times[80]. Coarse-grained computational models have several advantages such as they can investigate the full assembly process from peptide or protein monomers to the

formation of fibrils which is currently beyond the scope of atomistic simulations[80]. Coarse-grained computational models can also be used to study amyloid aggregation in presence of other biomolecules that are in present in the cell environments such as lipid structures[80,82] and membranes[80,83].

A wide range of coarse graining models are available starting from models where coarse-graining is performed on the molecular scale to the models where coarse graining is performed lightly, thus mostly retaining the atomic resolution of the protein or peptides involved[80]. It should be noted that the coarse-grained computational models come with the trade-off between computational efficiency and accuracy[80]. As a result, the degree of coarse graining must be chosen carefully depending on the time and length scale of the system under investigation, as well as the type of question that needs to be addressed.[80] Examples of coarse-grained computational models are outlined below in order of increasing resolution, starting from extreme coarse-graining to light coarse-graining.

Muthukumar and colleagues[84] presented a coarse-grained computational method where an extended peptide chain is considered as an elementary building block which is modeled as a cuboid. Each cuboid can have several orientations and two cuboids can aggregate when the orientation of both are same. Through the course of the simulation each cuboid undergoes a random walk and changes its orientation. The specific cuboid lattice coarse-grained computational simulation aimed to capture the essence of nucleation growth mechanism of the amyloid fibrils.[84] Vendruscolo and colleagues[85] developed a coarse-grained model where the elementary building block was represented as a tube and they aimed to capture the nucleation growth mechanism of amyloid fibrils as well. The results from these simulations such as the size of the fibril, lag time during the mechanism, solubility and elongation rate were in line with experimental

observations.[85] The two aforementioned computational models are highly coarse-grained resulting in lower atomic resolution but are useful when the length and time scales are in the order of μm and μs, respectively.

Additional coarse-grained computational methods have been presented with higher resolution than the ones that were described above. The specific type of coarse-grained computational methods could also be called as phenomenological methods where one or more beads are used to represent each amino acid in the peptide or protein.[80] For example, Caflisch and colleagues[86,87,88,89,90,91] developed a model where each amino acid is represented by two beads. Within this model, each peptide can possess can two possible configurations, a β-competent conformation or a folded conformation, with a dihedral potential that can bias both the conformations to different degrees. The specific model showed the capability of differentiating highly amyloidogenic proteins from the less amyloidogenic ones. Examples of the high amyloidogenic proteins include $A\beta_{40}$, GNNQQNY, Phe-Phe whereas the less amyloidogenic ones include $A\beta_{42}$, prion proteins, myoglobin and Sup35.[88] According to the model, fibril formation for the systems investigated is a single pathway process and does not involve the formation of intermediates such as protofilaments or micelles and starts with the formation of smaller nucleus.[88]

An additional example involves the studies of Shea and colleagues[92,93,94,95,96] who developed a coarse-grained computational model where each amino acid was represented by three beads (one for the side chain and two for the backbone). Similar to the Caflisch model[86,87,88,89,90,91], dihedral potential within the model can control the β-sheet propensity of the simulation system. The specific model indicated different aggregation pathways for peptides depending on the β-sheet propensity of the peptide; highly amyloidogenic proteins formed fibrils through an ordered β-sheet nucleus whereas lower amyloidogenic peptides formed disordered oligomers first through which

ordered β-sheets emerged.[80] The specific model was also used to fibrils growth from which they observed that fibrils grew both by elongating (standard or primary growth mechanism) and by lateral growth (secondary growth mechanism)[93,94,95,96].

Furthermore, Hall and colleagues[97,98,99,100,101,102,103] developed a coarse-grained computational model called PRIME. The specific model represents each amino acid as four beads and is effectively able to capture the hydrogen bonding energetics by employing a directional square-well hydrogen bond. The specific model aims to perform parametrization by attempting to minimize the number of interaction parameters subject to the constraints that the parameters are physically meaningful. Hall and colleagues[97,98,99,100,101,102,103] applied the PRIME model in combination with discontinuous molecular dynamics (DMD) where transmission/reflection of the beads are computed in discontinuity to apply discontinuous breaks in the energy function. They applied the specific method to study the fibrillization of 48 fragment peptides from Tau with sequence VQIVYK at different temperatures[99]. They observed that at higher temperatures, fibrils formed rapidly through a templated mechanism. In addition, Hall and colleagues[102] also investigated the fibril formation for the peptide fragment Aβ (16-22) similar to the tau fragment mentioned above. They observed that at lower temperatures, fibril formation was through nucleation growth mechanism and at higher temperatures through templated mechanism. They also noted the formation of structural details such as the formation of antiparallel β-sheets, intersheet distance, side chain interdigitation in consistency with the experiments[15,104,105,106].

Moreover, Derreumaux and colleagues[107,108] employed a higher resolution coarse-grained model called OPEP[109]. The specific model represents the backbone of an amino acid in full atomic resolution for heavy atoms and represents the sidechain as a single bead. Among others, they

performed a simulation with trimer structures of Aβ (17-42) using the specific model in conjunction with a thermodynamically enhanced sampling method, REMD (described below)[107].

Atomistic Computational Models

Atomistic computational models typically offer the most detail where each atom of a protein is explicitly represented. Atomistic models are generally used to investigate proteins aggregation problems where the length scales are in the range of nm to μm and the time scales are in the range of ns to μs.[80] Similar to coarse-grained computational models, molecular mechanics is employed to calculate the velocities and positions of atoms in simulations using atomistic models. In order to calculate the positions and velocities of atoms, the potential energy of the proteins needs to be represented appropriately, from which positions and velocities are calculated.[110] The representation of the potential energy function, often called "force field", is used to calculate the potential energy of the system. Currently, there are various all atom force fields such as CHARMM[110,111,112,113,114], AMBER[115,116], OPLS[117]. The potential energy function of the CHARMM force field is presented below in *Eq. 2.1*.

$$
\begin{aligned}
U(\vec{R}) = & \sum_{\text{bonds}} K_b(b-b_0)^2 + \sum_{\text{angles}} K_\theta(\theta-\theta_0)^2 \\
& + \sum_{\text{Urey-Bradley}} K_{UB}(S-S_0)^2 + \sum_{\text{dihedrals}} K_\varphi(1+\cos(n\varphi-\delta)) \\
& + \sum_{\text{impropers}} K_\omega(\omega-\omega_0)^2 \\
& + \sum_{\substack{non-bonded \\ pairs}} \left\{ \varepsilon_{ij}^{min} \left[\left(\frac{R_{ij}^{min}}{r_{ij}} \right)^{12} - 2\left(\frac{R_{ij}^{min}}{r_{ij}} \right)^{6} \right] + \frac{q_i q_j}{4\pi\varepsilon_0 \varepsilon r_{ij}} \right\} \\
& + \sum_{\text{residues}} U_{CMAP}(\varphi,\psi)
\end{aligned}
\qquad 2.1
$$

The potential energy is described by the sum of bonded and non-bonded energy contributions as a function of the coordinates of the atoms in the system. According to *Eq. 2.1*, bonded terms comprise of bond (b), valence angle (θ), Urey-Bradley (UB, S), dihedral angle (φ), improper angle (ω), and backbone torsional correction (CMAP,φ,ψ) contributions. K_b, K_φ, K_{UB}, K_θ and K_ω are parameters which describe the force constants, while variables containing subscript 0 represent the corresponding values of equilibrium. With the exception of the dihedral angle term, represented by sinusoidal expression (*n:* multiplicity or periodicity of the dihedral angle, δ*:*phase shift) all the rest of the bonded terms are described by harmonic potentials. All possible valence and dihedral angles were included for bonded atoms. About a given bond, the dihedral angle term is expanded using Fourier series and one dihedral angle term is the most common although it can be expanded up to six terms. A numerical correction term, CMAP, has also been implemented for the protein main chain. The bonded term, Urey-Bradley, is implemented for three bonded atoms A-B-C and is a quadratic function of distance between atoms A and C. The bonded term, improper

dihedral angles are implemented when an atom is bound to three other atoms and is also a quadratic function of the dihedral angle for the four atoms. The Urey-Bradley is generally only used in special cases where as the improper dihedral angle term is commonly used. Non-bonded terms comprise of the Lennard-Jones (LJ) 6–12 term representing attractive and repulsive van der Waals interactions and Coulombic interactions between the point charges (q_i and q_j). Both the non-bonded terms are calculated within a distance specified by the user (covalently bonded atoms are excluded).[110] From the potential energy function defined above, force and acceleration on each atom are calculated using the Newton's second law of motion. Several algorithms are available for the determination of positions and velocities of atoms as a function of time.[110]

As protein aggregation involves timescales that can exceed hours and also length scales varying from one to several hundred nanometers, atomistic models are typically used to investigate the initial aggregation stages where the structural details can even exceed the experimental capabilities[80]. Atomistic models are also used to study the stability of preformed fibrils or the structural characteristics of monomers. Specifically, simulations using atomistic models have been helpful in studying intrinsically disordered proteins like Aβ.[80] The simulations could provide the transient secondary structure of the protein which could suggest the protein regions that are responsible for aggregation[80].

The performance of the atomistic models to investigate protein aggregation can be enhanced by employing various sampling methods such as replica exchange molecular dynamics (REMD) and by representing the solvent molecules implicitly via a mean effective potential without losing significant accuracy. The implicit representation of solvent molecules with sufficient accuracy is computationally efficient thus being cost effective. Several implicit solvent models[118,119] were first introduced based on the assumption that the protein under investigation is

a low dielectric medium even though the dielectric constant varies significantly in the protein's interior[120,121]. The specific models also assume the solvent as a medium with high dielectric constant and charge spatially distributed[118,119]. However, more detailed models were eventually formulated of which some commonly used implicit solvation models are based on generalized born (GB) electrostatics[110,122] represented by the equation *Eq. 2.2.*

$$\Delta G^{elec}_{\varepsilon_p \to \varepsilon_w} = -\frac{1}{2}\left(\frac{1}{\varepsilon_p} - \frac{1}{\varepsilon_w}\right)\sum_{i,j}\frac{q_i q_j}{\sqrt{r_{ij}^2 + \alpha_i \alpha_j \exp\left(-r_{ij}^2 / F\alpha_i \alpha_j\right)}} \qquad 2.2$$

In the above equation, ε_w, and ε_p are the exterior and interior dielectric constants with r_{ij} being the distance between atoms *i* and *j*, and αi being the effective Born radius of the atom *i*. The effective born radius, α, varies with the position of the atoms. The empirical factor, *F*, modulates the length-scale of the Gaussian term. The specific generally ranges from 2 to 10, with 4 as the most commonly used value. The continuum dielectric models based on GB electrostatics can be efficient compared to the explicit implementation of solvent and it also often works well in describing the thermodynamic aspects of the water solvation[123]. The specific models, referred to as generalized Born (GB) models, are also one of the fastest implicit solvent models and they are widely used for MD simulations of proteins. In contrast to a lot of simple implicit solvent models[124,125,126] which evaluate an electrostatic interaction between atom pair by using just distance dependent function, GB models also take into account the solvent exposure of the charges[123,127,128,129]. GB models evaluate the degree of burial of individual charges by calculating effective Born radii. The specific calculation is used to correct the Coulomb law for each pair of

atoms. Fast analytical continuum treatment of solvation (FACTS)[123], is one of simple version of GB models that evaluates solvent accessible surface area (SASA) and self-electrostatic solvation energy of each atom using intuitive geometric properties of the protein. The evaluation of those geometric properties requires only protein interatomic vectors. The specific model was extensively validated against more accurate GB models and the simulations performed using the FACTS implicit model were only four times slower than the corresponding simulations performed in vacuum[123], and has been used in several studies investigating amyloid formation.[40,130,131,132]

As mentioned above, the performance of MD simulations can also be improved by augmenting the simulations with enhanced sampling methods[80]. Replica exchange molecular dynamics (REMD)[133,134,135,136,137,138] is one of such enhanced sampling methods and it has been used extensively to investigate the problem of protein aggregation[139,140]. The specific method enhances sampling by starting multiple parallel simulations which each simulation exploring a particular point in the parameter space[80,138]. Temperature is the most commonly employed parameter for the specific method. After launching multiple simulations (also referred as trajectories) of the same system at different temperatures, each trajectory has the opportunity to exchange with neighboring trajectories. The exchanging between trajectories is performed according to a Metropolis criterion which enforces the proper thermodynamics. The specific exchanging procedure results in a discontinuous trajectory at each temperature, thus implicating the loss of kinetics in a simulation[80].

Various computational techniques have been developed to investigate protein aggregation which employ a combination of implicit solvent models and sampling method to enhance the performance of atomistic models. Among others, the computational protocol developed by Tamamis and Archontis[130] is an example of one such computational technique which investigates

the self-assembly properties of short amyloid peptides. The specific computational protocol, which has been extensively used in the current dissertation, performs MD simulations where multiple peptides with same sequence are placed in a cubic periodic box with finite concentration. The simulations are performed using CHARMM[110] with the corresponding CHARMM force-field. Water molecules in the simulation are take into account implicitly using FACTS[123] implicit solvent model (briefly described above). The performance of the simulation is enhanced by employing replica-exchange method[133,134,135,136,137,138] (briefly described above). Upon completion of the REMD simulation for an amyloid peptide under investigation, simulation snapshots at a specific temperature (typically at 300 K) are collected and structural analysis and statistical free energy calculations are performed to elucidate the peptide's elementary β-sheet structural units. Specifically, structural and statistical free energy analyses were performed by their in-house FORTRAN programs which are used to (i) analyze the propensity of the peptides to form intermolecular β-sheet structures of high complexity, and (ii) elucidate the key amyloidogenic physicochemical properties and β-sheet structures. High complexity β-sheet structures with highly-ordered and well-aligned strands (using nematic-P_2 and polar-P_1 order parameters calculated using WORDOM[141,142]) are collected from free energy minima basins of free energy landscapes and are analyzed in detail, as they are expected to elementary β-sheet represent structural units in the naturally formed amyloids. The computational protocol is schematically represented in the Figure 2.2 and has been implemented in several studies.

Specifically, on the basis of the insights gained by the computational study of the self-assembly of the peptide with sequence NSGAITIG which indicated that the N-terminal residues are not participating in the β-sheets and are solvent exposed, a series of amyloid materials with have been produced with a series of potential applications.[25,26,28]

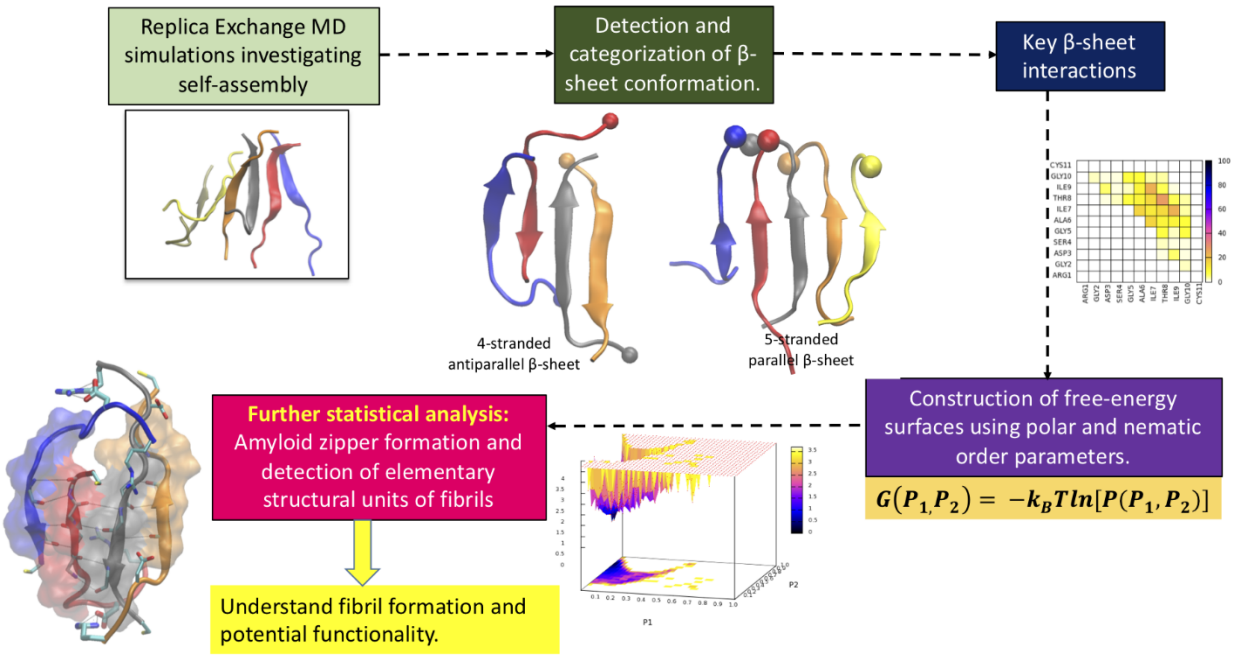


Figure 2.2 Schematic representation of the computational protocol developed by Tamamis and Archontis.[130]

2.3 Computational Protein Design and Functionalization, and their Application in the Design of Amyloid Materials

Computational protein design is generally defined as the design of new peptides or proteins that can fold into a specific target protein structure[143,144,145,146,147,148,149]. The ultimate goal of the protein design is to create a protein that can perform desired function and the function that could be performed by a protein is largely dictated by the structure of the specific protein. Thus, computational protein design uses *in silico* methods to design a target protein structure which encodes the function of the protein. Protein design is also referred as an "Inverse folding problem" (IFP) where protein folding problem is the determination of a protein structure given the sequence of the specific protein[143]. Thus, the goal in an IFP is to determine the protein sequence with lowest

free energy state that can form a target conformation. Traditionally, protein design was performed using experimental methods like site directed mutagenesis, rational design, and directed evolution[150]. However, the recent increase in the computational capacity prompted significant improvements of molecular force fields, structural bioinformatics, and protein design algorithms which enabled the advancement of computational protein design.

Anfinsen's thermodynamic hypothesis of protein folding states that protein should fold to a conformation or structure with lowest free energy. However, it should be noted that many of the proteins occurring in the nature are only marginally stable with energy difference between misfolded or unfolded states and native folded state as low as 5 kcal/mol[146]. This could be equivalent to just a few hydrogen bonds' contribution.[146] Many naturally occurring proteins could be marginally stable due to several reasons such as the functional activity of the protein could necessitate compromising structural features; marginally stable proteins are easy to degrade and clear by proteases as opposed to highly stable which are resistant to degradation; conformational entropy for protein folding is generally high as the misfolded or unfolded protein states outnumber the native folding states; proteins in nature are generally optimized only until the stage after which gains in organism fitness are hard to make.[146] Marginal stability generally sensitizes the protein to the changes in environment or mutations which easily lead to misfolding or aggregation of the protein.[146] Proteins *in vivo* overcome these limitations by employing a network of molecular chaperones called proteostatsis network (PN)[146]. These set of molecular chaperones help the proteins to fold and function efficiently under both normal and stress conditions. However, marginal stable proteins without the help of PN are going to be very sensitive to environment, have low expressibility rates due to higher chance of misfolding, have increased cost of production. This limits their use in many research applications.[146]

Thus, increasing the thermodynamic stability of a protein is the primary principle that drives the computational protein design research so that the specific highly stabilized can be used in research applications.[146] Factors contributing to the stability of a protein can be classified into four types.[146] First, hydrophobic effect which is the preference for the non-polar amino acids to isolate themselves from water which is a dominant force for protein stability.[146] This also necessitates the presence of polar and charged amino acid on the surface of the protein.[146] Second, van der Waals interactions where the hydrophobic amino acids generally pack themselves in protein cores where they form large number of favorable van der Waals contacts.[146] Third, hydrogen bonding where partial positive and negative charge atoms become hydrogen donor and acceptor, respectively and form favorable polar interactions.[146] It should be noted that conformational loss of entropy from hydrogen bond formation is relatively high since they are sensitive to orientation of acceptor and donor groups and also, they are formed at very close separation of the specific groups.[146] Although the energetic contribution from hydrogen bond formation is small, they are important in specifying the native structure of protein.[146] Fourth, salt bridges which are formed between opposite charged amino acids contribute to protein stability but only to a small extent.[146] Salt bridges are generally formed between amino acids that are surface exposed.[146] Although, buried salt bridges have large stabilizing effects due to the low dielectric environment at the core of the protein, they are rarely seen as the penalty to transfer the specific amino acid from polar to non-polar environment would be high.[146] Indeed, in the cases where buried salt bridges are observed, they are generally part of a network polar interactions which could provide additional stabilization.[146] Buried salt bridges are also important in specifying native protein structures as buried charged residues which do not form salt bridges in misfolded protein incur huge penalties.[146]

Computational protein design could also be broadly classified into four classes. In the first class, proteins are designed from scratch from the first principle of biophysics. This is generally referred as *de novo* protein design. In the second class, amino acid substitutions are performed on a protein with known structure and sequence. Here, amino acid at non mutable positions are occupied by the wild-type amino acids. This approach could also be referred to as protein redesign. In the third class, multi-component or self-assembling proteins are designed computationally using the principles of *de novo* protein design, which introduces a complexity that mutations need to be taken into account in a multimeric/self-assembly fashion. In the fourth class, protein design is performed computationally to a protein with the aim to produce a new protein with a specific functionality, such as the ability to bind to another substance (e.g., specific ion or compound).

2.3.1 Class 1: de novo Protein Design

Although there are various algorithms developed for *de novo* protein design, they typically follow an approach that can be divided into the following steps.[147] First, building the target protein structure which can be done using a known structure of a similar protein. However, if the target protein structure is not observed in nature, then the target structure can be built mimicking many defining characteristics of the proteins that occur in nature[147]. Some of characteristics are backbone polar amino acids should form hydrogen bonds with other backbone amino acids, backbone torsional angles should occupy the regions in the Ramachandran plot that are allowed, and there should be sufficient space between secondary structure units which allows tight packing between amino acid side chains. Top7 is an example novel protein that does not exist in nature which was computationally designed from scratch[151].

The second step in the *de novo* protein design is to account for the rigidity or flexibility of the backbone in the design template(s). Both flexible and rigid design templates have been used in

various *de novo* protein design studies[152]. *De novo* protein design efforts have begun by considering that backbone structure of a protein in three-dimensional space is fixed[153]. These rigid backbone templates are generally combined with a discrete set of side chain conformations that are frequently observed which are referred to as rotamers[154]. Examples of the successful studies which were performed with rigid backbone template include full sequence design of a protein that adopts zinc finger fold[155], β1 domain design in the protein G[156], and redesign of T4 Lysozyme's core[157]. Subsequently, experiments have noted that backbone shows flexibility by making adjustments when mutations are performed. Thus, *de novo* protein design efforts were focused on accounting for flexibility in the backbone design templates. Flexible design templates can be classified into three types[152]: (i) Multiple backbone templates with fixed backbone assumption on each template combined with discrete side chain rotamers. Examples of this class include the design of the novel 93 residue α/β protein Top7[151], redesign of β1 domain's core of the streptococcal protein[158]. (ii) Generation of a continuum backbone design template through algebraic parametrization combined with discrete side chain rotamers. Examples of this class include the design of a family of α-helical protein which have a right handed super helical twist[159]. (iii) Using continuum backbone template in combination with continuous ranges of backbone angles. Examples of this class include design of new compstatin variants with improved activity to inhibit complement protein 3 (C3)[160,161,162], redesign of the human β-defensin-2 peptide which a 41-residue antimicrobial peptide of the innate immune system[163].

The third step in the *de novo* protein design is to define the sequence space that is allowed to explore. Defining the sequence space generally includes specifying the set of amino acids that are allowed at each mutable residue position. The set of amino acids can be defined based on the four factors that were defined earlier which contribute to the stabilization of the protein namely

hydrophobic effect, van der Waals interactions, hydrogen bond formation, and buried salt bridge formation.[146]

The fourth step in the *de novo* protein design is the formulation of an energy function that will be able to discriminate sequences that are stable to form target protein structure from the sequences that are stable to form other competing protein structures. An ideal energy function in the computational protein design algorithms should be both simple and accurate to design protein successfully but achieving simplicity and accuracy at the same time is difficult[145]. Energy functions in various computational protein design programs generally consider all the four factors contributing to protein stability (discussed above) which are summed as contributions to molecular energy. However, different *de novo* design programs calculate the terms in energy function by employing different approaches[145]. RossettaDesign, a suite of computational tools to design proteins, uses a combination of knowledge-based terms calculated based on protein structure with physical-based calculations for hydrogen bonding and van der Waals terms to represent its energy function[164,165,166]. EGAD, a genetic algorithm for protein design, bases its energy function largely on the physical based molecular mechanics force field[167]. Liang-Grishin, a score function for effective protein design uses a combination of physical-based, geometrical-based and knowledge-based terms to construct its energy function[168].

2.3.2 Class 2: Computational Protein Redesign

The second class of computational protein design, protein redesign, follows the same approach as *de novo* protein design. However, there is no necessity to build the target protein structure and backbone design templates from scratch as the structure of the protein on which mutations are going to be performed is already available. The range of sequence space to be

explored is generally smaller than the corresponding *de novo* design as most of the amino acids are maintained as their wild-type amino acids.[147]

2.3.3 Class 3: Computational Design of Self-assembling Peptides

Computational protein design algorithms have also been employed to design self-assembling proteins or peptides[169,170,171,172,173,174,175]. Andre and colleagues[169] successfully designed a class of self-assembling peptides with βαβ fold by developing a computational protocol which employed the structure of seven-residue self-assembling peptide (GNNQQNY) from a prion-like protein, sup35, in microorganism yeast. Baker and colleagues developed computational methods to design self-assembling proteins[170,171,172,173,174]. For example, they designed co-assembling peptide system which organize themselves to highly ordered nanoscale architectures.[173] The co-assembling peptide system consist of 24 subunits with two distinct peptides which self-assemble into an architecture with tetrahedral point group symmetry.[173] Floudas and colleagues[175] developed a two-stage peptide design framework which can generate self-associating peptides. The first stage comprises of optimization-based sequence selection with simulated multimeric templates as inputs where the aim was to generate low energy sequences. The second stage comprises of computational validation where association affinity and fold specificity were evaluated. The specific peptide design framework was applied to design self-assembling tripeptides which were subsequently shown to have self-associating properties using experiments.

2.3.4 Class 4: Computational Protein Design for Functionalization

Computational protein design algorithms have been successfully employed to design functional proteins[151,164,165,166,176,177]. Yuan and colleagues[176] developed a computational protein

design program, metal search, which design tetrahedral binding sites for metals in known protein structures. The specific program was employed to introduce zinc binding sites in the β1 domain of streptococcal protein and also in the four-helix bundle protein α4. Richards and colleagues[177] developed a computational design program, DEZYMER, which constructs ligand binding sites in known protein structures. The specific program was employed to introduce copper binding site in the E. Coli protein thioredoxin. Rossetta, a computational protein design program developed by Baker and colleagues[151,164,165,166] has also been employed successfully in the design of numerous functional proteins.

2.4 Dissertation layout

The overall goal of this doctoral study is to computationally design novel functional amyloid materials rationally or through the use of rational and protocol-based approaches that have been developed within the study's framework. Chapter 3 will be dedicated to describing the design of a bifunctional self-assembling peptide RGDSGAITIGC which originated from a combination of rational computational and experimental methods. Chapter 4 will be dedicated to describing the design of the bifunctional self-assembling peptide RGDSGAITIGY which was designed using a combination of optimization and rational computational methods as well as experimental methods. Chapter 5 discusses the investigation of the amyloidogenic properties of a pentapeptide, GAIIG, as well as discusses the discovery of a novel amyloid designable scaffold encompassing GAIIG as a central amyloidogenic core. Chapter 6 discusses the computational design of functional amyloid materials with cesium binding, deposition and capture properties, and presents our first steps toward a computational protocol to design functional amyloid materials. Chapter 7 discusses our approach toward a generalized computational protocol for the design of functional amyloid materials binding to ions or compounds. Chapter 8 discusses the implementation of the protocol

to design functional amyloid materials binding to compounds involved in treating Alzheimer's disease. Chapter 9 discusses the implementation of the protocol to design functional amyloid materials with cell-penetration and DNA binding properties. Chapter 10 discusses the implementation of the protocol to design functional amyloid materials with cell-adhesion, calcium ion, strontium ion binding properties. Conclusion are finally presented in chapter 11, in addition to some insights into future work.

3. SELF-ASSEMBLED AMYLOID PEPTIDES WITH ARG-GLY-ASP (RGD) MOTIFS AND METAL BINDING RESIDUES AS SCAFFOLDS FOR TISSUE ENGINEERING*

3.1 Introduction

3.1.1 Overview and Significance

Self-assembling peptides and proteins are steadily gaining more interest for their potential use as scaffolds in tissue engineering applications, due to their ability to be self-organized into supramolecular structures, which can mimic the native extracellular matrix (ECM). In addition, as mentioned in the introduction, their properties can be simply tuned through changes at the sequence level, and they can be produced in sufficient quantities with recombinant technologies or chemical synthesis to allow them to address standardization and homogeneity issues required for applications. Thus, several studies focused on the exploitation of self-assembling proteins, peptides, and peptide amphiphiles and their potential application to form scaffolds suitable for tissue engineering.[10]

ECM macromolecules encompass bioactive signal sequences which are recognized by cells via cell transmembrane receptors called integrins. The interaction between integrins and bioactive epitopes of ECM activates signal transduction mechanisms. This in turn induces specific cellular functions including adhesion, migration, proliferation and differentiation. The RGD sequence motif is a paradigm of a bioactive epitope found in the structure of extracellular matrix proteins such as fibronectin.[178] Incorporation of this motif into peptide and protein scaffolds is used as a strategy in order to convey cell adhesion properties to them. Such scaffolds range from the simple

Fmoc-phenylalanine amino acid[179] to short peptides[180,181] and to recombinant proteins such as spider silk.[182] Nevertheless, the incorporation of RGD sequence motif within a self-assembled amyloid material needs to be done in suitably selected positions within the sequence of the peptide or protein scaffold in order to ensure proper surface accessibility of the motif.

In addition, a previous study demonstrated the engineering of proteins incorporating cysteine which allows site-specific functionalization.[183] Comparison of the engineered variants with the original cysteine-free silk protein revealed no apparent differences in solution and in the films.[183] Nevertheless, functionalization of the thiol groups of these silk protein-based films with molecules such as nanogold, dyes, biotin and β-galactosidase demonstrated the potential of such films for a broad range of applications, including tissue replacement[183].

3.1.2 Goal

In this work, we aimed to design a bifunctional self-assembling amyloid peptide, based on the NSGAITIG β-amyloid forming motif of the adenovirus fiber shaft that would encompass cell adhesion properties through the incorporation of an ArgGly-Asp (RGD) sequence motif and functionalization possibilities through a cysteine (C) residue, respectively.

3.1.3 Hypothesis

We hypothesized that can be achieved through the rational design of a novel peptide preserving the β-sheet core (S)GAITIG[131,132] of the NSGAITIG self- assembling sequence, and simultaneously incorporating the RGD sequence motif and a cysteine at suitably selected terminal peptide positions. To address this, we introduced computational and experimental methods investigating the incorporation of RGD and C (i) at the C- and N-terminal ends, and (ii) the N- and

C-terminal end, and thus we investigated peptides NH_3^+CSGAITIGRGD-$CONH_2$ (CSGAITIGRGD), and NH_3^+RGDSGAITIGC-$CONH_2$ (RGDSGAITIGC), respectively.

3.1.4 Objectives Accomplished

Insights from simulations verified that both peptides self-assemble into structures rich in β-sheet content. Computational studies comparing the two peptides indicated that RGDSGAITIGC possesses advantageous properties to serve as a bifunctional biomaterial compared to CSGAITIGRGD. Therefore, the RGDSGAITIG peptide was chosen for further experimental investigation. In addition, peptide NH_3^+RGESGAITIGC-$CONH_2$ (RGESGAITIGC), which contains a glutamate instead of aspartate at residue position 3, was computationally and experimentally investigated.

Additionally, according to experiments performed by our experimental collaborators (see below), on the basis of insights provided by the theoretical studies, our collaborators investigated the ability of the RGDSGAITIGC as well as the ability of the RGESGAITIGC to self-assemble into amyloid-like fibrils when suspended in physiological solutions. Different techniques such as transmission electron microscopy (TEM), field-emission scanning electron microscopy (FESEM), atomic force microscopy (AFM), and X-ray fiber diffraction analysis were utilized in order to experimentally confirm whether the original sequence could conserve the ability to self-assemble in the presence of the RGD motif in one terminus and the incorporation of an additional cysteine/alanine residue in the other terminus. Furthermore, our collaborators aimed to investigate whether thiols remain in their reduced state in cysteine-containing peptides externally from their fibrous network. In fact, the presence of exposed −SH functional groups is crucial for binding selectively to metals or surfaces and possibly to conjugate with signaling molecules. DTNB titration was used as a valid technique to quantify the degree of exposed sulfhydryl groups from

peptides in solution. In the experimental part, our collaborators also tested along a peptide containing an alanine at the C terminus (RGDSGAITIGA) as a non-cysteine analogue of the peptide RGDSGAITIGC. All peptides were tested for their ability to support adhesion and proliferation of a model fibroblast cell line.

This study resulted in a publication of Dr. Tamamis' computational lab in collaboration with Dr. Mitraki's experimental lab, and other experimentalists which their names are listed below in the complete publication reference: "Deidda, G.; Jonnalagadda, S.V.R.; Spies, J.W.; Ranella, A.; Mossou, E.; Forsyth V.T.; Mitchell, E.P.; Bowler, M.W.; Tamamis, P.; Mitraki, A. Self-Assembled Amyloid Peptides with Arg-Gly-Asp (RGD) Motifs as Scaffolds for Tissue Engineering. *ACS Biomater. Sci. Eng.* **2017**, *3*, 1404-1416."

G.D. and S.V.R.J. are equally contributing first authors.

3.2 Computational Methods

The computational protocol developed by Tamamis and Archontis[130] described in chapter 2 was introduced to investigate the self-assembly properties of the peptides with sequences CSGAITIGRGD, RGDSGAITIGC, and RGESGAITIGC.

3.2.1 "Self-Assembly" Replica Exchange MD Simulations

Initially we placed six copies of peptide sequences CSGAITIGRGD, RGDSGAITIGC, and RGESGAITIGC, independently, in a cubic periodic 148 Å box, which approximately corresponds to a 3 mg/ml concentration. The initial conformations of the six peptides were extracted using clustering analysis from replica exchange MD simulations investigating the isolated peptides in infinite dilution which varied and primarily contained random coil and β-turn structural elements. All six peptides were initially placed in the center of the box, and subsequently peptides 1 and 2 were translated by +25 Å and -25 Å in the x direction, peptides 3 and 4 were translated by +25 Å

and -25 Å in the y direction, and peptides 5 and 6 were translated by +25 Å and 25 Å in the z direction. We performed replica exchange MD simulations to investigate the self-assembly properties of the three peptides, independently, as in references[40,130, 131,132,184]. Water was modelled implicitly using the FACTS19[123] solvation model and the value of the surface tension coefficient equal to 0.015 kcal/(mol*Å^2). We used Langevin dynamics, and a 5ps^{-1} friction coefficient was introduced on all non-hydrogen atoms of each peptide. The duration of each replica exchange run was equal to 10 ps. In the replica exchange MD simulation runs, sixteen temperatures (290, 295, 300, 305, 310, 315, 321, 327, 333, 339, 345, 352, 359, 366, 373 and 380 K) were employed, and the total simulation time for all temperatures per system was equal to 8 μs. The simulations were performed using CHARMM[110]. For each peptide sequence, we collected the final conformations of each replica exchange run at 300 K. These conformations were combined in one trajectory per peptide sequence, which corresponds to 500 ns and contains 50,000 snapshots.

3.2.2 Categorization and Probability of β-Sheet Conformations

Following the computational protocol developed by Tamamis and Archontis[130] described in chapter 2, we used DSSP[185] and in-house FORTRAN programs to investigate the self-assembly properties of the three different peptides, CSGAITIGRGD, RGDSGAITIGC, and RGESGAITIGC, using a similar protocol to the one described in references[40,131,132]. Specifically, DSSP[185] was employed to identify the formation of intermolecular beta sheet structures, and FORTRAN programs categorized the intermolecular beta sheet patterns observed into antiparallel, parallel and mixed conformations comprising 2- to 6- peptide strands. This analysis provided us with the information on which configuration, parallel versus antiparallel, is the most dominant per peptide sequence, according to the self-assembly simulations.

3.2.3 β-Sheet Residue-Pairwise Interactions

The aforementioned simulations resulted in the self-assembly of all three peptide sequences investigated into β-sheet rich conformations consisting of two to six peptides. The 4-, 5-, and 6-stranded antiparallel or parallel β-sheets that were computationally investigated acquire higher complexity than 2- and 3-stranded β-sheets; hence, the β-sheet patterns observed in the former can potentially correspond to patterns which may exist in the naturally occurring amyloid fibrils. In what follows, we focused our analysis on the 4-, 5-, and 6- antiparallel and parallel β-sheets. With the exception of 2-stranded antiparallel β-sheets in RGDSGAITIGC, the previous analysis depicted that a parallel β-sheet arrangement is more favored than antiparallel in CSGAITIGRGD, whereas an antiparallel β-sheet arrangement is more favored than parallel for the amyloid fibrils formed by both RGDSGAITIGC, RGESGAITIGC peptides. Thus, in the subsequent analysis we focused our analysis on 4-, 5-, and 6-stranded parallel β-sheets formed by CSGAITIGRGD peptides, as well as 4-, 5-, and 6-stranded antiparallel β-sheets formed by peptides RGDSGAITIGC and RGESGAITIGC. We calculated the (%) probability of an intermolecular pair of residues, which belong to two neighboring β-interacting peptides, to be involved in a β-sheet (extended β-sheet or isolated β-bridge) formation. This analysis shows the predominant patterns of intermolecular residue-pairwise β-sheet interactions and provides insights into the expected amyloidogenic regions of each of the peptides. Furthermore, it reveals which residues are not involved in the formation of intermolecular β-sheets in each peptide sequence, which may be important for functionalization purposes and specific biomaterial applications.

3.2.4 Classification of β-sheet Conformations Using P_1 and P_2 Parameters and Identification of Well-Aligned and Well-Ordered β-Sheet Conformations

We examined the extent of peptide alignment of the 4- to 6- stranded β-sheet conformations via the polar order-parameter P_1 and the nematic order-parameter P_2, defined in *Eq. 3.1*. These parameters are used in the structural characterization of liquid crystals, and have been employed successfully in simulation studies of peptide aggregation.[40,130,132,131,186,187]

$$P_1 = \frac{1}{N}\sum_{i=1}^{N}\vec{z}_i\vec{d}, \quad P_2 = \frac{1}{N}\sum_{i=1}^{N}\frac{3}{2}\left(\vec{z}_i\vec{d}\right)^2 - \frac{1}{2}, \qquad 3.1$$

In *Eq. 3.1*, N is the number of molecules in the simulation and $\vec{z}_i$ is a unit vector along a suitably defined molecular direction; $\vec{d}$ is a unit vector along a preferred direction of alignment, which emerges from the properties of the system. For CSGAITIGRGD, we selected $\vec{z}_i$ to be defined by the segment spanning from Cα atom of residue 4 to the Cα atom of 9, while for RGDSGAITIGC and RGESGAITIGC we selected $\vec{z}_i$ to be defined by the segment spanning from Cα atom of residue 5 to the Cα atom of 10. The selection was primarily based on the key β-sheet regions of each peptide.

We computed P_1 and P_2 for the 4-, 5-, and 6- stranded parallel β-sheets formed by CSGAITIGRGD peptides, as well as the antiparallel β-sheets formed by peptides RGDSGAITIGC and RGESGAITIGC peptides using Wordom[141,142]. According to the results, among 4-, 5-, and 6-stranded β-sheets, highly populated and highly-ordered β-sheets are observed only in 4-stranded

β-sheet conformations across all systems. We focused our analysis on 4-stranded β-sheets and specifically aimed at examining the structural properties of highly-ordered and well-aligned conformations corresponding to the parallel β-sheets formed by CSGAITIGRGD, and the antiparallel β-sheets formed by RGDSGAITIGC and RGESGAITIGC. We constructed free-energy surfaces for the specific 4-stranded β-sheets, using the two dimensional probability $P(P_1, P_2)$ and the relation:

$$G(P_1, P_2) = -k_B T \ln\left[P(P_1, P_2)\right] \qquad 3.2$$

In each case we detected the free-energy basin which encompasses the majority of conformations and where the global minimum falls into, and subsequently we extracted the states with the highest polar order and highest nematic order. The selected highly-ordered and well-aligned 4-stranded conformations were utilized for all subsequent analyses which aimed at obtaining additional insights, primarily focusing on the peptides' structural properties, as well as the role of the RGD/RGE sequence motifs and residue C, which are the potential key functional groups of the peptides.

3.2.5 Side Chain Contacts and Solvent Accessible Surface Areas

We used in-house FORTRAN programs to investigate the propensity of intra- and inter-molecular contacts formed between the side chains of residues of peptides within the selected 4-stranded highly ordered and well-aligned β-sheet states. We used a 6.5 Å distance cutoff between the geometric centers of two side chains as a criterion to denote that they are in contact. In addition, we used Wordom[141,142] to investigate the degree of solvent accessibility of the RGD/RGE sequence

motifs and residue C within the 4-stranded highly ordered and well-aligned states. We calculated the ratios of solvent/total accessible surface areas for specific backbone and side chain moieties belonging to RGD or RGE sequence motifs and residue C, of the two central peptides in each 4-stranded highly ordered and well-aligned β-sheet states. The degree of solvent accessibility of the two outer peptides is artificially high owing to the absence of interacting peptides on both sides; thus, the values for the two outer peptides are not reported. The total accessible surface areas were approximated as the maximum solvent accessible surface area (SASA) and is computed by calculating the SASA of the residue or residue moiety in the structure when all other atoms are removed. While this approximation is expected to overestimate the total SASA of a residue compared to other studies (e.g., ref [188]), we used this definition as we considered it as an optimum way to perform a fair comparison of the investigated residue groups' solvent accessibility across the different peptides investigated in this study. A 1.4 Å sphere (approximating the radius of a water molecule) was introduced to "probe" the surface of the molecule.

3.2.6 Hydrogen Bonds and Salt Bridges Formed by the Positively Charged Group of Arginine

The positively charged groups of Arg9 in CSGAITIGRGD, as well as Arg1 in RGDSGAITIGC and RGESGAITIGC are expected to be critical for cell attachment properties. For the selected 4-stranded highly ordered and well-aligned β-sheet states, we used in-house FORTRAN programs and calculated the intra- and intermolecular (%) occupancies of hydrogen bonds formed by the charged amide group of Arginine with all possible oxygen donors or acceptors. Furthermore, we calculated the probability of Arg9 in CSGAITIGRGD, as well as Arg1 in RGDSGAITIGC and RGESGAITIGC to be involved in an intermolecular salt-bridge or intramolecular salt-bridge or inter/intra- molecular salt-bridge. A 3.7 Å distance cutoff between

two heavy atoms was used as a criterion to denote a hydrogen bond/salt-bridge formation in this analysis.

3.2.7 Geometrical Properties of the RGD motif in Fibronectin Compared to the Designed Amyloid Peptides

NMR structural analysis of recombinant polypeptide domains encompassing an RGD motif, such as the tenth type III repeat of fibronectin depict that the RGD motif is located at the apex of a solvent exposed loop formed by two antiparallel β-strands[189]. Apart from the critical role of the solvent accessibility of the motif in its functionality[178], we also examined additional geometrical features of the RGD motif in the NMR structure of the fibronectin[189] which may also contribute to its functionality. Specifically, we investigated the correlation between three geometrical features of the RGD motif in the ensemble of NMR structures[189]: (i) the solvent accessibility of the motif, (ii) the tendency for a salt-bridge formation between the positively and negatively charged groups, and (iii) the tendency for a bend formation in the backbone of the motif. The solvent accessibility was assessed by calculating the ratio of the SASA divided by the total SASA of motif, the tendency for a salt-bridge formation was assessed by calculating the distance between the CZ atom of positively charged group and CG atom of negatively charged group, and the tendency for a bend in the motif was assessed by calculating the Cα : Cα distance between the first and the last residue of the motif. The aforementioned geometrical features which were calculated for the RGD motif in the fibronectin[189] were compared to the corresponding features of the RGD/RGE motifs, calculated for the ensemble of highly ordered and well-aligned 4-stranded parallel β-sheets formed by the CSGAITIGRGD peptide, as well as the antiparallel β-sheets formed by the RGDSGAITIGC and RGESGAITIGC peptides.

3.3 Materials and Experimental Methods

All experiments described in 3.3 were performed by Dr. Mitraki's lab and by additional experimental collaborators, listed as authors in ref.[24]

3.3.1 Materials

The following synthetic peptides were studied: a) NH_3^+-RGDSGAITIGA-$CONH_2$, b) NH_3^+-RGDSGAITIGC-$CONH_2$, and c) NH_3^+-RGESGAITIGC-$CONH_2$. The peptides were purchased from Genecust (Luxemburg) and possessed a degree of purity higher than 95%. These synthetic lyophilized peptide powders were dissolved in sterile water pH 7.4 or in 10 mM phosphate buffer pH 7.4 at concentrations of 3 mg/ml.

3.3.2 Transmission Electron Microscopy (TEM)

Samples for TEM analysis were prepared after drying for 24h a battery of grids (300 square mesh copper, 3,05 mm, Agar Scientific) coated with formvar film. After that, 8 μl of the sample were applied on the shiny face of the grid, left aside for two minutes, dried with a filter paper and then the same procedure was repeated with the stain. The samples were negatively stained with 8 μl of uranyl acetate 1% for two minutes. The TEM experiments have been performed by using a JEOL JEM 2100, High Resolution microscope, operating at 80 kV (University of Crete, Biology Department).

3.3.3 DTNB titrations

The first titrations setting with Ellman's reagent (5,5'-dithiobis-2-nitrobenzoic acid) or commonly named DTNB, Sigma Aldrich) was prepared by following the experimental section of Hauser's group[181]. Further details are provided in the ref[24].

3.3.4 Cell Culture and MTT Test

Cell cultures were performed by using murine fibroblasts NIH/3T3 (ATCC American Type Culture Collection) which were incubated at a concentration of 10^5 cells/mL in Dulbecco's Modified Eagle's Medium (DMEM)/Ham's Nutrient Mixture supplemented with 10% fetal bovine serum (FBS, Sigma), 4 mmol/L L-glutamine (Sigma) and 1% antibiotic (Pen-Strep) solution (GIBCO, Invitrogen). The MTT assay measures the capacity of viable cells to metabolize a light-yellow water-soluble tetrazolium salt (3-4,5 dimethylthiazol-2,5 diphenyl tetrazolium bromide) into an insoluble purple formazan product[190]. The assay was carried out by referring to the Sigma-Aldrich protocol. Further details are provided in the ref[24].

3.4 Results and Discussion

3.4.1 β-sheet Conformations

All the three peptides which were computationally investigated are frequently arranged into β-sheet conformations, in line with their amyloidogenic propensities (Figure 3.1). The β-sheet conformations formed by the peptides include antiparallel/parallel 2-stranded β-sheets, as well as antiparallel/parallel/mixed 3-, 4-, 5-and 6- stranded β-sheets. For β-sheet conformations with high degree of complexity (e.g., 4-, 5- and 6-stranded β-sheets), the parallel arrangement is more favorable compared to the antiparallel one for the CSGAITIGRGD peptide, while the antiparallel arrangement is more favorable compared to the parallel one for RGDSGAITIGC and the RGESGAITIGC peptides. The predominant parallel arrangement in the CSGAITIGRGD peptide is in line with our previous study indicating that NSGAITIG peptides tend to be arranged in a parallel configuration[131]. The predominant antiparallel arrangement of the RGDSGAITIGC and the RGESGAITIGC peptides can, at least for the current peptide sequences, presumably be attributed to the fact that the parallel arrangement may be somewhat disfavored owing to the

energetically unfavorable accumulation of two positively charged groups (a positively charged N-terminal domain, and a positively charged arginine side chain) on one side of the fibril, which presumably it cannot be counterbalanced by the presence of highly favorable intermolecular non-covalent interactions between cysteine residues on the other side of the fibril.

3.4.2 β-Sheet Residue-Pairwise Interactions

We focused our analysis on β-sheet conformations containing 4-, 5-, or 6- strands of (i) peptide CSGAITIGRGD bonded in parallel fashion, (ii) peptide RGDSGAITIGC bonded in antiparallel fashion, and (iii) peptide RGESGAITIGC bonded in antiparallel fashion. The analysis focused on 4-, 5-, or 6- β-sheets as these conformations represent β-sheet structures with higher complexity which can potentially correspond to the naturally occurring amyloid fibrils. We calculated the propensity of two residues which belong to neighboring β-sheet interacting peptides to be involved in a β-sheet formation (β-bridge) and the results are presented in Figure 3.2. In all systems, the β-sheet rich region is primarily comprised by residues within the sequence segment GAITIG. Sequence segment RGD and residue C do not play a primary role in the stabilization of the β-sheet conformations, with the exception of residues Arg9 and Gly10 which may participate in β-sheet interactions in CSGAITIGRGD (Figure 3.2A, 3.2D, 3.2G).

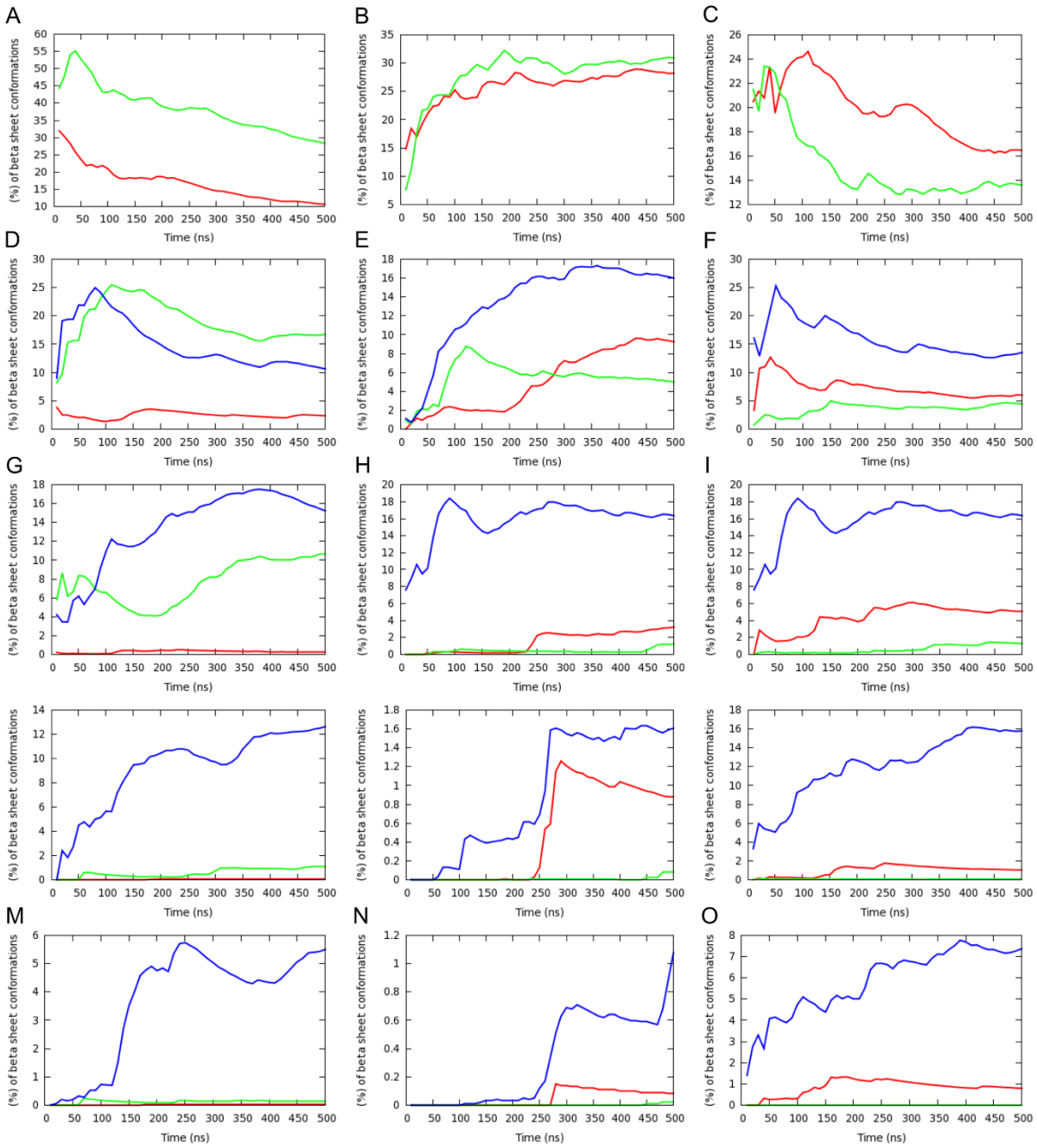


Figure 3.1 Moving averages of the fraction (%) of conformations with intermolecular β-sheets in the replica exchange MD simulations at 300 K for peptides with sequence CSGAITIGRGD, RGDSGAITIGC, and RGESGAITIGC. Figures A, D, G, J, M in the first column correspond to 2-, 3-, 4-, 5-, 6- stranded β-sheets of peptide CSGAITIGRGD, respectively. Figures B, E, H, K, N in the second column correspond to 2-, 3-, 4-, 5-, 6- stranded β-sheets of peptide RGDSGAITIGC, respectively. Figures C, F, I, L, O in the third column correspond to 2-, 3-, 4-, 5-, 6- stranded β-sheets of peptide RGESGAITIGC, respectively. Red color corresponds to antiparallel β-sheets, green color corresponds to parallel β-sheets, and blue color corresponds to mixed β-sheets. Reprinted with the permission from ref.[24]

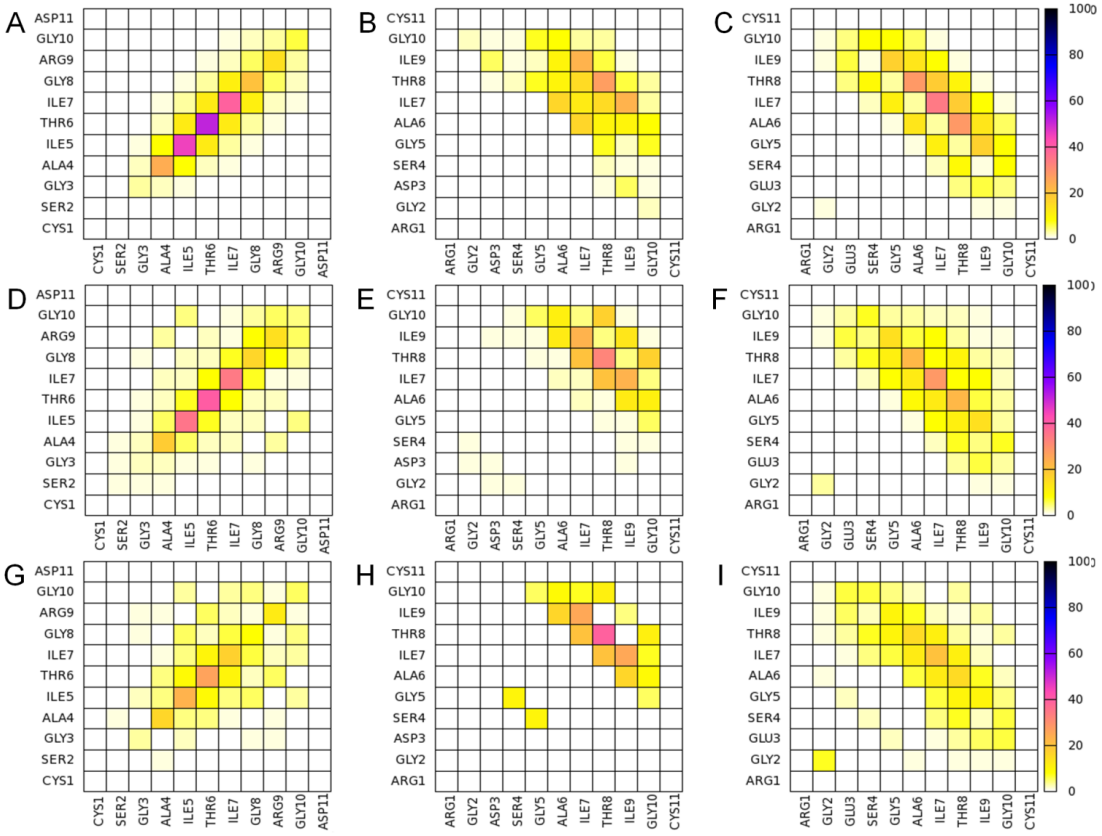


Figure 3.2 Density (%) maps of residue pairs forming intermolecular β-bridges for β-sheet conformations of CSGAITIGRGD, RGDSGAITIGC, and RGESGAITIGC. The pairs of residues belong to nearest neighboring peptides participating in an isolated β-bridge or extended β-sheet conformation in the replica exchange MD simulations at 300 K. Figures A, D, G correspond to the β-bridges in 4-, 5- and 6-stranded parallel conformations formed by peptide CSGAITIGRGD. Figures B, E, H correspond to the β-bridges in 4-, 5- and 6- stranded antiparallel conformations formed by peptide RGDSGAITIGC, respectively. Figures C, F, I correspond to the β-bridges in 4-, 5-and 6- stranded antiparallel conformations formed by peptide RGESGAITIGC, respectively. Reprinted with the permission from ref.[24]

3.4.3 Classification of β-sheet Conformations using P_1 and P_2 parameters and Identification of Well-Aligned and Highly-Ordered β-Sheet Conformations

We extracted the states containing 4-, 5- and 6-stranded β-sheet parallel conformations from peptide CSGAITIGRGD, as well as antiparallel conformations from peptides RGDSGAITIGC and RGESGAITIGC, and evaluated their degree of alignment and order by

plotting P_1 polar order parameter P_2 nematic order parameter (Figure 3.3). The results show that for all peptides which were investigated computationally, highly-ordered (high P_2 value) and well-aligned β-sheets (P_1 value depends on the number of peptides in the β-sheet) are predominantly observed in the 4-stranded rather than 5- and 6- stranded β-sheet conformations. Hence, since our aim was to identify β-sheet conformations containing high degree of order and well-aligned strands which can correspond to the naturally occurring β-sheet structures in the amyloid fibrils, we constructed 2D energy free surfaces, using the calculated P_1 and P_2 parameters for the 4-stranded parallel β-sheets formed by CSGAITIGRGD, as well as the 4-stranded antiparallel β-sheets formed by RGDSGAITIGC, RGESGAITIGC peptides (top panel of Figure 3.4). The global free energy minima in the plots are located in basins which are marked using black dashed lines, and within these basins we observe the presence of highly-ordered and well-aligned β-sheet conformations. As the specific β-sheet conformations are very likely to correspond to the naturally occurring conformations, we extracted a subset of 4-stranded highly-ordered and well-aligned β-sheet states for subsequent computational structural investigation. Representative conformations of the selected highly-ordered and well-aligned β-sheet states are shown in the bottom panel of Figure 3.4.

3.4.4 Amyloid-Forming Ability of the Designed Peptides

To determine whether the designed peptides conserved their ability to form amyloid fibrils, the peptide-lyophilized powders were dissolved in ultrapure water and followed by TEM over time by experimental collaborators listed as authors in ref. [24] Figure 3.5 shows TEM micrographs of the RGDSGAITIGA, RGDSGAITIGC, and RGESGAITIGC peptides after negative staining. For all peptides, straight, non-branched fibrils with diameters ranging from 10 to 50 nm and reaching several micrometers in length are observed. The TEM results confirm that the designed peptides

maintain the ability to self-assemble into amyloid-type fibers and the extra RGD or RGE residues can be accommodated at the N-terminus without hindering the self-assembling propensity of the amyloid-forming core.

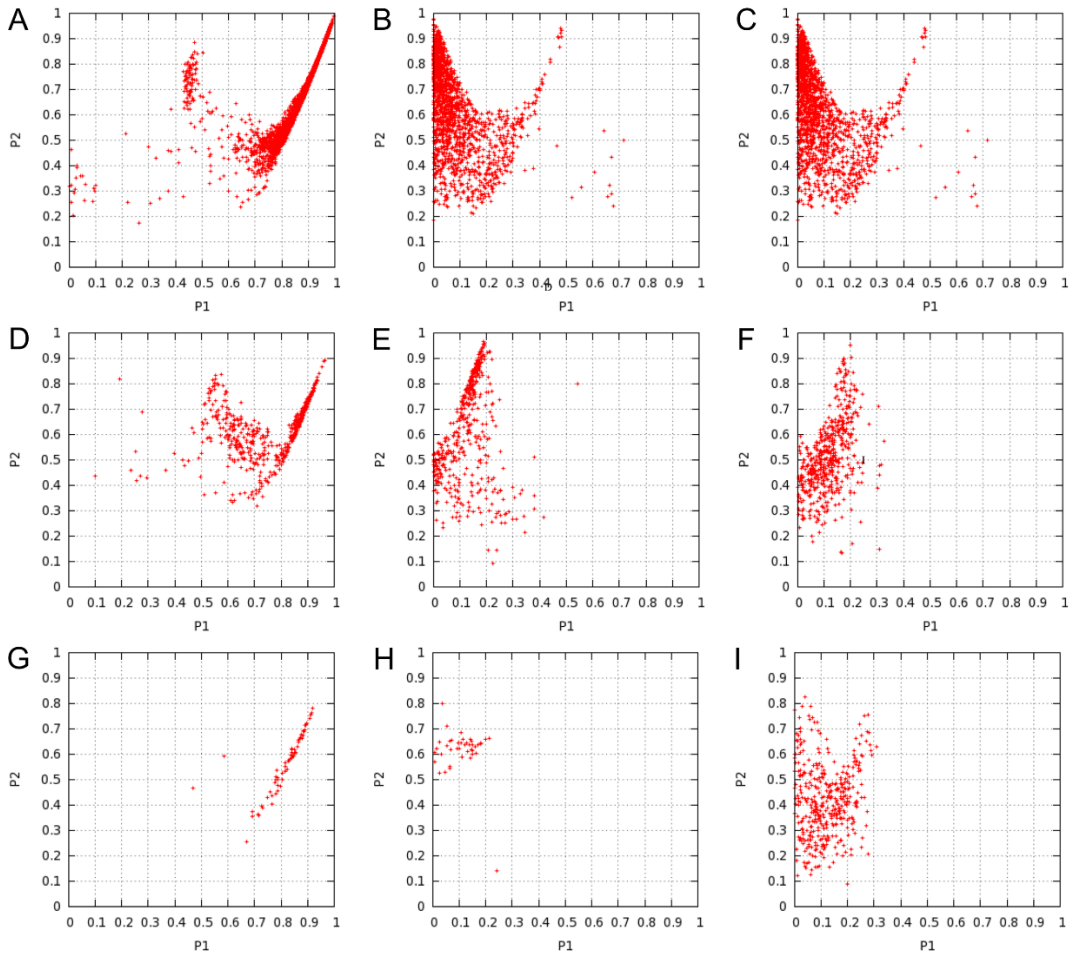


Figure 3.3 Plots of P_1 (x-axis) as a function of P_2 (y-axis) for the 4-, 5-, 6- stranded β-sheet conformations observed in the replica exchange MD simulations at 300 K of peptides with sequence CSGAITIGRGD, RGDSGAITIGC, and RGESGAITIGC. Panels A, D, G in the first column correspond to 4-, 5-, 6- stranded parallel configurations of peptide CSGAITIGRGD, respectively. Panels B, E, H in the second column correspond to 4-, 5-, 6- stranded anti-parallel configurations of peptide RGDSGAITIGC, respectively. Panels C, F, I in the fourth column correspond to 4-, 5-, 6- stranded anti- parallel configurations of peptide RGESGAITIGC, respectively. Reprinted with the permission from ref.[24]

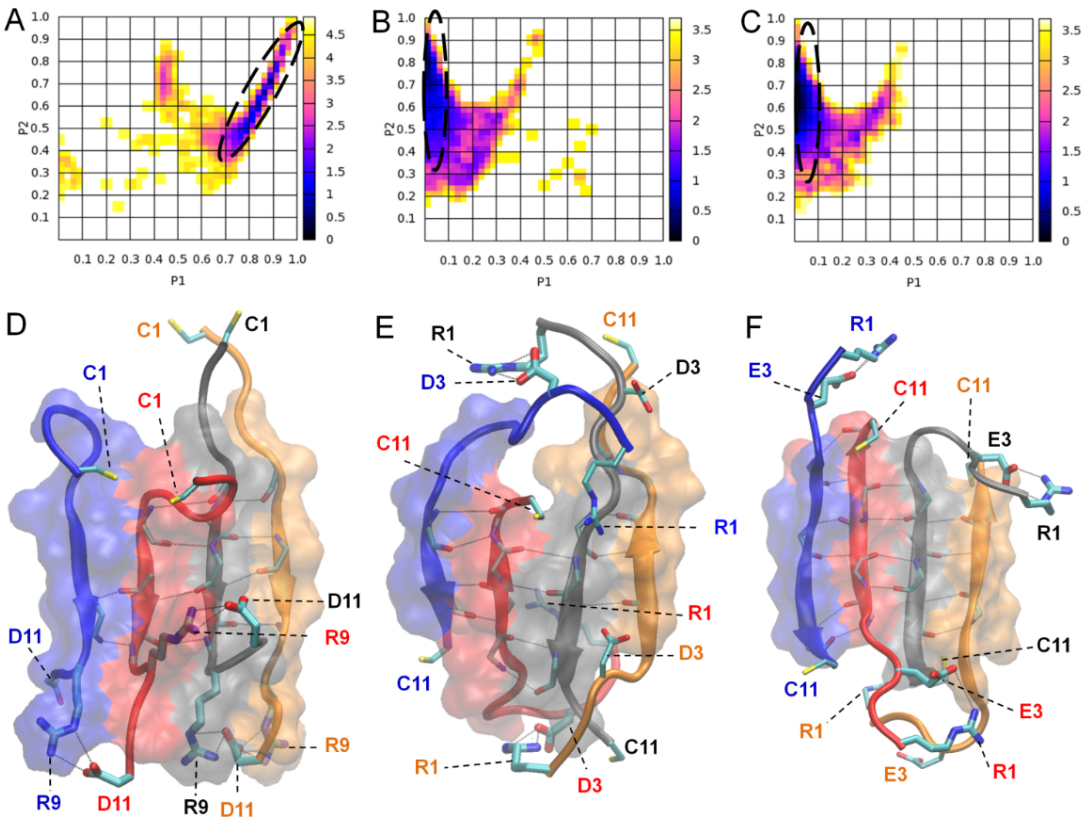


Figure 3.4 Free energy surfaces constructed from the 2-D probabilities of order parameters P_1 and P_2 and molecular graphics images of representative structures of CSGAITIGRGD, RGDSGAITIGC, and RGESGAITIGC peptides extracted from the free energy minima. Upper Panel: Free energy surfaces constructed from the 2-D probabilities of order parameters P_1 and P_2, calculated for the 4-stranded β-sheets observed in the replica exchange MD simulations at 300 K. Surface in panel A corresponds to 4-stranded parallel β-sheets formed by CSGAITIGRGD peptides, and surfaces in panels B and C correspond to 4-stranded antiparallel β-sheets formed by RGDSGAITIGC and RGESGAITIGC peptides, respectively. Bottom Panel: Simulation snapshots of representative highly-ordered and well-aligned β-sheet states of CSGAITIGRGD peptides in parallel arrangement (panel D), RGDSGAITIGC and RGESGAITIGC peptides in antiparallel arrangement (panels E and F). Reprinted with the permission from ref.[24]

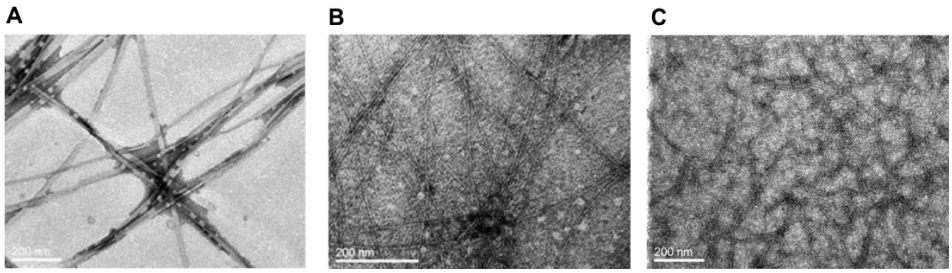


Figure 3.5 TEM micrographs of peptides RGDSGAITIGA (A), RGDSGAITIGC (B) and RGESGAITIGC (C) peptides at a concentration of 3mg/ml, aged for 96 hours at pH 7.4. The samples were diluted 1/3 prior to staining with uranyl acetate 1%. Figure was produced by our experimental collaborators and reprinted with the permission from ref.[24]

3.4.5 Side Chain Contacts and Solvent Accessibility of the Functional Groups

We performed a statistical analysis of the intra- and intermolecular side chain contacts within the highly ordered and well-aligned states of 4- stranded parallel β-sheets formed by CSGAITIGRGD, 4- stranded antiparallel β-sheets formed by RGDSGAITIGC and RGESGAITIGC peptides to gain additional insights on the role of side chain interactions in the stabilization of the β-sheets, as well as to investigate in-depth the side chain contacts formed by the key functional sequence motif RGD/RGE and residue C (Figure 3.6). The intermolecular contacts involve side chain interactions between nearest neighboring β-sheet interacting peptides only, whereas intramolecular contacts involve side chain interactions between residues within the same peptide. As for intermolecular side chain contacts, in all β-sheet states analyzed here, we observe a tendency for the side chains to cluster across the β-sheet region of each peptide (Figure 3.6) and form an amyloid steric zipper[15], (Figure 3.6). In CSGAITIGRGD the amyloid zipper formed involves partly the RGD sequence motif, in contrast to RGDSGAITIGC and RGESGAITIGC peptides (Figures 3.4 and 3.6). This result suggests that the RGD sequence motif is potentially more solvent exposed when it is part of the β-sheet states formed by RDGSGAITIGC

compared to CSGAITIGRGD (see below). As for intramolecular side chain contacts, apart from the expected tendency of nearest neighboring residues in the sequence to interact with each other, the RGD motif in the antiparallel RGDSGAITIGC β-sheet states forms contacts with the side chains of residue moiety 6−10. An in- depth investigation revealed that this is partly attributed to the fact that one of the two outer peptides tends to bend in approximately 1/3 of the β-sheet states, which is presumably an artifact of the limited number of peptides used in the simulations.

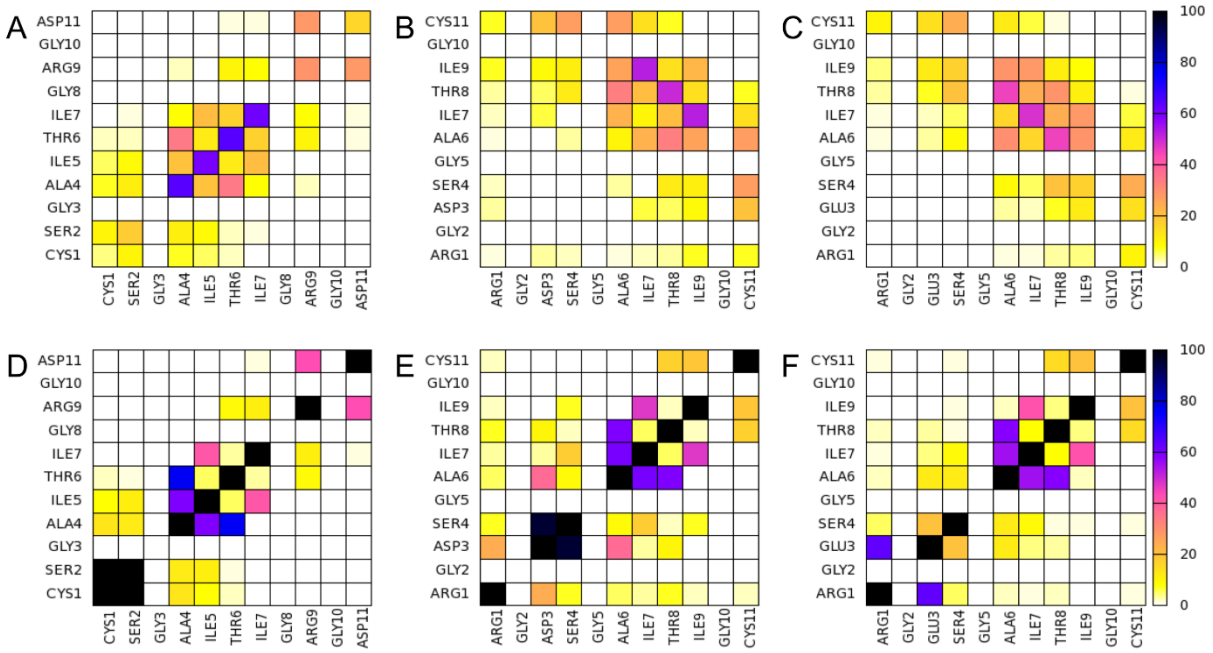


Figure 3.6 Percentage (%) probability density maps of side chain contacts for the highly-ordered and well-aligned 4-stranded β-sheets observed at 300 K for peptides with sequence CSGAITIGRGD, RGDSGAITIGC, and RGESGAITIGC. Maps A, B, C correspond to intermolecular side chain residue-residue contacts between nearest neighboring interacting peptides. Maps D, E, F correspond to intramolecular side chain residue-residue contacts. A contact is present when the two geometric centers of the two side chains are within 6.5 Å. Maps A and D correspond to parallel β-sheets formed by CSGAITIGRGD, maps B and E correspond to antiparallel β-sheets formed by RGDSGAITIGC, and maps C and F correspond to antiparallel β-sheets formed by RGESGAITIGC. Reprinted with the permission from ref.[24]

In addition, we performed a statistical analysis of the degree of exposure of the entire residues and side chains of the sequence motifs RGD/E, as well as the arginine and cysteine residues which belong to the two central peptides. The degree of solvent accessibility by RDG/E motif is higher in the antiparallel RGDSGAITIG (66±8%) and RGESGAITIGC (69±6%) β-sheet states compared to the parallel CSGAITIGRGD (41±7%) β-sheet states, and this can partly be attributed to the fact that the side chain group of the arginine residue is more solvent exposed in the RGDSGAITIG (55±11%) and RGESGAITIG (56±10%) β-sheet states compared to CSGAITIGRGD (35±8%) ones. The side chain group of cysteine residues is more solvent exposed in the CSGAITIGRGD (48±9%) parallel β-sheet states compared to the antiparallel RGDSGAITIGC (34±9%) and RGEGAITIGC (36±10%) β-sheet states. A visual inspection of the 4-stranded highly-ordered and well-aligned β-sheets formed by all peptides shows that the thiol group of cysteines is adequately solvent exposed to be functional with regard to metal-nanoparticle/surface-attachment (Figure 3.4).

3.4.6 Geometrical Properties of the RGD motif in Fibronectin Compared to the Designed Amyloid Peptides

The geometrical analysis of the RGD motif as part of the fibronectin NMR structures validates that high solvent accessibility can be considered as a factor correlating with functionality as in all structures the RGD motif experiences high exposure to the solvent (Figure 3.7)[178]. While the formation of a salt-bridge between the oppositely charged groups should not preclude functionality of the motif as it occurs with an approximately 16% in the NMR structures[189], it is plausible that a persistent salt-bridge formation would potentially limit arginine's accessibility, and thus weaken the motif's functionality. Also, the formation of a bend within the motif does not seem to correlate with the motif's functionality (Figure 3.7).

In the highly-ordered and well-aligned parallel β-sheets formed by CSGAITIGRGD and antiparallel β-sheets formed RGDSGAITIGC and RGESGAITIGC peptides we observe that the occurrence of an intramolecular salt-bridge between the oppositely charged groups in the RGD/RGE motifs (as the distance between Arg CZ : Asp/Glu CG can be below 6.0 Å), and that the formation of a bend within the RGD motif are possible (Figure 3.7); in line with the aforementioned salt-bridge analysis, the formation of an intramolecular salt-bridge between the oppositely charged groups within the RGD/RGE motif is significantly more favored in the antiparallel β-sheets formed by RGESGAITIGC peptides. While the solvent exposure of the RGD motif in the fibronectin is comparable to the solvent exposure of the RGD/RGE motifs in the highly ordered antiparallel RGDSGAITIGC and RGESGAITIGC β-sheets, it is interestingly significantly higher than the solvent exposure of the RGD motif in the highly ordered parallel CSGAITIGRDD β-sheets (Figure 3.7). As the solvent exposure of the RGD motif is essential for cell adhesion[178], the aforementioned analyses suggest that the fibrils formed by RGDSGAITIGC and RGESGAITIGC peptides are expected to possess the desired bifunctional properties, in contrast to the fibrils formed by CSGAITIGRGD peptides. The lower solvent accessibility in the latter is in correlation with the fact that in the parallel CSGAITIGRGD β-sheets the amyloid zipper involves residues 3-10 and includes partly the RGD sequence motif, in contrast to the antiparallel RGDSGAITIGC and RGESGAITIGC β-sheets, where the amyloid zipper does not overlap with the RGD motif. Thus, the amyloid fibrils formed by RGDSGAITIGC and RGEGAITIGC peptides are expected to simultaneously encompass cell-attachment and metal nanoparticle-attachment properties.

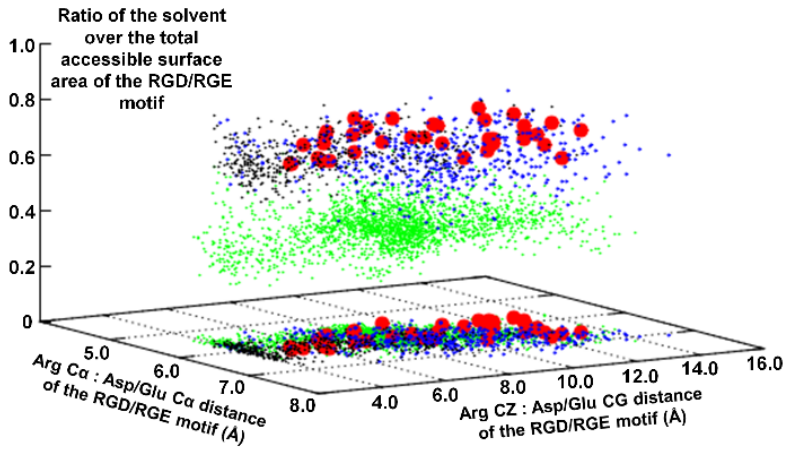


Figure 3.7 Three-dimensional plot of the ratio of the solvent over the total accessible surface area (Z-axis) with respect to the Cα : Cα distances between Arg and Asp/Glu (X-axis) and the distance between the Arg CZ atom and Asp/Glu CG atom (Y-axis) of the RGD/RGE motif in the Fibronectin protein [PDB code:1TTF[189]] (red dots) and the peptides CSGAITIGRGD (green dots), RGDSGAITIGC (blue dots), RGESGAITIGC (black dots). All the data points in the three-dimensional space are projected onto the XY plane as well. Reprinted with the permission from ref.[24]

3.4.7 Titration of exposed thiol groups with DTNB

The following experiments were performed by Drs. Mitraki's lab and by additional experimental collaborators, listed as authors in ref.[24] DTNB (5,5'-dithiobis-2-nitrobenzoic acid), or Ellmann's reagent produces a yellow color upon reaction with free thiols. To evaluate the percentage of free thiols remaining post-assembly, aliquots of the peptide solutions were titrated with DTNB at different time points after dissolving the peptide powder in water. The values at each time point, normalized to the value at time 0 to give the percentage of free thiols remaining, are presented in Figure 3.8. A control sample of peptide dissolved in 8 M Guanidine hydrochloride, where assembly cannot take place was carried out in parallel. After 96 hours in water nearly 80 % of thiols remain titratable by DTNB and therefore exposed; at this time point, assembly and fibril formation is essentially complete, as shown by the microscopy results. Following this time point,

the percentage of free thiols is slowly declining to reach a percentage of about 60% after two weeks. As the profile of free thiols in the water sample and in the guanidine sample are essentially the same, it can be assumed that the thiols are not buried during assembly and they remain exposed from the self-assembling core. The slow decline of free thiols in water following the first 96 hours could be attributed to post-assembly oxidation events.

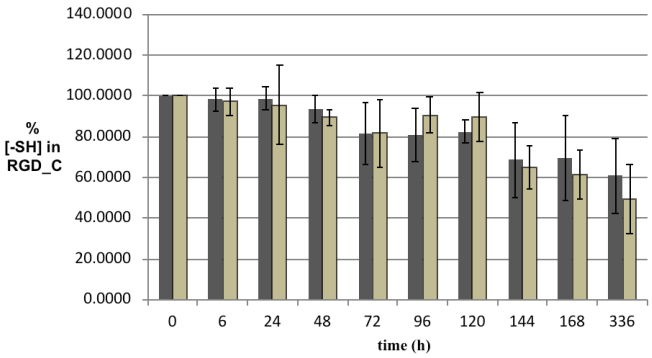


Figure 3.8 Free –SH concentrations monitored by DTNB titration of the peptide RGDSGAITIGC 3 mg.ml in pH 7.4 in water (black bars) and in guanidine hydrochloride 8M (light grey bars). Figure was produced by our experimental collaborators and is reprinted with the permission from ref.[24]

3.4.8 Fibroblast adhesion and proliferation

The following experiments were performed by Drs. Mitraki's lab and by additional experimental collaborators, listed as authors in ref.[24] The adhesion and proliferation of the fibroblast model cell NIH/3T3 was followed in cell culture polystyrene plates covered with tissue culture coverslips where the peptide networks were deposited. Cell viability was followed over time, from day 1 to day 7. Metabolic activity showed significant increase from day 1 to day 7 in

all substrates [F_{time} (2, 276) = 314.6, $P < 0.0001$; 2-way-ANOVA] (Figure 3.9). At days 1 and 4 cells appeared to attach and proliferate similarly in the coverslips and peptide networks ($P > 0.05$), however, at day 7, a significant increase in metabolic activity was observed in the peptide networks compared to coverslips. The statistically most significant difference was observed for the peptide RGDSGAITIGC ($P < 0.0001$), followed by the peptide RGESGAITIGC ($P < 0.01$) and last by the peptide RGDSGAITIGA ($P < 0.05$).

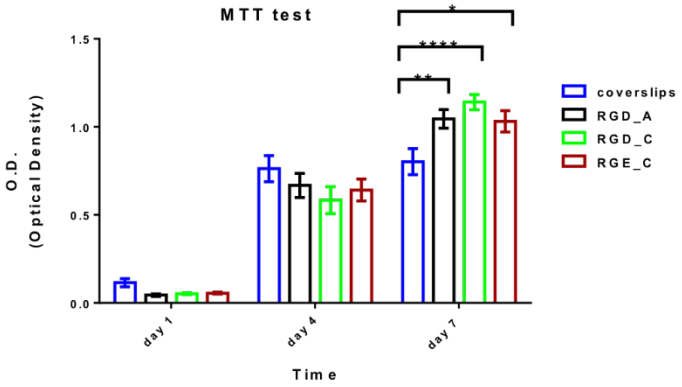


Figure 3.9 Survival and viability of NIH/3T3 murine fibroblasts on the peptide substrates over one-week period. The metabolic activity was measured by the MTT test and statistical analyses were performed by using grouped two-way analysis of variance (Two-way ANOVA) with Graph Pad Prism 6. Figure was produced by our experimental collaborators and reprinted with the permission from ref.[24]

3.5 Concluding Remarks

In this chapter, we rationally designed RGDSGAITIGC, a bifunctional self-assembling amyloid peptide which encompasses cell-attachment and potential metal nanoparticle/ surface attachment properties through the incorporation of an RGD sequence motif and a cysteine residue at the N- and C- terminal end, respectively. The novel peptide sequence preserves the β-sheet core (S)GAITIG of the NSGAITIG self-assembling sequence,[51,131] and simultaneously incorporates the

RGD sequence motif and a cysteine at suitably selected terminal peptide positions. We consider that the current bifunctional properties of the RGDSGAITIGC fibril-forming peptide can be exploited to fabricate novel biomaterials with promising biomedical applications[10]. Apart from the current biomedical applications of the RGDSGAITIGC fibrils, the comparison of solvent accessibility of the RGD motif in the RGDSGAITIGC peptide and the fibronectin protein was utilized by others to validate the solvent accessibility of the RGD motif in the peptide investigated by them[191]. Additionally, the computationally derived structural scaffolds of the highly-ordered and well-aligned antiparallel/parallel β-sheet states formed by the peptide were used as designable scaffolds to computationally design bifunctional self-assembling peptide RGDSGAITIGY[192] which encompasses cell-adhesion and cross-linking properties.

3.6 Innovation in Methods

In this study, we defined a systematic approach to calculate the ratio of solvent/total accessible surface area of amino acids in amyloid fibrils. It was applied to certain backbone and side chain functional motifs in order to determine the relative exposure of functional amino acid motifs in amyloid self-assembling peptide structures using Wordom[141,142]. In this approach, we calculated the ratio of solvent/total accessible surface area for specific backbone and side chain functional motifs. Recently, this approach has been utilized by others for calculating the solvent accessibility of the RGD functional motif for five different peptides.[191]

3.7 New Findings and Potential Applications

In collaboration with Dr. Mitraki's experimental lab and by additional experimental collaborators, listed as authors in ref[24], we successfully rationally designed the self-assembling peptide RGDSGAITIGC with cell-adhesion and metal binding properties which can be exploited to fabricate novel biomaterials with promising biomedical applications[10]. We also investigated the

role of key geometrical features of the RGD motif that determine its functionality. We observed that high solvent accessibility of the RGD motif is the key factor correlating to its functionality.

We also observed that the RGE sequence motif in the peptide RGESGAITIGC is less functional compared to the RGD sequence motif in the peptide RGDSGAITIGC with regard to cell-adhesion properties presumably owing to the fact that in the RGE sequence motif, the arginine may be partially "locked" in an intramolecular salt-bridge with its neighboring glutamate residue, and thus is less susceptible to interact with the cell. Since the Arg : Glu salt-bridge in the RGE sequence motif is facilitated to be formed intramolecularly between arginine and its i+2 nearest covalently bonded neighbor, in contrast to the Arg : Asp salt-bridge in the RGD sequence motif, it is possible that this observation can shed light into why in general the RGE sequence motif could be considered a negative control with regard to cell-adhesion properties compared to the RGD sequence motif.

4. COMPUTATIONAL DESIGN OF AMYLOID SELF-ASSEMBLING PEPTIDES BEARING AROMATIC RESIDUES AND THE CELL ADHESIVE MOTIF ARG-GLY-ASP*

4.1 Introduction

4.1.1 Overview and Significance

In chapter 3, we used rational computational and experimental design approaches to design a self-assembling peptide with sequence RGDSGAITIGC which can potentially be used as a novel sandwich self-assembly construct with multi-functional cell targeting and functionalization properties, toward geometrically and spatially directing the fate of a cell population. The peptide comprised of an RGD sequence motif which allows cell-attachment properties and a free-thiol group through the C-terminal cysteine, which could be used for metal-binding properties.

RGD, which was also presented in chapter 3, is the principal integrin-binding domain present within extracellular matrix (ECM) proteins such as fibronectin, vitronectin, fibrinogen, osteopontin, and bone sialoprotein; due to the capacity of the RGD sequence to bind to multiple integrin species, synthetic RGD peptides offer several advantages for biomaterial applications.[193] RGD bearing self-assembling peptide amphiphiles not only have cellular adhesion properties but also serve as ideal scaffolds to induce bone tissue growth and biomineralization, to induce osteogenic differentiation of rat mesenchymal stem cells (rMSCs), and can be used for the functionalization of bone implants.[194,195,196]

Apart from the importance of incorporating an RGD sequence motif in biomaterials, in some natural biomaterials, dityrosine residues are found, which are formed through covalent

* Reprinted with the permission from "Computational design of amyloid self-assembling peptides bearing aromatic residues and the cell adhesive motif Arg-Gly-Asp" by Jonnalagadda et al., 2017. *RSC Molecular Systems Design & Engineering*, 2, 321-335, Copyright 2017 by Royal Society of Chemistry.

crosslinking of tyrosines. For example, tyrosine–tyrosine crosslink networks found in resilin provide outstanding mechanical properties such as elasticity,[197] which were exploited for the formation of peptide hollow nanocapsules and peptide lamella films,[197] as well as to engineer small tyrosine-containing peptide hydrogels with enhanced mechanical properties.[198]

Specifically, dityrosine cross-linking of short peptide sequences was previously exploited for the fabrication of nanostructures[197,199,200] and to significantly strengthen peptide-based biomaterials.[198] For example, self-assembled, hollow, nanocapsules were synthesized using a one-step photopolymerization of a tyrosine-rich peptide exhibiting an elastic modulus of approximately 30 GPa and a hardness of 740 MPa.[197] Dityrosine-based gold nanoparticles have also been fabricated by cross-linking of Lys-co-Tyr polypeptides using either UV irradiation or chloroaurate reduction.[200] Moreover, design of a peptide incorporating tyrosine residues allowed for both self-assembly of the peptides and subsequent dityrosine cross-linking; in this case, incorporation of the covalent cross-links increased the modulus by 104-fold, significantly increasing the properties and potential applications of the material.[198] Thus, the incorporation of dityrosine cross-linking in short self-assembling peptides can provide advance biomaterials with significant advantageous properties for tissue engineering applications.[199]

4.1.2 Goal

In this work, we aimed to introduce a computational strategy to energetically stabilize and potentially functionalize an amyloid peptide fibril. The flexible structures used as designable scaffolds in the model comprised of two independent sets of computationally derived, parallel and antiparallel highly ordered and well-aligned β-sheet states formed by peptide RGDSGAITIGC, investigated in chapter 3. A computational design model was developed and introduced at the beginning of the strategy to minimize the energy of the designable scaffolds upon a single amino

acid mutation at position 11, so as to preserve the peptide's amyloidogenic and cell-adhesive properties. This enabled us to investigate the mutational effect on realistically modeled β-sheet arrangements of peptides. According to the results produced, the introduction of aromatic (Tyr, Trp, Phe) or imidazole ring (His) groups at residue position 11 were among the most energetically favorable mutations in parallel β-sheet bonded peptides.

4.1.3 Hypothesis

Based on the aforementioned results denoting that the introduction of aromatic (Tyr, Trp, Phe) or imidazole ring (His) groups at residue 11 were among the most energetically favorable mutations in parallel β-sheet bonded peptides, we hypothesized that the one of the designed peptides, with sequence RGDSGAITIGY, would be capable of possessing a crosslinking capacity through the tyrosine residues belonging to adjacent β-sheet pairs of peptide RGDSGAITIGY. Previous studies provided above show the significance of such dityrosine linking in biomaterials, thus, we aimed to provide additional insights into the potential cross-linking properties of the dityrosines in the designed peptide with sequence RGDSGAITIGY.

4.1.4 Objectives Accomplished

Following the results of the computational design model, an additional constraint was introduced to the computational design model to investigate the crosslinking capacity of tyrosine residues belonging to adjacent β-sheet pairs of peptide RGDSGAITIGY. Furthermore, we used replica exchange molecular dynamics (REMD) simulations and free energy calculations depicting that the designed peptides self-assemble into parallel β-sheets, and suggesting that the highly ordered and well-aligned amyloid fibrils formed by RGDSGAITIGY can combine potential cell adhesive and crosslinking properties. Experiments by our experimental collaborators confirmed

that designer peptides self-assemble into amyloid-type fibrils, and ongoing studies aim to exploit the predicted bifunctional properties of RGDSGAITIGY in biomaterial and tissue engineering applications.

This paper resulted in a publication of Dr. Tamamis' computational lab in collaboration with Dr. Mitraki's experimental lab, and other experimentalists which their names are listed below in the complete publication reference: "Jonnalagadda, S.V.R.; Ornithopoulou, E.; Orr, A.A.; Mossou, E.; Forsyth, V.T.; Mitchell, E.P.; Bowler, M.W.; Mitraki, A.; Tamamis, P. Computational Design of Amyloid Self-Assembling Peptides Bearing Aromatic Residues and the Cell Adhesive Motif Arg-Gly-Asp. *Mol. Syst. Des. Eng.* **2017***, 2*, 321-335."

S.V.R.J. and E.O. are equally contributing first authors.

4.2 Computational Methods

4.2.1 Computational Design – Modification of Position 11

We developed and introduced a computational design model to stabilize the self-assembling β-sheets formed by the peptide RGDSGAITIGC[24] by minimizing the energy of a short modeled amyloid fibril upon single or multiple amino acid substitutions. The computational design model was implemented using our in-house Fortran programs. The energy term under minimization in the objective function accounts for protein-protein interactions, analogously to references[201,202,203,204], as well as protein-water interactions, which are expected to play an important role in stabilizing the structure of a self-assembling peptide. The decomposition of energy into residue pairwise terms and the minimization of the total energy of a protein has been implemented as a strategy in several protein design[202,203,204] and protein folding studies[201].

$$\min\left(\frac{1}{f}\left(\sum_{a=1}^{p}\sum_{b=1}^{p}\sum_{i=1}^{n}\sum_{j=1}^{m_i}\sum_{\substack{k=1\\k\geq i}}^{n}\sum_{l=1}^{m_k}\sum_{t}^{f}E_{i,k,a,b}^{j,l}y_i^j y_k^l w(t)_{ik}^{jl}\right)+\frac{1}{f}\left(\sum_{a=1}^{p}\sum_{i=1}^{n}\sum_{j=1}^{m_i}\sum_{t}^{f}\gamma(\mathrm{SASA}(t)_{i,a}^{j})y_i^j\right)\right) \qquad 4.1$$

$$\text{Subject to} \qquad \sum_{j=1}^{m_i} y_i^j = 1 \forall i$$

$$y_i^j, y_k^l, w(t)_{ik}^{jl} = 0-1 \forall i,j,k,l$$

$$k \neq i \forall a = b$$

$$b > a \forall k = i$$

$$i \in \{\alpha_1, \alpha_2, ..., \alpha_v\}$$

$$l = j \forall (k \in \{\alpha_1, \alpha_2, ..., \alpha_v\}; k = i; a \neq b)$$

$$\left(\sum_{a=1}^{p}\sum_{b=1}^{p}\sum_{i=1}^{n}\sum_{j=1}^{m_i}\sum_{k=1}^{n}\sum_{l=1}^{m_k}\sum_{t}^{f}\left(7.0\ \text{Å} \leq D(t)_{i,a,k,b}^{j,l} \leq 10.65\ \text{Å}\right)\right)\ \forall\ (i=11, k=11, b=a+1) \qquad 4.2$$

The aforementioned model was inspired by late Dr. Floudas' studies[201], and simultaneously minimizes the pairwise interaction energy, $E_{i,a,k,b}^{j,l}$, accounting for protein-protein interactions using a coarse-grained surface residue pairwise force field, SIPPER Surface[205], and solvent accessible surface area, $\mathrm{SASA}(t)_{i,a}^{j}$, multiplied by the surface tension coefficient, γ, accounting for protein-water interactions. The specific force field is based on residue composition of protein surfaces and was trained to consider residue pairs with separated partners having relative Accessible Surface Area (ASA) ≥ 5% (with respect to that of the given residue type when it is exposed). We considered that the specific force field is more proper to estimate energetic effects of a mutated residue at the fibril's surface compared to other coarse-grained residue pairwise force fields trained to account for residue pairwise interactions in globular proteins in general (e.g., references[206,207,208]). The pairwise interaction energy, $E_{i,a,k,b}^{j,l}$, corresponds to the interaction energy between residue *j* at mutable position *i* in strand *a* and residue *l* at position *k* in strand *b*. The

positions *i* and *k* can belong to the same peptide strand (intra-molecular interactions) or different peptide strands (inter-molecular interactions). In this study, the total number of residue positions, *n*, is 11 per strand for the peptide RGDSGAITGC[24] and the total number of peptide strands, *p*, is 4. $\{\alpha_1,\alpha_2,...,\alpha_v\}$ is a set of residue positions that are amenable for modification in the peptide under the condition that $\alpha_1 < \alpha_2 < ... < \alpha_v$. In the present study, position 11 is mutable, i.e. $\{\alpha_1=11\}$. The average of the pairwise interaction energy was taken over the total number of designable scaffolds, *f*, for the peptide RGDSGAITGC[24]. In the present study, mutations at position 11 were introduced to two sets of 4-stranded highly ordered and well-aligned β-sheet fibrils, referred to as designable scaffolds. Of the two sets, one set comprised of highly ordered and well-aligned 4-stranded parallel β-sheet fibrils and the other set comprised of highly ordered and well-aligned 4-stranded antiparallel β-sheet fibrils. Both antiparallel and parallel designable scaffolds were used to investigate in which fashion, parallel or antiparallel, a specific mutation would be most favorable to be introduced. The number of designable scaffolds used, *f*, was equal to 50 for both parallel and antiparallel configurations, independently.

The computational design model was solved in Fortran using exhaustive enumeration independently for both antiparallel and parallel configurations of designable scaffolds; thus, there are two change in energy values from the objective function (*Eq.4.1*) per amino acid substitution. The binary variable y_i^j equals one if position *i* is occupied by amino acid *j*, and zero otherwise. The binary variable y_k^l equals one if position *k* is occupied by amino acid *l*, and zero otherwise. The binary variable $w(t)_{ik}^{jl}$ equals one if the center of mass of the residue side chain at position *k* is within a specific distance of the center of mass of the residue side chain at position *i.* The specific distance in our case is taken to be 6.5 Å, as this distance proved optimal to account interactions of

the mutated residues with its neighboring counterpart residues. The constraint, $l = j \forall (k \in \{\alpha_1, \alpha_2, ..., \alpha_v\}; k = i; a \neq b)$, defined under *Eq.4.1* is introduced in order to ensure that mutations occurring at a position in a peptide strand simultaneously occur at that position in each of the peptide strands of the self-assembled structures.

The second term of *Eq.4.1*, approximately estimates the non-polar solvation free energy change upon a mutation in the elementary structural unit of the designed fibrils. The non-polar solvation free energy is required for (1) the formation of a cavity in the solvent to accommodate the solute and (2) establishing solute-solvent dispersion interactions. The formation of the cavity requires entropic and solvent-reorganization energy, and due to the strong self-attraction of water stemming from its ability to form hydrogen bond networks with itself, energy is also lost due to the introduction of hydrophobic molecules, which cannot form hydrogen bonds and disrupt the hydrogen bond network[209,210,211]. Thus, the second term may also be considered as a penalty for the introduction of larger hydrophobic residues into the designable scaffold, which would result in the formation of a larger cavity and the disruption of more solvent-solvent interactions. To account for solute-solvent interactions, we additionally used a coarse-grained approach to approximate the contribution of each residue to the solvation free energy by assuming a linear dependence between the solvation free energy and the solvent accessible surface area (SASA)[212,213]. Under this assumption, the cavity formed to accommodate a solute is proportional to the SASA of the introduced solute, and the solute-solvent dispersion interaction energy also correlates with SASA as solute atoms at the solute-solvent interface interact more strongly with the solvent compared to buried solute atoms; in the present study. The $\mathrm{SASA}(t)_{i,a}^{j}$ term is the estimated solvent accessible surface area of the introduced amino acid *j* at position *i* in strand *a*. The surface tension coefficient,

γ used in *Eq.4.1* for this study is equal to 0.002 kcal/(mol·Å^2). The $\mathrm{SASA}(t)_{i,a}^{j}$ term is approximated through the following calculation:

$$\mathrm{SASA}(t)_{i,a}^{j} = \left(\text{theoretical SASA of the introduced residue}\right) - \left[\left(\text{theoretical SASA of the native residue}\right) - \left(\text{SASA of the native residue in scaffold(t)}\right)\right] \quad 4.3$$

where the second term, [(theoretical SASA of the native residue) – (SASA of the native residue in the scaffold(t))], represents the solvent-excluded surface, or how "buried" the residue is. The theoretical SASA values are ASA values of the whole residue (X) in the tripeptide, Ala-X-Ala, taken from Table 2 of reference[188] and the structural SASA values are calculated using the GEPOL[214] algorithm in Wordom[141,142] for the whole residue in the designable scaffold.

According to the results, apart from a cysteine placement at position 11 (investigated by us in chapter 3), aromatic residues (tyrosine, tryptophan and phenylalanine), methionine, histidine and arginine can additionally be considered energetically favorable substitutions at position 11 according to the energy term under minimization in *Eq.4.1*. Motivated by our findings in conjunction with previously published studies suggesting the role of aromatic residues with π-stacking properties in peptide self-assembly[184,215,216,217], we focused our investigation on residue substitutions containing an aromatic (tyrosine, tryptophan and phenylalanine) or an imidazole ring (histidine). According to the coarse-grained computational design, the specific residue substitutions were among the most energetically favorable mutations in parallel β-sheet bonded peptides. Thus, a tyrosine mutation at position 11 may lead to potential functionalization of the self-assembled peptides via tyrosine crosslinking between adjacent β-sheet peptides bonded in parallel upon fibril formation.

4.2.2 Constraint Introduced to Investigate the Potential Crosslinking Capacity between Tyrosine Residues upon Fibril Formation

The aforementioned results suggest that the placement of a tyrosine at position 11 stabilizes the self-assembling β-sheet peptide in a parallel configuration of peptides. It is possible that owing to the proximity of aromatic/imidazole-ring containing residues, π-π interactions formed between residues at position 11 belonging to neighboring peptide strands might play a key role in the fibrils' stabilization. Inspired by the fact that tyrosine-tyrosine crosslink networks found in resilin have outstanding mechanical properties such as elasticity[197], we introduced an additional constraint (*Eq.4.2*) during the solution of the computational design problem, to investigate the possibility of tyrosine residues in adjacent (neighboring) β-sheet forming peptides to be crosslinkable upon fibril formation, in the designed peptide RGDSGAITIGY. Through the constraint defined in *Eq.4.2* we investigated the probability of a Cα : Cα distance $D(t)_{i,a,k,b}^{j,l}$, between residues at position 11 belonging to parallel β-sheet forming adjacent peptides in the designable scaffolds to be less than 10.65 Å, which is the maximum Cα : Cα distance for covalently bonded dityrosine molecules deposited in the ZINC database[218]. Specifically, for the dityrosine compounds of the ZINC database[218] the Cα : Cα distances between the two covalently bonded tyrosines range from 7 Å to 10.65 Å.

4.2.3 Infinite dilution simulations

We first simulated an isolated peptide of each of the four designed peptides $NH3^{+}$-RGDSGAITIG(Y/W/H/F)-$CONH_2$ in aqueous solution. The aqueous solvent effects were modeled implicitly using the FACTS19[123] solvation model in conjunction with the CHARMM19 all-atom force field[111]. The value of surface tension coefficient in the solvation model is taken to

be 0.015 kcal/(mol*Å^2), which is in line with the CHARMM19 force-field for reversible folding simulations of the structured peptides. We performed replica exchange molecular dynamics (REMD) simulations,[133,134,135,136,137,138] employing a total of eight replicas with temperatures 283, 300, 318, 336, 356, 377, 403, and 432 K, and using Langevin dynamics with a friction coefficient of 5.0 ps^{-1} on heavy atoms and 0 ps^{-1} on hydrogen atoms. The total simulation time for all temperatures per designed peptide system was 1.6 μs. We analyzed 20,000 snapshots of the 300 K trajectory, extracted at 100 ps intervals. To extract the most representative conformations, we performed a root mean square deviation (RMSD)-based clustering analysis on the extracted snapshots. The clustering analysis was conducted via WORDOM[141,142] employing all backbone atoms and a clustering radius of 2 Å. The six most representative conformations were extracted from the six most populated clusters for each of the four designed peptides and were used as initial structures for simulations in finite dilution described below.

4.2.4 Self-assembly Simulations

We investigated the self-assembly properties of the four computationally designed peptides, RGDSGAITIGY, RGDSGAITIGW, RGDSGAITIGF, and RGDSGATIGH using REMD simulations analogous to the computational protocol developed by Tamamis and Archontis[130] as described in chapter 2. Using the replica-exchange method,[133,134,135,136,137,138] we performed MD simulations of the designed peptide systems. We placed six copies of each peptide sequence, with a positively charged N-terminal and an amidated C-terminal end, with different initial conformations in a cubic periodic 148 Å box, resulting in an approximate 3 mg ml^{-1} peptide concentration. The simulations were performed using CHARMM,[110] analogously to refs.[24,40,130,131,132]. Initially the six copies of each peptide sequence were placed at the center. After the initial placement, peptides 1, 3, and 5 were translated by +25 Å in the x, y, and z directions,

respectively, and peptides 2, 4, and 6 were translated by −25 Å in the x, y, and z directions, respectively. We modeled water implicitly using the FACTS19[123] solvation model, as in refs.[24,40,130,131,132], and the surface tension coefficient was set to 0.015 kcal/(mol*Å^2). For the purpose of Langevin dynamics, a 5.0 ps^{-1} friction coefficient was introduced on all non-hydrogen atoms of each peptide. We employed a total of sixteen temperatures (290, 295, 300, 310, 305, 315, 321, 327, 333, 339, 345, 352, 359, 366, 373 and 380 K) with each replica exchange run's duration being equal to 10 ps. The total simulation time for all temperatures per system was equal to 16 μs. We collected the final conformations of each replica exchange run at 300 K for each peptide sequence. These conformations were combined into one trajectory per peptide sequence, corresponding to a length of 1000 ns and 100,000 snapshots per self-assembling peptide.

4.2.5 Categorization of β-Sheet Conformations and Key β-Sheet Interactions

We used DSSP[185] to determine the secondary structure of the designed peptides within the simulations, and identified the formation of intermolecular β-sheet structures. We categorized the β-sheets into antiparallel, parallel and mixed conformations comprising 2 to 5 peptide strands and the categorization was performed following the computation protocol by Tamamis and Archontis.[130] The (%) moving averages of antiparallel, parallel and mixed β-sheet conformations, composed of 2-, 3-, 4-, and 5-stranded peptides depict that the designed peptides in the simulations have the tendency to self-assemble into β-sheet rich conformations. For all four computationally designed peptides, an overall tendency for the formation of parallel, rather than antiparallel β-sheets is observed. 4- and 5-stranded antiparallel and parallel β-sheets have higher complexity than 2- and 3-stranded β-sheets; hence, the structural patterns observed in the former can potentially correspond to patterns that may exist in naturally occurring fibrils formed by the peptides. Since parallel configuration is primarily the dominant configuration for the designed peptides, we

extracted the 4- and 5-stranded parallel β-sheets from the corresponding MD simulation snapshots for each designed peptide sequence. We calculated the percent probability of an intermolecular pair of residues, belonging to two adjacent β-sheet peptides in the 4- and 5-stranded parallel β-sheet states of the four designed peptides, to be involved in a β-sheet (or β-bridge) conformation. The propensities of β-intermolecular interactions between adjacent residue pairs are presented in two-dimensional maps (see *Results and Discussion*). Through this analysis we identified the predominant patterns of pairwise β-sheet interactions between intermolecular residues, and the key amyloidogenic regions of each peptide. In addition, this analysis can be used to indicate the residues that are unlikely to be involved in the formation of intermolecular β-sheets in each peptide sequence, and thus, can be mutated for functionalization purposes.

4.2.5 Identification of Well-Aligned and Well-Ordered β-Sheet Conformations using P_1 and P_2 Parameters

Using the polar order-parameter, P1, and the nematic order-parameter, P2, defined in *Eq.4.4*, we examined the extent of peptide alignment and order for the extracted 4- and 5-stranded parallel β-sheet conformations. These parameters are used in the structural characterization of liquid crystals, and have been employed successfully in simulation studies of peptide aggregation[40,130,132,131,219,220].

$$P_1 = \frac{1}{N}\sum_{i=1}^{N}\vec{z_i}\vec{d}, \quad P_2 = \frac{1}{N}\sum_{i=1}^{N}\frac{3}{2}\left(\vec{z_i}\vec{d}\right)^2 - \frac{1}{2}, \qquad 4.4$$

In *Eq. 4.4*, *N* is the number of molecules in the simulation and $\vec{z}_i$ is a unit vector along a suitably defined molecular direction; $\vec{d}$ is a unit vector along a preferred direction of alignment, which emerges from the properties of the system. For all four peptides, we selected $\vec{z}_i$ to be defined by the segment spanning from the Cα atom of residue 5 to the Cα atom of 10, as these residues primarily comprise the key β-sheet regions of each designed peptide.

Using WORDOM,[141,142] we computed P_1 and P_2 parameters for the 4- and 5-stranded parallel β-sheets of the designed peptides. The results revealed that among the 4- and 5-stranded parallel β-sheets across all systems, β-sheets that are both highly populated and highly ordered occur only in 4-stranded parallel β-sheet conformations. The lower population and lower degree of order observed in 5-stranded parallel β-sheets can partly be attributed to the limited number of peptides used in the self-assembly simulations. Thus, we focused our further analysis on 4-stranded parallel β-sheets and examined the structural properties of highly ordered and well-aligned conformations formed by the designed peptides. We constructed free energy landscapes for the specific 4-stranded parallel β-sheets, using the two dimensional probability *P(P1, P2)* and *Eq.4.5*.

$$G(P_1, P_2) = -k_B T \ln\left[P(P_1, P_2)\right] \qquad 4.5$$

In each case we identified the free energy basin that encompasses the majority of conformations and the location of the global minimum. We extracted the states with the highest polar order and highest nematic order, corresponding to P_1, P_2 regions [P_1 = 0.8:1.0, P_2 = 0.75:1.0] for parallel β-sheet conformations of the designed peptides. The selected highly ordered and well-

aligned 4-stranded parallel conformations were utilized for subsequent analysis focusing on the peptides' structural properties. Representative conformations of the selected highly ordered and well-aligned parallel β-sheet states are shown in the *Results and Discussion*.

4.2.6 Amyloid zipper formation, salt bridge formation and exposure of functional motifs

We investigated the propensity of intramolecular and intermolecular contacts formed between the residue side chains of peptides within the 4-stranded highly ordered and well-aligned parallel β-sheet states. We defined a distance threshold of 6.5 Å between geometric centers of two side chains as a criterion for a contact. Furthermore, we calculated the average propensity for Arg1 and Asp3 to be involved in the (i) intermolecular salt bridge, (ii) intramolecular salt bridge, and (iii) intra or intermolecular salt bridge for the selected 4-stranded highly ordered and well-aligned β-sheet states in each of the four designed peptides. A distance cutoff of 3.7 Å between the corresponding nitrogen: oxygen atoms of the oppositely charged groups was used as a criterion to denote a salt bridge formation between Arg1 and Asp3.

In addition, we investigated the degree of solvent accessibility of the RGD sequence motifs and residues at position 11 (Y/W/H/F) within the 4-stranded highly ordered and well-aligned parallel β-sheet states of the designed peptides. We calculated the solvent accessible surface areas for the backbone and side chain moieties belonging to the RGD sequence motif and the residue at position 11 (Y/W/H/F) of the two central peptides in each 4-stranded highly ordered and well-aligned parallel β-sheet state. A 1.4 Å radius sphere (approximating the radius of a water molecule) was introduced to “probe” the surface of the molecule for the solvent accessible and total accessible surface area calculations.[221] The total accessible surface area of a specific moiety in a peptide strand was approximated as the maximum solvent accessible surface area (SASA) and was calculated by assuming that the specific moiety in that strand is not part of the peptide system.

While this approximation is expected to overestimate the total solvent accessible surface of a specific moiety, we used this definition as we considered it an appropriate way to perform a fair comparison of the investigated moieties across the four designed peptides, as well as in the context of the entire fibronectin (see *Results and Discussion*).[24] Subsequently, we summed the total accessible surface area of a specific moiety in the two central peptides, and calculated the ratio of solvent accessible surface area/total accessible surface area for these specific moieties. The solvent accessible surface area and total accessible surface area values for the two outer peptides are not reported as the degree of solvent accessibility of the two outer peptides is artificially high due to the absence of interacting peptides on the ends.[24] Nevertheless, the two outer peptides were appropriately considered for the purpose of calculating the solvent accessibility areas of specific regions belonging to the central peptides.[24]

RGD sequence motif is known for its capacity to promote cell adhesive properties in extracellular matrix (ECM) proteins such as fibronectin, vitronectin, and fibrinogen.[178] The RGD sequences in the ECM proteins are solvent exposed, thus aiding the ECM proteins in fast recognition and binding to the cell adhesive receptors.[189] To uncover the potential cell adhesive properties of the RGD motif in our designed peptides, we compared their solvent accessibility properties to the NMR-resolved functional cell adhesive RGD motif of the fibronectin (PDB code: 1TTF[189]).

4.2.7 Potential of Crosslinking in the Computationally Designed Self-Assembling Peptide RGDSGAITIGY

We calculated the distance (D) and angle (A) between the tyrosine's aromatic rings of the adjacent interacting peptides in the highly ordered and well-aligned 4-stranded parallel β-sheet states of the computationally designed self-assembling peptide RGDSGAITIGY. To investigate

the peptides' crosslinking potential upon fibril formation, we extracted the distance and angle between the tyrosine's aromatic rings in the covalently bonded dityrosine compounds from the ZINC database,[218] and compared the values of these parameters with the ones calculated for the designed peptide RGDSGAITIGY.

We projected the 4-stranded highly ordered and well-aligned parallel β-sheet states of the computationally designed peptide RGDSGAITIGY on the free energy landscape using the aforementioned distances (D) and angles (A) as reaction coordinates. We divided the (D, A) subspace into grids of size (0.5 Å, 2.5°) and then we calculated the two-dimensional probability *P(D, A)*. Subsequently, the free energy landscape was constructed through *Eq.4.6*.

$$G(D,A) = -k_B T \ln\left[P(D,A)\right] \qquad 4.6$$

We identified two free energy minima regions and compared the distances (D) and angles (A) between the tyrosine's aromatic rings of the adjacent interacting peptides in highly ordered and well-aligned 4-stranded parallel β-sheets of the peptide RGDSGAITIGY with the distances (D) and angles (A) extracted from the covalently bonded dityrosine compounds in ZINC database.[218]

4.3 Experimental Materials and Methods

All experiments described in 4.3 were performed by Drs. Mitraki's lab and by additional experimental collaborators, listed as authors in ref.[192]

4.3.1 Materials

The peptides were purchased from Genecust (Luxemburg) and possessed a degree of purity higher than 95%. These synthetic lyophilized peptide powders were dissolved in sterile water pH 7 or in 10 mM phosphate buffer pH 7.4 at concentrations of 3 mg/ml.

4.3.2 Transmission Electron microscopy

Sample preparation included the deposition of an 8 microlitre droplet of the sample onto a formvar covered 300 mesh copper grid (used for TEM). After 2 minutes, the excess fluid was soaked by a filter paper, and an 8 microlitre droplet of a solution of 2% uranyl acetate was deposited onto the mesh. After that, the excess fluid was removed in the same way, and the grid placed in the grid holder. Experiments were performed on a JEOL JEM-2100 Microscope, at the Electron Microscopy Lab "Vassilis Galanopoulos", at the Department of Biology, University of Crete. For observation, a voltage of 100–120 kV and Ultra High Vacuum were used.

4.4 Results and Discussion

4.4.1 Computational Design – Modification of Position 11

We started from our computationally derived highly ordered and well-aligned antiparallel as well as parallel β-sheets formed by peptide RGDSGAITIGC.[24] We initially aimed to further stabilize the fibrils by not interfering with either the residues in the sequence motif RGD that encompass cell adhesive properties or with the residues of the amyloidogenic core (S)GAITIG of the peptide RGDSGAITIGC. Thus, we selected position 11 as a mutable position, and identified optimum mutations which could potentially lead to amyloid fibrils with improved properties (e.g., stability, rigidity, functionality). Apart from the placement of cysteine, additional energetically favored mutations were identified which primarily involved the presence of aromatic or imidazole

ring side chain groups at residue position 11. As mentioned in the *Computational Methods*, motivated by our findings in conjunction with previously published studies suggesting the role of aromatic residues with π-stacking properties in peptide self-assembly,[184,215,216,217] we focused our investigation on residue substitutions containing an aromatic (tyrosine, tryptophan and phenylalanine) or an imidazole ring (histidine). The introduction of aromatic (Tyr, Trp, Phe) or imidazole ring (His) groups at residue 11 were among the most energetically favorable mutations in a parallel arrangement of peptides, while the introduction of cysteine seems to favor an antiparallel arrangement, in line with our previous findings.[24] It is possible that the interactions between the non-covalently bonded cysteine residues in the parallel β-sheet arrangement of RGDSGAITIGC may not suffice to sufficiently counterbalance energetically the unfavorable presence of two-positively charged groups at the first N-terminal position (charged N-terminal group and charged arginine residue).[24]

4.4.2 Constraint Introduced to Investigate the Potential Crosslinking Capacity between Tyrosine Residues upon Fibril Formation

Upon the solution of the model, we introduced an additional constraint during the solution of the computational design problem to investigate the feasibility of crosslinking between adjacent tyrosine residues upon fibril formation, in the parallel β-sheet designable scaffolds of peptide RGDSGAITIGY. According to the ZINC database[218], the Cα : Cα distance between covalently bonded tyrosines in the dityrosine compounds is less than 10.65 Å, and thus, the specific distance was considered in the constraint. In 42 out of 50 parallel β-sheet designable scaffolds we used in the computational design, the Cα : Cα distance between residues at position 11 of adjacent β-sheet forming peptides is less than 10.65 Å. According to this preliminary constraint, tyrosine residues in the RGDGAITIGY fibrils can potentially be crosslinkable upon fibril formation.

4.4.3 Computational Categorization of β-sheet Conformations and Key β-sheet Interactions

Following the computational design results, we investigated the self-assembly properties of the four designed peptides using REMD simulations. Within the simulations, all peptides were frequently arranged into β-sheet conformations (Figure 4.1), which include 2-, 3-, 4-, and 5-stranded β-sheet states. 2-stranded β-sheets include parallel/antiparallel β-sheets (Figure 4.1A-4.1D); 3-, 4-, and 5-stranded include parallel/antiparallel and mixed β-sheets (Figure 4.1E-4.1P). Mixed conformations are composed by at least one pair of parallel and one pair of antiparallel β-sheets and, due to their high entropic favorability, have large populations. Non-symmetric β-sheets (containing strands in different orientations, such as mixed arrangements rather than entirely antiparallel or parallel) are entropically favored because of the larger number of possible configurations that the peptides can be arranged into.[130] These conformations are not investigated in this study as they are not expected to be formed naturally due to their asymmetry. Configurations in which one peptide is forming β-sheet interactions with more than two peptides at a specific instance were considered "complex" and were not investigated in detail either as they are not expected to be present in the naturally occurring amyloid fibrils. The analysis of the REMD simulations indicates that the parallel configuration is overall more favorable than the antiparallel one for all the computationally designed peptides (Figure 4.1). While all peptides have a predominantly higher tendency to form parallel β-sheets, it is worth noting that in specific cases, e.g., mainly involving 2-stranded or 5-stranded β-sheets, the population of antiparallel β-sheets can be larger. Yet, this can be attributed to the fact that 2-stranded β-sheets are of low complexity as only two peptides are involved in the β-sheet, while the population of 5-stranded β-sheets is low and the statistics is poorer. The overall higher tendency of parallel β-sheet formation suggests that the favorable intermolecular π-stacking interactions which can be formed by histidine, tyrosine,

tryptophan or phenylalanine residues in the parallel-fashion (see below), can counterbalance the unfavorable presence of two-positively charged groups at the first N-terminal position owing to a positively charged N-terminal and the positively charged group of arginine.

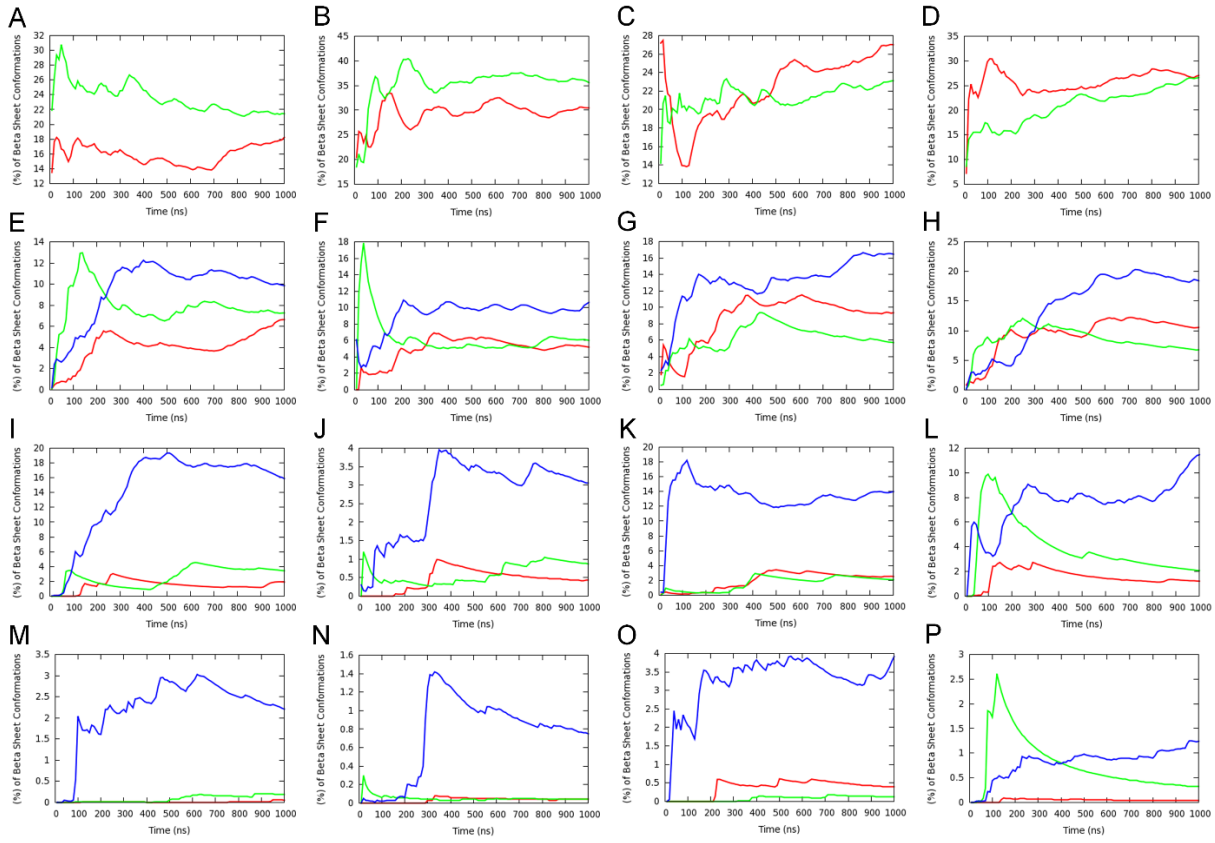


Figure 4.1 Moving averages of the fraction (%) of conformations with intermolecular β-sheets (Y-axis) in the REMD simulations at 300 K with respect to time (X-axis) for peptides RGDSGAITIGY/W/H/F. Figures A, E, I, and M correspond to 2-, 3-, 4-, and 5-stranded β-sheets of peptide RGDSGAITIGY, respectively. Figures B, F, J, and N correspond to 2-, 3-, 4-, and 5-stranded β-sheets of peptide RGDSGAITIGW, respectively. Figures C, G, K, and O correspond to 2-, 3-, 4-, and 5-stranded β-sheets of peptide RGDSGAITIGH, respectively. Figures D, H, L, and P correspond to 2-, 3-, 4-, and 5-stranded β-sheets of peptide RGDSGAITIGF, respectively. Red color corresponds to antiparallel β-sheets, green color corresponds to parallel β-sheets, and blue color corresponds to mixed β-sheets. Reprinted with the permission from ref.[192]

From the four computationally investigated peptides, we focused our analysis on 4-, 5-stranded parallel β-sheet conformations as they may represent to a higher extent the naturally occurring fibrils compared to lower complexity states, e.g., 2- and 3-stranded β-sheets. The propensities of two residues which belong to adjacent β-sheet interacting peptides to form intermolecular "β-bridge" interactions were calculated from the simulations at 300 K, and the results are presented in Figure 4.2. In all four designed peptides, the residues involved in the β-sheet interactions are primarily the residues in the GAITIG motif (see Figure 4.2A-4.2H). Intermolecular parallel β-sheets formed by the peptides RGDSGAITIGY and RGDSGAITIGH are primarily in-register 5–10:5–10 patterns. Both peptides in some cases also form off-register (e.g., 5–9:6–10) patterns. (Figure 4.2A and 4.2E for the peptide RGDSGAITIGY and Figure 4.2C and 4.2G for the peptide RGDSGAITIGH). Intermolecular parallel β-sheets formed by the peptides RGDSGAITIGW and RGDSGAITIGF are primarily in-register 6–10:6–10 patterns. (Figure 4.2B and 4.2F for the peptide RGDSGAITIGW and Figure 4.2D and 4.2H for the peptide RGDSGAITIGF). In all four of the designed peptides, both the RGD motif and the residue at position 11 (Tyr11/Trp11/His11/Phe11) are not involved in the β-sheet interactions, indicating that the RGD and Y/W/F/H motifs can be exposed at the surface of the fibril and possess functional properties, including cell adhesion and crosslinking as is the case for RGD and tyrosine, respectively.

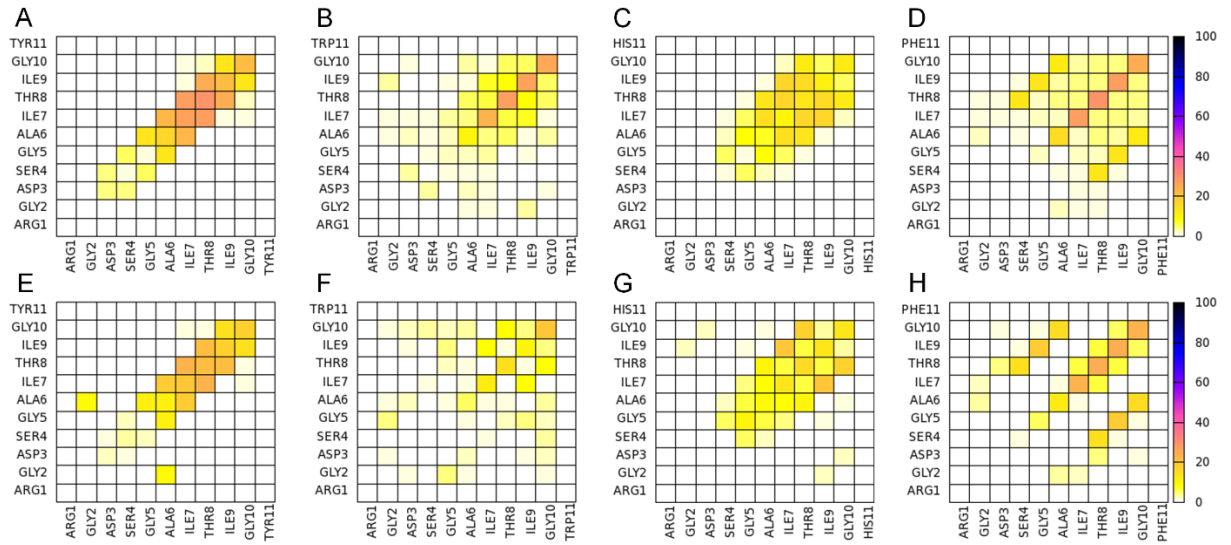


Figure 4.2 Density (%) maps of residue pairs forming intermolecular β-bridges for peptides RGDSGAITIGY/W/H/F. Figures B and F correspond to 4- and 5- stranded parallel configurations of peptide RGDSGAITIGW respectively. Figures C and G correspond to 4- and 5- stranded parallel configurations of peptide RGDSGAITIGH respectively. Figures D and H correspond to 4- and 5- stranded parallel configurations of peptide RGDSGAITIGF respectively. Reprinted with the permission from ref.[192]

4.4.4 Experimental Amyloid Fibril Formation

All experiments were performed by Drs. Mitraki's lab and by additional experimental collaborators, listed as authors in ref.[192] The lyophilized powders were dissolved in ultrapure water and followed by TEM over time in order to determine whether the designed peptides could self-assemble into amyloid fibrils. Figure 4.3 shows TEM micrographs of the RGDSGAITIGY, RGDSGAITIGW, RGDSGAITIGH and RGDSGAITIGF peptides after negative staining. For all peptides, straight, non-branched fibrils with diameters ranging from 10 to 50 nm and reaching several micrometers in length are observed. The TEM results confirm that the four designed peptides self-assemble into β-amyloid fibrils, validating that the *in silico* introduced mutations at

position 11, indeed can be accommodated in the amyloid-like β-sheet self-assembling peptide RGDSGAITIGC.

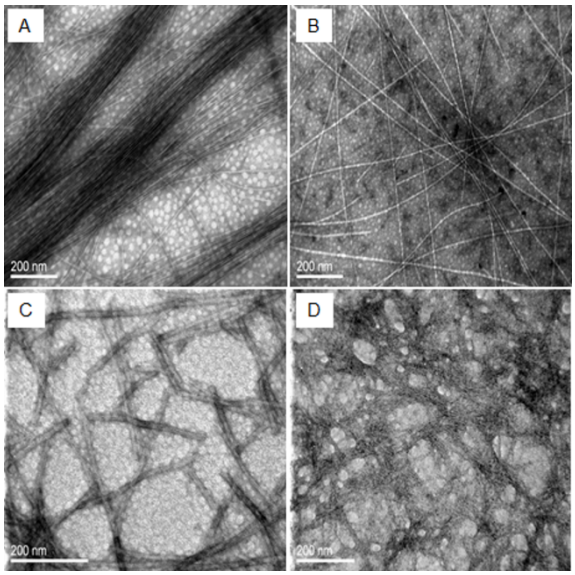


Figure 4.3 TEM Images of peptide solutions of 3mg/ml concentration. (A) RGDSGAITIGY after 24 hours in water. (B) RGDSGAITIGW after 48 hours in water. (C) RGDSGAITIGH after 24 hours in phosphate buffer (10mM-pH 7.4). (D) RGDSGAITIGF after 24 hours in water. Figure was produced by our experimental collaborators and is reprinted with the permission from ref.[192]

4.4.5 Computational Identification of Well-Aligned and Well-Ordered β-Sheet Conformations using P_1 and P_2 Parameters

We computationally evaluated the degree of alignment and order by plotting polar order parameter P_1 against nematic order parameter P_2 for the 4- and 5-stranded parallel β-sheets formed by the four designed peptides. For all the peptides that were investigated computationally, highly ordered and well-aligned states predominantly occur in 4-stranded rather than 5-stranded parallel β-sheet conformations (Figure 4.4). This can possibly be attributed to larger population of 4-stranded compared to 5-stranded parallel β-sheet conformations. We constructed two dimensional

free energy landscapes of 4-stranded parallel β-sheets formed by the designed peptides using the calculated P_1 and P_2 parameters as reaction coordinates (Figure 4.5A-4.5D), to investigate the degree of order and identify β-sheet conformations containing highly ordered and well-aligned strands that most likely correspond to naturally occurring amyloid fibrils. The global free energy minima in the plots are located in basins, marked with black dashed lines. From these basins we extracted the highly ordered and well-aligned β-sheet states for all subsequent computational analysis, as these conformations likely correspond to elementary β-sheet structural units of the naturally occurring fibrils. Representative conformations of the selected highly ordered and well-aligned β-sheet states are shown in the bottom panel of Figure 4.5 displaying the exposed residues and the amyloid zipper region of each peptide.

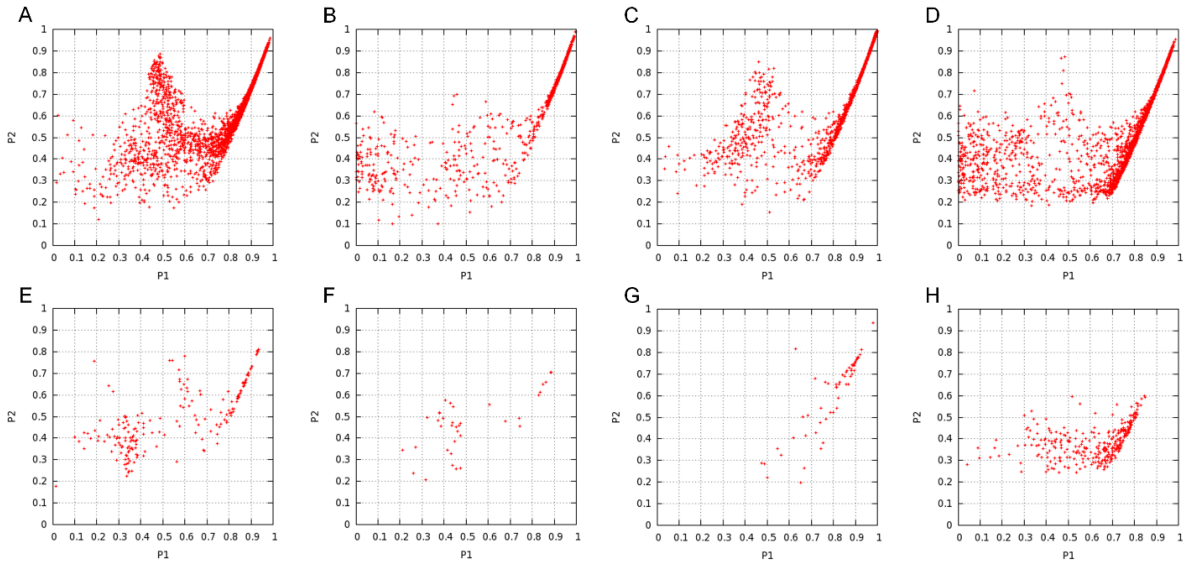


Figure 4.4 Plots of P_2 (Y-axis) as a function of P_1 (X-axis) for the 4- and 5- stranded parallel β-sheet conformations observed in the REMD simulations at 300 K for peptides RGDSGAITIGY/W/H/F. Figures A and E correspond to 4- and 5- stranded parallel configurations of peptide RGDSGAITIGY, respectively. Figures B and F correspond to 4- and 5- stranded parallel configurations of peptide RGDSGAITIGW, respectively. Figures C and G correspond to 4- and 5- stranded parallel configurations of peptide RGDSGAITIGH, respectively. Figures D and H correspond to 4- and 5- stranded parallel configurations of peptide RGDSGAITIGF, respectively. Reprinted with the permission from ref.[192]

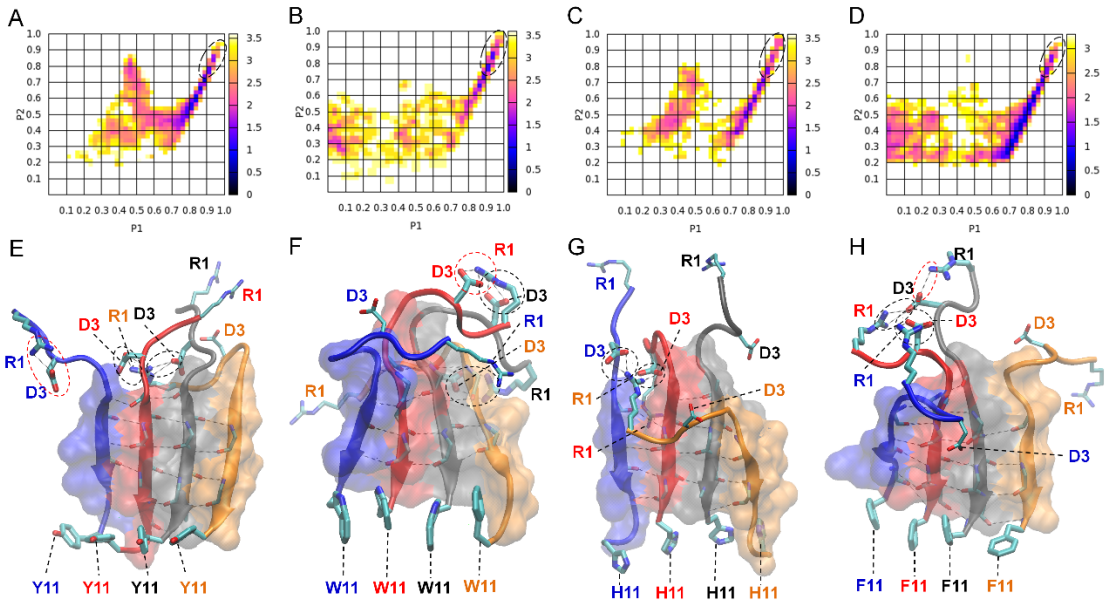


Figure 4.5 Free energy surfaces constructed from the 2-D probabilities of order parameters P_1 and P_2 and molecular graphics images of representative structures of RGDSGAITIGY/W/H/F extracted from the free energy minima. Upper panel: Free energy landscapes constructed from the 2-D probabilities of order parameters P_1 and P_2, calculated using the 4-stranded parallel β-sheets observed in the REMD simulations at 300 K. The surface in panels A, B, C and D correspond to 4-stranded parallel β-sheets formed by RGDSGAITIGY, RGDSGAITIGW, RGDSGAITIGH and RGDSGAITIGF respectively. Bottom panel: Molecular graphic images of representative highly ordered and well-aligned conformations of peptides RDSGAITIGY (E), RGDSGAITIGW (F), RGDSGAITIGH (G) and RGDSGAITIGF (H) in parallel arrangement. Reprinted with the permission from ref.[192]

4.4.6 Computational Analysis of Amyloid Zipper Formation, Salt Bridge Formation and Exposure of Functional Motifs

We performed a statistical analysis of the intra- and intermolecular side chain contacts within the highly ordered and well-aligned states of 4-stranded parallel β-sheets formed by the four computationally designed peptides (Figure 4.6). The intermolecular contacts involve side chain interactions between adjacent β-sheet interacting peptides only, while intramolecular contacts involve side chain interactions between residues within the same peptide. In general (i) the side

chains of aromatic residues at position 11 interact significantly with each other (Figure 4.6A-4.6D), indicating the presence of aromatic π–π interactions (aromatic residues are participating in such π-stacking in the bottom panels of Figure 4.5), and (ii) the side chains within the residue moiety GAITIG stabilize the β-sheet formation and tend to cluster across the β-sheet region forming an "amyloid zipper" (shown in surface representation in Figure 4.5E-4.5H). According to the contact maps, Arg1 and Asp3 of the RGD motif have the tendency to form both intra- and intermolecular contacts (Figure 4.6), as a result of intra- and intermolecular salt bridges formed by the two residues. The corresponding average propensities of salt bridge formation are similar for the four designed peptides. The average propensity for Arg1 and Asp3 to be involved in an (i) intermolecular salt bridge, (ii) intramolecular salt bridge, and (iii) intra or intermolecular salt bridge is 47%, 19%, 57% in the RGDSGAITIGY 4-stranded highly ordered and well-aligned parallel β-sheet states; 32%, 33%, 54% in the RGDSGAITIGW 4-stranded highly ordered and well-aligned parallel β-sheet states; 40%, 17%, 52% in the RGDSGAITIGH 4-stranded highly ordered and well-aligned parallel β-sheet states; 29%, 35%, 60% in the RGDSGAITIGF 4-stranded highly ordered and well-aligned parallel β-sheet states. Additionally, the probability of an arginine residue to form either an intra- or intermolecular salt bridge never exceeds 60%, indicating that arginine residues are sufficiently exposed. Figure 4.5E-4.5H show representative intramolecular salt bridges (shown in red dashed circles) and intermolecular salt bridges (shown in black dashed circles) formed by the four designed peptides. A detailed investigation of highly ordered and well-aligned β-sheet states formed by the four designed peptides within the REMD simulations, revealed the importance of intermolecular π–π interactions formed by the mutated residues at position 11 in the stabilization of the amyloid fibrils. This result complies with findings

of previous studies outlining the key role of π-stacking in peptide and protein self-assembly.[184,215,216,217]

We examined the degree of exposure of the entire functional motif RGD as well as the side chain of the designed residue at position 11 belonging to the two central peptides in the highly ordered and well-aligned states of 4-stranded parallel β-sheets formed by the four computationally designed peptides. The extent to which the RGD motif is solvent accessible in the four designed peptides is similar, and is equal to 58 ± 8%, 63 ± 7%, 66 ± 9%, 66 ± 7%, respectively for peptides RGDSGAITIG(Y/W/H/F)-CONH2. The degree of solvent accessibility of the RGD motif in the ensemble of NMR structures of the fibronectin protein102 is equal to 70 ± 5%, and in the selected 4-stranded highly ordered and well-aligned antiparallel β-sheet states of the self-assembling peptide RGDSGAITIGC,[24] which was experimentally verified for its cell adhesive properties, is equal to 66 ± 8%. Thus, a comparison between the degree of solvent accessibility of the RGD motif in our designed peptides and fibronectin as well as the cell adhesive fibrils formed by RGDSGAITIGC suggests that the RGD motif in the designed peptides of the present study is adequately solvent exposed to potentially possess cell adhesion properties. Despite the fact that the ratio is slightly smaller in the RGDSGAITIGY (58 ± 8%) peptide compared to fibronectin or the amyloid fibrils formed by RGDSGAITIGC,[24] the solvent accessibility of the RGD motif in 39% of the 4-stranded highly ordered and well-aligned parallel β-sheet formed by RGDSGAITIGY is within the range of solvent accessibility of the corresponding motif in the ensemble of NMR structures of fibronectin. Also, the solvent accessibility of the RGD motif in 90% of the 4-stranded highly ordered and well-aligned parallel β-sheets formed by RGDSGAITIGY is within the range of solvent accessibility of the corresponding motif in the 4-stranded highly ordered and well-aligned antiparallel β-sheet conformations formed by

RGDSGAITIGC.[24] Thus, the aforementioned results suggest that in the highly ordered and well-aligned parallel β-sheet states of the four designed peptides, the RGD motif is not part of the β-sheet core, it is exposed and it can possess functional cell adhesive properties within the fibrils formed.

The side chain moieties of residues Tyr11/Trp11/His11/Phe11 in the designed peptides are also adequately solvent exposed, and the ratios of solvent over total accessible surface areas ranges from 0.37 ± 0.09 (RGDSGAITIGH) to 0.46 ± 0.08 (RGDSGAITIGY). The significant degree of solvent accessibility of the tyrosine residues in the self-assembled structures is in agreement with the fact that the tyrosine residues are not part for the amyloid-like zipper formed. This further suggests that the tyrosine residues of adjacent β-sheet bonded peptides can be crosslinkable in the amyloids formed by the peptide with sequence RGDSGAITIGY, which is analyzed in detail below.

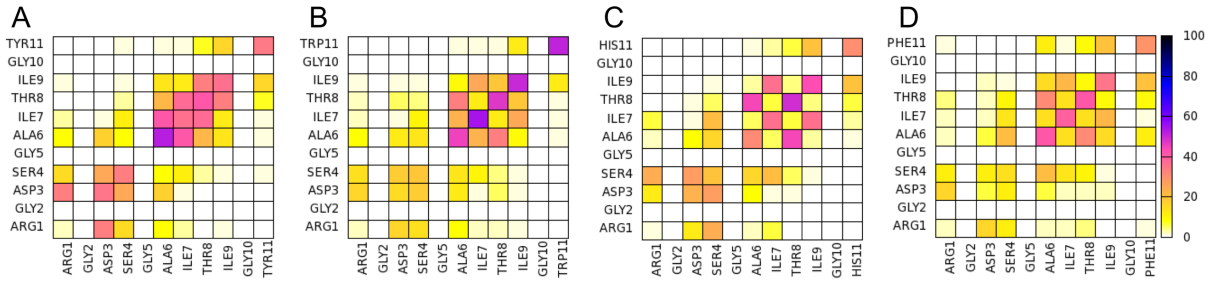
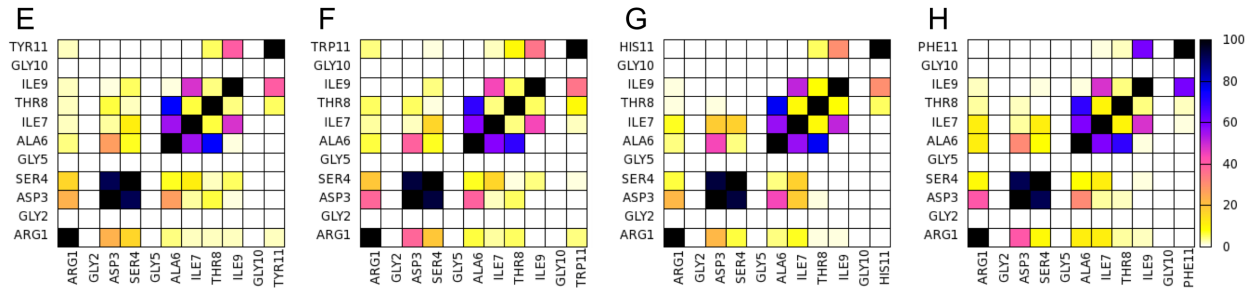


Figure 4.6 Percentage (%) probability density maps of side chain contacts of the highly ordered and well-aligned 4-stranded parallel β-sheets observed at 300 K for RGDSGAITIGY/W/H/F. Maps A, B, C and D correspond to intermolecular side chain residue–residue contacts between adjacent interacting peptides. Maps E, F, G and H correspond to intramolecular side chain residue–residue contacts. A contact is present when the two geometric centers of the two-side chain are within 6.5 Å. Figures A and E correspond to 4-stranded parallel configurations of peptide RGDSGAITIGY, respectively. Figures B and F correspond to 4-stranded parallel configurations of peptide RGDSGAITIGW, respectively. Figures C and G correspond to 4-stranded parallel configurations of peptide RGDSGAITIGH, respectively. Figures D and H correspond to 4-stranded parallel configurations of peptide RGDSGAITIGF, respectively. Reprinted with the permission from ref.[192]

Figure 4.6 Continued



4.4.7 Potential of Crosslinking in the Computationally Designed Self-Assembling Peptide RGDSGAITIGY

The aromatic residues at position 11 in each of the four designed peptides are not part of the amyloidogenic core and are accessible to the solvent, indicating that tyrosine residues within the designed peptide with sequence RGDSGAITIGY can presumably be crosslinked upon fibril formation. As tyrosine–tyrosine crosslink networks have outstanding mechanical properties[197,198] and a possible crosslinking between tyrosine residues within the RGDSGAITIGY fibrils can be of utmost importance for the designed fibril's stability, we focused our investigation on the designed peptide RGDSGAITIGY and examined its potential crosslinking properties. We calculated the distance (D) and angle (A) between the aromatic ring centers of tyrosines that belong to adjacent peptides of the 4-stranded highly ordered and well-aligned β-sheet states of RGDSGAITIGY. Using distance (D) and angle (A) as reaction coordinates (Figure 4.7) we constructed a free energy landscape to investigate the correlation between the distance and angle between tyrosine aromatic ring centers in the π–π interactions (e.g., when D is below approximately 7.5 Å). In the landscape, two free energy minima basins are observed in which the distance between the tyrosine aromatic ring centers is primarily within the range of 4–6 Å, and the angle between the tyrosine aromatic

ring centers is primarily within the range of 60–80° and 10–30° in the first and second free energy minima, respectively. Notably, the values of the distance and angle between tyrosine aromatic ring centers in the first minimum are close to corresponding distance and angle of the covalently bonded dityrosine compounds found in the ZINC database[218] (marked using a black "X" in Figure 4.7); the smaller value of the distance between the aromatic ring centers in the covalently bonded tyrosine compounds can be attributed to the fact that the two rings are covalently bonded and thus can be in closer proximity. Nevertheless, the relatively close proximity between the tyrosine aromatic ring centers in the modeled amyloid fibrils, in combination with the appropriate angle between the aromatic ring centers provide an additional indication that tyrosine residues can potentially be crosslinked upon fibril formation.

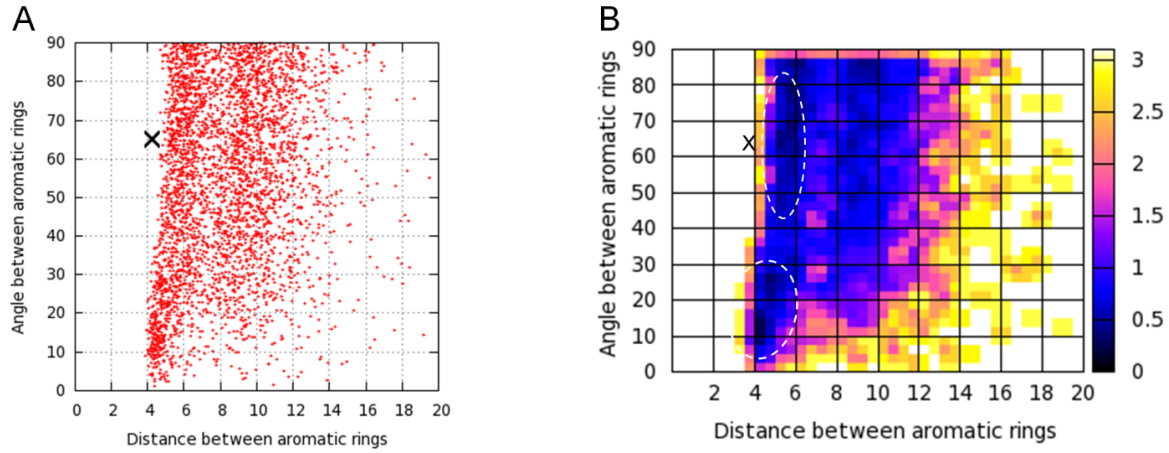


Figure 4.7 Angle between aromatic rings (Y-axis) as a function of distance between aromatic rings (X-axis). The angle and distance between the aromatic rings in the covalently bonded dityrosine compounds taken from the ZINC database is marked using a black "X". Panel A: The angle and the distance between the aromatic rings of tyrosine residues in adjacent interacting peptides of 4-stranded highly ordered and well- aligned parallel β-sheet states observed in the REMD simulations at 300 K of the peptide RGDSGAITIGY are shown in red dots. Panel B: Free energy landscape constructed from the 2-D probabilities, using as reaction coordinates, the distance and angle between the aromatic rings of tyrosine residues in adjacent interacting peptides of 4-stranded highly-ordered and well-aligned parallel β-sheet states observed in the REMD simulations at 300 K of the peptide RGDSGAITIGY. Reprinted with the permission from ref.[192]

4.5 Concluding Remarks

According to the computational design and analysis, a cysteine to tyrosine substitution from RGDSGAITIGC to RGDSGAITIGY "switches" the preferred β-sheet configuration from antiparallel to parallel, and the distance between adjacent tyrosine residues in the parallel β-sheet designable scaffolds is appropriate for crosslinking upon fibril formation. In addition, according to the highly ordered and well-aligned parallel β-sheet states formed by RGDSGAITIGY peptides within the simulations, the geometric properties of tyrosine residues belonging to adjacent β-sheet forming peptides are expected to be appropriate to accommodate crosslinking properties upon fibril formation. Preliminary experiments suggest the crosslinking ability of the specific peptide fibril scaffolds, which in combination with the findings of the current study, propose that the amyloid fibrils formed by the RGDSGAITIGY may pave the way for future biomaterials with tissue engineering applications.

4.6 Innovation in Methods

In the present work, we developed and introduced an in-house computational design strategy to energetically stabilize and potentially functionalize an amyloid peptide fibril. In summary, the following six features were innovatively combined and led to the successful design of a novel amyloid biomaterial with potential cell adhesive and crosslinking properties: (i) the minimized energy function in the computational design model, considered both protein-protein and protein-water interactions; (ii) the designable scaffolds (flexible templates) corresponded to an ensemble of computationally modeled highly ordered and well-aligned β-sheet states and not random peptide aggregates; (iii) the designable scaffolds comprised of both antiparallel and parallel modeled β-sheets independently, and thus, both possible configurations parallel versus antiparallel were energetically examined upon an amino acid substitution; (iv) the choice of the

coarse-grained force field used[205]; (v) the additional constraints introduced to verify the functionality of the designed peptide sequences for potential crosslinking applications; (vi) the use of REMD simulations and free energy calculations to extract highly-accurate structures as modeled fibrils validating that the computationally designed peptides (a) self-assemble into highly ordered and well-aligned parallel β-sheet states, (b) are arranged in line with the computational design results, and (c) possess potential bifunctional cell adhesion and crosslinking properties. Notably, according to points (ii) and (iii) the mutations during the design are introduced on realistically modeled β-sheet arrangements of peptides, parallel or antiparallel, which can constitute elementary structural units of the fibrils. Our study suggests that an analogous computational design strategy can be introduced for the design of novel functional biomaterials with diverse applications.

4.7 New Findings and Potential Applications

We observed that the introduction of the aromatic residues at position 11 stabilize the fibrils formed by the peptide RGDSGAITIGC along with a switch in the dominant configuration (from antiparallel to parallel) of the peptide. We also computationally predicted the potential cross-linking properties of the designed peptide RGDSGAITIGY by comparing the geometrical properties of the highly-ordered and well-aligned conformations formed by the peptide with the geometrical properties of the dityrosine compound in Zinc database. Preliminary experiments suggest the crosslinking ability of the specific peptide fibril scaffolds. Thus, the designed amyloid materials formed by the peptide can pave the way for the discovery of tissue engineering agents formed by self-assembling peptides, encompassing cell-adhesion and cross-linking properties, with potential superior mechanical properties.

5. A NOVEL AMYLOID DESIGNABLE SCAFFOLD INSPIRED BY GAIIG OF AMYLOID BETA AND THE HIV-1 V3 LOOP*

5.1 Introduction

5.1.1 Overview and Significance

As outlined in the Introduction, naturally occurring peptide sequences extracted from amyloid proteins or β-sheet protein regions can self-assemble outside the context of the entire sequence into amyloid β-sheets and can serve as scaffolds for novel materials.[25,26,27,28,29,36,222,223,224,225,226,227,228] GAIIGL[229] and NSGAITIG[131] are two peptide sequences similar in sequence which are part of the amyloid-β (Aβ) peptide, linked to Alzheimer's disease, and the adenovirus fiber shaft[230,231], respectively. Both GAIIGL and NSGAITIG form amyloid β-sheets outside the context of the entire peptide or protein. According to experimental X-ray and computational molecular dynamics (MD) simulation studies, in both peptides, residues outside the GAIIG or GAITIG sequences are not part of the amyloidogenic β-sheet core: the C-terminal leucine, and the N-terminal asparagine and serine residues in the two peptides, respectively, are exposed. The latter provided impetus[131] for the discovery of a series of amyloid materials with several applications[25,26,27,28,29] by modifying the NS-residues. While aromatic residues are key components of amyloid self-assembly (as pioneered by Gazit and colleagues[184,215,217]), patterns of aliphatic residues are also key self-assembly components[51,232233] , contributing to the amyloid properties of GAIIGL[229].

* Reprinted with the permission from "A novel amyloid designable scaffold and potential inhibitor inspired by GAIIG of amyloid beta and the HIV-1 V3 loop" by Kokotidou et al., 2018. *Federation of European Biochemical Societies Letters*, 2, 321-335, Copyright 2018 by Federation of European Biochemical Societies.

The recent experimentally resolved structures of entire Aβ fibrils[234,235,236] depicted that the last glycine of GAIIG could introduce a turn into the amyloid β-sheets. The aforementioned information suggests that GAIIG can be a sufficiently short amyloidogenic core of larger amyloid forming peptides containing additional (e.g., 2–3) residues at both termini, which can be outside the amyloid β-sheet, as glycine residues can act as β-turn promoters halting β-sheet elongation. The discovery of such amyloid designable scaffolds formed by relatively short peptides can serve as a source of inspiration for the discovery of amyloid materials with advanced properties, as the exposed residues can be modified accordingly depending on the desired application.

5.1.2 Goal

In this study, we carried out experimental and computational studies to investigate if the shorter (compared to GAIIGL) GAIIG peptide (amidated at the C-terminus) self-assembles into amyloid β-sheets. Following the positive outcome, we aimed to exploit the GAIIG as a core of longer self-assembling peptides which can constitute amyloid designable scaffolds (see below).

5.1.3 Hypothesis

Dr. Tamamis' postdoctoral studies in late Dr. Floudas lab[237,238] showed that the HIV-1 gp120 V3 loop adopts a structure in which the opposite stems of the loop form a β-sheet in its interaction with chemokine receptors CXCR4 and CCR5. Interestingly, in this study, we observed that a GAIIG sequence fragment (or other homologous fragments including GQIIG, GQIVG, etc.) is part of one of the two stems, comprising residues 24–28 in a typical 35-residue long V3 loop, according to the HIV sequence database (https://www.hiv.lanl.gov/). This observation led us to postulate that such variable sequences derived from the HIV-1 gp120 V3 loop containing GAIIG

as an amyloid core and additional (e.g., 2–3) residues at both termini could serve as a source of inspiration for novel amyloid material scaffolds.

5.1.4 Objectives Accomplished

We performed computational studies which combined with experiments showed that the GAIIG sequence, common to the amyloid beta peptide (residues 29–33) and to the HIV-1 gp120 (residues 24–28 in a typical V3 loop), self-assembles into amyloid fibrils, as suggested by theory and the experiments presented here. The studies also showed that the longer YATGAIIGNII sequence from the V3 loop also self-assembles into amyloid fibrils, of which the first three and the last two residues are outside the amyloid GAIIG core. We postulated that this sequence, with suitably selected modifications at the flexible positions, can serve as a designable scaffold for novel amyloid-based materials.

This study resulted in a publication of Dr. Tamamis' computational lab in collaboration with Dr. Mitraki's, and Dr. Llamas-Saiz's experimental labs, and other experimentalists which their names are listed below in the complete publication reference: "Kokotidou, C.; Jonnalagadda, S. V. R.; Orr, A. A.; Seoane-Blanco, M.; Apostolidou, C. P.; van Raaij, M. J.; Kotzabasaki, M.; Chatzoudis, A.; Jakubowski, J. M.; Mossou, E.; Forsyth, V.T.; Mitchell, E.P.; Bowler, M.W.; Llamas-Saiz, A.L.; Tamamis, P.; Mitraki, A. A novel amyloid designable scaffold and potential inhibitor inspired by GAIIG of amyloid beta and the HIV-1 V3 loop. *FEBS Lett*. **2018**, *592*, 1777-1788."

CK, SVRJ, AAO and MS-B are equally contributing first authors.

5.2 Computational Methods

The following two peptides, NH_3^+-GAIIG-$CONH_2$, and NH_3^+-YATGAIIGNII-COO^-, were studied using a combination of computational and experimental methods. The computational

methods were performed by following the computational protocol developed by Tamamis and Archontis[130] described in chapter 2. We first performed infinite dilution simulations to investigate the conformational properties of the isolated peptides, followed by finite dilution simulations, at which we initially placed the isolated peptides in a cubic periodic box to investigate the self-assembly properties of the three peptides. In both cases, we used replica exchange MD (REMD) simulations[133,134,135,136,137,138] to facilitate peptides' conformational sampling as well as to avoid the system of self-assembly peptides in finite dilution to being trapped in local energetic minima. After the completion of the self-assembly (finite dilution) simulations, we categorized the observed β-sheet content into 2-,3-,4-,5-, and 6-stranded β-sheets and identified the key β-sheet interactions. Subsequently, we investigated the presence of highly ordered and well-aligned β-sheets for the three peptides. In what follows, we summarize each step.

5.2.1 Infinite Dilution Simulations

We independently simulated an isolated peptide of NH_3^+-GAIIG-$CONH_2$, NH_3^+-YATGAIIGNII-COO^- in aqueous solution using REMD simulations in CHARMM[110]. A total of eight replicas with temperatures 283, 300, 318, 336, 356, 377, 403, and 432 K were employed. The simulations were performed using the FACTS19[123] implicit solvent model and a surface tension coefficient value of 0.015 kcal/(mol*Å^2). We employed Langevin dynamics and a friction coefficient of 5 ps^{-1} was introduced on all heavy atoms. The simulation time for each temperature per peptide system was 150 ns. Upon completion of the simulations, we analyzed 15000 snapshots of the 300 K trajectory, extracted at 100 ps intervals. Subsequently, we performed a root mean squared deviation (RMSD)-based clustering analysis using Wordom[141,142] on the extracted snapshots to obtain six representative conformations for each peptide. The clustering analysis was performed using all backbone atoms and a clustering radius of 2 Å. From each of the six most

populated clusters, we extracted the most representative conformation, resulting in a total of 6 conformations per peptide. These conformations were used as initial structures in the subsequent self-assembly simulations described below.

5.2.2 Self-Assembly Simulations

We performed two independent REMD simulations in CHARMM[110], analogously to refs[24,40,130,132,131,192] , to investigate the self-assembling properties of the three peptides in solution. We placed six copies of the peptide NH_3^+-GAIIG-$CONH_2$ in a 110 Å cubic periodic box, and six copies of the peptide NH_3^+-YATGAIIGNII-COO^- in a 148 Å cubic periodic box, resulting in an approximate 3 mg/ml concentration for the two peptide systems. All six conformations from each peptide were initially placed at the center of the box. Subsequently, for each of the peptide systems, the two first peptide conformations were translated by +25 Å and -25 Å in the x-direction respectively, the following two peptide conformations were translated by +25 Å and -25 Å in the y-direction respectively, and the last two peptide conformations were translated by +25 Å and -25 Å in the z-direction respectively. The simulations were performed using the FACTS19[123] implicit solvent model and a surface tension coefficient value of 0.015 kcal/(mol*$Å^2$). We employed Langevin dynamics and a friction coefficient of 5 ps^{-1} was introduced on all heavy atoms. In the simulations investigating the self-assembling properties of GAIIG, we employed sixteen replicas with temperatures 284, 292, 300, 309, 319, 330, 342, 355, 368, 382, 397, 413, 430, 448, 467, and 487 K. In the simulations investigating the self-assembling properties of YATGAIIGNII, we employed sixteen replicas with temperatures 290, 295, 300, 305, 310, 315, 321, 327, 333, 339, 345, 352, 359, 366, 373, and 380 K. The time duration for each of the three replica exchange runs was equal to 10 ps and the total simulation for all the temperatures per peptide system was equal to 16 μs. Upon completion of the simulations, we collected the final conformations at 300 K for

each of the self-assembling peptide sequences. These final conformations were combined into one trajectory per peptide sequence, which corresponded to 1000 ns and 100,000 snapshots.

5.2.3 Analysis of the Self-Assembly Simulation Trajectories

We identified the formation of intermolecular β-sheet structures and categorized the observed β-sheets into 2-, 3-, 4-, 5-, 6- stranded parallel, antiparallel, mixed, and complex β-sheet conformations similarly to the refs[24,40,130,132,131,192]. According to our definitions, complex conformations correspond to cases in which one peptide forms β-sheet interactions with more than two peptides at a specific instance[24,40,130,132,131,192]. This analysis showed that antiparallel is the dominant configuration for all of the three peptides. Subsequently, similar to our previous studies[24,40,130,132,131,192], we focused our further analysis on 4-, 5-, and 6-stranded antiparallel β-sheets which represent conformations of higher complexity competed to 2- and 3-stranded β-sheets. We extracted 4-, 5- and 6-stranded antiparallel β-sheets from the corresponding MD simulation snapshots for each peptide and calculated the (%) probability of an intermolecular pair of residues, belonging to two neighboring β-sheet interacting peptides, to be involved in a β-sheet (or β-bridge) conformation. This analysis indicated the predominant patterns of intermolecular residue-pairwise β-sheet interactions, depicting the key amyloidogenic regions of each peptide. Subsequently, we identified highly ordered and well-aligned β-sheets for the three peptides using P_1 *and* P_2 parameters. The analysis was performed similarly to refs[24,40,130,132,131,192] with the only difference being the unit vector, $\vec{z}_i$, for each peptide. For GAIIG, we selected unit vector $\vec{z}_i$ to be defined by the segment spanning from the Cα atom of residue 1 to the Cα atom of 5. For YATGAIIGNII, we selected unit vector $\vec{z}_i$ to be defined by the segment spanning from the Cα atom of residue 4 to the Cα atom of 9. We computed P_1 and P_2 parameters for the 4- to 6- stranded

antiparallel β-sheets of each of the three peptides as in ref[24]. We observed the existence of highly populated and highly ordered β-sheets in the 4-stranded antiparallel β-sheet conformations formed by GAIIG and YATGAIIGNII.

Subsequently, we focused our further analysis on 4-stranded antiparallel β-sheets formed by GAIIG and YATGAIIGNII as they are more populated than 5- and 6- stranded antiparallel β-sheets and encompass highly-ordered and well-aligned β-sheet structures. Specifically, we examined the structural properties of the highly ordered and well-aligned conformations formed by the two peptides. We constructed free energy landscapes for the specific 4-stranded antiparallel β-sheets, using the two dimensional probability $P(P_1, P_2)$ and *Eq. 5.1*:

$$G(P_1, P_2) = -k_B T \ln\left[P(P_1, P_2)\right] \qquad 5.1$$

For both GAIIG and YATGAIIGNII self-assembling peptide systems, we identified the free energy basin that encompasses the majority of conformations, from which we extracted the highly ordered and well-aligned states. For YATGAIIGNII, the selected highly ordered and well-aligned 4-stranded antiparallel conformations were utilized for subsequent analysis focusing on the peptides' structural properties, and similarly to our previous studies[24,192] , we investigated the solvent accessibility of the terminal residues in the peptide YATGAIIGNII. We observed that the first three N-terminal residues, Tyr1, Ala2, and Thr3 as well as the last two C-terminal residues Ile10 and Ile11 are solvent exposed, and thus they are amenable for modification for functionalization purposes.

5.3 Materials and Experimental Methods

All experiments described in 5.3 were performed by Drs. Mitraki's and Llamas-Saiz's labs and by additional experimental collaborators, listed as authors in ref.[23]

5.3.1 Peptides and chemicals

The two peptides NH_3^+-GAIIG-$CONH_2$, and NH_3^+-YATGAIIGNII-COO^- were purchased from Genecust (Luxemburg) and possessed a degree of purity higher than 95%.

5.3.2 Transmission Electron Microscopy (TEM)

Samples for TEM analysis were prepared by depositing 8 μl of the sample on carbon-coated formvar copper grids (Agar Scientific), left aside for two minutes, dried with a filter paper and then the same procedure was repeated with the stain. The samples were negatively stained with 8 μl 1% (w/v) phosphotungstic acid for two minutes for GAIIG and with 8 μl 2% (w/v) uranyl acetate for YATGAIIGNII. The TEM experiments were performed using a JEOL JEM 2100 High Resolution microscope, operating at 80 kV (University of Crete, Biology Department).

5.3.3 X-ray fiber diffraction

A droplet of a peptide fibril solution was placed between two glass rods that were supported by two plasticine balls and allowed to dry while pulling to induce shear alignment as previously described[239]. The X-rays were focused on the aligned fibers at right angles and the diffraction patterns were recorded.

For the GAIIG peptide fibrils, the diffraction patterns were recorded with a SIEMENS M18XHF Rotating anode generator equipped with a MarResearch 345 image detector system, at a wavelength of 1.541 Å (Cu Ka edge). The exposure times were 30 min per image.

For the YATGAIIGNII peptide fibrils, X-ray fiber diffraction experiments were carried out at the European Synchrotron Radiation Facility (ESRF in Grenoble) on the MASSIF-1 beamline[240] with a wavelength of 0.966 Å and a beam size of 15 lm. The exposure time was of the order of 3 s. The sample–detector distance was set to 435.14 mm for giving a resolution of 3.5 Å at the edge of the detector.

5.4 Results and Discussion

5.4.1 Self-Assembly of the GAIIG and YATGAIIGNII Sequences

According to 16 μs replica exchange MD simulations and a subsequent computational analysis performed analogously to previous studies,[24,40,130,132,131,192] GAIIG primarily self-assembles into antiparallel off-register β-sheets (Figure 5.1A, 5.1C) which possess a high degree of order and alignment of peptides (Figure 5.2A). The β-sheet core of the peptide is predominantly composed of the two isoleucine residues Ile3 and Ile4. A representative structure of a highly-ordered and well-aligned β-sheet composed by four peptide strands is presented in Figure 5.3A. Compared to the antiparallel β-sheets formed by GAIIGL, our analysis suggests that the presence of an additional leucine at the end of GAIIG is not necessary for self-assembly.

Experiments performed by Drs. Mitraki's and Llamas-Saiz's labs and by additional experimental collaborators, listed as authors in ref[23] validated the amyloidogenic properties of GAIIG, revealing that the peptide forms non-branched fibrils with diameters of around ten nm and lengths reaching the order of microns as revealed by Transmission Electron Microscopy (Figure 5.4A). Fibrous rods of the peptide display the characteristic cross- beta signature in X-ray fiber diffraction: a 4.6 Å meridional reflection that corresponds to the distance between b-strands and a 10.1 Å equatorial reflection that corresponds to the distance between b-sheets (Figure 5.4B).

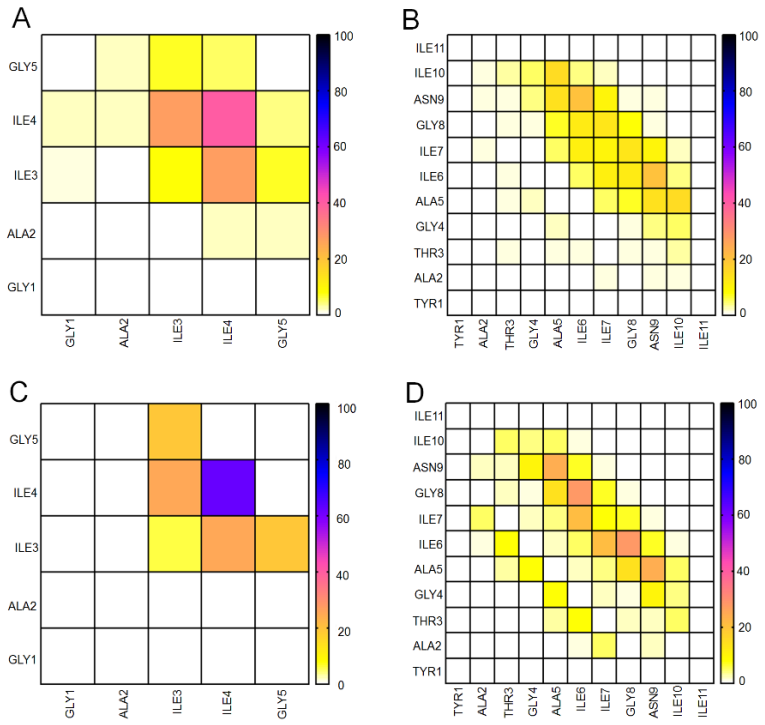


Figure 5.1 Density (%) maps of residue pairs forming intermolecular β-bridges for GAIIG and YATGAIIGNII. The pairs of residues belong to nearest neighboring peptides participating in an isolated β-bridge or extended β-sheet conformation in the REMD simulations at 300 K. Figures A, and C correspond to 4- and 5- stranded antiparallel configurations of peptide GAIIG respectively. Figures B, and D correspond to 4-, and 5-stranded antiparallel configurations of peptide YATGAIIGNII respectively. Reprinted with the permission from ref.[23]

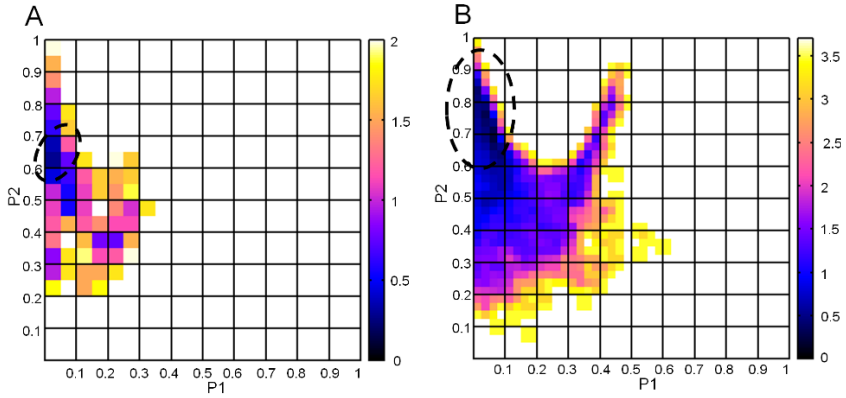


Figure 5.2 Free energy surfaces constructed from the 2-D probabilities of order parameters P_1 and P_2, calculated using the 4-stranded antiparallel β-sheets observed in the replica exchange MD simulations at 300 K for GAIIG and YATGAIIGNII. Surface in panel A, and B correspond to 4-stranded antiparallel β-sheets formed by GAIIG and YATGAIIGNII peptides respectively. Reprinted with the permission from ref.[23]

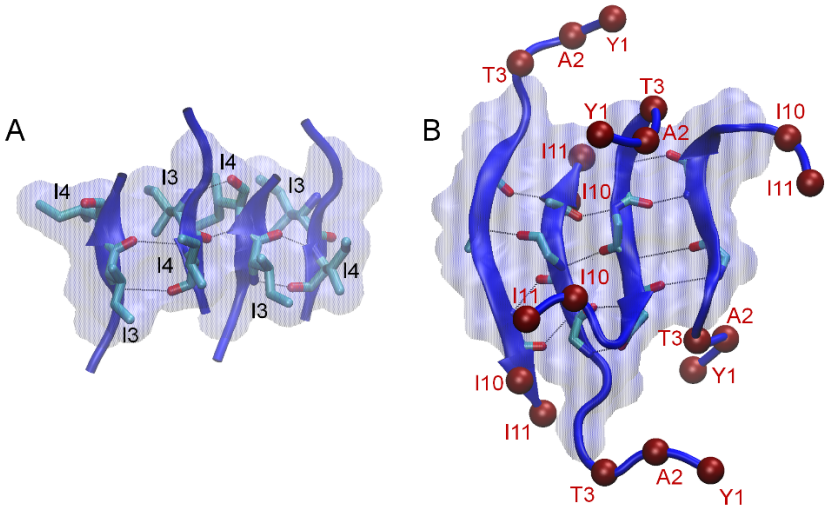


Figure 5.3 Molecular graphics images of representative highly ordered and well aligned 4-stranded β-sheet fibrils of A) GAIIG and B) YATGAIIGNII peptides. Reprinted with the permission from ref.[23]

The aforementioned information suggests that GAIIG can be a sufficiently short amyloidogenic core of larger amyloid forming peptides containing additional (e.g. 2-3) residues at both termini which can be outside the amyloid β-sheet, as glycine residues can act as β-turn promoters halting β-sheet elongation. The discovery of such amyloid peptide scaffolds can serve as a source of inspiration for the discovery of amyloid materials with advanced properties, as the exposed residues can be modified accordingly depending on the desired application.

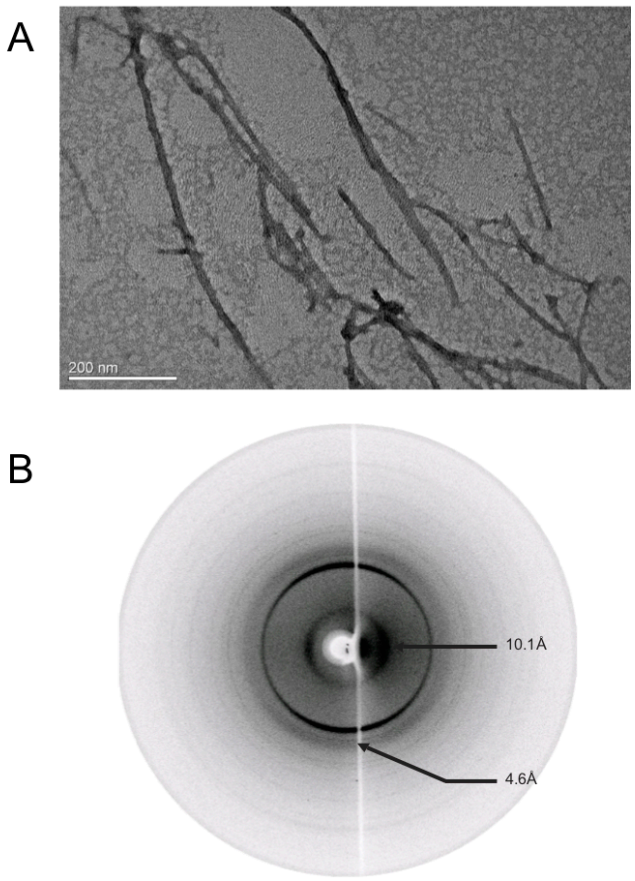


Figure 5.4 TEM micrograph and X-ray fiber diffraction images of GAIIG solutions. (A) TEM micrograph of a 20 mg•mL^{-1} solution of the GAIIG peptide following 62 days of incubation in phosphate buffer pH 7, negatively stained with phosphotungstic acid 1%. (B) X-ray fiber diffraction pattern of rods formed from a 7 mg•mL^{-1} solution in phosphate buffer following aging for 40 days. Figure was produced by our experimental collaborators and is reprinted with the permission from ref.[23]

In this study, we focused on one such sequence: YATGAIIGNII derived from a V3 loop[241] (without any modifications at the termini). Similarly to GAIIG, according to 16 μs replica exchange MD simulations and a subsequent computational analysis, which was performed analogously to previous studies,[24,40,130,132,131,192] the YATGAIIGNII peptide primarily self-assembles into antiparallel off-register β-sheets (Figure 5.1B, 5.1D), which possess a high degree of order and alignment of peptides (Figure 5.2B). The β-sheet core of the peptide is predominantly

composed of residues Gly4 to Asn9, encompassing the GAIIG domain. A representative structure of a highly-ordered and well-aligned β-sheet composed by four peptide strands is presented in Figure 5.3B. Interestingly, we observe that both glycine residues act as β-turn promoters, which halt the elongation of the β-sheet core outside the domain. Thus, N-terminal residues Tyr1, Ala2, Thr3, as well as C-terminal residues Ile10 and Ile11 are rarely involved in β-sheet formation and are outside the amyloid zipper-like region formed within the GAIIG domain and Asn9 (Figure 5.1B, 5.1D). As a result, the amyloid scaffolds formed by YATGAIIGNII can be considered as excellent designable scaffolds for the synthesis of functional amyloid materials. This can be achieved by introducing suitably selected mutations at the non β-sheet forming terminal residue positions 1, 2, 3, 10 and 11, which would not disrupt the amyloid self-assembly properties and at the same time would allow the newly designed amyloid fibrils to bind to ions, molecules or surfaces.

Experiments validated the amyloidogenic properties of YATGAIIGNII. TEM micrographs of YATGAIIGNII fibrils reveal a typical amyloid-type morphology (Figure 5.5A) and X-ray fiber diffraction of rods display the characteristic cross-b signature with a 4.67 Å reflection at the meridian and 9.92 Å at the equator (Figure 5.5B). Additional replica exchange MD simulation runs with a longer 13-residue peptide with sequence AFYATGAIIGNII extracted from the same HIV-1 gp120 V3 loop show that the inclusion of additional residues result in the formation of U-shaped b-sheets (preliminary results not shown) similarly to β-sheets formed by LSFDNSGAITIG[132]. Thus, 11-residue peptides containing three residues before and after the GAIIG domain can be optimal designable amyloid scaffolds containing the maximum number of mutable positions and at the same time comprising linear shaped peptides in which the non-b-sheet residues are exposed for functionalization purposes.

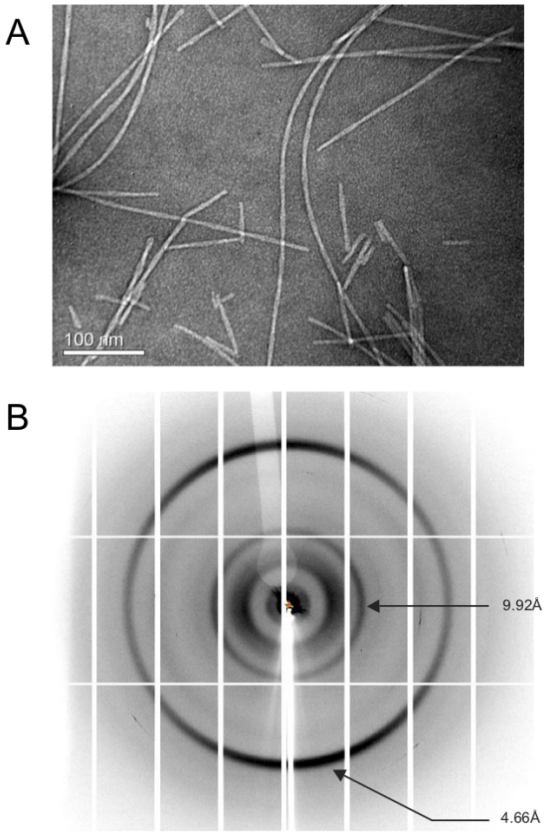


Figure 5.5 TEM micrograph and X-ray fiber diffraction images of YATGAIIGNII solutions. (A) TEM micrograph of a 5 mg•mL^{-1} solution of the YATGAIIGNII peptide following 5 days of incubation in water, negatively stained with uranyl acetate 2%. (B) X-ray fiber diffraction pattern of rods formed from a 3 mg•mL^{-1}solution in water after 2 h of incubation. Figure was produced by our experimental collaborators and is reprinted with the permission from ref.[23]

5.5 Concluding Remarks

Our computational studies in tandem with experimental studies performed by collaborators (see ref[23]) point to the GAIIG sequence as an amyloid-forming building block. Larger amyloid forming peptides like YATGAIIGNII have GAIIG as amyloidogenic core, and additional (e.g. 2-3) residues at both termini are outside the amyloid β-sheet, as glycine residues can act as β-turn promoters halting β-sheet elongation. By introducing suitably selected mutations at the non β-sheet forming terminal residue positions of the longer sequence YATGAIIGNII, novel materials could be designed. Additional peptide sequences containing the GAIIG amyloid core plus 2-3 residues

at both termini and inspired by either the HIV-1 V3 loop sequence variability or other naturally occurring proteins encompassing the amyloid GAIIG, may also be designable amyloid scaffolds that could be further investigated. Interestingly, this can be supported by the fact that GAIIG domain can be found in amyloid or β-sheet rich regions of proteins of known structure, including Aβ ($_{29}$GAIIG$_{33}$ in PDB ID: 5OQV235), α-tubulin acetyl transferase ($_{91}$GAIIG$_{95}$ in PDB ID: 4PK2[242] Chain A), *Mycobacterium smegmatis* alpha-ketoglutarate decarboxylase homodimers ($_{275}$GAIIG$_{279}$ in PDB ID: 2XT6[243]Chain A), and Japanese encephalitis virus non-structural protein 1' ($_{181}$GAIIG$_{185}$ in PDB ID: 5O36[244] Chain A).

5.6 Innovations in Methods

Using well-established computational methods, we have elucidated that peptide sequence YATGAIIGNII can represent an amyloid designable scaffold as GAIIG can act as amyloidogenic core, while additional (e.g. 2-3) residues at both termini are outside the amyloid β-sheet core; this could be attributed to the fact that glycine residues may act as β-turn promoters halting β-sheet elongation. Such variable sequences with GAIIG as an amyloid core and additional (e.g. 2-3) residues could serve as a source of inspiration for novel amyloid material scaffolds with variable geometries. Such designable scaffolds can be methodologically instrumental as templates for the computational design of amyloid functional materials (refer to chapters 6, 7, 8, 9, and 10) and can transform the way of functionalization of amyloid materials.

5.7 New Findings and Potential Applications

We have computationally elucidated the elementary β-sheet structural units of the peptide YATGAIIGNII using the computational protocol developed by Tamamis and Archontis.[130] Our experimental collaborators also showed that the specific peptide self-assembles into amyloid fibrils. The specific peptide forms highly ordered and well-aligned antiparallel β-sheet structures

predominantly although the peptide can form highly ordered and well-aligned parallel β-sheet structures as well. Both the glycine residues of the peptide act as β-turn promoters halting the elongation of the β-sheet core. As a result, the residue at the termini don't form β-sheet and are flexible. The non-β-sheet forming residue can be mutated to achieve desired functionality (achieved in the following chapters). Thus, the amyloid scaffolds formed by the specific peptide can be considered as excellent designable scaffolds for the synthesis of novel functional amyloid materials. Even though the peptide predominantly antiparallel designable scaffolds, both parallel and antiparallel designable scaffolds can be considered suitable for the design of functional amyloid materials to improve the success rates of the successfully designed amyloid materials), which can potentially be functional in either configurations (further elaborated in chapter 7 and 8).

6. COMPUTATIONAL DESIGN OF FUNCTIONAL AMYLOID MATERIALS WITH CESIUM BINDING, DEPOSITON AND CAPTURE PROPERTIES*

6.1 Introduction

6.1.1 Overview and Significance

Dr. Tamamis' doctoral studies on the amyloid self-assembly of peptides with sequence NSGAITIG extracted from the adenovirus fiber shaft showed that the NS domain is primarily exposed outside the β-sheet amyloid core;[131] this provided impetus for the design of novel amyloid materials, through the incorporation of suitably selected modifications at the non-β-sheet forming positions, with a series of applications in biomedicine and technology.[25,26,27,28,29] We recently utilized this result in conjunction with computational methods to design amyloid-forming peptides RGDSGAITIGC[24] presented in chapter 3 and RGDSGAITIGY[192] presented in chapter 4. The amyloid materials formed by both peptides were suggested to possess promising biomedical applications; the former supports cell adhesion and proliferation of a model cell line and contains free thiols for metal binding properties,[24] whereas the latter has potential dityrosine cross-linking properties according to preliminary experiments.[192] Dityrosine cross-linking can provide a means to control the mechanical properties of the materials, which is reflected by their inclusion in several natural and engineered materials (e.g., resilin, fibrinogen, silk, keratin, collagen, and elastin).[245]

Although functional amyloid materials have been produced for several applications including ion binding (e.g., calcium[25,40], arsenic[41], gold[42], mercury[42], lead[42], and palladium[42]) or compound binding (e.g., CO_2[32,33]), experiments have primarily relied on scientists' intuition to

* Reprinted with the permission from "Computational Design of Functional Amyloid Materials with Cesium Binding, Deposition, and Capture Properties" by Jonnalagadda et al., 2018. *The Journal of Physical Chemistry B*, 122, 30, 7555-7568, Copyright 2018 by American Chemical Society

introduce mutations to amyloid scaffolds yielding functional amyloid materials, which hinders their discovery. Therefore, if functionalization is not computationally guided (based on biophysical, mathematical, physicochemical principles), several mutants need to be generated and tested experimentally, often with low success, making the procedure inefficient or nearly impossible if the desired property is hard to achieve. As a result, the computational design of functional amyloid materials has been limited to rational or optimization-based approaches, which do not explicitly consider their functionalization during the design (e.g., our studies on RGDSGAITIGC,[24] presented in chapter 2 and RGDSGAITIGY,[192] presented in chapter 3 respectively).

In this chapter, we considered the computational design of functional amyloid materials with the potential applicability to remove cesium ions from water. During nuclear reactor accidents or minor nuclear power station accidents, cesium, a key uranium fission product, can easily dissolve in water,[246] posing a significant risk to human health.[247] With its long half-life (approx. 30 years) and high activity, volatility, and solubility in water, cesium-137 can easily enter the food chain to inflict radiological harm to humans, potentially causing diseases such as hematological abnormalities, morphological abnormalities, and cancers.[247] Technologies including natural inorganic cation exchangers (e.g., zeolites and clays) and synthetic inorganic materials (e.g., synthetic γ-zirconium phosphate, niobate molecular sieves, micas, and titanate nanomaterials) have been shown to capture cesium ions.[246,248,249,250,251] In addition, poly(ethylene glycol)-decorated Prussian blue magnetic nanoparticles have recently been synthesized for the removal of cesium ions from blood.[247] Thus, considering the potential biocompatibility of amyloid materials, the design of amyloid materials removing cesium ions from blood could be considered a promising step in the field.

6.1.2 Goal

In this work, our two-fold aim was to (i) develop the first computational protocol for the design of functional amyloid materials binding to an ion of interest which could constitute constituting a stepping-stone for the future design of amyloid materials with advanced applications potentially binding to any ion of interest, and (ii) implement-validate the protocol in a test case involving the design of peptides forming amyloid materials capable of binding to cesium ions, which can effectively be used as materials encompassing cesium deposition and capture properties. Specifically, our utmost goal was to design novel amyloid materials that bind and capture cesium ions at neutral and low pH conditions.

6.1.3 Hypothesis

Here, we hypothesized that our goal could achieved through the development of a computational protocol for the design of functional amyloid materials capable of binding to an ion of interest. The key postulation within the computational protocol involved the generation of an optimization-based design model that could introduce mutations at non-β-sheet residue positions of an amyloid designable scaffold in such a way that the designed amino acids introduced to the scaffold (i) mimic how amino acids bind to cesium ions according to experimentally resolved structures and, (ii) form energetically stable binding pockets binding to cesium ions.

6.1.4 Objectives Accomplished

We developed the first to our knowledge protocol for the computational design of functional amyloid materials which can summarized in the following steps: (i) The flexible structural templates representing the elementary β-sheet structural unit of an amyloid designable scaffold are analyzed, and the mutable positions are identified. (ii) Binding motifs of amino acids

in complex with the ion of interest deposited in the Protein Data Bank (PDB)[252] are analyzed to define materialphore models describing how amino acids bind to the ion. (iii) Mutations are introduced to the non-β-sheet residue positions of the amyloid designable scaffold's flexible structural templates, aiming to mimic how amino acids bind to the ion according to the defined materialphore models as well as to energetically stabilize the bound conformation of the pockets. (iv) Short MD simulations are performed to investigate the conformational properties of the most energetically favorable designed peptides determined in (iii), and a two-component validation procedure is applied to select the simulated designs with the highest probability to represent the corresponding materialphore models and thus form proper amino acid pockets capable of binding to the ion of interest. (v) REMD simulations according to Tamamis and Archontis' protocol[130] outlined in the Introduction are performed to investigate the top selected designed peptides from (iv) for their capacity to self-assemble into the anticipated elementary β-sheet structural units and re-assure, via the aforementioned two-component validation procedure, that the designed amino acid pockets are capable of binding to the ions of interest.

The developed computational protocol was successfully applied for the design of two peptide sequences, which according to experiments have cesium binding and deposition properties. The experiments were performed in Dr. Mitraki's lab and Dr. Jeong's experimental labs, and additional experimentalists which their names are listed below in the complete publication reference: "Jonnalagadda, S.V.R.; Kokotidou, C.; Orr, A.A.; Fotopoulou, E.; Henderson, K.J.; Choi, C.H.; Lim, W.T.; Choi, S.J.; Jeong, H.K.; Mitraki, A.; Tamamis, P. Computational Design of Functional Amyloid Materials with Cesium Binding, Deposition, and Capture Properties. *J. Phys. Chem. B* **2018**, *122(30)*, 7555-7568."

SVRJ and CK are equally contributing first authors.

6.2 Computational and Experimental Methods

A computational protocol was developed with the aim to design amyloid materials capable of binding to an ion of interest (i.e., cesium). The protocol consists of an optimization-based design model and computational validation procedure and was applied in the first stage of this study to functionalize amyloid materials for cesium ion binding. In the second stage of this study, the capacity of the designed peptides to form fibrils with cesium ion deposition properties was experimentally verified, followed by additional experiments investigating their ability to capture cesium ions in neutral and acidic aqueous conditions. Furthermore, additional simulations were performed investigating the cesium ion binding properties of modeled elementary β-sheet structural units of the designed amyloid materials. The computational and experimental methods are summarized below.

6.2.1 Elucidating the Elementary β-sheet Structural Unit of a Short Amyloid-Forming Peptide

We introduced the computational protocol developed by Tamamis and Archontis[130], described in chapter 2, analogously to the refs.[24,40,130,131,192], to investigate the amyloid peptide's self-assembly properties and elucidated its elementary β-sheet structural units the peptide with sequence NH_3^+-AGKGAIIGFIK-COO^-. We computationally investigated the self-assembly properties of the peptide by performing replica exchange molecular dynamics (REMD) simulations investigating 6 copies of a peptide in a 150 Å cubic periodic boundary condition box using CHARMM[110] and by performing a structural analysis of the simulated conformations to examine the gradual formation of β-sheet structures as in refs.[24,40,130,131,192]. For the REMD simulations, we employed a total of sixteen temperatures (290, 295, 300, 310, 305, 315, 321, 327, 333, 339, 345, 352, 359, 366, 373 and 380 K). The total simulation time for all temperatures combined was equal to 16 μs (1μs per temperature). We collected the final conformations produced

at 300 K, which were combined into one trajectory, corresponding to a length of 1μs, and 100,000 snapshots. We categorized the resulting β-sheet structures formed by the peptide into antiparallel, parallel and mixed conformations comprising 2- to 6- peptide strands and observed that the antiparallel arrangement is the most dominant. We focused our further analysis on 4-stranded antiparallel β-sheets as they are more highly populated and highly ordered than 5-, 6-stranded β-sheets and acquire higher complexity than 2-, 3-stranded β-sheets[24,131]. Using polar-P_1 and nematic-P_2 order parameters, we constructed a free energy landscape and extracted highly ordered and well-aligned antiparallel β-sheet structures from the global energy minima of the free energy landscape. Subsequently, we investigated the degree of solvent accessibility of the residues that are not involved in β-sheet interactions. The analysis showed that the 3 first N-terminal and 2 last C-terminal residue positions are exposed, and that the peptide forms linear-shaped β-sheets. Thus, we extracted the top ten highly ordered and well-aligned 4-stranded antiparallel β-sheet structures formed by the peptide, which are predicted to represent the peptide's self-assembled elementary β-sheet structural units. The ensemble of structures was used as input flexible structural templates of the amyloid designable scaffold formed by AGKGAIIGFIK in the optimization-based computational design model described below.

6.2.2 Uncovering How Amino Acid Motifs Recognize Cesium Ions Using Experimentally Resolved Protein Structures

The analysis of 3D structures of protein-ligand binding sites can provide valuable insights on how amino acids recognize ligands. The PDB[252] currently contains ~130,000 3D protein structures, providing an abundance of information on how amino acids can bind particular ions or compounds. This information can be used to functionalize proteins[253] or protein-based materials. We extracted, to the best of our knowledge, the complete set of X-ray resolved structures of protein

: cesium (ion) complexes from the PDB; date of extraction: 08/01/2016. Using FORTRAN programs, we collected amino acid : cesium ion binding motifs, which are defined here as a set of amino acids with side chain centers of mass within 8.5 Å of the resolved cesium ion, and their corresponding coordinates from each of the extracted X-ray resolved structures to analyze how amino acid motifs bind to cesium ions. Using the coordinates of the collected amino acid : cesium ion binding motifs, for each motif we stored the identities of the amino acids, and (i) their relative distances, as well as (ii) their distances to the cesium ion. Hereinafter, (i) and (ii) are referred to as primary and secondary materialphore models. Representation using the aforementioned distances can completely capture the geometry of an amino acid : cesium ion binding pocket. The primary materialphore model represents the geometry of the amino acids in the cesium ion binding site and was used for both computational optimization-based design modeling and validation purposes. The secondary materialphore model represents how a cesium ion interacts with the amino acids in the binding pocket and was used only for computational validation purposes. If, in an X-ray resolved structure, the number of amino acid : cesium binding motifs was more than one, then the different binding motifs were considered as independent materialphore models. For example, the structure of a thermotoga maritima CorA protein (PDB ID: 4EEB)[254] contains 9 cesium ions, resulting in 9 separate extracted materialphore models. In the computational design model all primary materialphore models were considered while searching for optimum solutions, while both the primary and secondary materialphore models were used to validate the computationally designed sequences. It is worth noting that the number of amino acids interacting with a cesium ion in each binding motif may vary.

The programs for extracting materialphore models for an ion (current study) or compound; (see further studies) of interest were initially developed by Chang-Hyun Choi, an undergraduate

student of Dr. Tamamis' lab, and were further advanced for the purpose of the current study and subsequent studies.

6.2.3 Optimization-Based Design Model Stabilizing and Functionalizing the Amyloid Designable Scaffold to Bind to Cesium Ions

A computational design model was developed that designs the sequence of β-sheet forming peptides, so that cesium ions can bind to pockets consisting of non-β-sheet forming amino acids belonging to a pair of β-sheet bonded peptides. The ten obtained flexible structural templates of the amyloid designable scaffold, and the primary materialphore models collected from the X-ray resolved amino acid : cesium ion complex structures deposited in the PDB[252] (Figure 1B) were used as inputs to the design model. The computational design model is formulated as follows:

$$\min_{y_i^j, y_k^l, g(s)^j_{\mathrm{PDB_{PMM}}}, w(s)^{jl}_{ik}} \frac{1}{f}\left(\left(\sum_{a=1}^{p}\sum_{b=1}^{p}\sum_{i=1}^{n}\sum_{j=1}^{m_i}\sum_{\substack{k=1\\k\geq i}}^{n}\sum_{l=1}^{m_k}\sum_{s=1}^{f} E^{jl}_{ikab}\, y_i^j y_k^l w(s)^{jl}_{ik}\right)+\left(\sum_{a=1}^{p}\sum_{i=1}^{n}\sum_{j=1}^{m_i}\sum_{s=1}^{f}\gamma\ (\mathrm{SASA}(s)^j_{ia}) y_i^j\right)\right) \quad 6.1$$

s.t

$$g(s)^j_{\mathrm{PDB_{PMM}}}\left|D(s)^{jl}_{ikab} - D(\mathrm{PDB_{PMM}})^{jl}\right| \leq \ \mathrm{cutoff}_i \ \forall\ i,j,k,l,a,b,\mathrm{PDB_{PMM}} \quad 6.2$$

$$\left|D(s)^{jl}_{ikab} - D(\mathrm{PDB_{PMM}})^{jl}\right| \geq \left(1 - g(s)^j_{\mathrm{PDB_{PMM}}}\right)\mathrm{cutoff}_i \ \forall\ i,j,k,l,a,b,\mathrm{PDB_{PMM}} \quad 6.3$$

$$\Psi_{ikab\mathrm{PDB_{PMM}}} = 1 \ \forall\ i,a,k,b \in (S_1 \cup S_2), \forall\ \mathrm{PDB_{PMM}} \quad 6.4$$

$$\Psi_{ikab\mathrm{PDB_{PMM}}} = 0 \ \forall\ i,a,k,b \notin (S_1 \cup S_2), \forall\ \mathrm{PDB_{PMM}} \quad 6.5$$

$$g(s)^j_{\mathrm{PDB_{PMM}}} \leq \Psi_{ikab\mathrm{PDB_{PMM}}} \ \forall\ i,j,k,l,a,b,\mathrm{PDB_{PMM}} \quad 6.6$$

$$d^{j}_{\mathrm{PDB_{PMM}}} = \sum_{s=1}^{f} g(s)^{j}_{\mathrm{PDB_{PMM}}},\ d^{j}_{\mathrm{PDB_{PMM}}} \geq \lambda \frac{p(p-1)}{2} f\ \ \forall\ \ j, \mathrm{PDB_{PMM}} \qquad 6.7$$

$$\mathrm{PDB_{PMM}} \in \{1,\ldots,t_{\mathrm{PMM}}\},\ \ g(s)^{j}_{\mathrm{PDB_{PMM}}} \in \{0,1\},\ \ \Psi_{ikab\mathrm{PDB_{PMM}}} \in \{0,1\}\ \ \forall\ \ i,j,k,l,a,b,\mathrm{PDB_{PMM}} \qquad 6.8$$

$$\sum_{i=1}^{v} \mathrm{cutoff}_i \leq 17,\ \ \mathrm{cutoff}_i \in \{2.5, 3.5\}\ \ \forall\ \ i \qquad 6.9$$

$$S_1 = \left\{ \begin{array}{l} a = b = c \forall \left(k, i \in \{\phi_1,..,\phi_r\}\right) \\ a = b = c + 1 \forall \left(k, i \in \{\phi_{r+1},..,\phi_v\}, c < p\right) \\ a = c, b = c + 1 \forall \left(k \in \{\phi_{r+1},..,\phi_v\}, i \in \{\phi_1,..,\phi_r\}, c < p\right) \end{array} \right\} \qquad 6.10$$

$$S_2 = \left\{ \begin{array}{l} a = b = c \forall \left(k, i \in \{\phi_1,..,\phi_r\}\right) \\ a = b = c - 1 \forall \left(k, i \in \{\phi_{r+1},..,\phi_v\}, c > 1\right) \\ a = c, b = c - 1 \forall \left(k \in \{\phi_{r+1},..,\phi_v\}, i \in \{\phi_1,..,\phi_r\}, c > 1\right) \end{array} \right\} \qquad 6.11$$

$$\sum_{j=1}^{m_i} y_i^j = 1 \forall i,\ \ y_i^j, y_k^l \in \{0,1\} \qquad 6.12$$

$$k \neq i\ \forall\ a = b,\ i \in \{\phi_1, \phi_2, ..., \phi_v\},\ l = j\ \forall\ (k \in \{\phi_1, \phi_2, ..., \phi_v\}, k = i, a \neq b),\ b > a \forall\ k = i,\ c \in \{1,..,p\} \qquad 6.13$$

$$w(s)^{jl}_{ik} D(s)^{jl}_{ikab} \leq \mathrm{tol},\ \ D(s)^{jl}_{ikab} \geq \left(1 - w(s)^{jl}_{ik}\right) \mathrm{tol},\ \ w(s)^{jl}_{ik} \in \{0,1\}\ \ \forall\ \ i,j,k,l,a,b \qquad 6.14$$

The objective function (*Eq. 6.1*) minimizes the energy of an ideal amyloid designable scaffold, represented by an ensemble of flexible structural templates (elementary β-sheet structural units), subject to the constraints (*Eq. 6.2-6.14*) by introducing amino acid substitutions at the designable positions (*Eq. 6.12-6.14*) in a coarse grained fashion, which simultaneously allow for the functionalization of cesium ion binding to be satisfied (*Eq. 6.2-6.11*); the energy accounts for amino-acid : amino-acid interactions and the desolvation free energy of nonpolar surfaces, which, is calculated from the change in solvent-accessible surface area (ΔSASA) as in ref.[188] The

constraints introduced for functionalization purposes guarantee that all possible combinations of distances between the centers of mass of the designed amino acid side chains in the flexible structural templates can be matched with any combination of distances between the centers of mass of the same amino acid side chains in any of the primary materialphore models (*Eq. 6.2-6.11*). The matching between the distances of the introduced amino acids with reference to the corresponding distances of the primary materialphore models is carried out by *Eq. 6.2-6.3*. These equations allow matching to occur only if the relative difference between the two sets of the compared distances is below certain cutoff values in at least a fraction λ (=0.05) of the total number if designable binding pockets available (six pockets per flexible structural template) within the 10 flexible structural templates representing the amyloid designable scaffold. When defining the cutoff values, we considered (i) the large flexibility of the designable binding pocket, (ii) the fact that the aforementioned compared distances cannot match exactly, and (iii) the fact that the sizes of the introduced amino acids which are placed on the flexible structural templates can also be different from the sizes of the original amino acids of the amyloid designable scaffold. Thus, we allowed a 2.5 Å maximum deviation to occur for four out of the six compared distances, and a more generous 3.5 Å maximum deviation to occur for any of the two remaining compared distances; this is defined by *Eq. 6.9*. In addition, upon tuning and testing higher and lower values of λ, the specific value was chosen to account for the significantly large conformational variability of the designable amino acid side chains (belonging to non-β-sheet forming residues) in the flexible structural templates (≈8.2 Å); larger values of λ were not preferred as they yielded a significantly lower number of designed peptides to be further investigated.

The term $E^{jl}{}_{iakb}$ represents the pairwise interaction energies between amino acid j at position i and amino acid l at position k (taken from the SIPPER force field[205]). Position i is the

modified residue position within strand *a*. Amino acid *j* is one of 20 amino acids, m_i, corresponding to the amino acid occupying position *i*. Position *k* is the residue position within strand *b* interacting with any residue position *i* ($k \neq i$ if $a=b$). Amino acid *l* is one of the 20 amino acids, m_k, corresponding to the amino acid occupying position *k*. Positions *i* and *k* may belong to the same peptide strand (representing intra-molecular interactions) or different peptide strands (representing inter-molecular interactions). The term γ ($SASA(s)^j_{ia}$) accounts for protein-water interactions, where $SASA(s)^j_{ia}$ is the solvent accessible surface area of amino acid *j* at position *i* in strand *a* at structure *s* and γ (=0.001 kcal/mol·A^2) is the surface tension coefficient; this term was used analogously to ref.[192] Structure *s* corresponds to one of the ten flexible structural templates of AGKGAIIGFIK, comprising elementary β-sheet structural units of the amyloid designable scaffold. $SASA(s)^j_{ia}$ is calculated as the theoretical SASA of the introduced amino acid[188] minus the buried surface area of the native amino acid in the original flexible structural template *s* (as in ref.[192]). *n* (=11) is the total number of residue positions per strand and *p* (=4) is the total number of peptide strands. The pairwise interaction energy calculated for each structure, *s*, was summed and averaged over the total number of flexible structural templates, *f* (=10). Binary variables y^j_i and y^l_k are equal to one if position *i* and *k* are occupied by amino acids *j*, *l*, respectively, and zero otherwise. The binary variable $w(s)^{jl}_{ik}$ equals one if $D(s)^{jl}_{iakb}$, the distance between side chain center of mass of amino acid *j* at position *i* in strand *a* and amino acid *l* at position *k* in strand *b* of structure *s,* is less than a specific distance, *tol* (=6.5 Å), and zero otherwise. d^j_{PDBPMM} is an integer variable indicating if a primary materialphore model can be satisfied in at least λ (=0.05) fraction of the total number of binding pockets in the flexible structural templates.

$D(\mathrm{PDB_{PMM}})^{jl}$ is the distance between the side chain center of mass of amino acids *j* and *l* in a primary materialphore model, $\mathrm{PDB_{PMM}}$. Amino acids *j* and *l* in the designed peptide are

identical to the amino acids *j* and *l* of the corresponding amino acid : cesium ion binding motif. *c* is a dummy index that indicates the strand which residue positions *i* or *k* belong to. *v* (=4) is the total number of residue positions amenable for modification. *r* (=2) is the number of designable positions at the N-terminal end of the flexible structural templates, while *v-r* is the number of designable positions at the C-terminal end. $\{\phi_1, \phi_2, \ldots, \phi_v\}$ is a set of residue positions that are amenable for modification in the peptide under the condition that $\phi_1<\phi_2<\ldots<\phi_v$. In the present study, positions 2, 3, 10, 11 are designable, i.e. $\{\phi_1=2, \phi_2=3, \phi_3=10, \phi_4=11\}$. Alanine at position 1 of the amyloid designable scaffold was replaced with an acetylation in the designed peptides, which aimed to enhance the designed peptides' self-assembly properties[32]. $\Psi_{iakb\text{PDBPMM}}$ is a binary variable ensuring that a designed binding pocket is formed by amino acids belonging to sets S_1 and S_2. Sets S_1 and S_2 define the combination of amino acids at designable positions that make up a designed binding pocket. $g(s)^j{}_{\text{PDBPMM}}$ is a binary variable indicating if a primary materialphore model can be satisfied in a designed binding pocket of a structure *s*. t_{MM}(=436) is the total number of primary materialphore models.

The computational design model was solved in Fortran. Of the 20^4 possible designs, the functionalization constraints (*Eq. 6.2-6.11*) reduced the feasible set of designs to 89442. For the resulting designs, the energy defined in the objective function (*Eq. 6.1*) was calculated and the designs were rank ordered based on this energy. Imposing the additional constraints (described in *Results and Discussion*) specific to cesium ion capture resulted in a total number of 1331 of possible designs. From these reduced set of designs, we selected the top 10% (total of 134 sequences) designs for further investigation.

6.2.4 Modeling, Simulation and Computational Validation of Selected Designed Peptides Predicted to Form Functional Amyloid Fibrils Binding to Cesium Ions

MD Simulation-Based Modeling of the Designed Amino Acids on the Original Designable Scaffold and Computational Screening-Validation of the Selected Designed Peptides

The initial structures of the elementary β-sheet structural units of the selected designed peptides were modeled using one of the ten flexible structural templates as the initial conformation with the designed amino acids introduced using CHARMM[110]. Water was modeled implicitly using the FACTS19 solvation model[123], and the value of the surface tension coefficient was equal to 0.015 kcal/(mol·Å^2). The elementary β-sheet structural units of each of the selected designed peptides were subjected to 100 steps of steepest descent minimization with 100 kcal/(mol·Å^2) harmonic constraints on all non-hydrogen atoms and 50 steps of Adopted Basis Newton-Raphson minimization with 30 kcal/(mol·Å^2) harmonic constraints on all non-hydrogen atoms. An additional 50 steps of steepest descent minimization with 10 kcal/(mol·Å^2) harmonic constraints on all backbone atoms and subsequently 50 steps of steepest descent minimization with 5 kcal/(mol·Å^2) harmonic constraints on all backbone atoms were performed.

The simulation systems for each of the designed peptides were subsequently equilibrated in two stages. In the first stage, each designed peptide system was simulated for 0.04 ns with 5 kcal/(mol·Å^2) harmonic constraints on the backbone atoms of the amyloidogenic core (GAIIG motif). In the second stage, each designed peptide system was simulated for 0.04 ns with 3 kcal/(mol·Å^2) harmonic constraints on the backbone atoms of the amyloidogenic core. Subsequently, the designed peptide systems entered production stage with 1.5 kcal/(mol·Å^2) harmonic constraints on the backbone atoms of the amyloidogenic core. Each of the designed peptide systems were simulated for 20 ns with snapshots extracted every 10 ps. All constraints

imposed during the equilibration and production stages were applied using the bestfit module in CHARMM[110]. We used Langevin dynamics with a 5 ps^{-1} friction coefficient introduced on all non-hydrogen atoms of each peptide and a temperature of the heat bath set to 300 K. All simulations were performed using CHARMM[110]. Upon the completion of the MD simulation runs, the 20 ns simulations for each designed peptide, containing 2,000 snapshots each, was subjected to validation procedure described below.

Upon the completion of the MD simulation runs, we computationally determined whether the designed peptides can form binding pockets consisting of non-β-sheet forming amino acids which are part of a pair of β-sheet bonded peptides resembling the geometry of the amino acid : cesium ion binding motifs using a two component validation procedure. In the first validation component, we determined whether the center of mass of the designed amino acid side chains form appropriate side chain conformations to form pockets with geometries resembling the primary materialphore model(s). Each elementary β-sheet structural unit of a designed peptide comprises four peptide strands forming six, non-identical designed pockets. Then, we compared the distances between the four centers of mass of the designed amino acid side chains to the corresponding distances of the same amino acids within the primary materialphore model, from which the design originated from, for each of the 2000 simulation snapshots produced in each simulation. Designed pockets with four out of six distances falling within 2.0 Å and two out of six distances within 3.0 Å of the corresponding distances of the amino acid : cesium ion binding motif were considered primary matches. If a primary match with any of the six pockets occurred, then the snapshot and its corresponding pocket was saved for further analysis. In the second component of the validation procedure, the ability of the designed amino acid side chain pockets to encapsulate a cesium ion in the same fashion as the secondary materialphore model(s) from which the designed amino acids

originated was evaluated. The evaluation was done only for the pockets containing primary matches. We used the distances between the cesium ion and the centers of mass of the amino acid side chains in the secondary materialphore model from which each design originated from as a guide to place the cesium ion in the designed binding pocket. Specifically, the cesium ion was placed in such a way to minimize the difference between the distances of the cesium ion to the centers of mass of the designed amino acid side chains in comparison to the corresponding distances in the secondary materialphore model it originated from. If the difference between each of the compared distances was within 2.0 Å, then the designed cavity was considered a secondary match. To avoid steric clashes of the placed cesium ion and any other atom of the designed peptides, pockets containing secondary matches in which the distance between the cesium ion and any peptide in the snapshot under investigation less than 2.0 Å were not considered as successful validations. The designed peptides were ranked based on their first validation stage percentages and the designed peptides with second validation stage percentages less than 40% were removed from the rank-ordered list of designs. From the reduced list of designs, the designed peptides with the top 10% highest first validation stage percentages (13 designed peptides) underwent further analysis through REMD simulations investigating their self-assembly properties and subsequent structural analysis described below.

MD Simulations Investigating the Self-Assembly Properties of the Designed Peptides and Final Computational Screening-Validation

The selected designed peptides were subsequently investigated through REMD simulations to examine their self-assembly properties and determine if the designed peptides were likely to form the desired antiparallel β-sheet fibrils while forming designed binding pockets that match both primary and secondary materialphore models for cesium ion binding properties. The REMD

simulations were performed in CHARMM[110] through the Tamamis and Archontis protocol[130] described in chapter 2 which was also applied in this study to elucidate the β-sheet elementary structural unit of the amyloid designable scaffold, AGKGAIIGFIK. We performed 13 independent REMD simulations for each of the 13 designed peptides. Analogously to the simulations performed initially to investigate the self-assembly properties of AGKGAIIGFIK, the total duration of each REMD run was equal to 16 μs (which corresponds to 16 replicas of 1 μs duration each per investigated peptide). The corresponding 1 μs trajectories collected at 300 K for each of the 13 designed peptides were collected for further analysis. β-sheet conformations comprising 2- to 6- peptides strands were categorized into antiparallel, parallel, and mixed β-sheet structures. Designed peptides predicted to favor primarily parallel β-sheet formation over antiparallel β-sheet formation were immediately discarded from further investigation. The 500 most highly ordered and well-aligned antiparallel β-sheet structures from the remaining designed peptides were extracted, and the same two-component validation procedure described above was performed on the extracted structures.

6.2.5 Experimental Testing, Verification of Amyloid Self-Assembly, and Determination of Cesium Ion Capture Properties by the Designed Amyloid Materials

All microscopic analyses to verify the self-assembly properties and cesium deposition properties of the designed peptides were performed by Dr. Mitraki's lab. Elemental analysis experiments were done by Dr. Bryan Tomlin through the elemental analysis facilities of Texas A&M University using amyloid materials and solutions prepared by the student and Asuka A. Orr prepared in Dr. Tamamis' and Dr. Karim's labs under the guidance of Dr. Hae-Kwon Jeong (see ref[22]).

Amyloid Materials Synthesis

The peptides NH_3^+-AGKGAIIGFIK-COO^-, Ace-FQGAIIGFNE-$CONH_2$, and Ace-FNGAIIGFQE-$CONH_2$ were custom-synthesized by WuXi AppTec with purity over 95%. Each synthetic lyophilized peptide powder was dissolved in double distilled water (pH 7) to a concentration of 12 mg/ml and incubated at room temperature for 3 days. Fiber formation was confirmed with Field Emission Scanning Electron Microscopy (FESEM) and Transmission Electron Microscopy (TEM). The complete solubilization of the peptide powders into aqueous solution for increased concentrations up to 20 mg/ml was observed. For all tested concentrations the peptide powder becomes clear instantaneously upon addition to water, indicating the complete dissolving of the peptide powder. Increased blurring of the dissolved peptide solution was observed indicating the immediate fibril formation. At the concentrations (12 mg/ml) used for our cesium capture experiments, only fibril formation, and not amorphous aggregation, was observed after careful scanning of the area of the sample with the use of electron microscopy.

Transmission Electron Microscopy (TEM)

8μl of each diluted 1:10 sample was deposited directly onto formvar/carbon coated electron microscopy grids (Agar Scientific) for 2 min. After 2 min incubation, the excess was removed with filter paper, an additional wash with 8 μl of double distilled water was performed and the sample was stained with 2% wt/vol uranyl acetate for 2 min. For the samples treated with CsCl no negative staining was used, therefore the contrast afforded is only due to the cesium deposition. Specimens were examined in a JEOL JEM-2100 Transmission Electron Microscope at an accelerating voltage of 80 kV and 200kV.

Field Emission Scanning Electron Microscopy (FESEM)

A sample solution (10 μL, diluted 1:10) was deposited on a circular cover glass and left to dry in air overnight. Samples were subsequently covered with 10 nm Au/Pd sputtering and observed directly. Experiments were performed using a JEOL JSM-7000F microscope operating at 15 kV.

Congo Red Staining

Each solution of the fibrils was thoroughly mixed with the Congo Red assay solution (10 mM Congo Red, 2.5 mM NaOH in 50% ethanol) in a ratio of 6:1. 10 μL of the mixture was deposited on a glass coverslip to be dried at room temperature. Stains were examined with a Zeiss Stemi 2000-C microscope with and without the use of a crossed polarizer.

Elemental Analysis

Using the same procedure described above (see *Amyloid Materials Synthesis*) we synthesized the amyloid materials for the investigated peptides at pH 7. After 3 days of incubation allowing the peptides to self-assemble, the amyloid fibrils were collected by centrifugation at 13300 rpm for 30 min at 25ºC. The collected amyloid fibrils were subsequently resuspended in a 10 ppm cesium chloride aqueous solution (pH 7) to a concentration of 12 mg/ml. The same procedure was also conducted using a 10 ppm cesium chloride, 10 mM boric acid solution of pH 4.5 to resuspend the amyloid fibrils. The resuspended amyloid fibrils were subjected to vigorous shaking for 15 hours and subsequently allowed to settle for 6 hours. The amyloid fibrils were separated from the supernatant through a total of 4 centrifugation cycles at 13300 rpm for 30 min at 25ºC. During the last centrifugation cycle, 0.22 μm nylon filters were used to further improve the separation of the fibrils from the supernatant solution. The supernatant solution was collected

and submitted for cesium elemental analysis. The elemental analysis was conducted on a PerkinElmer NexION 300D ICP-MS instrument which analyzed the concentration of cesium in the supernatant solution. The percentage of cesium ions removed from the 10 ppm cesium chloride solution by the amyloid fibrils were calculated using *Eq. 6.15*:

$$(\%)\ \text{Cs ions removed} = \frac{C_i V_i - C_f V_f}{C_i V_i} \times 100 \qquad 6.15$$

C_i and V_i are the initial concentration (10 ppm) and initial volume of the cesium chloride solution in which the amyloid fibrils were resuspended. C_f and V_f are the final concentration and the final volume of the supernatant collected.

6.2.6 Explicit-Solvent MD Simulations Investigating the Cesium Ion Binding Properties of the Designed Amyloid Materials

Explicit-solvent, all-atom MD simulations in presence of cesium ions were performed using CHARMM[110] to investigate cesium ion interactions with an elementary structural β-sheet of the designed peptides FNGAIIGFQE and FQGAIIGFNE, as well as peptide AGKGAIIGFIK, which served as an amyloid designable scaffold to derive the first two peptides. We performed five independent runs of 10 ns MD simulations for each of the three systems under investigation, with snapshots collected every 2.5 ps. Thus, the total number of simulation snapshots in each of the three systems was equal to 20,000. The designed peptides N- and C- termini were acetylated and amidated, whereas the amyloid designable scaffold peptide's termini were not modified, in line with the previous analysis.

The initial structure of the explicit-solvent MD simulations for each of the investigated peptides was extracted from the ensemble of highly ordered and well-aligned antiparallel β-sheet structures produced by the REMD simulations described above. All explicit-solvent MD simulations were performed using CHARMM36 topology and parameters[255]. The MD simulations were performed using the Drude polarizable force field[256,257,258] and the velocity verlet 2 (VV2) integrator. Each of the peptide fibril systems was solvated in pre-equilibrated water box with sides 17.0 Å greater than the maximum dimension of the isolated fibril. The cesium chloride concentration was set to 0.5 M for each of the peptide fibril systems with additional cesium ions introduced to neutralize the systems. We used a higher concentration to enable the computational study of the cesium ion binding mechanism of the designed fibrils. The cesium and chloride ions were introduced through 2000 steps of Monte Carlo simulations. Solvent molecules underwent energy minimizations through 100 steps of steepest descent, 100 steps of Adopted Basis Newton-Raphson, and 100 steps of steepest descent minimization. An additional 100 steps of steepest descent, 100 steps of Adopted Basis Newton-Raphson, and 100 steps of steepest descent minimization were performed on each system with all backbone atoms constrained under 1.0 kcal/(mol·Å^2) and all non-hydrogen side chain atoms constrained under 0.1 kcal/(mol·Å^2). The simulation systems were equilibrated for 0.25 ns, during which all protein backbone atoms were constrained under 1.0 kcal/(mol·Å^2) and all side chain atoms were constrained under 0.5 kcal/(mol·Å^2). Subsequently, all constraints were released, and the systems were simulated for 10 ns with simulation snapshots extracted every 2.5 ps. The simulations were performed in five independent runs using different initial seeds for reproducibility purposes. The dual Nosé-Hoover thermostat was used to maintain the temperature of the systems at 300K and the Andersen-Hoover barostat was used to maintain the systems at a constant pressure of 1.0 atm. All simulations were

performed using periodic boundary conditions and the SHAKE algorithm to constrain covalently bonded hydrogens.

In the simulations investigating cesium ions binding to AGKGAIIGFIK, we considered that a cesium ion or a chloride ion is interacting with an amino acid if the distance between the center of mass of the amino acid side chain and the cesium ion or chloride ion is within 6.5 Å. In the simulations investigating cesium ions binding to FQGAIIGFNE and FNGAIIGFQE, we considered that a cesium ion is interacting with four amino acid side chains simultaneously if the four distances between the center of mass of the amino acid side chains and a cesium ion are within 6.5 Å. In the analysis of all the three simulations, we neglected glycine amino acids which do not contain any side chain atoms.

6.3 Results and Discussion

6.3.1 Elucidating the Elementary β-sheet Structural Unit of a Short Amyloid-Forming Peptide

An important factor for the design of functional amyloid materials is the choice of an amyloid scaffold based on which mutations will be introduced to yield functionalization. As mutations functionalizing an amyloid scaffold should not interfere with its self-assembling properties, confining the mutations to residue positions outside its β-sheet core is considered an attractive solution. This can be approached by adding extensions to an amyloid scaffold, which can serve as functional motifs. Yet, attention should be paid as such extensions to the termini may elongate the β-sheet domain of the peptide, disrupt self-assembly or result in a change of conformation (e.g., U-shaped β-sheets). We consider that an ideal "I"-shaped amyloid scaffold for the design of functional amyloid materials should encompass (i) a key central β-sheet region, and (ii) at least 2 non-β-sheet residue positions at both termini which can be mutated to yield a certain function (Figure 6.1A).

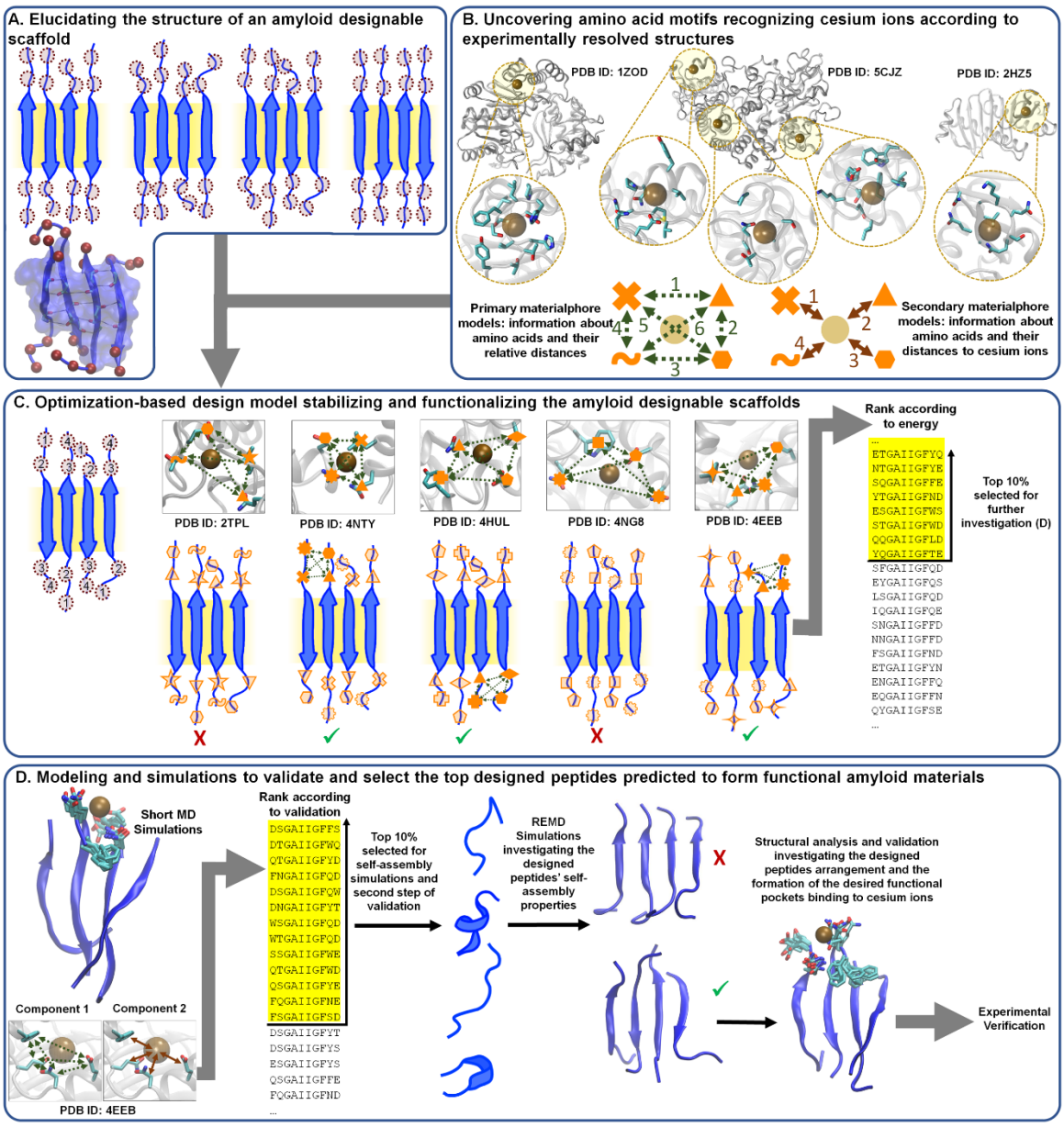


Figure 6.1 Overview of the computational protocol used for the design of amyloid materials capable of binding to cesium ions. (A) Illustrations of highly ordered and well-aligned structures extracted from the replica exchange molecular dynamics (REMD) simulations of the amyloid designable scaffold and a molecular graphics image of a representative elementary β-sheet structural unit representing a flexible structural template of AGKGAIIGFIK. In the top section of panel (A), the backbone of the peptides is shown as cartoons; the key central β-sheet region is highlighted in yellow; the first and last two mutable positions are shown using maroon dotted circles. In the bottom section of panel (A), the peptides are shown in blue cartoon representation with the backbone atoms forming β-bridges shown in licorice representation; β-bridge-associated hydrogen bonds are shown using black dashed lines. Residues 4–9 form amyloid-zipper-like patterns and are shown in transparent surface representation. Residues 1, 2, 3, 10, and 11 are not involved in β-sheet formation and their Cα atoms are shown as maroon spheres; these residue

Figure 6.1 Continued

positions are deemed as mutable positions. (B) Examples of amino acid–cesium ion binding motifs from experimentally resolved structures and a schematic representation of the extracted materialphore models. Each protein–cesium ion complex is shown in a transparent silver cartoon representation with bound cesium ions shown in opaque ochre spheres. Amino acids in the extracted amino acid–cesium ion binding motifs, encircled by dotted ochre lines, are shown in a licorice representation. The bottom section of panel (B) shows an illustration of the primary and secondary materialphore models. Distances, indicated using green dotted arrows, between the extracted amino acids, indicated using orange symbols, constitute information encapsulated in primary materialphore models. Distances, indicated using maroon arrows, between the extracted cesium ion, indicated by a beige circle, and each of the amino acids, indicated using orange symbols, constitute information encapsulated in secondary materialphore models. The structures produced in (A) and the materialphore models produced in (B) are provided as input to the optimization-based design model stabilizing and functionalizing the amyloid designable scaffold to bind to cesium ions, which is presented in panel (C). On the far-left section of panel (C), a designable amyloid scaffold is presented with its four mutable positions marked in maroon dotted circles. If the geometry of certain amino acid combinations in the primary materialphore models, indicated using orange symbols in the molecular graphics images of the Protein Data Bank (PDB) structures, can be matched to the geometry of a flexible structural template, indicated by solid orange symbols and green dotted arrows, then a green checkmark is indicated; otherwise, red crosses are indicated. For the matched amino acid combinations, the introduction of the amino acids to the scaffold is evaluated by the energy defined in the objective function. The designed peptides are ranked according to their energy, and the top 10%, highlighted in yellow, are selected for computational validation (shown in panel (D)). (D) Two-step, two-component computational validation procedure comprising two steps of simulations and structural analysis. In the first step, the designed amino acids are introduced to an elementary structural β-sheet unit and are sampled using short MD simulations. In the far-left section of (D), sampled conformations of the designed amino acids from the short MD simulation snapshots are shown; the designed amino acids are shown in licorice representation, and peptides are shown in blue cartoon representation. The two-component validation procedure comparing the geometry of the designed amino acids to that of the primary (green dotted arrows) and secondary (solid maroon arrows) materialphore models from the PDB (for example, as shown, PDBID: 4EEB) evaluates the designs. The top designed peptides based on a two-component validation procedure are selected. In the second step, the selected top designs are investigated further using longer REMD simulations assessing the peptides' self-assembly properties. In the far-right section of (D), sampled conformations of the designed amino acids collected from the REMD simulations are shown. Subsequent structural analysis is used to assess the designed peptide ability to (i) self-assemble into the proper arrangement that they were designed based on (antiparallel is check marked in green, whereas parallel is crossed-out in red), and (ii) form of the desired binding pockets for cesium ion binding. Selected top designed peptides are chosen for experimental verification. In both steps (short MD simulations and REMD simulations investigating the peptides' self-assembly properties), a two-component validation procedure is performed. In the first component (bottom left section of panel (D)), the distances between the modeled designed amino acids are compared to the distances of the primary materialphore models they originated from, indicated by the green dotted arrows in

Figure 6.1 Continued

the molecular graphics image of the amino acid–cesium ion binding motif. In the second component (bottom left section of panel (D)), the distances between the cesium ion and the modeled designed amino acids are compared to the distances of the secondary materialphore models they originated from, indicated by maroon arrows in the molecular graphics image of the amino acid–cesium ion binding motif. Reprinted with the permission from ref.[22]

Our studies suggest that naturally occurring sequences comprising a central amyloid GAIIG motif extracted from amyloid-forming peptides (e.g., Aβ) or proteins with regions rich in β-sheet content (e.g., α-tubulin acetyl transferase[242] or the HIV-1 gp120 V3 loop) are a promising starting point to extract amyloid designable scaffolds. The selection of such peptide sequences with 2-3 amino acids outside the GAIIG amyloid core at both termini can potentially result in the formation of amyloid designable scaffolds, as these amino acids can maintain the linear conformation of the peptide while not participating in β-sheets (i.e., not elongating the peptide's amyloid core). This can be attributed to the fact that glycine amino acids of GAIIG may act as β-turn promoters[234,235,236]. Motivated by our recently published study showing that the peptide with sequence YATGAIIGNII[23], extracted from an HIV-1 gp120 V3 loop, self-assembles into amyloid designable scaffolds, in this study we selected an additional peptide with sequence NH_3^+-AGKGAIIGFIK-COO^- (AGKGAIIGFIK), extracted from the β-sheet domain (residues 88 to 98) of the α-tubulin acetyl transferase protein (αTAT1 with PDB code: 4PK2)[242], to examine its capacity to self-assemble into amyloid designable scaffolds using computational and experimental methods. We considered that the presence of a phenylalanine residue sequentially after the GAIIG motif can potentially further enhance self-assembly[31] and fibril stability[192]. The analysis depicted that AGKGAIIGFIK self-assembled into highly-ordered and well-aligned antiparallel, "I"-shaped (non-U-shaped) β-sheets with the first 3 N-terminal and last 2 C-terminal residue positions being

outside the amyloid core and thus exposed and amenable for further functionalization (Figure 6.1A). Thus, the peptide can self-assemble into amyloid designable scaffolds, of which residue positions 1, 2, 3, 10, and 11 are mutable and can be modified to enhance functionalization. Subsequent experiments showed that the AGKGAIIGFIK peptide forms distinctive fibrillar morphologies with widths in the range of 10-20 nm as observed under FESEM (Figure 6.2). Thus, from the 16 μs REMD simulations, we extracted the top ten highly-ordered and well-aligned 4-stranded antiparallel β-sheets formed by the peptide from the 1 μs replicas sampled at 300 K, representing flexible structural templates of its elementary β-sheet structural unit (similarly to refs[23,24,40,131,192]). The selected flexible structural templates were subsequently used as inputs to the optimization-based design model that was developed in this study for the functionalization of amyloid materials. A representative flexible structural template of the amyloid designable scaffold is presented in Figure 6.1A.

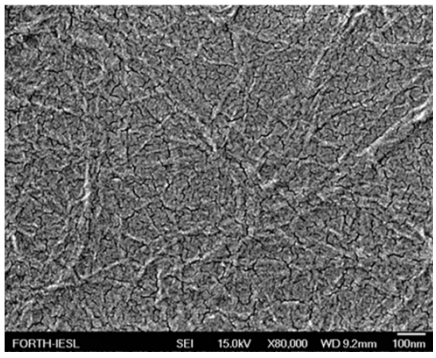


Figure 6.2 FESEM picture of the self-assembled fibrils of the peptide AGKGAIIGFIK after incubation in double distilled water for 3 days. Figure was produced by our experimental collaborators and is reprinted with the permission from ref.[22]

6.3.2 Uncovering How Amino Acid Motifs Recognize Cesium Ions Using Experimentally Resolved Protein Structures

We extracted all available 106 X-ray resolved protein : cesium ion complexes containing at least one cesium ion deposited in the PDB[252] (date extracted: 08/01/2016). From the extracted protein structures, we collected amino acid : cesium ion binding motifs, defined here as a set of amino acids with side chain centers of mass within 8.5 Å of the resolved cesium ion. Examples of experimentally resolved protein : cesium ion complex structures and their amino acid : cesium ion binding motifs are presented in Figure 6.1B. Each of the amino acid binding motifs has a minimum of four amino acids as there are four mutable positions within each of the peptide strands of the flexible structural templates (described in the following section). Some of the extracted protein : cesium ion complexes have multiple cesium ions bound within the resolved protein, resulting in multiple amino acid : cesium ion binding motifs from the resolved complex. For example, the protein structure with PDB ID: 5CJZ, presented in Figure 6.1B, contains three cesium ions, thus three amino acid : cesium ion binding motifs were extracted from this structure. As a result, a total of 436 amino acid : cesium ion binding motifs were extracted from the 106 X-ray resolved structures. For each of the collected amino acid : cesium ion binding motifs, we stored the identities of the amino acids as well as their (i) distances to other amino acids within the motif and (ii) distances to the resolved cesium ion. (i) and (ii) are indicated in dotted green double headed arrows and solid maroon double headed arrows, respectively, in Figure 6.1B. Similar to pharmacophore models describing the structural features necessary for the molecular recognition of a ligand by a macromolecule in drug design[259,260,261,262], here we refer to (i) and (ii) as primary and secondary materialphore (material + gr. phérō φέρω) models respectively, which combined can capture the geometric properties of amino acid : cesium ion binding motifs. Thus, 436 primary and secondary

materialphore models were extracted to be subsequently used for design purposes under the postulation that amyloid materials bearing mutations, which encompass the geometric properties captured by both primary and secondary materialphore models, can be functional to bind to cesium ions. Based on our definitions, primary materialphore models represent the geometry of the amino acid binding site and are used in both the optimization-based design model and computational validation procedure. Secondary materialphore models represent the distances of cesium ions with the interacting amino acids, reflecting how amino acids interact with cesium ions, and are used only in the computational validation procedure.

The programs for extracting materialphore models for an ion (current study) or compound; (see further studies) of interest were initially developed by Chang-Hyun Choi, an undergraduate student of Dr. Tamamis' lab, and were further advanced for the purpose of the current study and subsequent studies.

6.3.3 Optimization-Based Design Model Stabilizing and Functionalizing the Amyloid Designable Scaffold to Bind to Cesium Ions

Using the designable scaffolds (Figure 2.1A) and primary materialphore models collected from the X-ray resolved protein : cesium ion complex structures deposited in the PDB[252] (Figure 2.1B) as inputs, we developed and introduced the optimization-based design model described in *Methods*. The model was developed to functionalize the amyloid designable scaffold to bind a specific ion (cesium) by introducing mutations at the mutable positions of the amyloid designable scaffold's flexible structural templates while considering the energetic stability of the mutations introduced in the scaffold. Functionalization is achieved through the introduction of mutations at pockets consisting of non-β-sheet forming amino acids which are part of a pair of β-sheet bonded peptides. Suitably selected constraints are introduced to allow only the introduction of

combinations of amino acids which, according to primary materialphore models, bind to cesium ions and conform to geometries that match, with some deviation, the corresponding geometry of the residue positions in the flexible structural templates of the amyloid designable scaffold. Specifically, the constraints introduced guarantee that all possible combinations of distances between the centers of mass of the designed amino acid side chains can be matched with any combination of distances between the centers of mass of the same amino acid side chains in any of the primary materialphore models in a portion of the flexible structural templates. As illustrated in Figure 6.1C, matching is indicated with green checkmarks, whereas failure in matching is indicated with red crosses. Stabilization of the introduced mutations at the mutable positions of the flexible structural templates is represented by an objective function minimizing the energy to identify designed amino acids that yield the most stable conformations, thereby increasing the probability that the resulting designed peptides achieve the proper fold or geometry for functionalization. This term can additionally serve as a means to avoid the destabilization of the amyloid-forming properties of the designed peptides. In the present study, residue positions 2, 3, 10, and 11 were amenable for modification while Ala1 was treated as an acetylation in the designed peptides, which we introduced to further enhance the self-assembly properties of the designed peptides similarly we considered an amidated C-terminal end modification. The modifiable residue positions are represented as dotted circles labeled 1, 2, 3, and 4 in Figure 21C of which designed amino acids in the first two positions of one peptide and the last two positions of an adjacent peptide form a designed pocket. In the computational design model, we use a coarse-grained approach where each of the amino acid side chains are treated as one united point (e.g., see ref.[192]).

Out of the 20^4 total possible theoretical designed peptides, our optimization-based computational model produced 89442 designed peptides that could match the primary

materialphore models in different combinations of amino acids. The model was solved to global optimality and the produced sequences were rank ordered based on the energy defined in the objective function. We imposed additional constraints specific to cesium ion binding to guide the computational design procedure. Specifically, we extracted the designs that meet the following criteria: (i) an aspartic acid or glutamic acid is present at first or last designable position, (ii) a polar amino acid (serine, threonine, asparagine, glutamine) is present in at least two designable positions, (iii) the total charge of amino acids in the designable positions is equal to -1, and (iv) the exclusion of glycine, proline, cysteine, methionine, or histidine amino acids in the designable positions. As for (i), an aspartic acid or glutamic acid mutation at first or last designable position is desirable as these positions are expected to be the most solvent accessible positions; thus an aspartic acid or glutamic acid at those positions maximizes the probability of the negatively charged amino acids directly interacting with positively charged cesium ions. As for (ii), we performed an additional bioinformatics analysis on the propensities of amino acids that are present in the collected materialphore models. The analysis showed that polar amino acids occur most frequently: polar amino acids (serine, threonine, asparagine, and glutamine) make up 31.3% of all amino acid : cesium ion binding motifs, non-polar amino acids (alanine, valine, leucine, and isoleucine) contribute 27.9%, negatively charged amino acids (aspartic acid and glutamic acid) contribute 20.3%, positively charged amino acids (lysine and arginine) contribute 10.5%, and aromatic amino acids (tyrosine, phenylalanine, and tryptophan) contribute 9.9%. Furthermore, according to a study by Armentrout et al.[263], polar amino acids have been shown to bind cesium ions with enhanced affinity[263]. Thus, we imposed an additional constraint according to which a polar amino acid should be included in any two designable positions, while in the remaining positions, any allowable amino acid except the ones listed in (iv) was allowed as other amino acids,

such as aromatic amino acids[263], may be beneficial for cesium ions binding. As for (iii), a charge of -1 was imposed for charge complementarity to the positively charged cesium ion. As for (iv), glycine, proline and cysteine amino acids were excluded as glycine may contribute minimally to functionality, proline can bias the backbone conformation of the pocket, and cysteine can, under specific conditions, result in unintended intermolecular disulfide bridges. In addition, methionine and histidine residues were excluded as both amino acids are not abundant in the materialphore models, and histidine could bias the net charge of the pocket due to slight pH changes, and thus have a negative effect on cesium ion binding due to small pH variations. From the reduced rank ordered list of sequences, we selected the top 10% (total of 134 sequences) of the designed peptide sequences for further investigation.

6.3.4 Modeling, Simulation and Computational Validation of Selected Designed Peptides Predicted to Form Functional Amyloid Fibrils Binding to Cesium Ions

MD Simulation-Based Modeling of the Designed Amino Acids on the Original Designable Scaffold and Computational Screening-Validation of the Selected Designed Peptides

The ability of the elementary β-sheet structural units of the designed peptides to form binding pockets resembling the materialphore models from which their design originated was investigated through short MD simulations modeling the designed amino acids and subsequent computational validation. Using one of the ten flexible structural templates belonging to the amyloid designable scaffold AGKGAIIGFIK as an initial starting conformation, we modeled the elementary β-sheet structural units of the selected designed peptides. Appropriate mutations were introduced using CHARMM[110]. We performed short 20 ns implicit-solvent MD simulations for the selected designed peptides, in the absence of cesium ions, and extracted the snapshots for each

of the designed peptides to explore the conformational variability of the binding pockets comprising the designed amino acids in their β-sheet conformations. The absence of cesium ions served as an ultimate test to check the capacity of the designed peptides to form amyloid structures in which the designed pocket of the elementary β-sheet structural unit can adopt the proper geometry to bind to a cesium ion in its absence. Throughout the simulations, the core of the elementary β-sheet structural units of the designed peptides was subjected to light constraints to maintain the β-sheet interactions, thus preserving the integrity of the elementary β-sheet structural units, and the four designed amino acids in each peptide were unconstrained and flexible.

Visual inspection of the MD simulation snapshots indicated that, with the β-sheet forming amino acids constrained during the simulations, the short duration of the MD simulations was sufficient to allow the conformational exploration of the binding pocket and to screen out the designed peptides incapable of forming the desired amino acid pockets required for functionalization. The choice of short rather than long simulations aimed at a fast screening procedure that can efficiently be used to sample the conformational space of the designed amino acids and could facilitate its implementation for hundreds of selected designed peptides. Considering that the backbone of each of the first two and last two unrestrained amino acids has two degrees of freedom in the φ-ψ space (helical or β-sheet), the total degrees of freedom for backbone conformations is equal to 2^4. Thus, such short simulation durations can be attractive in their application in a future implementation of our protocol as a computational package to design functional amyloid materials on demand. It is worth noting that longer and more highly-accurate simulations investigating the top designed peptides' self-assembly properties are additionally performed afterwards to guarantee higher precision in validation. From this perspective, this step can be considered as a tool screening out designs that are not highly promising.

Upon the completion of the MD simulations, we introduced a two-component computational validation procedure to evaluate the ability of the designed peptides' elementary β-sheet structural units to form binding pockets resembling the geometry of the amino acid : cesium ion binding motifs (far left section of Figure 6.1D). In the first component of the validation procedure, we evaluated the ability of the designed amino acid side chain binding pockets, per pair of two β-sheet bonded peptides, to acquire geometries similar to the primary materialphore models they originated from. In the second component of the validation procedure, we evaluated the ability of the designed peptides' amino acid side chain pockets to bind a cesium ion in the same fashion as the secondary materialphore model(s) from which the designed amino acids originated. Specifically, in the first validation component, the distances between the centers of mass of the four designed amino acid side chains within the MD simulation were compared to the corresponding distances encompassed in the primary materialphore model(s) each design originated from. An example of a primary materialphore model is presented in bottom left section of Figure 6.1D. If the two sets of distances were similar for any of the designed pockets within a given simulation snapshot, then the snapshot and its corresponding pocket was saved for further analysis and the pocket was considered a primary match. The percentage of snapshots containing primary matches for the first validation component for the simulated designed peptides is presented in Table 6.1. In the second validation component, a cesium ion was placed in the designed pockets comprising primary matches, which formed appropriate side chain conformations according to the primary materialphore models. Specifically, the second validation component assures that cesium ions can be placed successfully in the designed pockets containing primary matches and also guarantees that the placed cesium ion does not have a steric hindrance with any atoms, belonging to both designed and unmodified amino acids. If the distances between the centers of mass of the

four designed amino acid side chains and the placed cesium ion within the MD simulations were similar to the corresponding distances encompassed in the secondary materialphore model(s) each design originated from designed pocket was considered a secondary match. An example of a secondary materialphore model is presented in bottom left section of Figure 6.1D. The percentage of pockets containing secondary matches in the snapshots used in the second validation stage for the selected designed peptides is presented in Table 6.1. The designed peptides were ranked based on their first validation component percentages. We removed any designed peptides with second validation component percentages less than 40% from the rank-ordered list of designs. From the reduced list of designed peptides, we selected the top 10% (13 designed peptides) of peptides with the highest first validation component percentages for further investigation described below.

Table 6.1 Designed peptide sequences predicted to bind to cesium ions and their corresponding first and second validation component percentage statistics calculated over the snapshots extracted from short MD simulations. The listed percentages for the first validation component correspond to the percentage of snapshots containing primary matches. The listed percentages for the second validation component correspond to the percentage of pockets containing secondary matches within the snapshots containing primary matches, which were identified from the first validation component. The designed peptide sequences were ranked according to their first validation component percentages. The designed peptide sequences shown in boldface text were selected for further investigation of their self-assembly properties using REMD simulations, and re-validation (see Table 6.2). Reprinted with the permission from ref.[22]

Designed peptide sequences	First validation component (%)	Second validation component (%)
DSGAIIGFFS	**100.00**	**93.67**
DTGAIIGFWQ	**99.95**	**93.07**
QTGAIIGFYD	**99.70**	**91.58**
FNGAIIGFQD	**99.70**	**65.08**
DSGAIIGFQW	**99.65**	**87.65**
DNGAIIGFYT	**99.65**	**99.33**

Table 6.1 Continued

Designed peptide sequences	First validation component (%)	Second validation component (%)
WSGAIIGFQD	**99.60**	**90.83**
WTGAIIGFQD	**99.40**	**94.92**
SSGAIIGFWE	**99.35**	**91.03**
QTGAIIGFWD	**99.35**	**67.40**
QSGAIIGFYE	**99.10**	**43.55**
FQGAIIGFNE	**99.10**	**53.51**
ESGAIIGFWS	97.00	
DTGAIIGFWS	96.90	
YTGAIIGFQD	96.65	
DSGAIIGFFN	96.25	
NQGAIIGFFE	96.15	
YSGAIIGFSE	95.55	
SQGAIIGFFE	94.70	
STGAIIGFWD	94.65	
SSGAIIGFFD	94.40	
DQGAIIGFWT	93.25	
EQGAIIGFYS	92.40	
FQGAIIGFQD	91.50	
STGAIIGFYD	91.30	
QYGAIIGFSD	90.85	
DTGAIIGFYQ	90.50	
SWGAIIGFQD	90.10	
DQGAIIGFWS	88.80	
NQGAIIGFFD	87.75	
QNGAIIGFFD	86.15	
DQGAIIGFFN	84.60	
SNGAIIGFYD	84.45	
NNGAIIGFYE	83.95	
DSGAIIGFFQ	83.25	

Table 6.1 Continued

Designed peptide sequences	First validation component (%)	Second validation component (%)
SNGAIIGFYE	83.10	
SYGAIIGFQE	82.05	
FNGAIIGFQE	81.60	
ESGAIIGFYN	80.45	
SSGAIIGFYD	79.75	
DYGAIIGFQS	78.75	
SQGAIIGFWD	78.15	
SQGAIIGFYE	77.05	
ESGAIIGFYQ	76.30	
QSGAIIGFFD	76.25	
NSGAIIGFYE	74.80	
DQGAIIGFFQ	74.45	
YSGAIIGFSD	74.40	
QQGAIIGFFD	74.20	
DSGAIIGFWQ	73.80	
SSGAIIGFWD	72.70	
DYGAIIGFNS	72.10	
DYGAIIGFSN	71.25	
DYGAIIGFSS	70.60	
DNGAIIGFFQ	70.25	
NTGAIIGFYE	70.10	
DQGAIIGFYS	67.85	
QSGAIIGFWD	67.65	
DSGAIIGFYQ	66.60	
YQGAIIGFTD	65.80	
DNGAIIGFYS	64.75	
DTGAIIGFYS	64.50	
FSGAIIGFQE	63.85	
QQGAIIGFLD	63.80	

Table 6.1 Continued

Designed peptide sequences	First validation component (%)	Second validation component (%)
ETGAIIGFYQ	61.95	
DYGAIIGFQT	61.90	
DSGAIIGFWS	61.30	
QNGAIIGFFE	53.05	
YQGAIIGFSE	51.35	
YSGAIIGFND	51.10	
FQGAIIGFSE	50.95	
DTGAIIGFYN	48.90	
YSGAIIGFQE	47.40	
QSGAIIGFYD	45.55	
WSGAIIGFSD	44.45	
DQGAIIGFYT	43.95	
YQGAIIGFSD	41.40	
DNGAIIGFFS	41.20	
ENGAIIGFYS	36.90	
DSGAIIGFQY	36.50	
SQGAIIGFYD	35.40	
DSGAIIGFWT	34.60	
DYGAIIGFNQ	32.20	
SSGAIIGFYE	29.60	
QTGAIIGFYE	29.60	
ENGAIIGFWS	28.35	
EQGAIIGFYN	28.25	
LQGAIIGFQE	28.10	
EQGAIIGFFQ	26.30	
DYGAIIGFSQ	24.70	
FQGAIIGFQE	24.10	
YQGAIIGFNE	23.55	
DSGAIIGFYN	23.15	

Table 6.1 Continued

Designed peptide sequences	First validation component (%)	Second validation component (%)
NSGAIIGFYD	22.70	
YSGAIIGFQD	20.75	
EQGAIIGFYT	20.60	
YTGAIIGFND	19.25	
YSGAIIGFNE	18.60	
QYGAIIGFNE	11.25	
SQGAIIGFFD	10.00	
YNGAIIGFSE	9.65	
NQGAIIGFYE	9.40	
SYGAIIGFQD	9.25	
YQGAIIGFND	9.15	
ESGAIIGFWN	7.75	
EYGAIIGFQN	7.45	
DQGAIIGFYN	6.35	
FSGAIIGFQD	5.90	
QQGAIIGFFE	5.80	
YNGAIIGFTD	5.40	
QFGAIIGFQD	4.10	
NTGAIIGFYD	3.80	
FQGAIIGFSD	3.60	
QNGAIIGFYE	3.55	
ENGAIIGFYN	3.50	
DNGAIIGFFN	3.10	
ENGAIIGFYT	3.05	
QNGAIIGFYD	2.60	
YNGAIIGFSD	2.05	
YNGAIIGFQE	2.05	
YQGAIIGFTE	1.95	
NYGAIIGFQE	0.80	

Table 6.1 Continued

Designed peptide sequences	First validation component (%)	Second validation component (%)
NQGAIIGFYD	0.75	
ENGAIIGFYQ	0.60	
DNGAIIGFWN	0.60	
YTGAIIGFQE	0.25	

MD Simulations Investigating the Self-Assembly Properties of the Designed Peptides and Final Computational Screening-Validation

In the computational validation using short MD simulations-based modeling described above, we assumed that the designed peptides self-assemble into β-sheet structures with elementary structural units resembling the amyloid designable scaffold's flexible structural templates on which design was performed. Here, we computationally investigated the selected 13 peptides' capacity to self-assemble into amyloid β-sheet structures containing designed pockets appropriate for functionalization. In contrast to the previous simulations, in this step, the β-sheet structure of the 13 designed peptides is not assumed to be identical to the structure of the elementary β-sheet structural unit of the designable scaffold it originated from. Thus, to investigate the peptides' self-assembly properties in a first-principles-based approach, we employed REMD simulations (16 μs each for each of the 13 deigned peptides) followed by structural analysis, analogously to our aforementioned investigation on the designable amyloid scaffolds formed by peptide sequence AGKGAIIGFIK, and in line with the computational protocol published in refs[23,24,40,131,192]. This analysis was performed in order to verify that the 13 selected designed peptides indeed self-assemble into the desired functional structures without any prior assumption,

and in the absence of cesium ions. Similar to the previous section, the absence of cesium ions served as an ultimate test to check the capacity of the designed peptides to form amyloid structures in which the designed pocket of the elementary β-sheet structural unit can adopt the proper geometry to bind to a cesium ion in its absence. Upon completion, our results suggested that 11 out of the 13 designed peptides showed a tendency to favor parallel β-sheet formation over antiparallel β-sheet formation and were discarded from further investigation since the initial amyloid designable scaffold that the designed peptides originated from were antiparallel and the formation of parallel β-sheets rather than antiparallel β-sheets would disallow the formation of the intended designed amino acid functional pockets capable of binding to cesium ions. An example of a discarded designed peptide favoring parallel configuration is presented in right section of Figure 6.1D. For the remaining 2 designed peptides, with sequences FQGAIIGFNE (Ace-FQGAIIGFNE-$CONH_2$) and QTGAIIGFYD (Ace-QTGAIIGFYD-$CONH_2$), favoring antiparallel β-sheet formation, we extracted their highly ordered and well-aligned antiparallel β-sheet structures. Using the extracted structures, we performed a second additional step of the two-component validation procedure (far right section of Figure 6.1D) described above, which identified the former, FQGAIIGFNE, as the most promising designed peptide for experimental testing and further computational analysis. The results of the two-component validation procedure are presented in Table 6.2.

Table 6.2 Designed peptide sequences predicted to bind to cesium ions and their corresponding first and second validation component percentage statistics calculated over the highly ordered and well-aligned snapshots extracted from the REMD simulations investigating the peptides' self-assembly properties. The listed percentages for the first validation component correspond to the percentage of snapshots containing primary matches. The listed percentages for the second validation component correspond to the percentage of pockets containing secondary matches within the snapshots containing primary matches, which were identified from the first validation component. Reprinted with the permission from ref.[22]

Designed peptide sequences	First validation component (%)	Second validation component (%)
FQGAIIGFNE	81.80	87.33
QTGAIIGFYD	52.80	80.70
FNGAIIGFQE	78.40	61.80

6.3.5 Designs Selected for Experimental Testing and Further Computational Analysis

We applied the computational protocol for the design of amyloid materials and identified the designed peptide with sequence FQGAIIGFNE as the most promising peptide for cesium ion binding. The designed amino acids (F, Q, N, E), originated from three materialphore models, corresponding to experimentally resolved protein : cesium ion complex structures: (1) the Thermotoga maritima CorA coiled-coil mutant (PDB ID: 4EEB[264]), (2) the acid-sensing ion channel in complex with snake toxin (PDB ID: 4NTY[265]), and (3) the Lactococcus lactis cell wall pellicle by the phage 1358 receptor binding protein (PDB ID: 4L9B[266]). According to further analysis (not shown) on the two-component validation procedure performed using the highly ordered and well-aligned antiparallel β-sheet structures of FQGAIIGFNE, the simulated conformations of the four designed amino acids resemble to a higher degree the first of the three aforementioned materialphore models[264]. This is well represented in Figures 6.3A and 6.3C which

present the mimicry between the four designed amino acids in the designed β-sheet structures to the corresponding amino acids of the experimentally resolved structure[264].

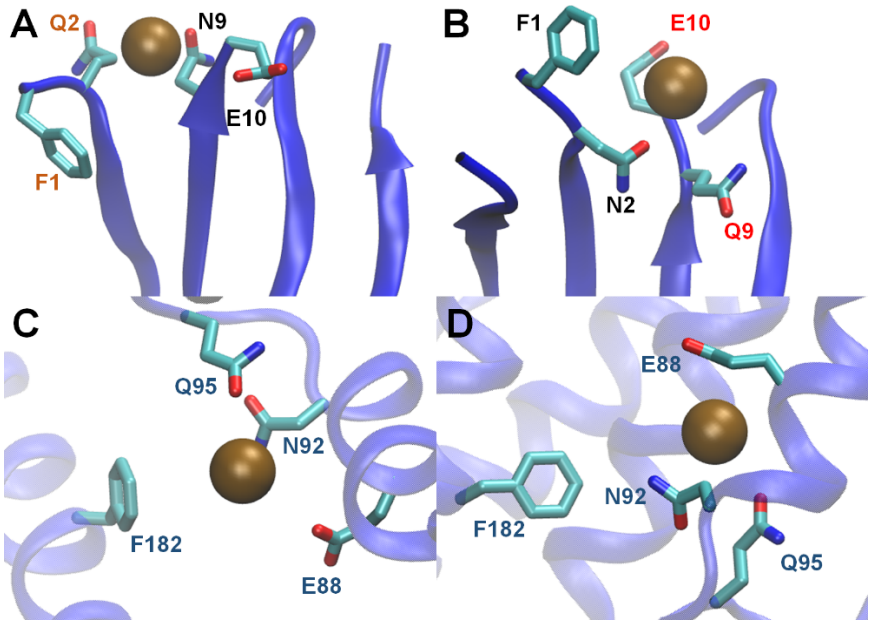


Figure 6.3 Molecular graphic images showing the mimicry of the designed pockets formed by the peptides FQGAIIGFNE, FNGAIIGFQE to the materialphore model comprising the amino acid : cesium ion binding motif from the structure of a Thermotoga maritima CorA coiled-coil mutant (PDB ID: 4EEB). Panels A, and B show the computationally predicted cesium binding modes by the designed peptides FQGAIIGFNE, and FNGAIIGFQE, respectively. Panels C, and D show the amino acid : cesium ion binding motif from which the designed peptides FQGAIIGFNE, and FNGAIIGFQE originated from. Reprinted with the permission from ref.[22]

Given that FQGAIIGFNE was identified as the most promising designed peptide, we returned to the reduced rank ordered list of designed sequences and searched for additional designed peptides with the same amino acid composition (F, Q, N, and E in the designable positions). From this search, we additionally selected the designed peptide with sequence FNGAIIGFQE (Ace-FNGAIIGFQE-$CONH_2$), of which the asparagine and glutamine amino acids are switched compared to FQGAIIGFNE. The designed peptide, FNGAIIGFQE, was actually

ranked higher in the rank ordered list of sequences from the design model than FQGAIIGFNE. Although the designed peptide FNGAIIGFQE was not among the top 10% of designed peptides ranked using the first validation procedure step based on short MD simulations-based modeling, its first and second component validation percentages were decent. Thus, we additionally investigated the self-assembly properties of FNGAIIGFQE using REMD simulations, followed by further structural and validation analysis, as detailed in the aforementioned section for FQGAIIGFNE. Similarly to FQGAIIGFNE, FNGAIIGFQE favors antiparallel β-sheet structural formation, its two-component validation statistics are also decently high, and its four designed amino acids (F, N, Q, E) originated from same three materialphore models that FQGAIIGFNE originated from, with the higher degree of resemblance corresponding to the same materialphore model[264]. This is also well represented in Figures 6.3B, 6.3D which present the mimicry between the four designed amino acids in the designed β-sheet structures to the corresponding amino acids of the experimentally resolved structure[264].

Thus, both FQGAIIGFNE and FNGAIIGFQE were selected for experimental testing and further computational analysis. As mentioned above, the designed amino acids (F, Q, N, E) of peptides FQGAIIGFNE and FNGAIIGFQE originate from the same materialphore model, and correspond to the amino acids Phe182 of chain B, Glu88 of chain C, Asn92 of chain C, and Gln95 of chain C in the structure of a CorA coiled-coil mutant (PDB ID: 4EEB[264]). The amino acids Glu88 of chain C, Asn92 of chain C, and Gln95 of chain C that were included in the designed pocket are interacting directly with the resolved cesium ion and Phe182 of chain B forms long-range cation-π interactions[267,268] with the resolved cesium ion. The fact that only four residue positions are available for modification limited the capacity of the designed peptides to completely mimic the binding of cesium ions by proteins shown in X-ray resolved structure, which may

include more than four amino acids interacting with a bound cesium ion. Despite this limitation, the design procedure was not significantly affected as the net charge of the vast majority of the neglected amino acids binding to the specific cesium ion that were not included in the computationally designed pockets are serendipitously counterbalanced. Of the excluded amino acids, Asp253 of chain B interacts with His257 of chain B and Asp89 of chain C interacts with Lys98 of chain C in the X-ray resolved structure[264].

6.3.6 Experimental Testing, Verification of Amyloid Self-Assembly, and Determination of Cesium Ion Capture Properties by the Designed Amyloid Materials

All microscopic analyses to verify the self-assembly properties and cesium deposition properties of the designed peptides were by Dr. Mitraki's lab. Elemental analysis experiments were performed by Dr. Bryan Tomlin through the elemental analysis facilities of Texas A&M University using amyloid materials and solutions prepared by the student and Asuka A. Orr prepared in Dr. Tamamis' and Dr. Karim's labs under the guidance of Dr. Hae-Kwon Jeong (see ref[22]).

Amyloid Fiber Formation and Fiber Stability

The ability of the computationally designed peptides, Ace-FQGAIIGFNE-$CONH_2$ and Ace-FNGAIIGFQE-$CONH_2$, to self-assemble into amyloid fibrils was verified through FESEM observations and CR staining. Both the designed peptides form distinctive fibrillar morphologies with width in the range of 10-20 nm as observed under FESEM.

Cesium Ion Binding and Capture Capacity of the Designed Amyloid Fibers

We observed the fibrils formed by FQGAIIGFNE and FNGAIIGFQE under TEM before and after exposure to cesium to visually verify that the fibrils bind cesium (Figures 6.4A, 6.4B, 6.4C, 6.4D). The fibrils that were exposed to cesium chloride solution were washed and were not negatively stained before observation under TEM. Since cesium has a high atomic number (Z=55) which leads to an increased electron density in the area of the fibrils bound with cesium, it can contribute to the detection of the fibrils via transmission electron microscopy, as compared to organic moieties consisting of carbon, hydrogen, and oxygen[269]. The detection of the fibrils exposed to cesium using TEM indicates that the designed peptides form amyloid materials functional to capture cesium. TEM studies also confirmed that the fibril morphology is maintained after uptake of cesium ions.

We subsequently performed elemental analysis in triplicates to quantify the amount of cesium ions captured by the amyloid materials formed by designed peptides, FQGAIIGFNE and FNGAIIGFQE. To form the fibrils, 6 mg of each designed peptide was self-assembled in 0.5 mL of water. The fibrils were subsequently separated from the water and exposed to 0.5 mL of 10 ppm cesium chloride solution. At the neutral pH 7, the amyloid materials formed by FQGAIIGFNE and FNGAIIGFQE exhibited high cesium binding capacities with 70.9 ± 2.2% and 60.4 ± 0.8% of cesium ions removed from a 10 ppm cesium chloride solution, respectively. We additionally investigated the ability of the amyloid materials to capture cesium ions in acidic conditions. At pH 4.5, the amyloid materials formed by FQGAIIGFNE and FNGAIIGFQE maintain their ability to capture cesium, removing 64.9 ± 1.9% and 59.3 ± 0.4% of the cesium ions from a 10 ppm cesium chloride solution, respectively. The maintained capacity of the designed amyloid materials in acidic conditions indicates that amyloid materials may be promising for further improvement (e.g.,

re-design) to remove cesium from radioactive acidic wastewater conditions. We also performed the same analysis for the amyloids formed by the amyloid designable scaffold, NH_3^+-AGKGAIIGFIK-COO^-; at neutral pH 7, the amyloid fibrils formed by AGKGAIIGFIK removed 16.2 ± 3.3% of cesium ions from a 10 ppm cesium chloride solution. At pH 4.5, the amyloid fibrils formed by AGKGAIIGFIK removed 18.1 ± 6.1% of cesium ions from a 10 ppm cesium chloride solution. Further investigations were performed to delineate the non-specific cesium capture properties of AGKGAIIGFIK. The percent cesium ions captured by the amyloid materials formed by FQGAIIGFNE and FNGAIIGFQE as well as the amyloid fibrils formed by AGKGAIIGFIK are plotted in Figure 6.4E.

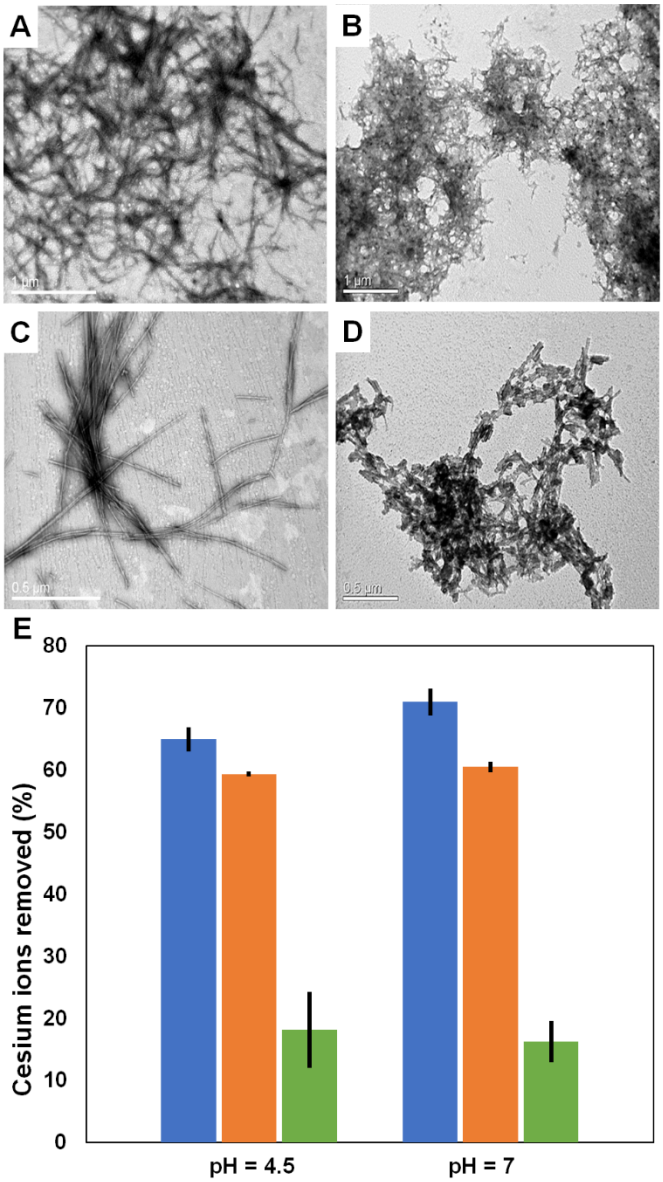


Figure 6.4 TEM micrographs of peptide fibrils before and after cesium deposition and histogram showing percentage (%) of cesium ions removed from a 10 ppm cesium chloride solution. Panels A and C show fibrils formed after 3 days of incubation by the peptides FQGAIIGFNE and FNGAIIGFQE, respectively. Panels B and D show fibrils after cesium capture for peptides FQGAIIGFNE and FNGAIIGFQE, respectively. Panel E shows the percentage (%) of cesium ions removed from a 10 ppm cesium chloride solution by the amyloid material formed by the peptide FQGAIIGFNE (blue), amyloid material formed by the peptide FNGAIIGFQE (orange), and fibrils formed by the peptide AGKGAIIGFIK (green) under two different pH conditions, pH 4.5 and pH 7. Figures A-D were produced by our experimental collaborators. Raw values are available in an excel spreadsheet of the student's lab book. Reprinted with the permission from ref.[22]

6.3.7 Explicit-Solvent MD Simulations Investigating the Cesium Ion Binding Properties of the Designed Amyloid Materials

Explicit-solvent, all-atom MD simulations were performed using the Drude polarizable force field[256,257,258] in CHARMM[110] to delineate why the amyloid fibrils formed by AGKGAIIGFIK exhibited a low cesium ion capture capacity, as well as to provide insights into how the amyloid material formed by the designed peptides FQGAIIGFNE and FNGAIIGFQE bind to and capture cesium ions. Specifically, we aimed to investigate whether the capture can primarily be attributed to the computationally designed mechanism. For this purpose, we performed 5 independent MD simulations runs, 10 ns each, investigating the 4-stranded elementary β-sheet structural units of each of the three peptides in the presence of cesium chloride (0.5 M).

According to the MD simulations cesium ions interact with all the amino acids of the modeled AGKGAIIGFIK fibrils nearly uniformly, with the exception of Ala1, which has an elongated backbone owing to its acetylation (Figure 6.5A), potentially suggesting a partly non-specific binding mechanism. Upon visual inspection we observed that cesium ions frequently interact with the C-terminal end of the peptide, which is not amidated. We quantified the specific interactions and observed that these interactions occur in approximately 20% of the snapshots. Furthermore, we have also observed that in 25% of the snapshots in which this interaction occurs, chloride ions interact with positively charged group of Lys11. This suggests that a portion of the capture can be associated with cesium chloride salts being "trapped" in such a way that their positive and negative ions simultaneously interact with the negatively charged C-terminus and the positively charged side chain group of lysine residues (Figure 6.5B).

On the contrary, according to the MD simulations, cesium ions interact primarily with the designed residues located at the termini of the modeled FQGAIIGFNE and FNGAIIGFQE fibrils

(results not shown). In the simulations performed to investigate cesium ion binding to the elementary β-sheet structural units formed by peptides FQGAIIGFNE and FNGAIIGFQE, our structural analysis focused on detecting the binding pockets formed by four amino acid side chains that can simultaneously interact with a cesium ion during the simulation trajectory. From this analysis, we considered only binding pockets formed by four amino acids which belong to two different peptide strands; this consideration neglects any potential binding pockets formed by amino acids which belong only to any of the two outer peptides; such pockets, if formed, are simply an artifact of our simulations lacking a nearly infinite array of peptides in an amyloid fibril. According to the analysis, in the FQGAIIGFNE simulations, the probability of a cesium ion to interact with all four designed amino acids is equal to 24%, which is also the most populated cluster among all other possible binding pockets comprising four amino acids (Figure 6.5C), while the corresponding probability of a cesium ion to interact with at least three of the four designed amino acids is equal to 43%. Importantly, this suggests that the binding and related capture mechanism of the FQGAIIGFNE amyloid fibrils can be attributed to the designed amino acids (Figure 6.5C). In the FNGAIIGFQE simulations, the probability of a cesium ion to interact with all four designed amino acids is equal to 3% (Figure 6.5D), while the corresponding probability of a cesium ion to interact with at least three of the four designed amino acids is equal to 82%; the latter is attributed to the fact that in the majority of cases, one out of the four designed amino acids is not participating in interactions with a cesium ion, and instead another nearby amino acid does. For example, in the top two populated clusters (31% and 27% respectively), Ala4 of the first strand interacts with a cesium ion instead of Gln9 of the second strand (Figure 6.5E), and Phe8 of the first strand interacts with a cesium ion instead of Phe1 of the second strand (Figure 6.5F). Thus, the binding and related capture mechanism of the FNGAIIGFQE amyloid fibrils can be primarily attributed to the

designed amino acids. Quite interestingly, we observe that Phe8 which was originally part of the amyloid designable scaffold AGKGAIIFIK, can in place of designed amino acid Phe1, bind to the cesium ion resembling the materialphore model from which FNGAIIGFQE originated from (Figure 6.5F, Figure 6.3D). While the aforementioned simulations can provide insights into the cesium ion binding and capture mechanisms of the designed peptides, longer simulations are needed to provide more in-depth insights, especially in the case of FNGAIIGFQE for which the instances of pockets formed comprising four amino acid side chains simultaneously interacting with a cesium ion was smaller (170) compared to FQGAIIGFNE (773). Yet, this investigation was not a primary scope of the current study, which focused mainly on the computational functionalization design aspect.

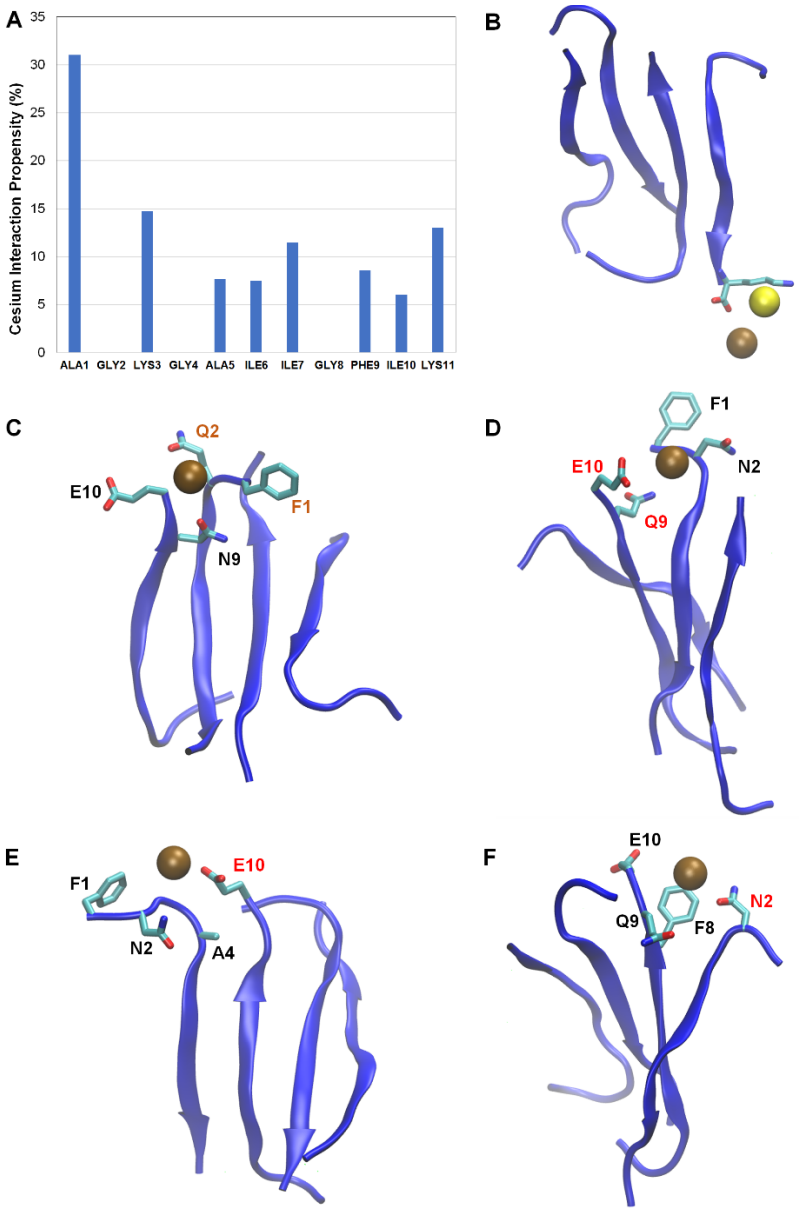


Figure 6.5 Interactions of the designed amyloid peptide materials to cesium ions. A: Histogram showing the percent propensities for each of the amino acid side chains of the AGKGAIIGFIK amyloid peptide to interact with a cesium ion in the explicit-solvent simulations. B: Molecular graphics images depicting the interaction of negative charged C-terminus of the AGKGAIIGFIK amyloid peptide with a cesium ion and positively charged side chain group of lysine residues with chloride ion. C: Molecular graphic image depicting the cesium ion binding to the designed amino acids Phe1, Gln2, Asn9, and Glu10 for the designed amyloid peptide FQGAIIGFNE (1st most populated cluster). D, E, F: Molecular graphic images depicting the cesium ion binding to: the designed amino acids Phe1, Asn2, Gln9, and Glu10 (D; 9th most populated cluster), the designed amino acids Phe1, Asn2, and Glu10 and additionally amino acid Ala4 (E; 1st most populated cluster), and the designed amino acids Asn2, Gln9, and Glu10 and additionally amino acid Phe8 (F; 2nd most populated cluster), for the designed amyloid peptide FNGAIIGFQE. Reprinted with the permission from ref.[22]

6.4 Concluding Remarks

Here, we developed and applied a computational protocol for the design of functional amyloid materials functional for cesium ion capture. Considering the increasing demand for novel amyloid materials that can be designed "on demand" to bind (and either capture, or to be used for the sustained release of) specific ions or compounds for biomedical, environmental and technological applications, we suggest that such a protocol can be advanced to tackle a series of problems associated with, but not limited to, the application of amyloid materials in tissue engineering, separations and water treatment, as well as drug delivery. An advanced protocol with improvements on the capacity and specificity of the designed amyloid materials could provide a possible solution to weaknesses associated with existing rational approaches functionalizing amyloid materials, which rely on scientists' intuition and may yield low success rates or unsuccessful designs if the desired functional properties are hard to achieve. Therefore, it can constitute a stepping stone toward a generalized protocol for the computational design of functional amyloid materials in the future.

6.5 Innovation in Methods

We developed the first computational design protocol functionalizing amyloid materials to bind a specific ion (cesium in the current study). According to the protocol's strategy, an amyloid material can be functionalized to bind a specific ion by introducing mutations at suitably selected non-β-sheet residue positions at the termini of an amyloid scaffold. The introduced mutations at the selected positions mimic how amino acids bind to the ion according to experimentally resolved structures, through choosing the amino acid combinations and geometries yielding the most stable bound conformations for functionalization.

The computational design is operated through an optimization-based model, which minimizes the energy of designed amino acids outside the amyloid core of a designable scaffold. The designed amino acids are introduced in non-β-sheet residue positions of a pair of two β-sheet bonded peptides to spatially reproduce, and thus mimic, any of the extracted materialphore models, which represent how proteins bind to the ion according to available experimentally resolved structures. The mimicry is innovatively handled by constraints, which are introduced such that the optimum solutions, among all possible materialphore models considered, and the lowest energy solutions are selected for further investigation. The selection of lowest energy solutions aims to both stabilize the bound conformation of the designed amino acids, as well as to avoid destabilizing the self-assembly properties of the designed peptides. Additional constraints during the solution of the design model are used to ensure that mutations are introduced at non-β-sheet forming positions belonging to a pair of non-covalently bonded β-sheet peptides, considering the peptides' fibrillar structural arrangement (e.g., antiparallel), and also considering the necessary condition that the designed peptide sequences must be identical. Simulations and structural analysis programs are subsequently used to screen, computationally validate and select top performing designs for further experimental validation. All aforementioned components were crucial and essential pieces that were strategically assembled in the computational protocol to overcome limitations of current computational tools[253], which in general can be used to functionalize a protein but not necessarily an amyloid material. An inherent limitation of existing computational tools to solve the problem addressed in this study is that they are developed to functionalize proteins in which the designed amino acid pockets consist of amino acids covalently bonded in a macromolecular protein structure[22]. On the contrary, in our approach, functionalizing mutations (i) are introduced to all peptides (specifically to binding pockets formed by two adjacent β-sheet

bonded peptides) so that the designed peptides are identical in sequence and can form functional binding pockets when bonded in β-sheets, (ii) are guided by the properties of the amyloid scaffold, and attention is payed so as not to disrupt self-assembly, (iii) are checked to ensure their introduction does not lead to an unanticipated structural arrangement of the peptides for functionalization (e.g., switch from parallel to antiparallel structure or vice versa). While the aforementioned points impose additional difficulty in the computational design of functional amyloid materials, our approach was significantly empowered by the use of multiple materialphore models from existing X-ray resolved PDB structures which were all simultaneously considered during the search for optimal solutions[22].

6.6 New Findings and Potential Applications

Through our innovative computational protocol, we designed amyloid materials capturing cesium ions at physiological and acidic conditions. Importantly, materials with these properties can be useful in the development of technologies to capture radioactive cesium ions from acidic wastewater conditions or as biocompatible materials removing cesium from blood. The current protocol also has specific limitations, which include the fact that key amino acids for binding according to materialphore model(s) may be neglected during the optimization-based design due to the limited number of mutable positions on the designed peptide elementary structural units and the fact that the interaction energy between the ions and the designed amino acids has not been considered during both the computation design and validation. Features selecting and placing the most critical amino acids for functionality and incorporating ion-amino acid interactions can potentially be added in future studies of the computational protocol aiming at improving the protocol's accuracy and the designed amyloid material's capacity and specificity; notably, the latter component was not investigated in this study and requires further investigation. Although

the materials were computationally designed to bind cesium ions, our study and analysis do not preclude binding to other ions (e.g., ions of the same group, such as rubidium). Overall, the capacity and specificity of future amyloid materials can be benefited from both future computational advancements improving our protocol combined with experimental studies tuning and optimizing the conditions of the amyloid materials preparation and testing (e.g., filtration in the present study). The difference between the percentage of cesium ions captured on the basis of the elemental analysis of the supernatant between the computationally designed peptides and the amyloid designable scaffold, the experiments showing the cesium deposition properties of the amyloid materials, and the explicit-solvent simulations investigating the cesium ion binding properties of a modeled amyloid fibril formed by FQGAIIGFNE strongly suggest that the cesium ion capture properties can be principally attributed to the computationally designed mechanism. While in the two designed peptides, mimicry is primarily attributed to the designed residues N, Q, E, and to a lesser extent F, which has weaker long-range cation- π interactions of which the latter is observed in the X-ray structure,[254] the formation of such interactions of the designed fibrils with cesium ions proved important in our explicit solvent MD simulations. In addition, the methodology applied aimed at identifying the most promising designed materials following a specific strategy, which was not engineered for optimum outcome. Thus, it is possible that due to user-defined parameters a portion of the discarded designed sequences could prove successful upon experimental testing.

7. TOWARD A GENERALIZED COMPUTATIONAL PROTOCOL FOR THE DESIGN OF FUNCTIONAL AMYLOID MATERIALS

7.1 Introduction

7.1.1 Overview and Significance

In chapters 3, 4, and 5, we investigated peptide sequences containing GAIIG or GAITIG as amyloidogenic cores and showed that additional residues at their termini are predominantly not forming β-sheet interactions in their self-assembling structures. Thus, the additional (e.g. 2-3) residues at both termini could serve as a source of inspiration for novel amyloid material scaffolds. In chapter 6, we developed a computational protocol for the design of functional amyloid materials capable of binding to cesium ions, at which mutations were introduced at the termini of an amyloid designable scaffold with sequence AGKGAIIGFIK. We postulated that our introduced protocol could constitute a stepping stone toward a generalized computational protocol for the design of functional amyloid materials.

There are various available computational design tools, protocols or algorithms to functionalize proteins which among others include Metal Search[270], ORBIT[271], and Rosetta[272]. Metal Search[270] is a computational program to design tetrahedrally coordinated metal binding sites on a protein with known structure. The input to the specific computational program is the backbone atomic coordinates of the scaffold protein structure with the output being the positions of amino acids that can form tetrahedrally coordinated binding sites if the wild type amino acids were replaced with histidines and cysteines. In addition, the program also output the projected coordinates of the metal ion along with the dihedral angles for the amino acid side chains forming the binding pocket. The specific program employs simple geometrical criteria that constitute the

tetrahedral coordinated binding from the existing literature for the specific ion. The specific program starts by dividing the entire protein space into cubic lattices and subsequently reduces the search space by identifying four amino acid sets that are part of a specific lattice. Subsequently, each set is divided into subsets where each subset is a different combination of side chain dihedral angles and rotamers and the position of the metal ion is evaluated for each subset. Furthermore, the subsets to subjected to tetrahedral geometric criteria to evaluate the feasible designed sequences that could potentially bind to metal ions. The specific program has been used to introduce zinc binding sites in the β1 domain of streptococcal protein and also in the four-helix bundle protein α4.[270]

Optimization of rotamers by iterative techniques, ORBIT[271], is a cyclical protein design strategy consisting of four components such as a design module, a simulation module, experimental verification, and data analysis. The design module uses the fixed backbone as design template and describes the side chains using rotamers. The specific design module employs Dead End Elimination (DEE) algorithm, a fast-discrete search algorithm, which guarantees optimal global packing in the protein if the algorithm converges. The specific algorithm can reduce the sequence search space by eliminating the rotamer choices that are probably not part of the Global Minimum and employing the DEE algorithm reduces the search space by a combinatorial factor.[271]

Rosetta[272] is a suite of widely used protein design computational tools developed by David Baker's lab. Rosetta *de novo* enzyme design protocol, a module in Rosetta[272], has been developed to design enzymes that catalyze chemical reactions. Enzymes, typically proteins, are biomolecules that catalyze a chemical reaction by stabilizing the intermediate transition state. Thus, specific interactions between the amino acids in the enzymes and the chemical reactant that stabilize the transition state need to be known to design enzymes with improved stability. The design enzyme

protocol starts by taking the specific interactions as inputs in the form of geometric constraints which are fed into a file called “cstfile”. For each interaction between an amino acid’s side chain atom belonging to enzyme and compound’s atom there are six degrees of freedom: one distance, two angles, and three dihedrals. Thus, these parameters need to be defined in order to specify an interaction. Once the interactions are specified, the design protocol tries to identify the sites on a scaffold protein where the active sites can be realized. The identification is done by a module in Rosetta called RosettaMatch which designs the specific interactions by matching the geometrical constraints in the “cstfile” onto the residue positions in the scaffold which are specified by the user. Subsequently, the identities of the amino acids the surrounding the amino acids are mutated so that the catalytic interactions are optimal, and the resulting designed sequences are evaluated to identify the most stable designed enzymes. The specific design protocol could also be applied to design proteins that can bind to ion or compound of interest as specifying the interactions between the ion or compound of interest and mimicking the interactions on a target protein could potentially functionalize the protein to bind the specific ion or compound of interest.[272]

7.1.2 Considerations and Challenges Associated with the Design of Functional Amyloid Materials

Importantly, there are various challenges associated with designing functional amyloid materials binding to an ion or compound of interest which could impede the immediate use of existing computational protein design tools, as is, to address the problem: The following considerations need to be taken into account: (1) amyloid materials are composed of identical β-sheet strands, thus mutations made on a β-sheet strand should be appropriately made on all other β-sheet strands; (2) if design by mimicry is considered, the set of amino acids designed to bind a specific ion or compound of interest can be part of two or more identical β-sheet strands with the

orientation of the β-sheet strands being antiparallel or parallel configuration, and thus the designed peptide strands should retain the same configuration for the peptide to be functional; (3) mutations introduced to functionalize such a material should not interfere with the amyloid peptide's self-assembly properties, and if possible, they should provide further stabilization, rather than destabilizing the amyloid fibril. All these considerations impose challenges in the design of functional amyloid materials using existing computational protein design tools[270,271,272]. The considerations provided above:

- may significantly limit the number of possible feasible solutions, thus, when designing by mimicry, one should ideally consider a big-data approach, including a rich representation of models of amino acids that can be designed through mimicry on the amyloid fibril. These models are defined by us as materialphore models.

- suggest the importance of the use of methods taking into account the energy of the designed fibril to avoid destabilization of the fibril during matching of the materialphore models.

- suggest the importance of the use of multi-scale simulations, from short simulations investigating the capacity of the designed amino acids to form binding pockets resembling the materialphore models, and long simulations investigating the designed peptides self-assembly properties and their ability to form binding pockets appropriate for functionalization.

- could also require the use of simulations of the designed fibril in solution with the ion or compound of interest to investigate how the amyloid materials formed by the designed peptides bind to or capture the ion or compound of interest.

7.1.3 Goal

Accounting for the challenges described above, this chapter focuses toward a generalized protocol to design amyloid capturing an ion or compound of interest. As part of the protocol, an

optimization-based design model is used to introduce mutations at non-β-sheet residue positions of an amyloid designable scaffold. The designed amino acids are introduced to the scaffold mimicking how amino acids bind to ion or compound of interest according to experimentally resolved structures and also aim at energetically stabilizing the bound conformation of the pockets. The optimum designs are computationally validated using a series of simulations and structural analysis to select the top designed peptides, which are predicted to form fibrils with ion or compound of interest binding properties for experimental testing.

7.1.4 Objectives Accomplished

The computational protocol presented in the previous chapter was generalized to design amyloid materials that can bind to an ion or compound of interest. Additionally, the computational protocol can also be used in combination with rational design or intuition such as introducing the motif RGD for cell adhesion property. The protocol subsequently has been used in three example cases for the design of amyloid materials as drug carriers of Alzheimer's disease drugs which is presented in chapter 8. The protocol has also been used in tandem with rational approaches to design amyloid materials that can be used as DNA carriers and penetrate the cell (chapter 9). Additionally, the protocol has been used for the design of amyloid materials combining cell adhesion properties with binding properties to calcium and strontium ions (chapter 10).

In the current chapter, the computational protocol is divided into three modules. Module I presents the extraction and generation of materialphore models for an ion or compound of interest. This module also generates additional statistical analysis on the collected materialphore models for an ion or compound of interest (detailed below). Module II involves the selection of an amyloid designable scaffold from a set of previous computationally investigated and elucidated amyloid designable scaffolds,[22,23,24] and subsequently involves the execution of computational design.

After design is performed, the top ranked designed sequences are selected for further investigation. Module III involves the execution of short MD simulations to computationally validate the computationally designed peptides of Module II. Subsequently, the top ranked design sequences according to the computational validation are selected for experimental investigation. The overall schematic of the computational protocol is presented in Figure 7.1.

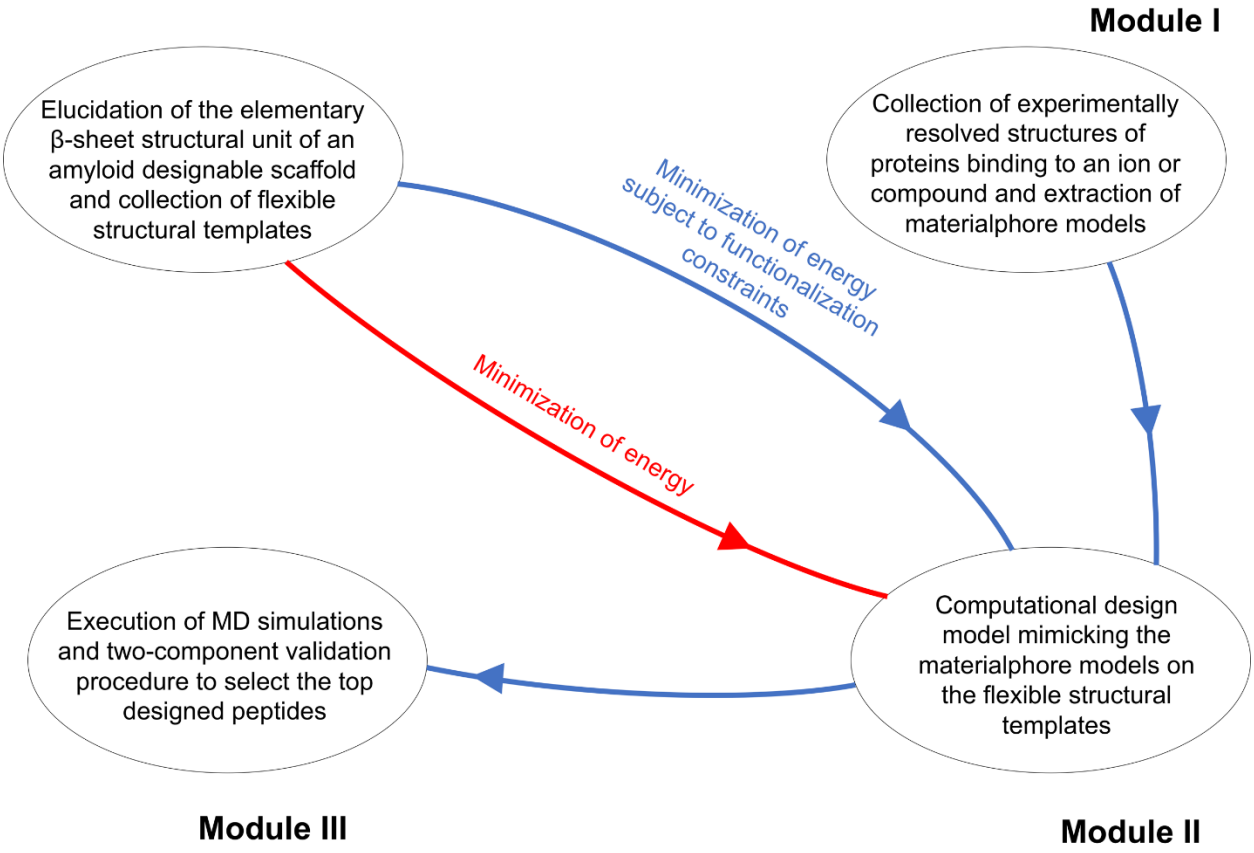


Figure 7.1 Overall schematic of the computational protocol. Designing amyloid materials using materialphore models is shown using blue color. Designing amyloid materials just by minimizing the energy due to mutations is shown in red color.

In what follows, each module is described in detail including the purpose and the execution of computational tasks performed. This also includes the inputs required for each module, the

calculations being performed in each module and the outputs from each module along with the interpretation about the generated outputs. Additionally, details required to execute each module is also presented. The specific modules are available in the shared folder of Dr. Tamamis' lab within the address: */mnt/research/Tamamis_Phanourios/Shared/ComDesFun*.

7.1.5 Necessary Resources

- A computer with Linux, Unix or Mac OS X operating systems or a Windows operating system with a bash terminal
- At least 8 GB RAM and 10 GB free disk space.

7.2 Module I: Generating Materialphore Models for an Ion or Compound of Interest

The programs for extracting materialphore models for an ion (current study) or compound; (see further studies) of interest were initially developed by Chang-Hyun Choi, an undergraduate student of Dr. Tamamis' lab, and were further advanced.

7.2.1 Purpose

The main aim of this module is to extract primary and secondary materialphore models which combined incorporate the information of how amino acids bind to an ion or compound of interest according to the experimentally resolved structures from the PDB.[252] Specifically, binding motifs of amino acids in complex with an ion or compound of interest are collected from the experimentally resolved structures, and (i) the distances between the pairs of amino acids in the binding motifs, as well as (ii) the distances between the ion or compound of interest and amino acids in the binding motifs are subsequently calculated. The set of distances described in (i) along with the identities of the amino acids involved is designated as a primary materialphore model and

the set of distances described in (ii) along with the identity of the amino acid involved is designated as a corresponding secondary materialphore model.

7.2.2 Input

The experimentally resolved protein structures containing amino acids binding to an ion or compound of interest in the PDB[252] format constitute the first source of input. Details on how to download the protein structures is presented in the file README_moduleI.txt. According to the developed programs, the downloaded protein structures should be placed in the same folder as the modules.

7.2.3 Input Parameter Specifications

1. The standard name code of the ion or compound of interest according to the PDB[252]. For example, CS for cesium ion (see chapter 6), or E20 for the compound donepezil (see chapter 8).
2. Type of ion or compound of interest. Each ion or compound of interest is classified into three types. A monoatomic ion of interest is classified as type 1, a polyatomic compound of interest for which the analysis needs to be done with respect to the center of mass is classified as type 2, and a polyatomic compound of interest for which the analysis needs to be done with respect to a specific-selected atom is classified as type 3 (the numerical number 1, 2 or 3 is provided as input).
3. Maximum distance in Å between the ion or compound of interest and the side chain center of mass of an amino acid that is considered to be interacting. Inputs 2 to 4 should be provided in the file ModuleI.in according to the programs.

7.2.4 Description

In-house Fortran programs were developed to execute the tasks described in this module. The specific module starts by preparing the experimentally resolved protein structures by removing any other information in the structure other than the information about the coordinates of the atoms. This processing to be done is written to a script file *mk.prep* created by the program *mkcreator.f*. In addition to the processing script file, the program also creates a script file *mk.organize* which organizes all the outputs produced in this module (described below) into different folders. After the processing is done, each processed experimentally resolved protein structure is read by the program *asteria.f.* The specific program reads each processed protein structure and identifies side chain atoms for each amino acid present in the structure. Subsequently, the program separately identifies all the ligands present in the structure. If the ligand is polyatomic and the analysis needs to be done with respect to its center of mass, the program calculates the center of mass coordinates for all the ligands present. The program then calculates the side chain center of mass coordinates for all the amino acids present in the protein structure and estimates the distance between each ion or compound of interest and each interacting amino acid. If the distance between the ligand and an amino acid is less than the specified distance (input parameter specification 3), the program saves the distance and the identity of the amino acid into a secondary materialphore model for that particular ligand. The number of materialphore models resulting from one protein structure corresponds to the total number of the same type of ligands present in the protein structure. The total number of materialphore models from all available structures in the PDB can be constituted to be the summation of number of ligands in each structure from the PDB under investigation over the total number of experimentally resolved protein structure for that ligand. Each ligand in the selected protein structures is designated with a unique code. The unique

code consists of the four letter PDB code, the chain name to which the ligand belongs in that protein structure, and the number of the ligand on that specific chain. The number for the specific ligand is assigned based on the order in which ligands are encountered on that specific chain in the protein structure. For example, the unique code for the compound donepezil belonging to the chain A in the experimentally resolved Acetylcholinesterase protein structure with PDB ID: 4EY7 is 1A4EY7. Subsequently, the programs calculate the distances between the amino acids in a specific binding motif in complex with an ion or compound of interest using the center of mass coordinates calculated above. The values of the distances between pairs of amino acids along with their identities are saved into primary materialphore model for each ligand. The secondary materialphore model is extracted first and prior to the primary materialphore model. This facilitates the identification of amino acids in the binding motifs, which subsequently allows the calculation of both amino acid to ion or compound of interest distances (secondary materialphore models) and amino acid to amino acid distances (primary materialphore models). In addition, the programs identify the number of amino acids in each binding motif and estimates the size distribution of binding motifs based on the number of amino acids in each binding motif. Subsequently, the module is organized into different folders which are detailed below. The overall schematic of the files and programs is presented in Figure 7.2.

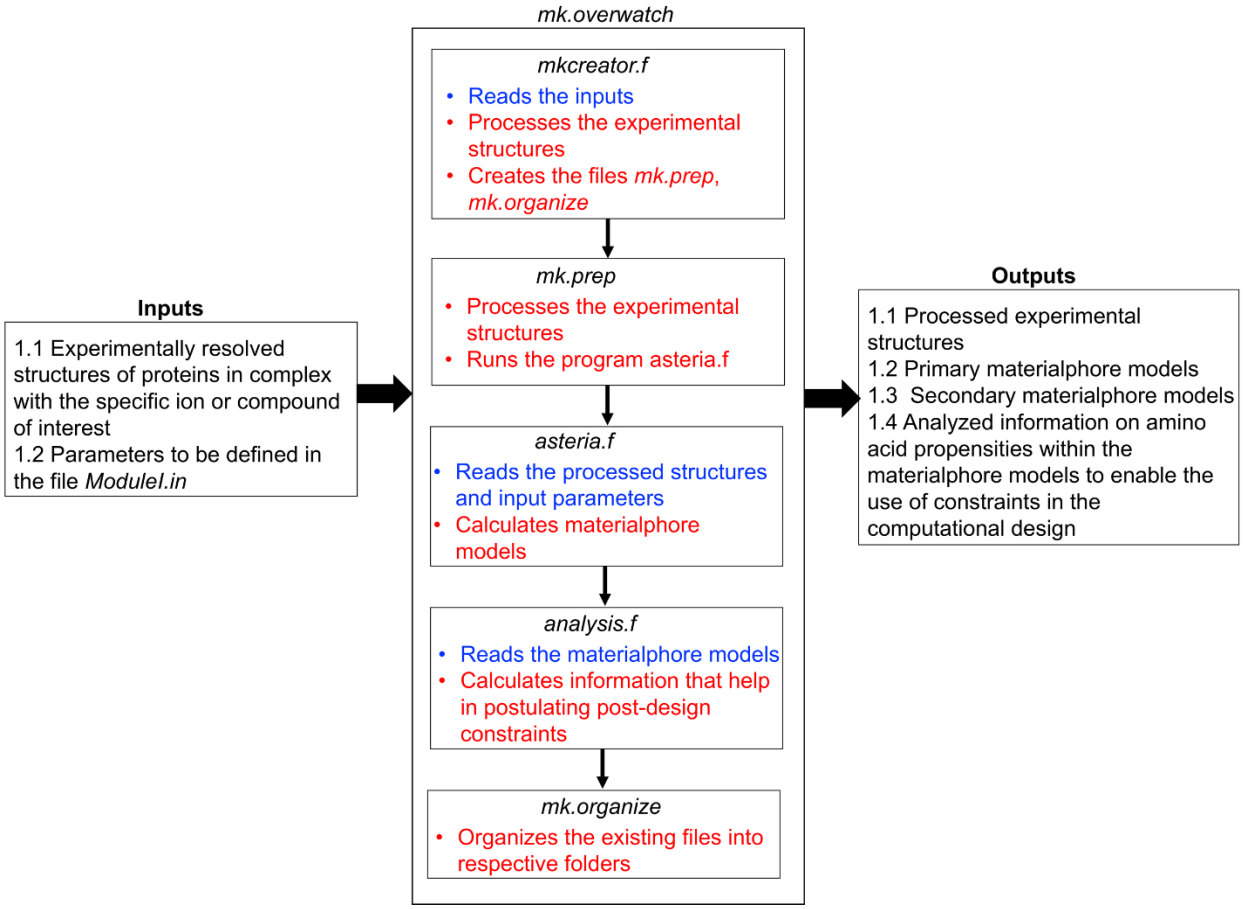


Figure 7.2 Overall schematic of programs and files in Module I. Inputs to a program or a script file is presented in blue color and calculations or outputs from a program or script files is present in red color.

7.2.5 Outputs

- The generated primary and secondary materialphore models are located in the folder: *~/ModuleI/ lig_codeoutputs* where lig_code is the input 2 in this module. The generated primary materialphore models are located in the file *lig_codereldist.txt* with the format: distance between a pair of amino acids followed by the identities of the amino acids followed by the unique code for a ligand (described above) and followed by the PDB ID of the experimentally resolved protein structure. The identities of the amino acids contain

both the standard three letter code for an amino acid as well as the residue number of the amino acid in the protein structure. Since some protein structures have same residue numbers for amino acids belonging to different chains, the chain name is appended to the residue number of an amino acid. The generated secondary materialphore models are located in the file *lig_codepocket.txt* with the format: distance between the ligand and an amino acid followed by the identity of the amino acid followed by the unique code for the ligand.

- The downloaded experimentally resolved protein structures from the PDB containing the ion or compound of interest are present in the folder: *~/ModuleI/ lig_codepdb*
- The processed experimentally resolved protein structures from the PDB without any remarks and only the coordinates of the atoms are present in the folder: *~/ModuleI/ lig_codecoord*
- Inputs used by various programs are present in the folder: *~/ModuleI/ lig_codeinputs.*
- Additionally, the folder *~/ModuleI/ lig_codepockets* contains the files *pocketstats.txt* and *stats.txt* which result from a statistical analysis on materialphore models.
    - *pocketstats.txt*: It contains the number of amino acids in each materialphore model and the frequency of materialphore models containing specific number of amino acids. The specific statistics can be used to reevaluate the maximum distance in Å between the ion or compound of interest and the side chain center of mass of an amino acid that is considered to be interacting (input parameter specification 3). If the number of amino acids of the most representative materialphore models is relatively high, the specific input parameter specification (i.e., the maximum distance in Å between the ion or compound of interest and the amino acid in a

binding motif) could be readjusted to include the nearest neighboring amino acids only.

- *stats.txt*: It contains information that can aid in determining post-design constraints, including amino acid propensities across all extracted and analyzed materialphore models, and providing the amino acid propensity by categorizing amino acids into families of polar, non-polar, aromatic and charged (further elaborated in the following module).

7.3 Module II: Optimization-Based Computational Design Model to Design Functional Amyloid Materials

7.3.1 Purpose

The main aim of this module is to design amyloid-forming peptides that can potentially capture the ion or compound of interest. For this purpose, mutations are introduced on the terminal non-β-sheet forming amino acids of amyloid designable scaffolds (such as the ones elucidated in the chapters 3, 5, and 6). Mutations are introduced according to the primary materialphore models calculated in module I to mimic how ion or compound of interest binds according to the experimentally resolved protein structures. Following the functionalization, the energy due to mutations is calculated for each designed peptide represented by an ensemble of flexible structural templates of elementary β-sheet structural units, and the designs are ranked in ascending order according to the energy. Energy minimization in the objective function aims to both stabilize the bound conformation of the designed amino acids within the β-sheets, as well as to avoid destabilizing the self-assembly properties of the designed peptides.

7.3.2 Input

The following amyloid designable scaffolds can be used as an input on which the computational design can be performed.

- Antiparallel or parallel flexible structural templates of the amyloid designable scaffold with sequence YATGAIIGNII[23] (elucidated in chapter 5)
- Antiparallel or parallel flexible structural templates of the amyloid designable scaffold with sequence RGDSGAITIGC[24] (elucidated in chapters 3, and 4)
- Antiparallel flexible structural templates of the amyloid designable scaffold with sequence AGKGAIIGFIK[22] (elucidated in the chapter 6).

  The flexible structural templates for the three aforementioned amyloid designable scaffolds are available in the shared folder of Dr. Tamamis' lab with the address: */mnt/research/Tamamis_Phanourios/Shared/ComDesFun/ModuleII/Scaffolds*.

7.3.3 Input Parameter Specifications

1. Inputs for module II can be provided in two ways. In the first way, information for inputs 2 to 11 below can be provided at the bash terminal. In the second way, information can be provided within the file *ModuleII.in*, and in this case, the default number 0 in the first line of the specific file should be changed to 1. The second input in the first line of the specific file refers to whether the design needs to be done based on primary materialphore models (described below) or by exploring all possible combinations of mutations without any specifically determined functionalization associated with the binding of specific ion or compound of interest. This option is currently applicable only for the design of antiparallel flexible structural templates (see below). The input in the former case would be "con" while in the latter case it would be "uncon".

2. The standard code for the ion or compound of interest according to the PDB[252]. For example, CS for cesium ion, E20 for the compound donepezil (same as input parameter specification 2 in module I).
3. The name of the designable scaffold on which design will be performed. Currently, there are three designable scaffolds available in the module. If the design needs to be performed on the scaffold AGKGAIIGFIK, answer as AGK, if the design needs to be performed on the scaffold YATGAIIGNII, answer as YAT, if the design needs to be performed on the scaffold RGDSGAITIGC, answer as RGD.
4. The number of flexible structural templates to be used or that are available for each of the three amyloid designable scaffolds. This depends on the designable scaffolds on which designs are going to be performed. For the three scaffolds, the maximum number of flexible structural templates provides is 50. They correspond to the top 50 most highly-ordered and well-aligned β-sheet conformations produced in the corresponding studies at which they were derived.
5. The number of peptide strands in a flexible structural template. For the three designable scaffolds, the number of peptide strands in a flexible structural template is 4.
6. If the number of peptide strands in a flexible structural template for the designable scaffolds is more than 1, then the configuration of the flexible structural templates, antiparallel or parallel should be provided as anti or para respectively. If the method of input is through moduleII.in file, then the specific entry should be placed at the second input of line 5.
7. The total number of residues per strand: the number of the residues per strand for all available amyloid designable scaffolds are 11 each.

8. The type of design to be performed depending on if the mutable positions belong to one or two adjacent β-sheet bonded peptide strands. If the amino acid positions forming the designed binding pocket are on the same strand, then the input entry should be 1; if the amino acid positions forming the designed binding pocket are on two adjacent β-sheet bonded peptide strands, then the input entry should be 2.
9. The number of mutable positions comprising the designed binding pocket in a flexible structural template of the designable scaffold. For example, in the design of amyloid materials binding to cesium (chapter 6)[22], the number of mutable positions were 4. Currently, the programs can handle from 3 to 5 mutable positions for antiparallel designs and 4 mutable positions only for parallel.
10. The maximum value of the distance between two to be designed amino acids in a designed binding pocket which can deviate from the corresponding distance in a primary materialphore model. For example, if the number of mutable positions is 4, then the number of distances between the mutable positions is equal to the number of combinations of 4 per 2, which is equal to 6. In the design of amyloid material binding to cesium ions (chapter 6)[22], 4 of the distances were allowed to deviate 2.5 Å and 2 of the distances were allowed to deviate 3.5 Å. The specific input is given in a single line where all the deviations are written separated by a space. For example, the input for the design of amyloid material binding to cesium ions was given as 2.5 2.5 2.5 2.5 3.5 3.5.
11. The positions of each designable position on a peptide strand. For example, in the design of amyloid materials binding to cesium ions (chapter 6)[22], the mutable positions were 2, 3, 10, and 11. The specific input is given in a single line where all the designable positions

are written separated by a space. For example, the input for the design of amyloid material binding to cesium ions was given as 2 3 10 11.

12. If the configuration of the flexible structural templates is selected to be antiparallel (input 6) and the type of design to be performed is type 2 (input 8), then the first mutable position on the C-terminal end. For example, in the design of amyloid materials binding to cesium ions (chapter 6)[22], this was determined to be residue position 10.
13. The fixed mutable positions if there are any. For example, in the design of amyloid materials binding to cesium ions (chapter 6)[22], residue position 1 was considered a fixed mutation to alanine. For example, to introduce this, two new lines should be introduced at which in the first line the number of mutable positions is specified (for example 1 2 3 if positions 1, 2 and 3 will be fixed), and in the second line the single letter code of the amino acids is provided (for example R G D if positions 1, 2 and 3 will be fixed).
14. The fraction of designed binding pockets in which the functionalization criteria should occur. For example, in the design of amyloid materials binding to cesium (chapter 6)[22], there were 10 flexible structural templates and each flexible structural template contained 6 binding pockets. The functional criterion was required to be satisfied in at least 5% (0.05 fraction value is given as input) of the 60 designed binding pockets. The specific fraction was denoted by λ in the *Eq. 6.7* of the previous chapter and the published study.[22]

7.3.4 Description

In-house Fortran programs were developed to execute the module containing the computational design model. The whole module can be executed by running the script file *mk.design* which in turn runs a set of programs and script files described below. Before running the module, output files from the folder *~/ModuleI/lig_codeoutputs* should be manually copied

into the current folder. Specifically, Module II requires the primary materialphore model informed which is present in the file *lig_codereldist.txt*. The inputs described above for this module are read by the program *mkgepol.f*. Depending on the input 1 choice, the program reads either input file method or command-based method. After processing the inputs, the program writes output files containing required information for the subsequent programs in the module. The program then creates a script file *mk.sasa* which evaluates the solvent accessibility surface area (SASA) of each of the designed amino acids in the flexible structural templates. The program *mkgepol.f* then creates a script file *mk.execdes* which copies files containing the structural information of the scaffolds from the folder *~/ModuleII/Scaffolds*. The script file *mk.execdes* also runs the script file *mk.sasa* to calculate the SASA and subsequently runs the programs *design_bridge_anti4.f* (or *design_bridge_para4.f* if the flexible structural templates have parallel configuration) and *overall_anti.f* (or *overall_para.f* if flexible structural templates have parallel configuration). The aforementioned programs are described in detail below.

The design program *design_bridge_anti4.f* (or *design_bridge_para4.f)* reads the required input information from the file *ligdesigninfo.txt* created by the program *mkgepol.f* and reads the primary materialphore models from the file *lig_codereldist.txt* provided as output in module I. The design program then calculates the distances between the amino acids that form the designed binding pocket in a template. The aforementioned distances are calculated for each designed binding pocket in all flexible structural templates of the designable scaffold. Functionalizing selected amyloid designable scaffold based on the primary materialphore models is executed in a pyramid style and the program currently can functionalize designed binding pockets having 3, 4, or 5 mutable positions. Let the number of mutable positions be n_{mut} and $x_1, x_2, \ldots, x_{nmut}$ be the mutable positions on a peptide strand under the condition that $x_1 < x_2 < \ldots < x_{nmut}$. The design

program is executed for each primary materialphore model, each flexible structural template, each designed binding pocket in a flexible structural template independently. Specifically, if the total number of primary materialphore models are p, the number of flexible structural templates within an amyloid designable scaffold are n, and the number of designed binding pockets in a flexible structural template are d (i.e., 6 in an elementary β-sheet structural unit composed of 4 peptide strands), then the total number of iterations of the design program core algorithm (described below) are $p*n*d$. For an iteration i, the core algorithm of the design program is executed in stages.

In the first stage, each distance between two amino acids in a designed binding pocket is compared with the corresponding distance in a primary materialphore model. The comparison is done separately for each distance in each designed binding pocket with every corresponding distance in the primary materialphore model. During comparison, each distance in the designed binding pocket is allowed to deviate between 1.5 Å and 3.5 Å with 0.5 Å increase from a distance in the primary materialphore model. If a match occurs between a distance in the designed binding pocket and a distance in the primary materialphore model, then the identities of the amino acids involved with the specific distance in the primary materialphore model are saved. The set of amino acid combinations, denoted via symbol "-", which are generated at this stage are $\{x_1\text{-}x_2,\ x_1\text{-}x_3,\ x_1\text{-}x_{nmut-1},\ x_1\text{-}x_{nmut},\ x_2\text{-}x_3,\ x_2\text{-}x_{nmut-1},\ x_2\text{-}x_{nmut},\ \ldots,\ x_{nmut-1}\text{-}x_{nmut}\}$. It should be noted that at this stage only the identities of amino acids involved in the pair are known. For example, the identities of the amino acids involved in the pair x_1-x_2 are known initially but the identity of amino acid at positions x_1, or x_2 is at this point not known. If the number of distances of the total amino acid pairs in a primary materialphore model is n_p (e.g., for a materialphore model comprising of 10 amino acids n_p is number of combinations of 10 per 2) then the possible number of matches of a distance between a pair of two amino acids in the designed binding pocket is between 0 and n_p. Thus,

matching between all possible pairs of distances in a primary materialphore model and a designed pocket is performed exhaustively, and all matches are equally considered for further investigation. In this stage, the matching of distances in the designed binding pocket is considered to be independent of each other. In the special case where the number of matches for a distance in the designed binding pocket is 0, then the evaluation for iteration i is skipped and the program continues to the next possible iteration, $i+1$.

In the second stage, matching pairs of two are aimed to be expanded into matching pairs of three. If the set of amino acid combinations generated in the first stage are $\{x_1\text{-}x_2,\ x_1\text{-}x_3,\ x_1\text{-}x_{nmut-1},\ x_1\text{-}x_{nmut},\ x_2\text{-}x_3,\ x_2\text{-}x_{nmut-1},\ x_2\text{-}x_{nmut},\ldots,\ x_{nmut-1}\text{-}x_{nmut}\}$, then amino acid combinations $\{x_1\text{-}x_2\text{-}x_3,\ x_1\text{-}x_{nmut-1}\text{-}x_{nmut},\ x_2\text{-}\ x_3\text{-}x_{nmut-1},\ldots,\ x_{nmut-2}\text{-}\ x_{nmut-1}\text{-}x_{nmut}\}$ are generated and examined. While combining the amino acid combinations, attention is paid so that there is at least one amino acid common between the combinations that are combined, otherwise matching is rejected. For example, the pairs $x_1\text{-}x_2$, and $x_1\text{-}x_3$ since x_1 is common. While combining the pairs, as the identities of amino acids at individual positions are still not known, each amino acid identity in a combination of two is compared with each amino acid identity in a combination of other two. If the identity of amino acids between the combinations match, then the combinations are further evaluated to make sure that the identity of other two amino acids across the two combinations are not same. Subsequently, the identities of amino acids involved in each combination are stored for further stages evaluation. Additionally, the deviation of the distance in the designed binding pocket from the corresponding primary materialphore model distance assigned in the first stage is also stored. For example, if the $x_1\text{-}x_2$, and $x_1\text{-}x_3$ are to be combined, then each amino acid identity involved in the combination $x_1\text{-}x_2$ is compared with each identity of amino acid in the combination $x_1\text{-}x_3$ which results in a total of four comparisons (i.e., x_1 of the first pair to x_1 of the second pair, x_1 of the first pair to x_3 of the

second pair, x_2 of the first pair to x_1 of the second pair, x_2 of the first pair to x_3 of the second pair; x_1 of the first pair to x_1 of the second pair is compared because the individual identify of x_1 is not known at this point). The identities of other two amino acids, supposedly x_2 and x_3, are evaluated to ensure they are not same. The identities of amino acids at positions x_1, x_2, and x_3 are stored along with deviations associated with the combinations x_1-x_2, and x_1-x_3. The identities of amino acids at different positions are fixed in each combination of three after this stage. For example, the identities of amino acids x_1, x_2, and x_3 in the combination x_1-x_2-x_3 are known after this stage unlike first stage where individual identity at positions x_1 and x_2 are known in the combination x_1-x_2. If the number of combinations after the first stage is odd, the remaining combination in the second stage is combined with a combination generated in the subsequent stage (see below).

In the subsequent stages, two combinations of all possible matching three entities (initially from the previous stage), or more (progressively) are generated and combined. This is done in such a way that there are as many amino acid positions common between the combinations as possible, so that all combinations are considered as equal feasible solutions. The number of stages (progressively) depends on the number of mutable positions constituting the designed binding pocket and the stages are executed until the combination x_1-x_2-x_3-….-x_{nmut-1}-x_{nmut} is generated, which constitutes the final stage in the progressive search. At any stage in the progressive search, if the number of matches for a specific combination are 0, then the evaluation of further stages for the iteration i is not performed and the iteration value is increased to $i+1$. While progressively searching, deviations between the distances in the primary materialphore models and the distances of amino acid combinations placed at different positions are considered, guaranteeing that distance matching is within the maximum allowed values. For the final stage, there are number of combinations of n_{mut} per 2 deviations which were originally assigned during the first stage. The

specific set of deviations are compared with the set of deviations which can be provided through input parameter specification 10 (see above). For the comparison, both the sets of deviations are arranged from highest value to lowest, and one on one comparison is done. If the deviation is greater than the specified deviation, then the whole combination generated after the final stage is discarded. The above generated successful combinations of amino acids represent "slices" of a materialphore model. For a given materialphore model, the slices represent combination of n_{mut} (input parameter specification 9) amino acids interacting with the ion or compound of interest. While the notion of slices was not explicitly mentioned in chapter 6, we clarify here that such slices were used as number of mutable positions in the designed binding pocket (4) and the number of amino acids in the primary materialphore models did not necessarily match.

The core algorithm of the design program *design_bridge_para4.f* has additional constraints during second stage to final stage. This is because the type of amino acids present at some positions should be same in parallel flexible structural templates. For example, if the set of mutable positions are $\{x_1=1, x_2=2, x_3=10, x_4=11\}$, then x_1, x_2 belonging to a strand and x_1, x_2 belonging to the adjacent strand form a designed binding pocket. It should be noted that in a design involving two strands and antiparallel flexible structural templates, all the designed binding pockets in a flexible structural template are identical whereas in a design involving two strands and parallel flexible structural templates, there are two sets of identical designed binding pockets. In case of single stranded designs, the configuration (antiparallel or parallel) of the flexible structural templates is irrelevant as the mutable positions constituting the designed binding pocket belong to the same peptide strand. The number of designed binding pockets in case of a design involving two strands is number of combinations of n_{str} per 2 where n_{str} is number of identical peptide strands in a flexible structural template (input 5). The number of designed binding pockets in case of a design involving

one strand is n_{str}. The specific style of execution performed here ensures that all the distance constraints are simultaneously satisfied.

The generated combinations which qualified all the above stages described constitute a new designed peptide sequence. Subsequently, for each design sequence, energy due to mutations is evaluated by introducing the identities of amino acids involved in the combination using a coarse-grained fashion. The evaluation of energy is done using the SIPPER surface force field[205] and the values of interaction energy between any two of the twenty of amino acids is represented as a single value. The interaction energy values between the amino acids is presented in the file *sippersurface.txt* in the folder *~/ModuleII/*. The specific file contains a total of 400 interaction energy values in the form of a 20X20 matrix and amino acids are arranged in the alphabetical order as follows: alanine, arginine, asparagine, aspartic acid, cysteine, glutamine, glutamic acid, glycine, histidine, isoleucine, leucine, lysine, methionine, phenylalanine, proline, serine, threonine, tryptophan, tyrosine, valine. For example, value at the position 1X1 represents the interaction energy between two alanines; values at the position 1X2 and 2X1 represents the interaction energy between an alanine, and an arginine. The energy term that is being evaluated has two components: 1) the pairwise interaction energy between amino acids 2) desolvation free energy of nonpolar surfaces, which is calculated from the change in solvent-accessible surface area. The details about the evaluation of energy term is presented in the ref[22]. It should be noted that the interaction energy between two amino acids is evaluated only when at least one of amino acids is at one of the positions under modification, and that the interaction energy values between two interaction amino acids are considered when two interacting amino acids are within 6.5 Å in a given flexible structural template. The evaluation is done by assuming that the designed amino acids are present on all the identical strands in a flexible structural template and the energy is averaged over all the

flexible structural templates. It should also be noted that all the above described energy evaluation is done for the iteration *i* (refer above), that is for a specific combination generated from a specific primary materialphore model, specific flexible structural template, and a specific designed binding pocket in that template. However, the energy calculations do include substitutions to all pockets. The identities of the amino acids along with the evaluated energy are written to the output file *energies_anti.dat* (or *energies_para.dat*). Subsequently the iteration *i* is increased to *i+1* and the above described design algorithm is executed again. Following the completion of all iterations, all the design sequences are sorted according to their energy value from low to high following the objective function under minimization. The specific program *design_bridge_anti4.f* (or *design_bridge_para4.f*) is schematically presented in the Figure 7.3.

The specific program *design_bridge_anti4.f* can also perform design by exploring all possible combinations of mutations without any specifically determined functionalization associated with the binding of specific ion or compound of interest. In this case, the program skips reading the information associated with primary materialphore models, and the model is solved only for any constraints provided (associated with sequence or the peptide arrangement into antiparallel β-sheets). Subsequently, the stages associated with functionalization using primary materialphore models are also skipped. As described above, for each design sequence generated by exploring all possible combinations, energy due to mutations is evaluated by introducing the identities of amino acids involved in the combination using a coarse-grained fashion. Following the completion of exploration of all possible combinations, the design sequences are sorted according to their energy value from low to high following the objective function under minimization. Currently this option is unavailable for the design of peptides bonded in parallel β-sheets. While this is a direction currently considered in the lab, the minimalistic approach to design

both antiparallel and parallel β-sheets without the introduction of materialphore constraints has been accomplished in chapter 4. The design program for antiparallel and parallel β-sheets is *design_anti_final.f* and *design_para_final.f*, respectively available in the shared folder of Dr. Tamamis' lab within the address: */mnt/research/Tamamis_Phanourios/Shared/Example-cases/Minimalistic_approach_chapter4.* The analyzed highly-ordered and well-aligned β-sheets formed by the designed peptides are provided in */mnt/research/Tamamis_Phanourios/Shared/Example-cases/chapter4*, and could be additionally used as amyloid designable scaffolds in future studies.

While performing the design for functionalization using primary materialphore models, it is possible that identical sequences can be produced in the following four ways: 1) two designed sequences could originate from two different primary materialphore model but result in the same designed peptide, 2) two designed sequences could originate from different slices of the same primary materialphore model and result in same designed peptide, 3) two designed sequences could originate from the same slice of a primary materialphore model but the identities of amino acids at the mutable positions are different and result in same designed peptide, and 4) two designed sequences could be identical with respect to the identities of amino acids at all the designable positions but they originate from different designed binding pockets or different flexible structural templates. All the designed sequences generated are subjected to a screening where the designed sequences that don't occur at least in a user-specified fraction of designed binding pockets (input parameter specification 14) are discarded. The above evaluation is done by counting the identical designed sequences (fourth category described above) and applying the specified fraction (denoted by λ in the *Eq. 6.7* of chapter 6 and the published study.[22]). The specific evaluation is done in the program *overall_anti.f* (or *overall_para.f*). The specific program also

produces a file *sequencelist.txt* that lists the unique peptide designs and their evaluated energy. Two designed sequences are considered different if they don't fall into either of the four categories described above. The program also produces a file *presequence.txt* that lists all the non-identical designed sequences and their evaluated energy value.

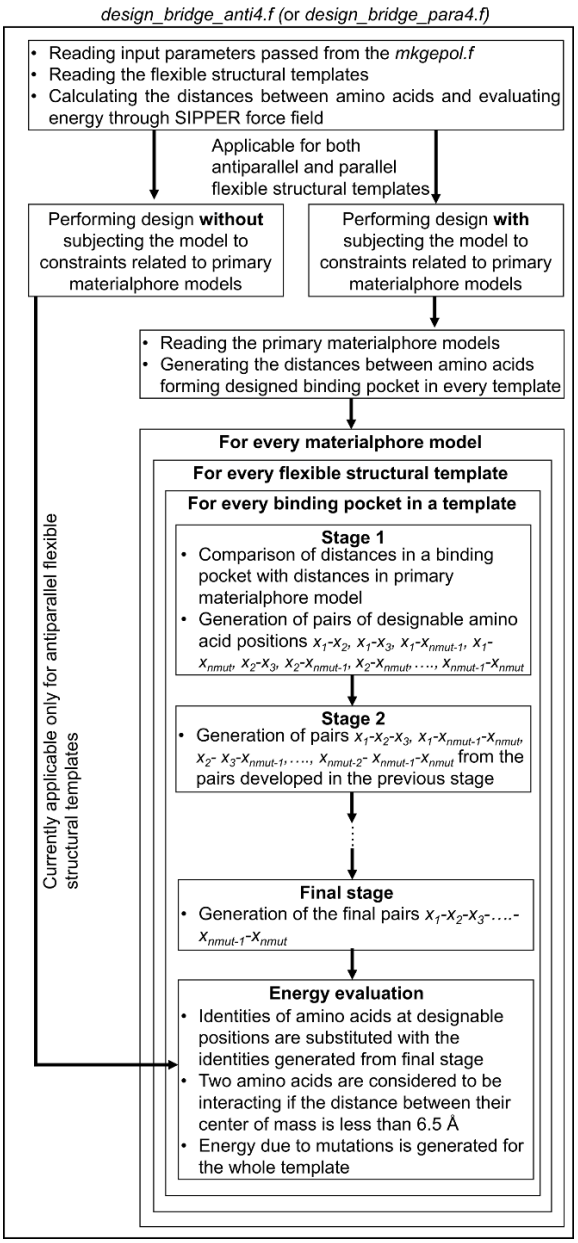


Figure 7.3 Schematic of the design program *design_bridge_anti4.f*.

Extra ion or compound of interest specific constraints can be applied if necessary, by running the program *custom_constraints.f*. The following type of constraints can be introduced: 1) Amino acid types that may/may not be allowed to appear at the designable positions and the number of instances an amino acid can occur out of the total designable positions. 2) Charge constraints over the total designable positions. For example, the following can be specified: the net charge of the designable positions can be between -2 and -1. 3) Constraints at specific designable positions to a set of amino acids. For example, the following constraint can be specified: the first and last designable positions to aspartic acid and glutamic acid. 4) Constraints based on the amino acid types (aromatic, charged, polar, and non-polar).

The optimization-model is constructed in such a way so that the lowest in energy solution represents the global optimum given the introduced constraints. Nevertheless, apart from the global optimum solution, additional solutions ranked low in energy can be evaluated-validated using subsequent simulations. Validating additional solutions with low energies (including the global optimum) can be considered a useful strategy given the coarse-grained energy representation used to solve the problem. The number or percentage of designs to be selected for further investigation could be adjusted based on the number of solutions produced, and depends on the ion or compound of interest. Selecting too few sequences will considerably lower the chances of getting a successful design to capture the ion or compound of interest. Selecting too many sequences requires high computational and experimental resources. Thus, in addition to the points referred above 5) one can select sequences by using an energy value threshold, and 6) one can select a specific subset (number) of sequences from the global minimum solution.

Outputs from Module I, particularly the files in the folder *~/ModuleI/ lig_codepockets* could be helpful in formulating the constraints and the applying ion or compound of interest

specific constraints can help achieve successful designs with less resources. All or any of the constraints can be introduced by modifying the file *custom_constraints.in*. The format of the specific file is described in the file *custom_constraints_input_format.txt*. Designs selected after introducing any or all of the constraints-selections 1 to 5 are written into the file *sequencelist_cc.txt* whereas the designs selected after exercising constraint 6 on the designs in the file *sequencelist_cc.txt* are written into the file *sequencelist_cc_top.txt*. The overall schematic of the programs and files for the specific module is presented in Figure 7.4.

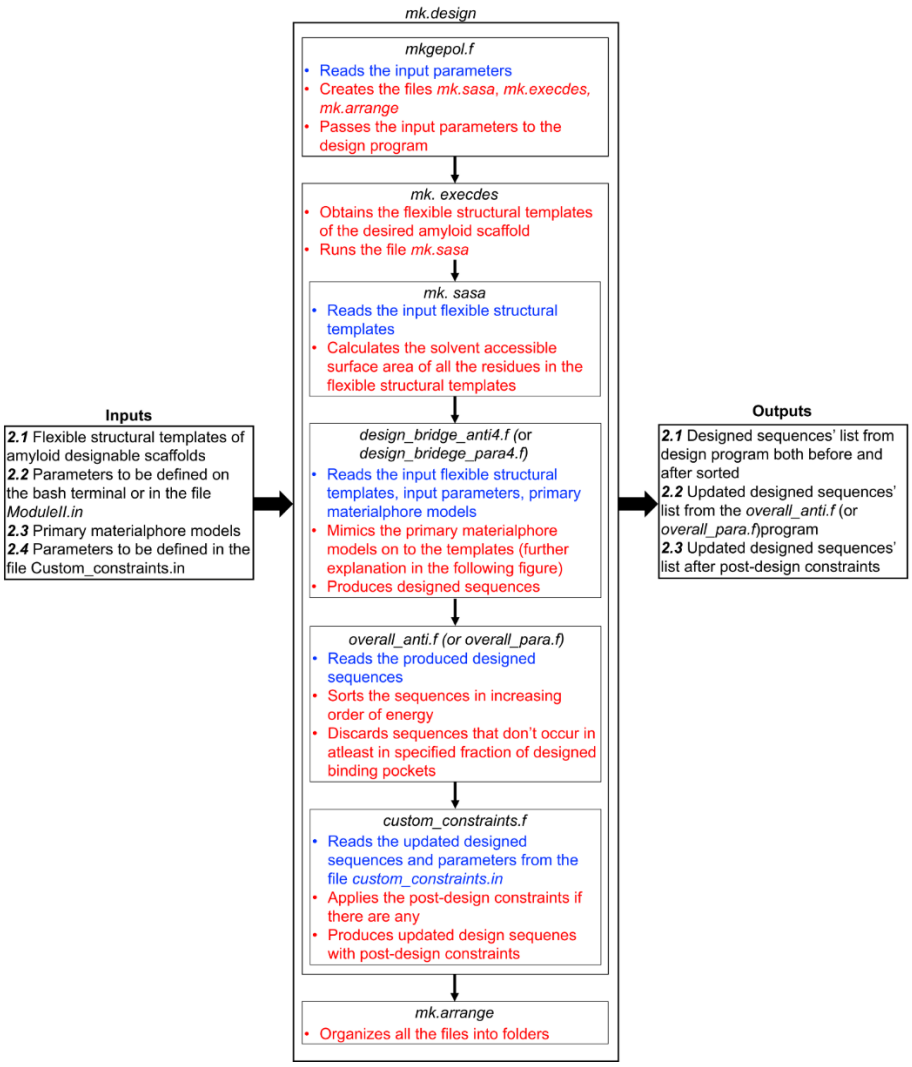


Figure 7.4 Overall schematic of programs and files in Module II. Inputs to a program or a script file is presented in blue color and calculations or outputs from a program or script files is presented in red color.

7.3.5 Commands

To run Module I and Module II, execute the following the commands:

- chmod 755 mk.execute
- ./mk.execute

7.4 Module III: Performing MD simulations and Two-Component Validation to Select Promising Designed Amyloid Materials

7.4.1 Purpose

The main aim of this module is to verify whether the binding pockets in the designed peptides resemble the materialphore model from which they are derived. For this purpose, designed amino acids are introduced at the mutable positions on the designable scaffold and short MD simulations are performed. During the simulations, the amyloidogenic core of the designed peptides is constrained so that the simulating structure maintains the β-sheet interactions. However, the designed amino acids are unconstrained and flexible. The simulations are performed in the absence of ion or compound of interest which serves as ultimate test to check whether the designed binding pockets in the simulated snapshots can adopt proper geometry to bind to an ion or compound of interest in its absence. Inspection of the MD simulation snapshots is recommended to ensure that the duration of the MD simulations is adequate to allow the mutated side chains to explore efficiently the conformational space.

Upon the completion of the short MD simulation runs, the designed peptides are computationally investigated to check the formation of pockets resembling the geometry of the amino acid : ion or compound of interest binding motifs using a two component validation procedure. In the first validation component, the center of mass of the designed amino acid side

chains are determined to verify whether they form appropriate side chain conformations to form pockets with geometries resembling the primary materialphore model(s). In the second component of the validation procedure for an ion, the ability of the designed amino acid side chain pockets to encapsulate the ion in the same fashion as the secondary materialphore model(s) from which the designed amino acids originated is evaluated; for a compound where the materialphore models are generated with respect to the center of mass (COM) of the compound, the ability of the designed amino acid side chain pockets to encapsulate the COM of the compound in the same fashion as the secondary materialphore model(s) from which the designed amino acids originated is evaluated; for a compound where the materialphore models are generated with respect to a specific atom of the compound, the ability of the designed amino acid side chain pockets to encapsulate the specific atom of the compound in the same fashion as the secondary materialphore model(s) from which the designed amino acids originated is evaluated. The evaluation is done only for the pockets containing primary matches. It should be noted that in the computational validation using the aforementioned short MD simulations, it is assumed that the designed peptides self-assemble into β-sheet structures with elementary structural units resembling the amyloid designable scaffold's flexible structural templates on which design was performed.

7.4.2 Input

A dcd trajectory converted to PDB containing short MD simulation snapshots for each of the selected peptides designed to bind the specific ion or compound of interest.

7.4.3 Input Parameter Specifications

1. N-terminal and C-terminal ends for the designed peptide. The N-terminal end can be natural (designated as NTER) or acetylated (designated as ACE) and the C-terminal end can be natural (designated as CTER) or amidated (designated as CONH2).
2. Job name when submitted to the supercomputing system.
3. Number of short MD simulation snapshots for each designed peptide on which validation is going to be performed.
4. Tolerances for the relative distances between the designed amino acid positions. Each relative distance can have its own tolerance and the tolerance can range from 1.0 to 3.5 Å.
5. Tolerances for the distances between the designed amino acid and the element or compound of interest. Again, each distance can have its own tolerance value and the value can range from 1.0 to 3.5 Å.

7.4.4 Description

The lists containing the designed peptide sequences selected at the end of Module II (*sequencelist.txt* or *sequencelist_cc.txt* or *sequencelist_cc_top.txt* and *presequence.txt*) should be copied into Module III along with file containing the information about mutable positions (*ligdesigninfo.txt*). Short MD simulations are performed for each selected designed peptide by introducing mutations at the designable positions on a flexible structural template on the designable scaffold. The flexible structural template selected is one of the flexible structural templates used for computational design in ModuleII. The selected flexible structural template with the name *fibril.4.pdb* is automatically copied into ModuleIII at the end of ModuleII. The simulations are performed using CHARMM[110] with CHARMM19[111] forcefield and FACTS19[123] implicit solvent model. Input simulation file for each designed peptide named as *md.inp* is created

by the program *inpcreator4.f*; the right input file (*sequencelist.txt* or *sequencelist_cc.txt* or *sequencelist_cc_top.txt*) should be specified. The specific program takes the terminal ends of the designed peptides as input (input 1) and constrains the amyloidogenic core of the designed peptides under light constraints so that the designed peptides maintain structural integrity. Subsequently, a folder is created for each designed peptide with the single letter code of the designed amino acids as the name of the folder. Program *mkcreator4.f* creates a script file *mk.direc* which copies the created simulation input file *md.inp* into the folder for each designed peptide; similarly, the right input file (*sequencelist.txt* or *sequencelist_cc.txt* or *sequencelist_cc_top.txt*) should be specified. The script file also copies the flexible structural template, *fibril.4.pdb*, topology, and parameter files required for simulation into the folder for each designed peptide. Subsequently, all the designed peptide folders are moved into a folder, *Simulation*. The script file also runs a program *jobcreator.f* which creates script files to run the simulations on the supercomputing system, Ada. The script file *mk.direc* also copies the files *lig_codereldist.txt*, *lig_codepocket.txt* from Module I and *presequence.txt* from Module II which are required to perform validation after the short MD simulations. The folder, *Simulation*, which contains all the files required for performing short MD simulation for each designed peptide needs to be moved to the supercomputing system. After moving the folder onto supercomputing machine the script file, *mk.bsub*, can submit the MD simulation for each designed peptide. The simulation time for each designed peptide is 20 ns and can be changed accordingly in the CHARMM input file *md.inp* which generates the simulation trajectory. The overall schematic of the programs and files involved in performing MD simulations in Module III is presented in Figure 7.5.

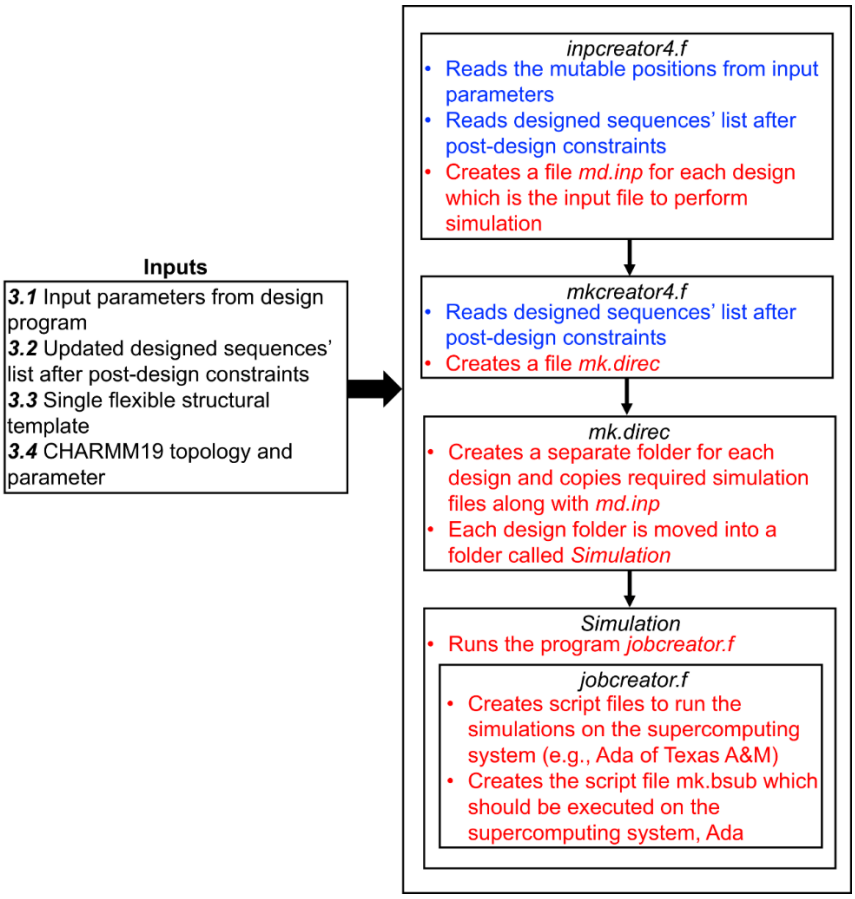


Figure 7.5 Overall schematic of programs and files involved in performing MD simulation in Module III. Inputs to a program or a script file is presented in blue color and calculations or outputs from a program or script files is presented in red color.

Upon completion of the short MD simulations, the script file *mk.head* should be executed. The script file in turn runs a program, *validation_prep.f*, which prepares files need to run the two-step validation procedure. The specific program requires as input information the number of simulation snapshots (input parameter specification 3), tolerances for first step (input parameter specification 4) and second step (input parameter specification 5) of the validation procedure (described below) as inputs, and also reads the file, *ligdesigninfo.txt*, which contains information about the mutable positions. The specific program creates the file, *validinfo.txt*, which takes the information required for validation from the file *ligdesigninfo.txt*. The program, *validation_prep.f*,

reads the file, *presequence.txt*, which contains the list of designed sequences for each type of designed peptide. The specific program creates a file, *sequence.txt*, for each design which contains the design sequences. If the N-terminal end of the designed peptides is acetylated, CHARMM19[111] force field considers acetylation as an amino acid during the simulation. Thus the program, *validation_prep.f*, creates a script file, *mk.renumber*, which removes the acetylation as an amino acid, integrates acetylation as a part of next amino acid, and renumbers all the amino acids of a designed peptide. The program subsequently creates a script file, *mk.executeval*, which converts the produced simulation output file from dcd format to pdb format so that validation can be performed. The script file, *mk.executeval*, for each designed peptide is run in parallel by the scrip file, *mk.mastexe1*, which uses the help of *tamulauncher*. The created pdb file containing the simulation snapshots is under the name, *foldername_ligcode_fin.pdb*, where foldername is the single letter designed amino acid codes using each designed peptide's folder is named. The program, *validation_prep.f*, also creates the script file, *mk.mastexe2*, which runs the two-step validation program, *validation_anti.f* (or *validation_para.f*), described in detail below. Finally, the program, *validation_prep.f*, creates a bash file, *validscript.sh*, which contains the information required to run the validation batch job on the supercomputing system ada. The bash file, *validscript.sh*, first runs the script file, *mk.mastexe1*, then runs a script file, *mk.intermediate*, which copies the files *ligcode_reldist.txt*, *ligcode_pocket.txt*, *validinfo.txt*, *sequence.txt* which are required for the validation into each designed folder. The bash file, *validscript.sh*, finally runs the script file, *mk.mastexe2*.

The program, *validation_anti.f* (or *validation_para.f*) performs the two-step validation procedure for each designed peptide independently. The program starts by reading required information to perform the design from the file *validinfo.txt*. The program then reads the pdb file

containing the simulation snapshots and calculates the distances the between designed amino acids in each designed binding pocket of each short MD simulation snapshot. The program then reads the list of the designed sequences from which the designed peptide originated from the file *sequence.txt*. Based on the identities of the amino acids involved in a design sequence, the program assigns the distances between two amino acids in the design sequence. The specific distances are extracted from the file, *ligcode_reldist.txt*, which contains the primary materialphore models. During the first step of the validation, the total number of iterations performed are *s*n*d*, where *s* is the number of the design sequences for a designed peptide, *n* is the total number of simulation snapshots for each designed peptide, *d* is the number of designed binding pockets in a simulation snapshot. For an iteration *i*, the distances between a pair of amino acids in a designed binding pocket of a MD simulation snapshot are compared with the distances between a pair of amino acids in a design sequence. While comparison, each distance in the designed binding pocket is allowed to deviate between 1.0 Å and 3.5 Å with 0.5 Å increase from a corresponding distance belonging to the design sequence. The specific set of deviations are compared with the set of deviations given by the user (input parameter specification 4). For the comparison, both the sets of deviations are arranged from highest value to lowest value and one on one comparison is done. If a specific deviation is greater than the user specified deviation, then the iteration *i* is increased to *i+1* without saving anything. On the contrary, if the two set of distances match within the user specified deviation, then the designed binding pocket is considered as a primary match. For a designed peptide, theoretically, there can be *s*n*d* number of primary matches. Subsequently, we calculate the percentage of snapshots containing at least one primary match.

For the second step of the validation procedure, the program assigns the distances between an amino acid in the design sequence to its corresponding ion or compound of interest. The specific

distances are extracted from the file, *ligcode_pocket.txt*, which contains the secondary materialphore models. In contrast to the first step of the validation procedure, the set of distances assigned from the materialphore model (secondary in this case) cannot be directly compared to a set of distances of the designed binding pockets as the ion or compound of interest is not actually present in the designed binding pockets of the MD simulation snapshots. Instead, we hypothetically attempt to place the ion or compound of interest in the designed binding pockets to check the feasible points that can match the corresponding distances of the designed sequence from the secondary materialphore models. The volume around each designed binding pockets is divided into spherical grids with each sphere having a radius of 0.5 Å. We assume that each of these spherical grids represents the ion or compound of interest and calculate the distance between the amino acid and a specific spherical grid. If these distances match the corresponding distances for a designed sequence from a secondary materialphore model, then the specific designed binding pocket is considered as a secondary match. It should be noted that the second step of validation is performed only for the designed binding pockets which are primary matches. Subsequently, we calculate the percentage of designed binding pockets containing secondary matches given that the designed binding matches are already primary matches. The overall schematic of programs and files involved in performing two-component validation of Module III is presented in Figure 7.6.

7.4.5 Commands

To run Module III, execute the following the commands:

- chmod 755 mk.head
- ./mk.head

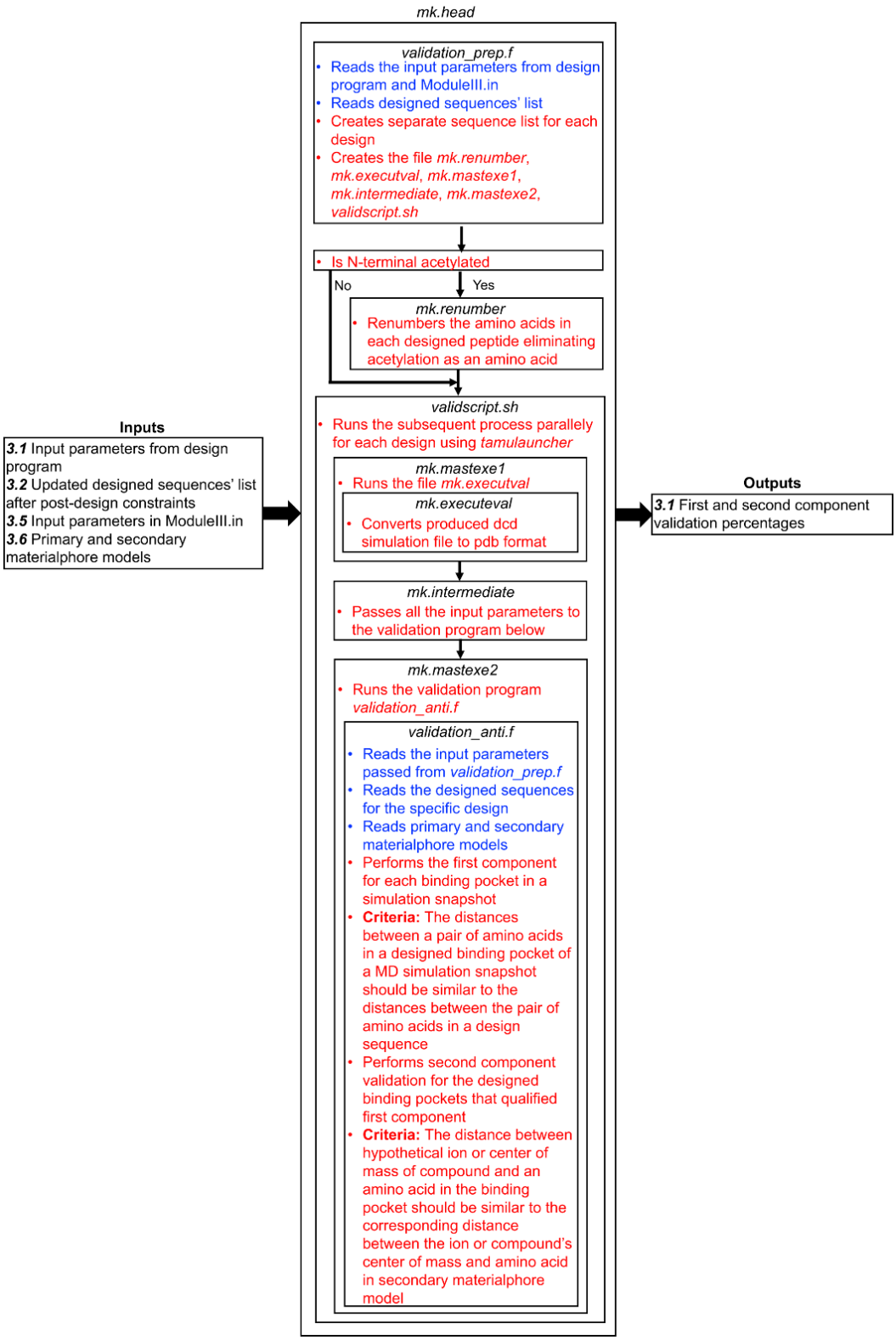


Figure 7.6 Overall schematic of programs and files involved in performing two-component validation in Module III. Inputs to a program or a script file is presented in blue color and calculations or outputs from a program or script files is presented in red color.

7.4.6 Output

Upon completion of the two-component computational validation for each of the designed peptides, the percentage of snapshots containing a primary match is written into the file *cutoff1.txt* in each designed peptides' folder. The percentage of pockets containing the secondary matches within the snapshots containing primary is written into the file *cutoff2.txt* in each designed peptides' folder. The percentage of primary and secondary matches of all the designed peptides are rank ordered according to the percentage of primary matches and are used as a basis to select designed peptides for further evaluation.

The designed peptides selected after performing Module III could be investigated using simulations investigating the designed peptides' self-assembly properties. The specific simulations performed can computationally investigate the designed peptides' capacity to self-assemble into amyloid β-sheet structures containing designed pockets appropriate for functionalization. The simulations and along with the elucidation of the elementary β-sheet structural unit elucidation are performed by implementing the computational protocol developed by Tamamis and Archontis[130] described in chapter 2. The simulations performed here are also in the absence of the ion or compound of interest which could serve as an ultimate check the capacity of the designed peptides to form amyloid structures in which the designed pocket of the elementary β-sheet structural unit can adopt the proper geometry to bind to the ion or compound of interest in its absence. Designed peptides should have the predominant configuration same as the configuration of the flexible structural templates of amyloid designable scaffolds they originated from as otherwise the formation of opposite configuration elementary β-sheet structural units would disallow the formation of the intended designed amino acid functional pockets capable of binding to the ion or compound of interest. Implementing the same protocol[130], highly ordered and well-aligned β-sheet

structures for each designed peptide. The extracted β-sheet structures the two-component validation can be performed again to identify the most promising peptides.

The protocol presented above was already executed in chapter 6 for the design of amyloid materials binding to cesium ions; the files are located in *mnt/research/Tamamis_Phanourios/Shared/Example-cases/CS.* In addition, the protocol as described in chapter 7 has been additionally implemented in three example cases. In the first case, using the protocol functional amyloid materials have been computationally designed to capture Alzheimer's disease drugs donepezil, and tacrine (presented in chapter 8; located in *mnt/research/Tamamis_Phanourios/Shared/Example-cases/AD*). In the second case, the protocol has been used to facilitate the rational design of amyloid materials that are used as DNA carriers and can penetrate cell (presented in chapter 9; located in *mnt/research/Tamamis_Phanourios/Shared/Example-cases/positive*). In the third case, the protocol has been used to design amyloid materials with cell adhesion properties as well as binding properties to calcium and strontium ions (presented in chapter 10; located in *mnt/research/Tamamis_Phanourios/Shared/Example-cases/CASR*).

In addition, the protocol was implemented for two test cases involving the design of amyloid materials binding to gallium ions and quercetin, with mutations introduced to the antiparallel β-sheet flexible template structures of YATGAIIGNII (mutated positions are underlined). The test cases were performed upto the stage of Module II and are located in *mnt/research/Tamamis_Phanourios/Shared/Test-cases/GA* and *mnt/research/Tamamis_Phanourios/Shared/Test-cases/QUE*, respectively. The test cases aimed to show the versatility of the protocol, and thus, no detailed analysis was performed on the results produced.

8. AMYLOID PEPTIDE SCAFFOLDS COORDINATE WITH ALZHEIMER'S DISEASE DRUGS

8.1 Introduction

8.1.1 Overview and Significance

Alzheimer's disease (AD) is an irreversible neurodegenerative disorder that develops slowly and worsens over time,[273] characterized by progressive deterioration of cognitive behavior and functionality that impairs significantly the activities of daily life.[274,275] AD is strongly related with acetylcholine (ACh), a neurotransmitter that is found in the brain and blood,[276,277] which is eleased by nerve cells to transfer signals to other cells, related to memory, motivation, language, and muscle contraction.[278] A characteristic of AD is the low concentration of ACh between cholinergic synapses in AD patients that makes neuroconnection extremely difficult. Current approved AD drugs, apart from Memantine which is an N-methyl-D-aspartate (NMDA) receptor antagonist, are acetylcholinesterase (AChE) inhibitors that block the enzyme in the cholinergic system (regulating the levels of ACh for postsynaptic nerve stimulation), allowing for higher accumulation of the neurotransmitter ACh in synapses, and therefore higher cognition. Donepezil, Galantamine, Rivastigmine, Memantine, and a Donepezil/Memantine combination are currently administered to patients depending on the progression stage. The first three comprise Cholinesterase inhibitors, while Memantine acts as an antagonist that blocks the NMDA receptors (glutamate receptors and ion channel proteins) found in nerve cells, reducing the glutamate neuroactivity. Though the aforementioned 4+1 AD drugs have been approved by FDA and used clinically, modest and transient therapeutic effects have been witnessed so far,[279] while minimal-to-negligible cognition is confessed by neurologists, caregivers, and primary care providers, most

likely due to the short half-life of the inhibitors.[280,281,282,283,284] Formulations that could enhance the effectivity of current marketed AD drugs and their delivery to the target enzyme seem to be the most optimal option in order to stabilize or even enhance cognition.

Recent studies showed the potential application of amyloid- or amyloid-like materials as carriers or compounds (including drugs and antimicrobials, e.g., ciprofloxacin, rhodamine B, and polyphenols,[43,44,285]). In addition, a most recent study showed that amyloid materials bind to iron with high affinity for the therapeutic release of iron ions to treat iron deficiency anemia in rodents,[45] and suggested the importance of amyloid materials as carriers of ions or compounds for their therapeutic release.

As outlined in chapter 1, the functionalization of amyloid materials to bind to specific ions or compounds using an experimentalists' intuition is not straightforward and has proven challenging when it is hard to achieve, relying solely on intuition on which modifications can transform an amyloid into a functional material. As a result, the exploitation of amyloid materials which can be used as carriers for the therapeutic-sustained release of drugs has been significantly limited due to (i) the scarcity of studies on amyloid peptide fibrils with exposed non-β-sheet forming mutable residues at the termini, and (ii) the lack of computational methods for designing such functional amyloid materials tailored to bind to certain compounds.

However, the solution to the design of amyloid materials binding to the aforementioned AD drugs can be provided through the computational protocol for the design of functional amyloid materials capable of binding to ions/compounds, demonstrated in chapter 7.

8.1.2 Goal

In this work, we aimed to implement and advance the computational design protocol for the design of functional amyloid materials binding to compounds, specifically AD drugs

Donepezil, Tacrine, Galantamine, and Memantine, acting as novel drug delivery carriers, targeting to enhance the half-life and effectivity of current AD drugs and potentially the stabilization of cognition (enhancement of current AD treatments). The advancement of the protocol was made by Joseph Jakubowski of Dr. Tamamis' lab who developed a method which can be executed after module III of computational protocol presented in chapter 7. The specific method is used to investigate the binding properties of the compounds in complex with the designed peptide fibrils.

8.1.3 Hypothesis

In previous chapters 5 and 6 and refs[23,24], inspired by the presence of the GAIIG amyloidogenic core, extracted from an HIV-1 gp120 V3 loop[237,238], and the α-tubulin acetyl transferase,[242] we showed that the peptide sequences YATGAIIGNII[23] and AGKGAIIGFIK,[22] can self-assemble into amyloid fibrils which can constitute amyloid designable scaffolds, incorporating mutable terminal non-β-sheet forming residues.[22,23] Specifically, flexible template structures of highly-ordered and well-aligned β-sheets formed by AGKGAIIGFIK used as input in our recently published computational protocol for the design of functional amyloid materials that bind cesium ions.[22] Here, we hypothesized that the same protocol can be further advanced for the design of amyloid materials binding to compounds of medical interest as well, such as the AD drugs, which is the point of interest in the current chapter. For this purpose, we used the amyloid designable scaffolds formed by YATGAIIGNII as a starting point to design novel amyloid materials with desired properties.

8.1.4 Objectives Accomplished

In this chapter, we suggest the design of two specific sequences which can constitute functional amyloid materials with the potential capacity to bind to AD drugs, acting as novel drug

delivery carriers, targeting to enhance the half-life and effectivity of current AD drugs and potentially the stabilization of cognition (enhancement of current AD treatments). Motivated by amyloids biocompatibility[36,286,287,288], and their ability to be modified at the sequence level, so as they can bind to one or combinations of AD drugs[289] with high affinity, such potentially successful drug-carrier amyloid materials could pave the way for a novel direction of delivering AD drugs.

Experiments evaluating the computationally designed materials binding properties to the AD drugs are performed in Dr. Gkikas' lab, and by additional experimentalists at the University of Massachusetts Lowell.

8.2 Computational Methods

In this study, we aimed to computationally design amyloid materials that can potentially bind to the four compounds representing AD drugs, Donepezil, Tacrine, Galantamine, and Memantine, through implementing the generalized computational protocol presented in the chapter 7. According to the protocol, amino acid mutations were optimally introduced at non-β-sheet residue positions of an amyloid designable scaffold, aiming to transform it into a functional amyloid material binding to Donepezil, Tacrine, Galantamine, and Memantine. In summary, the protocol as implemented in this study consists of the following steps executed successively: (1) The elementary β-sheet structural units (represented by flexible structural templates) of both antiparallel and parallel configurations formed by peptide YATGAIIGNII, previously elucidated by us in chapter 5 and ref[23], were further analyzed and were used as flexible structural templates for subsequent design, with the first and last two residues considered amenable for modification. (2) Primary and secondary materialphore models, which geometrically describe in a coarse-grained representation how amino acids bind to the four compounds, were independently extracted from experimentally resolved structures deposited in the PDB[252]. (3) An optimization-based model

was used to introduce combinations of mutations to the non β-sheet residue positions of the flexible structural templates of the amyloid designable scaffold aiming to mimic how amino acids bind to the four AD drugs according to slices of the primary materialphore models of the compounds. The designed binding pockets consist of amino acids belonging to the termini of two adjacent β-sheet bonded peptides. (4) The two designed peptides, FYTGAIIGNYF, YFTGAIIGNFY, which could be potentially functional in either antiparallel or parallel configurations in binding to all four AD drugs, and which possessed the lowest energy according to the optimization-based model's objective function were selected for further investigation. (5) The two selected designed peptides were further investigated using simulations, structural and binding free energy analyses; the analyses scope was to computationally validate and examine whether the designed peptides can self-assemble and form pockets capable to accommodate the binding of the four AD drugs, with respect to both primary and secondary materialphore models.

Additionally, we investigated whether the designed peptides can bind to the four AD drugs with sufficiently low binding free energy, and interact with the drugs similarly to the experimentally resolved complex structures they were derived from. This investigation was performed through programs developed and executed by Joseph Jakubowski, a student of Dr. Tamamis' lab. Specifically, the mimicry was examined and assisted through the additional definitions of tertiary and quaternary materialphore models, an important advancement to our computational protocol described in chapter 6, 7 and ref[22], which allowed us to accurately represent interactions between the designed amino acids and the compounds of interest. It is worth noting that one of the challenging aspects of the computational design-based functionalization is that the designed binding pockets consist of amino acids belonging to pairs of two adjacent β-sheet bonded peptides which are not covalently bonded.

8.2.1 Structural Analysis of an Amyloid Designable Scaffold and Determination of its Mutable Positions

In chapter 5, we have elucidated the amyloid self-assembly properties of the peptide with sequence YATGAIIGNII[23]. In summary, our analysis showed that self-assembled structures of YATGAIIGNII comprise a central β-sheet amyloidogenic core and terminal non β-sheet forming amino acids at which mutations can be introduced to achieve desired functionality; thus according to our definition it constituted an amyloid designable scaffold. Within the simulations, the peptide showed a higher propensity to form antiparallel rather than parallel β-sheets, and therefore our previous analysis focused on analyzing the antiparallel β-sheets which were more predominant.[23]

Here, we examined and verified that both configurations, antiparallel and parallel, can be defined as amyloid designable scaffolds. Initially, we analyzed the propensities of β-bridges formation between per pairs of residues in the β-sheets, which showed that the first three and last two residues were primarily involved in β-sheets in neither antiparallel nor parallel configurations. Additionally, we analyzed the degree of solvent exposure of the terminal residues using the same procedure as described in ref.[24], which verified that the first and last two residues are adequately solvent exposed in both configurations. Thus, there positions were deemed suitable for modifications to yield functionality for binding to the four compounds. As a result, the top 50 highly-ordered and well-aligned antiparallel and parallel 4-stranded β-sheet conformations, constituting the peptide's elementary β-sheet structural units in the two configurations, were collected independently and were used as flexible structural templates for subsequent design purposes, at which the designed amino acids were independently introduced in the first and last two terminal positions in both configurations (see below). Our rationale to consider both configurations as potential amyloid designable scaffolds was to improve the success rates of the

successfully designed amyloid materials (see below), which can potentially be functional in either configurations.

8.2.2 Uncovering How Amino Acid Motifs Bind to Donepezil, Tacrine, Galantamine, and Memantine According to Experimentally Resolved Protein Structures

As inputs, the computational protocol (described in detail in chapter 7) requires both flexible structural templates of the amyloid designable scaffold (determined above), and a library of primary and secondary materialphore models, which represent how amino acids bind to the compounds of interest (i.e., Donepezil, Tacrine, Galantamine and Memantine) according to experimentally resolved protein structures. The Protein Data Bank (PDB)[252] is a repository for three dimensional structures of biological molecules such as proteins and nucleic acids and contains an abundance of information on how ions or compounds bind to amino acids. From the PDB,[252] we collected all the experimentally resolved X-ray structures of protein complexes with the AD drugs Donepezil, Tacrine, Galantamine, and Memantine. From the collected protein structures, we defined and extracted materialphore models describing the identity and geometry of the amino acids interacting with a compound and located within an 8.5 Å cutoff radius from its center of mass. A relatively large cutoff distance was chosen due to the relatively large size of the four compounds. As described in chapter 6, 7, and ref,[22] primary materialphore models represent the distances between the centers of mass of all combinations of amino acids within the cutoff distance, and secondary materialphore models represent the distances between the centers of mass of the amino acids and the center of mass of a compound.

Primary materialphore models were used as input in the optimization-based design model, and contain all information needed for the design of the amyloid scaffold's mutable residue positions so that the newly placed amino acids geometrically mimic how amino acids bind to a

compound of interest (e.g., Donepezil, Tacrine, Galantamine, and Memantine) according to the PDB.[252] Both primary and secondary materialphore models were used to examine if the designed peptides' elementary β-sheet structural units can, within the simulations, form binding pockets resembling the corresponding materialphore model they were derived from. In both primary and secondary materialphore models, the compound and each of the amino acids are independently represented as single points corresponding to their centers of mass. Additional types of materialphore models, named as tertiary and quaternary, detailing specific interactions between amino acids and the compounds were defined below to examine the structural and energetic binding properties of the compounds in complex with the two designed peptides in docking studies.

8.2.3 Optimization-Based Design of the Functional Amyloid Materials Aiming to Bind to Donepezil, Tacrine, Galantamine, and Memantine

We introduced an optimization-based design model detailed in the chapter 7. The optimization model aimed to introduce amino acids at the mutable, underlined residue positions of both antiparallel and parallel flexible structural templates, independently, of the amyloid designable scaffold YATGAIIGNII to design amyloid materials binding to the four AD drugs Donepezil, Tacrine, Galantamine, and Memantine, independently. Thus, eight different design problems were solved in total. Within the design, we aimed that the substituted amino acids could be placed to reproduce the geometry of a slice of primary materialphore for each of the four compounds, independently, at least in a portion λ of the fifty antiparallel or parallel flexible structural templates. As in chapter 6, both the small value of λ (10% in this case) and the fact that we allowed a specific deviation for the matching between the distances of introduced amino acids in the flexible structural template and the corresponding distances in the primary materialphore model, enabled our design and accounted for the difference in distances that could originate due

to the different size of the substituted amino acids compared to the original amino acids Tyr1, Ala2, Ile10 and Ile11.

The optimization model was solved in such a way that primary materialphore model constraints were introduced first to eliminate number of substitutions and result in a total number of feasible number of solutions, for antiparallel and parallel flexible structural templates, independently. Due to the complexity of the current problem, aiming to identify designed peptides capable of potentially binding to all four AD drugs, and our additional condition on the functionality of the designed peptides to not depend on their finally predicted and potentially naturally occurring arrangement, parallel or antiparallel, only three peptides FYTGAIIGNYF, YFTGAIIGNFY, and YFTGAIIGNYF were capable to meet all aforementioned constraints. FYTGAIIGNYF, and YFTGAIIGNFY were selected for further investigation as the ones with lowest energy defined by the objective function of the optimization model.[22]

The optimization model's objective function aims to identify combinations of amino acid substitutions that yield the most energetically favorable selections by taking into account residue-residue energy interactions represented by a coarse-grained force field[205], and a term accounting for non-polar solvation free energy due to amino acid substitution.[22] It can be considered to act as a criterion to select designs of which the mutated residues' binding pocket is energetically stabilized to bind to each of the compounds investigated, and also act as a means to disallow the destabilization of the amyloid fibrils due to the designed amino acid substitutions.

8.2.4 MD Simulations Investigating the Self-assembly Properties of the Designed Peptides and Computational Validation against the Primary and Secondary Materialphore Models

The two selected peptides were investigated using triplicate implicit solvent REMD simulations according to the protocol developed by Tamamis and Archontis[130] to examine their

conformational self-assembly properties, and more importantly whether within the simulations the two peptides can form the desired binding pockets, mimicking the corresponding primary and secondary materialphore models, capable of binding to the four compounds. Each designed peptide was subjected to three independent REMD simulations which resulted in a total of six independent REMD simulations. The simulations were performed using CHARMM.[110] We used FACTS19[123] implicit solvent model in the simulations with the surface tension coefficient value of 0.015 kcal mol^{-1} $Å^{-2}$. Langevin dynamics were used with a 5.0 ps^{-1} friction coefficient on all non-hydrogen atoms of both designed peptides. A total of sixteen temperatures (290, 295, 300, 310, 305, 315, 321, 327, 333, 339, 345, 352, 359, 366, 373 and 380 K) were employed with each replica exchange run's duration equal to 10 ps.

The collected snapshots were categorized into antiparallel, parallel and mixed 2-, 3-, 4-, 5-, and 6-stranded β-sheet conformations. The categorization was done according to the computational protocol described in chapter 2 developed by Tamamis and Archontis.[130] Analogously to the refs[22,23,24,40,130,131,132], in order to uncover the most favorable configuration, antiparallel or parallel of each the two designed peptides, we focused our analysis on 4-, 5-, 6-stranded β-sheet conformations due to their higher complexity compared to 2-, 3-stranded β-sheet conformations. In both peptides both parallel, antiparallel and mixed configurations were depicted. We observed that the designed peptide YFTGAIIGNFY acquires a higher tendency to form antiparallel conformations, whereas the designed peptide FYTGAIIGNYF acquires a clear tendency towards parallel conformations. Subsequently, we combined all the 4-stranded β-sheet conformations from the three independent simulations for each peptide and used polar and nematic order parameters to examine the extent of peptide alignment and order. We identified and collected the highly ordered and well-aligned 4-stranded β-sheet conformations and from the collected

conformations, we discarded the structures which either had partial helical content or a bend in any of the outer peptide strands. As a result, we extracted the top 500 refined highly ordered and well-aligned 4-stranded antiparallel β-sheet structures for the designed peptide YFTGAIIGNFY and parallel β-sheet structures for the designed peptide FYTGAIIGNYF. These structures were used as input in the two-component computational validation to examine whether the two peptides within the β-sheet structures can form functional pockets to bind to the compounds under investigation. It is worth noting that the REMD simulations were performed in the absence of the compounds which served as an ultimate test to check whether proper binding pockets are formed in the presence of the compounds.

In the first component of the validation, we evaluated whether the distances between the centers of mass of the designed amino acids' side chains, per pair of two β-sheet bonded peptides in a given structure, are similar to the corresponding distances within the slice of the primary materialphore model(s) they originated from; the similarity criteria are detailed in the ref[22]. The designed binding pockets complying with the above criteria were considered primary matches. Given a primary match, in the second component of the validation, we determined whether the hypothetical center of mass of a compound (Donepezil, Tacrine, Galantamine, or Memantine) can be positioned in such a way that the distances between the hypothetical center of mass of a compound to the centers of mass of each designed amino acid are similar to the corresponding distances within the slice of the secondary materialphore model(s) they originated from; the placement of the compound's hypothetical center of mass as well as the similarity criteria are detailed in the ref[22]. These were considered secondary matches. The designed binding pockets which complied with both the validation components were selected for subsequent docking and energetic calculations.

8.2.6 Docking Studies Investigating the Structure and Binding Free-Energy of the Four AD Drugs to the Designed Amyloid Peptides

The methodology described in 8.2.6 was performed through programs developed and executed by Joseph Jakubowski, a student of Dr. Tamamis' lab.

In the two-component computational validation procedure described previously, we identified a set of selected structures with qualifying primary and secondary matches. For the specific subset of structures encompassing such binding pockets potentially capable to accommodate the four compounds, we performed docking studies to assess the compounds' (1) ability to be inserted in the designed binding pocket and to be oriented such that they interact with the designed amino acids, in line with their corresponding tertiary and quaternary materialphore models, and (2) binding free energy, evaluating their energetic favorability to bind to the designed peptides' elementary β-sheet structural units.

The docking studies were significantly enabled by defining and extracting additional sets of materialphore models, which we named as tertiary and quaternary materialphore models, for the four AD drugs under investigation. While primary and secondary materialphore models contain sufficient information to study the insertion of an ion in a designed binding pocket, the aforementioned types of materialphore models cannot fully construct and examine the insertion of a compound in a designed binding pocket, accompanied by its relative orientation and the formation of interactions with the designed amino acids. Undeniably, a compound's orientation within the designed binding pocket as well as its specific interactions with the designed amino acids are important for both the correct docking of a compound in the binding pocket and for the assessment of interactions formed between the compound and the designed amino acids. For this purpose, we classified the amino acids and compounds (encompassed within the primary and

secondary materialphore models) into a set of chemical features such as charged, hydrogen bond donor or acceptor, aromatic, and hydrophobic groups, and following this, we defined ionic, hydrogen bond, cation-π interactions as tertiary materialphore models, and π-π and hydrophobic interactions as quaternary materialphore models. Following definitions used by Piovesan et al[290]., ionic interactions were defined as interactions between oppositely charged groups within 5.0 Å. Hydrogen bonds were defined as interactions between hydrogen bond donors and acceptors within a relaxed 5.5 Å distance cutoff. Cation-π interactions were defined as interactions between the closest ring atom to the charged group within 7.0 Å. π-π interactions were defined as interactions between aromatic rings of which the nearest atomic distance is within 7.0 Å. Finally, we defined, π-hydrophobic or hydrophobic-hydrophobic interactions as interactions between the nearest atomic distance of an aromatic ring and a hydrophobic group, or the nearest atomic distance between two hydrophobic groups, respectively, within a relaxed criterion of 7.5 Å. These tertiary and quaternary materialphore models were used in the following section in order to enable us to dock the compounds in the designed binding pockets. The programs for defining and extracting tertiary and quaternary materialphore models was developed by Joseph M. Jakubowski, Dr. Tamamis' lab, Texas A&M University.

The docking was performed manually through an in-house docking algorithm written in CHARMM[110]. Specifically, we docked each AD drug at its designed binding pocket of four amino acids, guided by, and thus mimicking its interactions with the interacting amino acids to the tertiary and quaternary materialphore models it was derived from. The docking aimed to provide the designed amino acids, extracted from simulations performed in the absence of the compounds, the ability to adjust their conformations and accommodate the compounds binding, as would be the case in a naturally occurring process. Specifically, our in-house CHARMM[110] programs developed

for this purpose allowed the side chains of the designed amino acid in the binding pocket to refine their positions and relative orientations to mimic within docking the distances of interactions present in the tertiary and quaternary materialphore models; if multiple interactions formed between a specific amino acid and compound were present, then distances of interactions between their constituting heavy atom chemical groups were prioritized in the following order: ionic (between oppositely charged groups) > hydrogen bond > cation-π > π-π > hydrophobic-hydrophobic, and only one interaction between an amino acid and a compound was taken into account.

Within docking, the corresponding distances of the nearest neighboring heavy-atom interactions within the tertiary and quaternary materialphore models were introduced as constraints during a three-stage minimization in CHARMM,[110] while the compounds' center of mass initial placement would coincide with its defined position according to the previous validation stages. During minimization, residues not participating in binding were fixed to their initial positions. The energy minimization (of which its stage comprised of steepest descent and adopted Newton-Rapshon minimization) occurred in stages at which the harmonic force constants, imposed to enable mimicry between the amino acid-compounds' interactions in the tertiary and quaternary materialphore models and the docked structure, were gradually reduced from 5 kcal/(mol*Å^2) to 0.1 kcal/(mol.Å^2).

At the end the minimization-based docking procedure, the final potential energy of the constraint terms was recorded, and was used as a first metric to ensure within the docked output structure, a sufficiently high level of mimicry was achieved; docked structures accompanied by a constraint energy above a certain cutoff (chosen to be equal to 1 kcal/mol by visual inspection) were discarded and considered as infeasible solutions. This value was used to ensure that imposed

distance constraints were met to a sufficiently reasonable extent for the structures chosen to be analyzed in the last step of energy computations. At the last step a slight energy minimization of 50 steps was introduced and the final output structure was extracted to evaluate the compounds' binding free energy according the semi-empirical energy function, AutoDock Vina.[291] Finally, we introduced, a second metric, an posteriori criterion to ensure that after docking and minimization, the minimum distance between a designed amino acid and a heavy atom of a compound does not exceed 6 Å, which constitutes an additional check to verify the participation of all designed amino acids in the binding.

It is worth noting that AutoDock Vina[291] was not used for docking purposes and was used only to calculate the binding affinity of a compound to the entire elementary structural unit, represented by flexible structural templates of the designed peptides produced in the simulations. Also, non-designed amino acid side chains (e.g., which are part of the amyloid scaffold) may also contribute to the binding affinity in addition to the designed amino acid side chains.

The above docking procedure served as an ultimate test to select the structures that had already been validated by primary and secondary matches, for their ability to form the proper binding pockets so that the compounds can bind to the designed amino acids similarly to their binding to amino acids within the experimentally resolved structures. the value of the final potential energy of the constraint terms. Specifically, the final potential energy of the constraint terms calculated above for each docked structure indicated to which extent the mimicry could be achieved, while the binding free energies estimated between the docked compounds and the designed peptides enabled us to estimate the binding affinity of each of the four compounds to the designed amyloid materials (represented computationally by their elementary structural β-sheet units), enabling us to investigate the relative potency of the four compounds to the designed

amyloid peptides. Nevertheless, the constraints were not fully met (i.e., the value of the final potential energy of the constraint terms ranged between 0 to 1 kcal/mol) and thus the degree of mimicry between the geometry of the designed amino acids and the materialphore model they originated from was acceptable but not necessarily perfect. The in-house CHARMM programs for docking studies were developed by Joseph M. Jakubowski, Dr. Tamamis' lab, Texas A&M University.

8.3 Results and Discussion

8.3.1 Structural Analysis of an Amyloid Designable Scaffold and Determination of its Mutable Positions

In chapter 5, we showed that the peptide with sequence YATGAIIGNII self-assembles into amyloid fibrils which can be considered amyloid designable scaffolds due to the fact that the first three and last two residues are primarily not involved in β-sheets.[23] According to our statistical analysis, we observed that the peptide can self-assemble into both β-sheet antiparallel and parallel conformations, with higher probability for antiparallel conformations. Within the simulations, we observed the formation of antiparallel or parallel β-sheets involving 2-6 peptide strands, as well as cases at which the peptides were bonded in mixed antiparallel/parallel fashion.[23] In order to elucidate the elementary β-sheet structural units formed by the peptide in both the antiparallel and parallel configurations, we focused our analysis on 4-stranded antiparallel or parallel β-sheets. This is because they can comprise elementary β-sheet structural units, they possess higher complexity than 2- and 3-stranded β-sheets, and their population within the simulations is significantly higher than that of 5- or 6- stranded β-sheet, which allows richer statistical analysis. The analysis is performed according to the computational protocol developed by Tamamis and Archontis[130] and is described in detail in refs.[24,40,131,192]

We used polar (P_1) and nematic (P_2) order parameters, describing the degree of polarity and order across a specific unit vector of the self-assembled peptide structures, respectively, to create free energy landscapes through which we identified and extracted 4-stranded highly ordered and well-aligned β-sheet conformations (antiparallel or parallel), which can constitute flexible structural templates of the amyloid designable scaffold YATGAIIGNII. From the extracted β-sheet conformations, we discarded ones containing a partially helical or bend conformation within the outer peptide structures as a result of the absence of a nearly infinite array of peptides. For the remaining β-sheet conformations, which are considered to be the elementary β-sheet structural units of the peptide representing antiparallel and parallel configurations, we calculated the propensity for an intermolecular pair of residues that belong to adjacent β-sheets to be involved in extended β-sheet or isolated β-bridge formation (Figure 8.1A, 8.2B). This analysis verified that in both the antiparallel and parallel elementary β-sheet structural units of the peptide, the first three and last two residues (YATGAIIGNII) are predominantly not participating in the β-sheets, and thus can be considered mutable positions to yield functionalization. To further support this notion, we calculated the average degree of solvent accessibility of the first three and last two residues in both configurations. The minimum value of solvent accessibility probability, calculated for all five residues in both configurations, was estimated 0.44, which was above the minimum solvent accessibility of the functional amino acids (e.g., cysteine residues) as estimated in our previous study,[24] where the solvent accessibility was validated experimentally. This additionally supported that the first and last two residues (underlined residue positions YATGAIIGNII) are solvent exposed in both configurations, and thus they are amenable for modification. The specific residues could be mutated to form functional binding pockets that can bind to the compounds. Subsequently, we independently selected the top 50 highly ordered and well-aligned antiparallel

β-sheet conformations and top 50 highly ordered and well-aligned parallel β-sheet conformations to represent the flexible structural templates of the amyloid designable scaffold in both configurations, independently. In this study, the centers of mass positions of the sidechains within the flexible structural templates were used as input in the optimization-based design model. A schematic representation of antiparallel and parallel flexible structural templates is presented in Figure 8.2A.

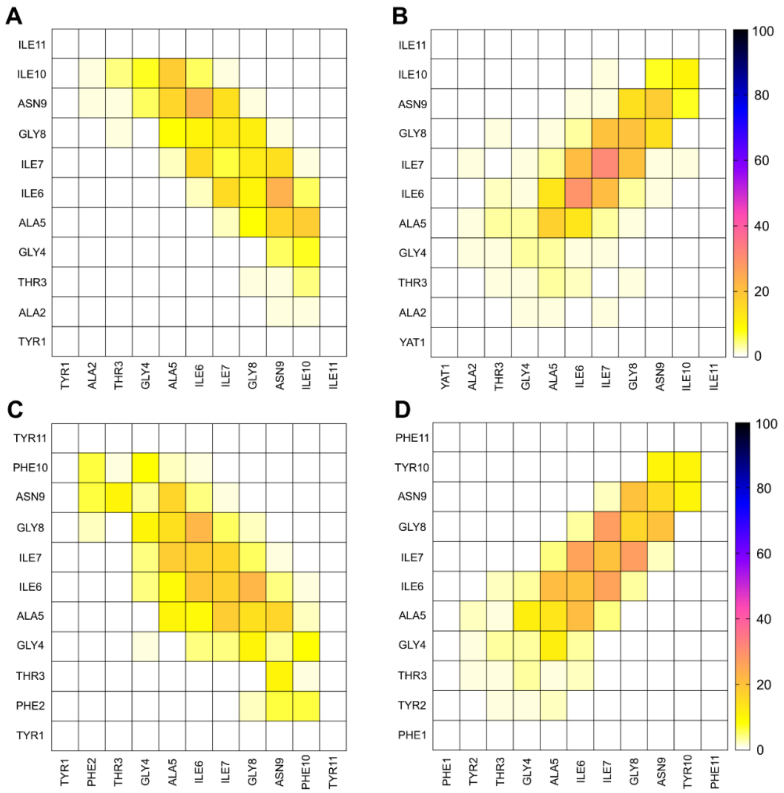


Figure 8.1 Density (%) maps of β-bridge interacting residue pairs formed between nearest neighboring peptides participating in a highly ordered and well-aligned 4-stranded β-sheet conformation observed at 300 K for YATGAIIGNII and YFTGAIIGNFY. Panels A, and C correspond to antiparallel conformations formed by the peptides YATGAIIGNII, and YFTGAIIGNFY, respectively. Panels B, and D correspond to parallel conformations formed by the peptides YATGAIIGNII, and FYTGAIIGNYF, respectively. Highly ordered and well-aligned 4-stranded β-sheet conformations were extracted from the corresponding 4-stranded β-sheet conformations for a specific peptide by using polar (P_1) and nematic (P_2) order parameters which examine the extent of peptide alignment and order.

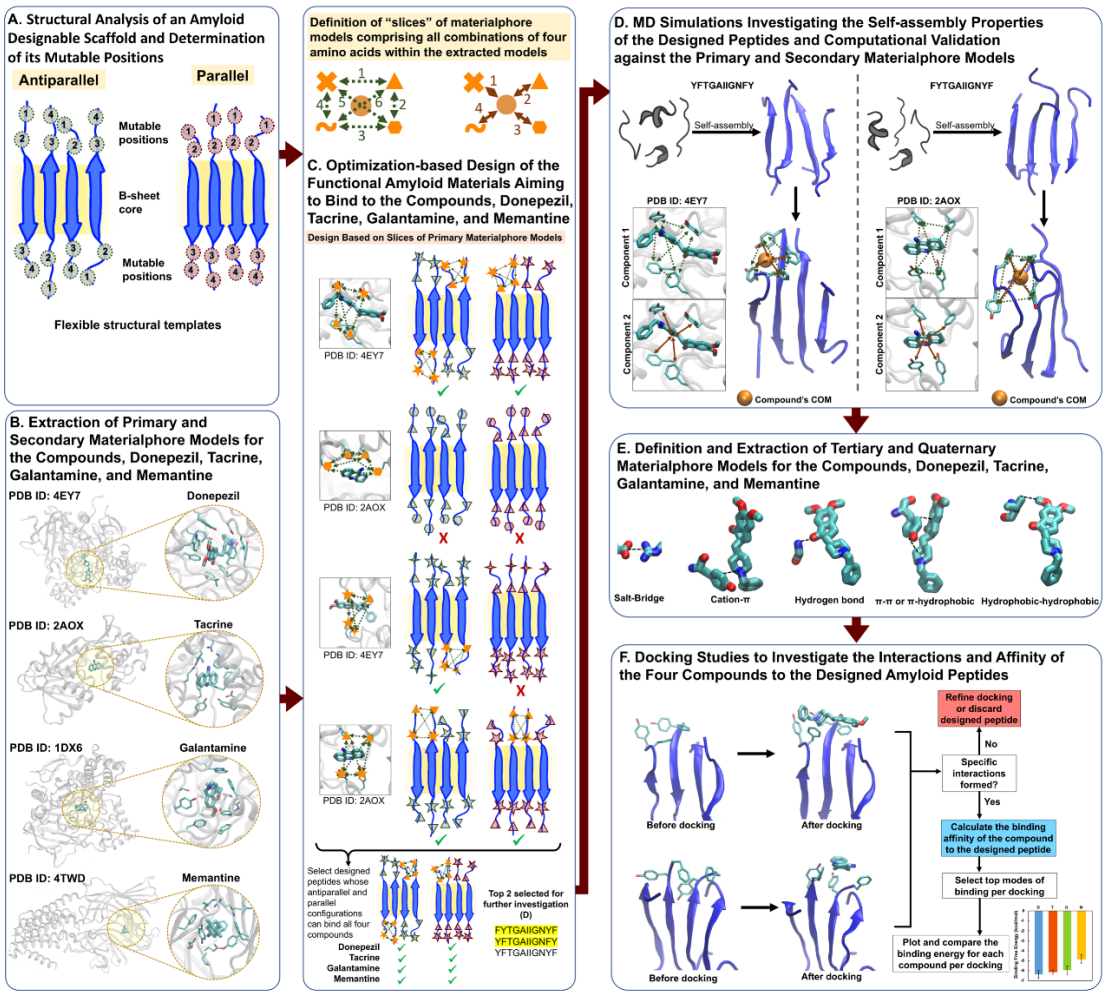


Figure 8.2 Overview of the steps used for the computational design and study of the two designed amyloid peptides in complex with the four compounds under investigation. In panel A, we show a schematic representation of antiparallel and parallel flexible structural templates of which their mutable residue positions are shown in green and maroon dotted circles, respectively. In panel B, we show examples of materialphore models extracted from experimentally resolved structures for each compound, and also describe the slices of materialphore models comprising combinations of four amino acids, which are represented by different orange symbols. In panel C, we show a schematic representation of the optimization-based procedure used to design the amyloid peptides, by considering functionality and energetic calculations in the objective function. Functionality is considered by evaluating if the geometry of slices in the primary materialphore models, indicated using orange symbols in the molecular graphics images of the Protein Data Bank (PDB) structures, can be matched to the geometry of a flexible structural template, indicated by solid orange symbols and green dotted arrows. Successful match is indicated using a green checkmark and unsuccessful match using red cross. In panel D, we show the formation of β-sheet structures by the two selected peptides, FYTGAITIGNYF and YFTGAIIGNFY, accompanied by the two-component computational validation procedure identifying primary and secondary matches. In panel E, we define and schematically represent the tertiary and quaternary materialphore models. In panel F, we show docking studies performed to investigate the interactions and affinity of the four compounds to the designed amyloid peptides.

8.3.2 Uncovering How Amino Acid Motifs Bind to Donepezil, Tacrine, Galantamine, and Memantine According to Experimentally Resolved Protein Structures

We collected all the experimentally resolved structures of proteins in complex with any of the compounds Donepezil, Tacrine, Galantamine, and Memantine, and their Protein Data Bank[252] (PDB) IDs are presented in the second column of Table 8.1. From the collected protein structures, we extracted materialphore models for each of the four compounds independently, mapping the geometries of amino acids and the compounds in experimentally resolved protein structures. Within a binding pocket of a compound interacting with amino acids in a specific PDB structures, primary materialphore models represent the all possible relative distances of pairs of interacting amino acids, and secondary materialphore models represent all distances between the amino acids and the compounds. Amino acids and compounds in the aforementioned definitions are geometrically described by their centers of mass, and an amino acid is considered to be within a compound's binding pocket if their distance is below 8.5 Å. An example of a materialphore model for each of the AD drugs is presented in Figure 8.2B.

Our analysis resulted in the extraction of multiple materialphore models for each of the four AD drugs. Given the fact that, in each experimentally resolved structure, a certain compound may have been resolved to bind to multiple either nearly identical (e.g., materialphore models derived for Donepezil binding to chains A and B of Acetylcholinesterase in the PDB ID: 4EY7[292]) or diverse binding pockets (e.g. materialphore models derived for Memantine binding to chains E and F of pentameric ligand gated ion channels in the PDB ID: 4TWD[293]), the total number of materialphore models extracted for each compound is equal to the sum of the number of experimentally resolved proteins multiplied by the number of complexed compounds with each proteins and is provided Table 8.1. Nearly identical binding pockets were automatically considered

individually to avoid exclusion of any possible binding mode, as this would not add to the complexity of solving the optimization-based model design model, described below, due to limited number of materialphore models that could be extracted (as the number of experimentally resolved structures of proteins complexed with the compounds is low).

In the present study, the number of designable residue positions is equal to two per peptide per binding pocket, and thus upon β-sheet formation, the total number of residue positions amenable for subsequent design in each binding pocket is four (green and marron dotted circles in Figure 8.2A). Since the primary and secondary materialphore models can contain more than four amino acids interacting with the compounds (as the number of amino acids interacting with the compound can be greater than four), we additionally considered "slices" of materialphore models to be used for subsequent design purposes. For a given materialphore model, the slices represent combinations of four amino acids interacting with the compounds and are used as input information in the optimization-based design. A schematic representation of slices originating from materialphore models is presented above in Figure 8.2B. For example, in the materialphore model shown in the Figure 8.2B for Donepezil, the number of amino acids in the binding pocket is 14, and thus the number of slices of materialphore models which can be extracted for design, is equivalent to the number of combinations of 14 per 4, which is equal to 1001. Nevertheless, constraints introduced in the optimization-based model, can reduce the number of possible combinations considered feasible during design (see below).

Primary materialphore models were used as input in the optimization-based design model, and contain all information needed for the design of the amyloid scaffold's mutable residue positions so that the newly placed amino acids geometrically mimic how amino acids bind to a compound of interest (e.g., Donepezil, Tacrine, Galantamine, and Memantine) according to

experimentally resolved structures in PDB. Both primary and secondary materialphore models were used to examine if the designed peptides' elementary β-sheet structural units can within the simulations form binding pockets resembling the corresponding materialphore model they were derived from (see below).

Table 8.1 Experimentally X-ray resolved structures of proteins in complex with the four compounds, their corresponding PDB IDs, and number of materialphore models extracted for each compound. The total number of materialphore models extracted for each compound is equal to the sum of the number of experimentally resolved proteins multiplied by the number of complexed compounds within each protein.

Compound	**Experimentally resolved protein structures**	**Number of Compounds complexed in each of the structures**	**Total number of Materialphore models extracted**
Donepezil	4EY7[292]	2	3
	1EVE[294]	1	
Tacrine	4BDS[295]	1	12
	2AOW[296]	2	
	2AOX[296]	2	
	1MX1[297]	6	
	1ACJ[298]	1	
Galantamine	4EY6[292]	2	9
	2PH9[299]	2	
	1W6R[300]	1	
	1W76[300]	2	
	1DX6[301]	1	
	1QTI[302]	1	
Memantine	4TWD[293]	10	10

8.3.3 Optimization-Based Design of the Functional Amyloid Materials Aiming to Bind to Donepezil, Tacrine, Galantamine, and Memantine

The optimization-based design model described in chapter 7 is introduced here for the design of amyloid materials binding to the four AD drugs of interest. The top fifty highly-ordered and well-aligned flexible templates of both antiparallel and parallel β-sheets were introduced as input to the model, and mutations were attempted and performed at the two first and last residue positions, guided by the collected primary materialphore models of each of the four AD drugs, independently. Within the flexible structural templates, represented by elementary β-sheet structural units of both 4-stranded antiparallel and parallel β-sheets, a total of six identical in sequence binding pockets were considered during design, composed by four amino acids in pairs of two-adjacent β-sheet bonded peptides (Figure 8.2A).

We independently solved an optimization-based design model with the aim to design amyloid materials binding to the four compounds. The optimization-based design model was solved for both antiparallel and parallel flexible structural templates of the amyloid designable scaffold. This resulted in a total of eight independent problems. Out of the 20^4 (20: total number of natural amino acids, 4=2+2: number of mutable positions per peptide) theoretical number of possible designed peptides, the optimization model produced 644, 7841, 1899, and 1703 different designed peptides that could potentially bind to Donepezil, Tacrine, Galantamine, and Memantine respectively on antiparallel flexible structural templates, as well as 18, 1258, 165, and 420 different designed peptides that could potentially bind those drugs on parallel flexible structural templates. The total number of designed peptides (originating from antiparallel and parallel flexible structural templates) is low for Donepezil because it has the lowest number of materialphore models (Table 8.1).

Upon solution of the problem, we selected peptide sequences which can potentially bind to all the four compounds, and thus we identified the designed peptides which were common for all the four compounds (Figure 8.2C). Furthermore, we introduced an additional constraint to identify designed peptide sequences which could be functional in both antiparallel and parallel configurations; through this, we aimed to improve the success rates as the finally produced designed materials could potentially be functional irrespective of the dominant configuration adopted by each peptide. The introduction of the additional selection feature and constraint yielded a limited number of three solutions, FYTGAIIGNYF, YFTGAIIGNFY, and YFTGAIIGNYF, which were common across all the eight set of design and possessed low energies as estimated within the objective function. The two lowest in energy (defined in by the minimized term in the objective function) were selected for further investigation comprising simulations, structural and energetic calculations. Through these analyses, we evaluated the capacity of the two peptides to self-assemble into highly ordered and well-aligned β-sheets and form binding pockets which possess the expected geometry (in line with the materialphore models they were derived from) to bind to the compounds in the absence of the compounds, and subsequently, their ability to bind to the four compounds as well as their affinity to the compounds, all independently.

8.3.4 MD Simulations Investigating the Self-assembly Properties of the Designed Peptides and Computational Validation against the Primary and Secondary Materialphore Models

We performed three independent REMD simulation runs to investigate each of the two peptides' self-assembly structural properties, which were analyzed according to the protocol of Tamamis and Archontis.[130] The simulation runs were performed for each of the two peptides independently in order to increase the statistical sampling of β-sheet conformations used in our analysis. The propensities of β-sheet conformations were analyzed as described in refs[24,40,130,131,192]

and the statistical convergence of the highly populated β-sheet conformations was verified as a function of time. Within the simulations, YFTGAIIGNFY acquired an overall higher tendency for antiparallel compared to parallel configurations, whereas the designed peptide FYTGAIIGNYF acquired a dominant tendency for parallel configurations. Subsequently we extracted the highly ordered and well-aligned 4-stranded β-sheet conformations of the two peptides in their most dominant configuration, antiparallel for YFTGAIIGNFY and parallel for FYTGAIIGNYF (Figure 8.1C and 8.1D). Representative conformations for both the designed peptides are shown in Figure 8.2D. First, we analyzed the propensities of β-bridge formation between per pairs of residues in the β-sheets, which showed that the designed amino acids were predominantly not involved in β-sheets. The calculated propensities of the two peptides were highly similar to the corresponding propensities of YATGAIINII which served as the amyloid designable scaffold they were derived from, while the main difference between the two is associated with the fact that Phe2 and Thr3 in the designed peptide YFTGAIIGNFY can be involved in the β-sheets to a small extent (Figure 8.1C). Subsequently, we performed a two-component computational validation procedure (Figure 8.2D), comparing the geometries of the binding pockets formed by the designed amino acids, in comparison to their corresponding geometries in the primary and secondary materialphore models, derived from experimentally resolved structures. Within the comparative analysis, we identified primary and secondary matches, which are detailed in the *Materials and Methods* section. The former denotes the capacity of the designed amino acids in the two peptides to form binding pockets such that the distances between their centers of mass in the extracted β-sheet conformations are similar to the corresponding distances of the same amino acids within the primary materialphore models from which the designed peptides originated. A representative primary match for each of the designed peptides is shown in Figure 8.2D where the distances

between the amino acids are indicated by green dotted lines. If a primary match occurs, then the latter denotes the capacity of the designed amino acids in the two peptides to form binding pockets such that the distances between their centers of mass and the hypothetical center of mass of the four compounds are similar to the corresponding distances between the same amino acids and the compounds within the secondary materialphore models. Representative primary match and secondary match for each of the designed peptides is shown in Figure 8.2D where the distances between the amino acids are indicated by green dotted lines and the distances between the amino acids and the compound are indicated by solid marron lines. According to the results presented in Table 8.2, the percentage values for the first and second validation components are relatively high. This is an indication that even in the absence of compounds, the two peptides can adopt proper binding pockets at which the designed amino acids belonging to two adjacent β-sheet bonded peptides can form binding pockets resembling the experimentally resolved ones they originated from.

Table 8.2 Designed peptide sequences for the four compounds and their corresponding first and second validation component percentage statistics. The validation statistics were calculated over the highly ordered and well-aligned snapshots extracted from the REMD simulations investigating the peptides' self-assembly properties. The listed percentages for the first and second validation components correspond to the percentage of snapshots containing primary and secondary matches, within the snapshots containing primary matches, which were identified from the first validation component.

Compound	**YF**TGAIIGN**FY**		**FY**TGAIIGN**YF**	
	First validation component (%)	Second validation component (%)	First validation component (%)	Second validation component (%)
Donepezil	96.00	80.05	98.00	61.92
Tacrine	89.40	78.97	79.60	60.90

Table 8.2 Continued

Compound	**YF**TGAIIGN**FY**		**FY**TGAIIGN**YF**	
	First validation component (%)	Second validation component (%)	First validation component (%)	Second validation component (%)
Galantamine	83.80	79.52	90.80	70.40
Memantine	62.80	63.09	88.60	46.20

8.3.5 Docking Studies Investigating the Structure and Binding Free-Energy of the Four AD Drugs to the Designed Amyloid Peptides

The analysis described in 8.3.5 was performed through programs developed and executed by Joseph Jakubowski, a student of Dr. Tamamis' lab.

As described in the *Materials and Methods* section, docking of the AD drugs to the binding pockets of the two designed peptides, with primary and secondary materialphore model matches, was performed to assess the compounds' (1) ability to be inserted in the designed binding pocket and to be oriented such that they interact with the designed amino acids, in line with their corresponding interactions to the amino acids within the PDB structures they were derived from, and (2) binding free energy by evaluating their energetic favorability to bind to the two designed peptides' extracted β-sheet structures, represented by their elementary β-sheet structural units. Instead of using available standard docking algorithms which could randomly place each of the fours compounds in complex with the β-sheet structures formed by the two peptides (within or without the binding pocket composed of the designed amino acids) without any guidance on the key expected interactions between them and the designed amino acids, we introduced an in-house docking algorithm written in CHARMM[110]. The in-house algorithm used a manually-constrained

docking procedure which aimed to provide the designed amino acids with the ability to adjust in the presence the compounds, as well as accommodate the compounds similarly to how the same amino acids bind to the compounds in the corresponding experimentally resolved structures that each design originated from, ultimately aiming to mimic the naturally occurring process (Figure 8.2F). For this purpose, we expanded the definitions of our previously defined materialphore models, to additionally include tertiary and quaternary materialphore models, entailing for each corresponding (slice of a) materialphore model key additional information on distances corresponding to any type of potential electrostatic, hydrogen-bond, cation-π, π-π, hydrophobic-π or simply hydrophobic interactions between the amino acids and the compounds (see *Materials and Methods*). The distances of these interactions included in the tertiary and quaternary materialphore models were used as constraints during energy minimization which allowed docking and refinement of interactions of the four compounds to the designed amino acids within both peptides' 4-stranded β-sheet structures.

Upon docking the four AD drugs to the extracted β-sheet structures of both YFTGAIIGNFY and FYTGAIIGNYF, we selected the top ten binding modes with the highest affinity (i.e., lowest binding free energy assessed with Autodock Vina's scoring function[291]) for both designed peptides. Any docked conformations failing to reproduce the constraints imposed by the tertiary or quaternary materialphore models were not considered in the selection. The selection of the top ten binding modes was based on the observation that overall the lowest free energy modes of each of the four AD drugs in complex with the two peptides correspond to binding modes with not necessarily very high structural similarity but with alike interactions and molecular recognition properties (see below). The presence of alike interactions rather than identical interactions across different binding modes is attributed to the variability of the designed amino

acids' geometries in the β-sheet structures used for docking. Due to the variability, the distances of interactions within the binding modes were not necessarily identical to the imposed distance constraints within the tertiary and quaternary materialphore models. Importantly, the energy minimization used in the presence of constraints allowed the compounds to optimize their interactions with the designed amino acids.

Subsequently, we performed a statistical analysis of the binding free energies of the four AD drugs in complex with the two peptides' β-sheet structures, presented in Figure 8.3. The binding affinities of the selected top ten structures per compound were subjected to a single factor ANOVA test, independently, for both designed peptides YFTGAIIGNFY (Figure 8.3A), and FYTGAIIGNYF (Figure 8.3B). Each test yielded a *p*-value of less than 0.0001 with $\alpha = 0.05$, suggesting with 95% confidence that the means values between the four compounds were different, in both cases (Figure 8.3A and 8.3B). The statistical analyses were performed one-way analysis of variance (One-way ANOVA) in Microsoft Excel. The overall strategy for the docking is presented in Figure 8.2F.

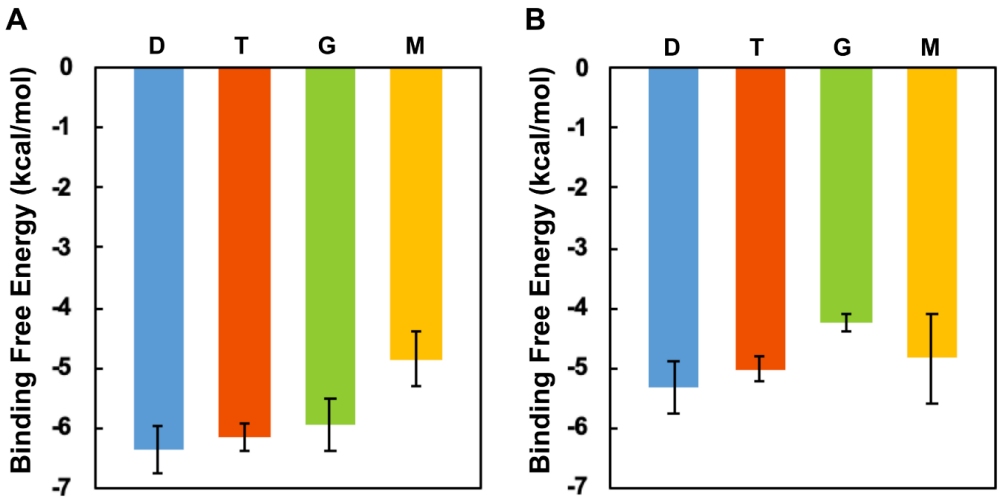


Figure 8.3 Binding affinity of the designed peptides for the four compounds Donepezil, Tacrine, Galantamine, and Memantine. Panels A, and B represent the binding capacities for the four compounds by the designed peptides YFTGAIIGNFY, and FYTGAIIGNFY, respectively. Statistical analyses were performed one-way analysis of variance (One-way ANOVA) in Microsoft excel.

Our results suggest that Donepezil and Tacrine have relatively higher affinity to the two designed peptides, YFTGAIIGNFY and FYTGAIIGNYF, than Galantamine, and Memantine. To confirm that the trend (observed in Figure 8.3) is not an artifact, the statistical analysis was performed on the top 25 and 50 structures, and it was observed that the trend remained the same, while the overall p-values remained less than 0.0001. It is worth noting that the computational predicted affinities suggest a trend of energetic favorability of the designed peptides to the four AD drugs rather than the ability of a designed peptide to bind to a compound or not. The relatively low free energies suggest that the two designed peptides have the potency to bind to all four compounds, achieving our original goal, while the binding free energy to Donepezil and Tacrine is predicted to be lower (and more energetically favorable) to Galantamine and Memantine.

Figures 8.4 and 8.5 present the representative mimicry between the docked AD drugs binding to the four designed (two initial and two terminal) amino acids in the β-sheet structures of

the two designed amyloid peptides, and the corresponding amino acids in the experimentally resolved structures from which materialphore models have been extracted. Interestingly, our designed peptides and their corresponding designed amino acids at the termini can bind to Donepezil and Galantamine based on how the two drugs bind to their target protein AChE as part of their inhibition mechanism;[292] this depicts the first to our knowledge amyloid materials suggested to bind to a compound by mimicking the mechanism according to which the same compound could act as an inhibitor. Analogously, our designed peptides and their corresponding designed amino acids at the termini can bind to Tacrine based on how the AD drug binds to Histamine N-methyltransferase[296] and Memantine based on how the AD drug binds to Erwinia chrysanthemi (ELIC).[293]

We performed a visual inspection of the interaction formed by the four AD drugs in complex with the designed amino acids of the extracted β-sheets of the two designed peptides. Donepezil primarily adheres to the binding pocket formed by the four designed amino acids of the two peptides. In the case of designed peptide YFTGAIIGNFY, Donepezil adheres to the designed binding pocket primarily through its amine containing moiety (Figure 8.4A). Alternatively, in the case of designed peptide FYTGAIIGNYF, the entire compound adheres to the binding pocket formed by the four designed amino acids (Figure 8.4B). For both designed peptides, Donepezil forms rich π-π and cation-π interactions with designed amino acids (Figure 8.4A and 8.4B). Additionally, for both the designed peptides, the amide group of Donepezil forms a hydrogen bond with the hydroxyl group of the designed tyrosines of the elementary β-sheet structural units formed by both peptides in nearly half of the top ten selected structures (Figure 8.4A and 8.4B).

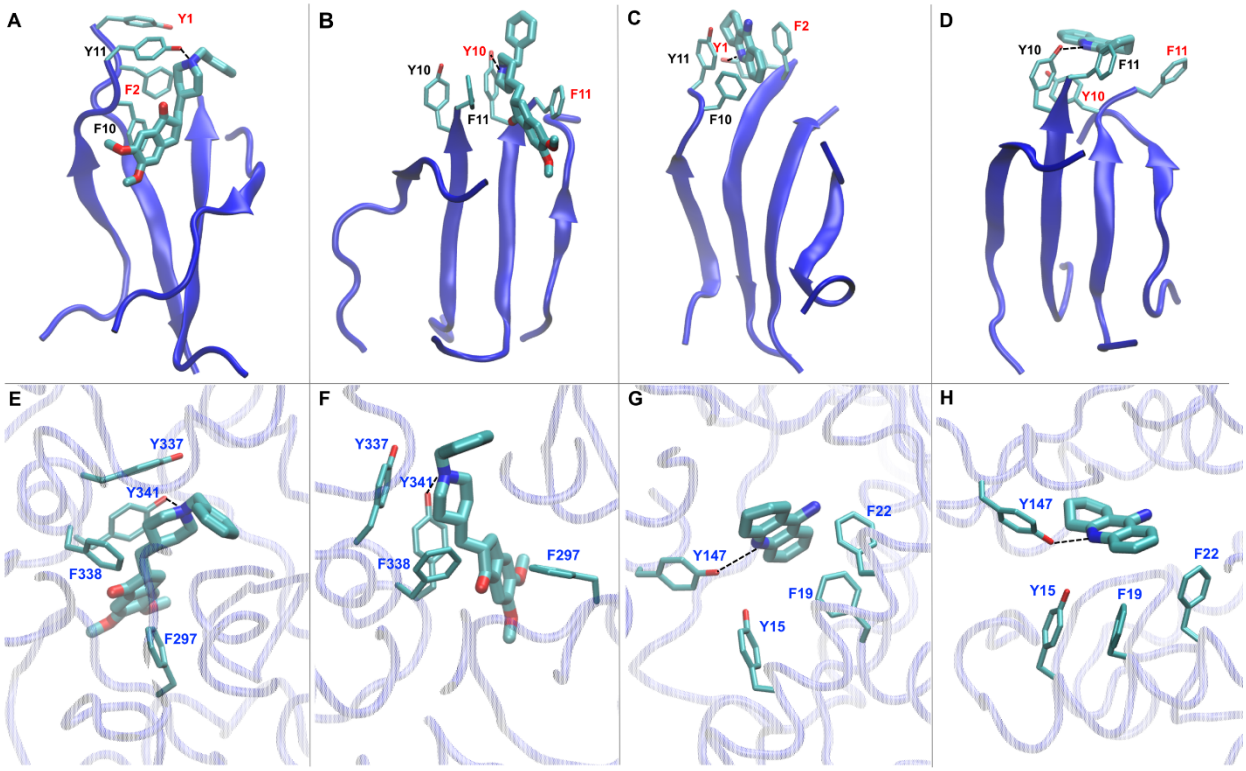


Figure 8.4. Molecular graphic images showing the mimicry between the binding modes of Donepezil and Tacrine binding to designed peptides, YFTGAIIGNFY and FYTGAIIGNYF, and experimentally resolved structures of Donepezil and Tacrine in complex with proteins human acetylcholinesterase and human methyltransferase. (A) Computationally docked binding mode of Donepezil binding to YFTGAIIGNFY (top), designed and validated against (E) a materialphore model extracted from PDB ID: 4EY7[292]. (B) Computationally docked binding mode of Donepezil binding to FYTGAIIGINYF (top), designed and validated against (F) a materialphore model extracted from PDB ID: 4EY7[292]. (C) Computationally docked binding mode of Tacrine binding to YFTGAIIGNFY (top), validated against (G) a materialphore model extracted from PDB ID: 2AOX[296]. (D) Computationally docked binding mode of Tacrine binding to FYTGAIIGNYF (top), validated against (H) a materialphore model extracted from PDB ID: 2AOW[296]. Panels E and H represent different slices of same materialphore model derived from the protein structure with PDB ID: 4EY7[292]. The backbone of the designed peptides and the specific amino acid−compound binding motif is shown in blue cartoon representation. The amino acid side chains interacting with a compound are shown in licorice representation; the compounds are shown in thick licorice representation; hydrogen formed between a tyrosine and compound is indicated by black dotted lines.

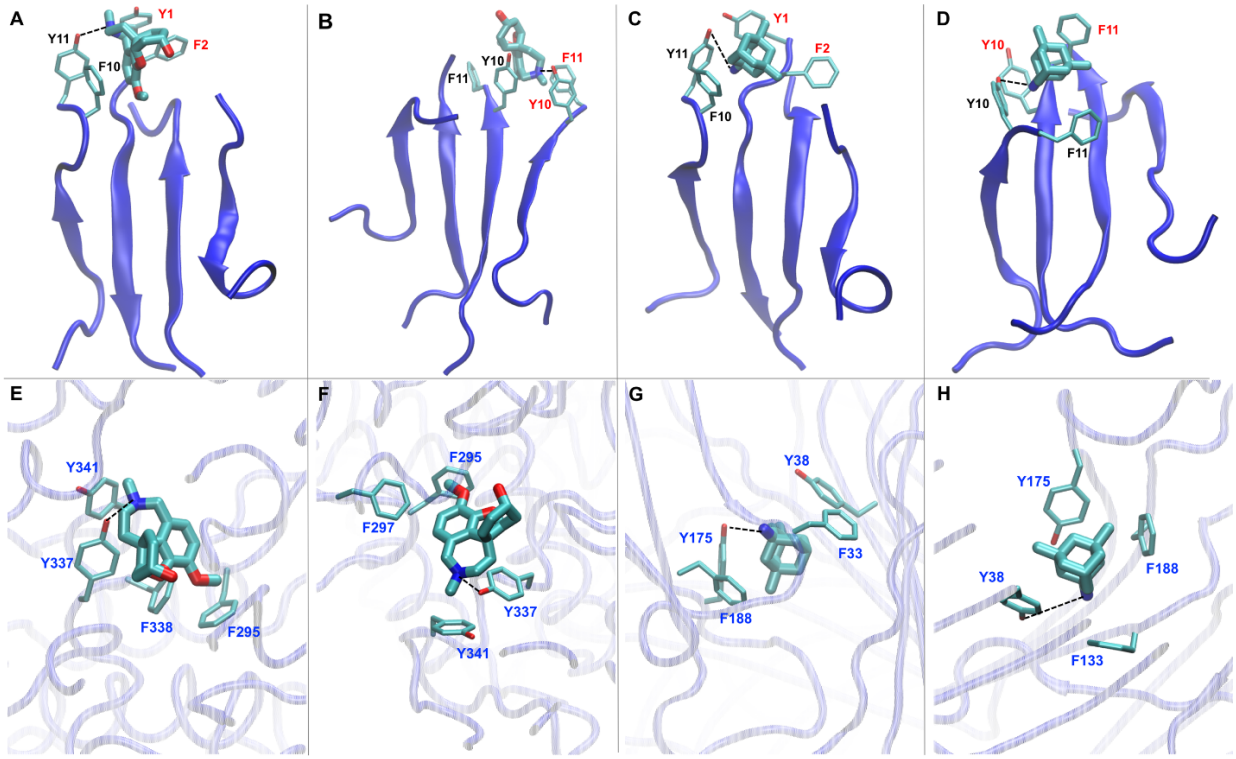


Figure 8.5. Molecular graphic images showing the mimicry between the designed peptides, YFTGAIIGNFY and FYTGAIIGNYF, binding modes to Galantamine and Memantine, and experimentally resolved structures of proteins human Acetylcholinesterase and Erwinia chrysanthemi binding to Galantamine and Memantine. (A) Computationally docked binding mode of Galantamine binding to YFTGAIIGNFY (top), designed and validated against (E) a materialphore model extracted from PDB ID: 4EY6[292]. (B) Computationally docked binding mode of Galantamine to FYTGAIIGINYF (top), designed and validated against (F) a materialphore model extracted from PDB ID: 4EY6[292]. (C) Computationally docked binding mode of Memantine to YFTGAIIGNFY (top), validated against (G) a materialphore model extracted from PDB ID: 4TWD[293]. (D) Computationally docked binding mode of Memantine to FYTGAIIGNYF (top), validated against (H) a materialphore model extracted from PDB ID: 4TWD[293]. Panels E and F represent different slices of same materialphore model derived from the protein structure with PDB ID: 4EY6[292]. The backbone of the designed peptides and the specific amino acid−compound binding motif is shown in blue cartoon representation. The amino acid side chains interacting with a compound are shown in licorice representation; the compounds are shown in thick licorice representation; hydrogen formed between a tyrosine and compound is indicated by black dotted lines.

Tacrine is primarily “wrapped” by the designed amino acids of the designed peptide YFTGAIIGNYF clustering around the entire compound (Figure 8.4C), whereas Tacrine primarily adheres the binding pocket formed by the four designed amino acids of the designed peptide

FYTGAIIGNYF (Figure 8.4D). For both designed peptides, Tacrine forms rich π-π and cation-π interactions with the designed amino acids (Figure 8.4C and 8.4D). Additionally, for both designed peptides, Tacrine forms a hydrogen bond with the hydroxyl group of the designed tyrosines through both its charged and uncharged amine groups of the respective designed peptides in nearly half of the top ten selected structures (Figure 8.4C and 8.4D).

The central core of Galantamine is primarily "wrapped" by the designed amino acids of the designed peptide YFTGAIIGNFY (Figure 8.5A), whereas Galantamine primarily adheres to the binding groove formed by the designed amino acids of the designed peptide FYTGAIIGNYF (Figure 8.5B). For both designed peptides, Galantamine forms rich π-π and cation-π interactions with the designed amino acids (Figure 8.5A and 8.5B). For the designed peptide YFTGAIIGNFY, Galantamine forms a hydrogen bond with the hydroxyl group of the designed tyrosines through its charged amine group in nearly half of the top ten selected structures as well as through its hydroxyl group in three of the top ten selected structures (Figure 8.5A). For the designed peptide FYTGAIIGNYF, Galantamine forms a hydrogen bond with the hydroxyl group of the designed tyrosines through its charged amine group in three of the top ten selected structures as well as through its hydroxyl group in two of the top ten selected structures (Figure 8.5B).

Memantine is primarily loosely "wrapped" by the designed amino acids of both designed peptides YFTGAIIGNYF and FYTGAIIGNYF, clustering around the entire compound (Figure 8.5C and 8.5D). Memantine is not tightly "wrapped", as is the case for Tacrine, presumably due to the bulkier shape of Memantine. For both designed peptides, Memantine forms rich hydrophobic-π and cation-π interactions with the designed amino acids (Figure 8.5C and 8.5D). Additionally, for both designed peptides, Memantine forms hydrogen bonds with the hydroxyl

group of the designed tyrosines through its amine group in nearly half of the top ten selected structures for each of the designed peptides (Figure 8.5C and 8.5D).

8.4 Concluding Remarks

The study presents the first, to our knowledge, functional amyloid materials that can potentially bind AD drugs by mimicking the mechanism according to which the same AD drugs bind to proteins according to experimentally resolved structures, including the target enzyme AChE, which is part of the inhibition mechanism for three of the four AD drugs investigated. While Tacrine was withdrawn from the market owing to hepatotoxicity, the computationally designed amyloid materials could be very promising carriers for Donepezil, constitute as starting points for the design of additional drug carriers that could bind to combinations of NMDA and Cholinesterase inhibitors. The importance of the former lies within the fact that among all the Cholinesterase inhibitors, Donepezil is the most superior due to its high potency and selectivity for AChE. In addition, a combination of Memantine and Donepezil (Namzaric) is administered to patients with moderate-to-severe AD. Especially after the announcement of the majority of AD companies that they will stop research and investment on new AD drugs due to failure in stage III clinical trials, our amyloid peptide drug delivery carriers could be very promising in enhancing the effectivity of current marketed drugs.

8.5 Innovation in Methods

In this study, we showed that the computational protocol as presented in chapter 7 with an additionally component developed and executed by Joseph Jakubowski, a graduate student of Dr. Tamamis' lab, can be successfully implemented to design functional amyloid materials binding to compound (AD drugs in this case). In addition to implementing the protocol of chapter 7 to design amyloid materials for the AD drugs, the following additional investigations were performed: (i)

We considered both antiparallel and parallel configurations as potential amyloid designable scaffolds to improve the success rates of the successfully designed amyloid materials, which can potentially be functional in either configurations. (ii) As our goal here was to design amyloid materials binding to compounds, advancement was made in the computational validation to take into account specific interactions (e.g., electrostatic interactions, hydrogen bonds, cation-π, aromatic and hydrophobic interactions) between the compounds' group and the amino acids. In this perspective, we additionally performed a structural and energetic analysis investigating the compounds' interactions with the designed amino acids and their binding affinity to the computationally designed amyloid materials. As mentioned above, the latter (i.e ii) was developed and executed by Joseph Jakubowski, a graduate student of Dr. Tamamis' lab.

8.6 New Findings and Potential Applications

In this study, we have explored the possibility of amyloid materials as carriers for the therapeutic-sustained release of AD drugs such as donepezil and tacrine. The amyloid materials designed here if proven successful can potentially act as novel drug delivery carriers which can enhance the half-life and effectivity of current AD drugs and potentially the stabilization of cognition (enhancement of current AD treatments). Motivated by the ability to design amyloid materials to bind multiple AD drugs, their ability to be modified at the sequence level and their biocompatibility, such potentially successful drug-carrier amyloid materials could pave the way for a novel direction of delivering the specifically studied AD drugs.

9. DESIGN OF AMYLOID MATERIALS WITH CELL PENETRATING AND DNA BINDING PROPERTIES

9.1 Introduction

9.1.1 Overview and Significance

The delivery of nucleic acids into cells to restore a deficient gene to normal expression level is an important is a powerful tool in the field of gene therapy. An ideal delivery agent should be able to bind to nucleic acids effectively, access cells, avoid degradation of the carrying nucleic acids and deliver them in cytosol for gene regulation or to the nucleus for gene expression.[303,304,305] Yet, the delivery of nucleic acids into cell can be challenging due to their strong negative charge that inhibits their internalization, in addition to their susceptibility to nucleolytic enzymes. Thus, there has been significant interest recently to develop stable, efficient, and biocompatible delivery agents which led to the development of a variety of non-viral vectors that can incorporate genetic material and efficiently deliver it into the cells.[306]

Cell-penetrating peptides (CPPs) are one such group of non-virial vectors that recently emerged to deliver nucleic acids. The specific group of vectors are generally short cationic peptides with 5-30 amino acids and have positively charged amino acids arginine and lysine amino acids in abundance. They can effectively cross cellular membranes, they have limited toxicity and could function as transfection carriers for nucleic acid cargos including siRNAs and plasmids.[307] The cellular uptake of the CPPs can occur through various energy-independent[308] or energy-dependent internalization mechanisms[309] such as the interaction of the positively charged residues with the negatively charged phospholipids of the cell membrane to facilitate direct internalization and endosomal uptake via endocytic pathways[310,311]. Due to their ability of accessing cellular

membranes, CPPs could enhance the transportation of conjugated bioactive cargos, which could initiate the expression or function of specific intracellular targets. Biomolecules can be conjugated with the CPPs through a covalent bond or by a non-covalent complex formation.[306]

CPPs act as delivering agents of nucleic acids by forming non-covalent complex which involves the electrostatic interaction of the positively charged residues of the peptide with the negatively charged nucleic acids. Yet, there are challenges such as DNA condensation and enzymatic degradation protection through the formation of the peptide-DNA complex.[312] Examples of the peptide delivery agents or enhancers of the expression of a gene of interest into cells include PepFect14 peptide vector,[313] amphipathic α-helical peptide NF55[314,315], and the highly cationic TAT peptide that can directly penetrate the plasma membrane as polyelectrolytic complex upon interacting with plasmid DNA[312]. Amyloid forming peptides are considered also as promising nanomaterials in boosting gene transduction by utilizing the positive charges on the fibrillar nanosheet to capture nucleic acids and virion particles and subsequently increase their cell attachment and fusion.[33,316,317,318]

9.1.2 Goal

Motivated by the fact that (i) cell penetrating peptides have a majority of positively charges residues, (ii) cell penetrating peptides can act as DNA delivering agents, and (iii) amyloid forming peptides can act as DNA delivering agents, we aimed to design amyloid materials combining cell penetrating and DNA binding properties.

9.1.3 Hypothesis

We hypothesized that our goal could be achieved through the use of a combination of computational and rational design approaches, starting from amyloid designable scaffolds

RGDSGAITIGC and YATGAIIGNII (presented in chapters 3 and 5), and the introduction of mutations at their non-β-sheet forming residues via the use of an amalgam of sequence-based information associated with amino acid motifs leading to CPPs and DNA binding properties.

9.1.4 Objectives Accomplished

In this chapter, we suggest the design of two specific sequences which can constitute functional amyloid materials with the potential capacity to bind to promote cell-penetration and DNA binding. The design was performed through a combination of computational and rational design approaches and was complemented by Dr. Tamamis' and Archontis' computational protocol for the elucidation of structures of self-assembling peptides[130]. Overall, our results indicate that the designed peptides could be promising amyloid materials promote cell-penetration and DNA binding, as the designed amino acids in the highly-ordered and well-aligned structures of the amyloid fibrils are sufficiently solvent exposed for functionalization purposes.

Experiments evaluating the computationally designed materials functional properties are currently being performed in Dr. Mitraki's lab, and by additional experimentalists at the University of Crete.

9.2 Computational Methods

In summary we used data-driven design approaches (in two directions), and rationally designed two peptides, with sequence, NH_3^+-KYRSGAITIGY-$CONH_2$ and NH_3^+-KYKGAIIGNIK-$CONH_2$. Upon selection of the two peptide sequences, we computationally investigated the two designed peptides using Replica Exchange Molecular Dynamics (REMD) simulations. Firstly, we performed simulations of the two peptides in infinite dilution to investigate their conformational properties, while subsequently, we performed simulations of copies of the two peptides in a cubic box to investigate their self-assembly properties. Using the structures

extracted from the finite dilution simulations, we categorized the β-sheet structures formed into 2-,3-,4-,5-, and 6-stranded antiparallel, parallel, or mixed β-sheet structures and identified the predominant configuration (parallel or antiparallel) of each of the two peptides in the β-sheets as well as the key β-sheet interactions formed in the self-assembled structures of the two peptides. Subsequently, we extracted the highly ordered and well-aligned β-sheet structures, investigated the key interactions formed between the amino acids within the structures, calculated the solvent accessibility of the designed residues to assess the peptides' functionality. The simulations and analysis were performed in line with a protocol developed by Tamamis and Archontis[130] which was widely used by our group to shed light into the amyloid self-assembly properties of short peptides[22,23,24,192]. In what follows, we provided a detailed description of the computational methods employed to design and study the two peptides.

9.2.1 Design of Amyloid Forming Peptides with Cell Penetrating and DNA Binding Properties

We aimed to design functional amyloid peptides which possess cell penetrating and DNA binding properties through two separate approaches:

In the first approach, we used the computationally elucidated structures of elementary antiparallel β-sheet structural units of the amyloid designable scaffold formed by the peptide RGDSGAITIGC[24] (elucidated in chapter 3 and ref[24]) as flexible structural templates which is an input in the Module II of the computational protocol (presented in chapter 7). The computational design was performed by introducing mutations to designable (underlined) positions 1, 2, 3, and 11 of <u>RGD</u>SGAITIG<u>C</u>[24]. The computational design in Module II of the protocol was performed without subjecting the model to the constraints related to the primary materialphore models. Upon solution of the model, a total of 20^4 set of mutations were evaluated and they were ranked according to the energy defined in the objective function. Subsequently, for the post design

constraints, we hypothesized that a combination of positively charged residues (Arg, and Lys) and tyrosine residues at the designable positions would be desirable as (i) positively charged residues are known to penetrate cell membranes[319], and (ii) aromatic residues in the order Y>W>F were popular in the proteins interacting with DNA[320]. To introduce cell penetrating motifs at the N-terminal of the designed sequence, we performed a bioinformatics analysis on the motifs with sequence XXXS in cell penetrating peptides deposited in the Database of Cell-Penetrating Peptides (CPPsite2.0[321]). From the bioinformatics search, we observed that only RYYS and KYRS sequence motifs occurred as a part of any cell penetrating peptides. We selected the sequence motif KYRS as we considered that tyrosine at position 3 in the motif RYYS could extend the β-sheet core, probably negatively affect its functional properties. After imposing the above post-design constraints, KYRSGAITIGY was the top ranked designed peptide, and was selected for further investigation. It is worth noting that the selected designed peptide sequence is among the top 10% of all 20^4 possible sequences according to the energy defined in the objective function prior to the introduction of any constraints.

In the second approach, we aimed at rationally designing a functional amyloid peptide, which possesses cell penetrating and DNA binding properties, by introducing mutations to designable (underlined) positions 1, 2, 3, and 11 of YATGAIIGNII. Similarly to the rational design describe above, we performed a bioinformatics analysis for sequence motifs containing positively charged residues or tyrosines in cell penetrating peptides deposited in the Database of Cell-Penetrating Peptides[321]. In the analysis, we disallowed tyrosine at position 3 to avoid the extension of the β-sheet core of the self-assembled designed peptides. Based on the bioinformatics analysis, the motifs KYK, KYR, RYK, and RYR are possible amino acid replacements at the designable positions 1, 2, and 3 with the motif KYK occurring most frequently. Thus, we mutated the residue

positions 1, 2, and 3 to lysine, tyrosine, and lysine, respectively. In contrast to the computational design described above in the first approach, in this approach, motivated by the fact that positively charged residues are beneficial for both cell-penetration and DNA binding functionality, we rationally substituted position 11 with lysine. Thus, the aforementioned direction resulted in the designed peptide KYKGAIIGNIK, which was selected for further investigation.

The first designed peptide was engineered on the basis of NH_3^+-RGDSGAITIGC-$CONH_2$. For the current study, we investigated both selected peptides with the same NH_3^+- and -$CONH_2$ terminal ends. Thus, the selected designed peptides with sequence NH_3^+-KYRSGAITIGY-$CONH_2$ and NH_3^+-KYKGAIIGNIK-$CONH_2$ were subsequently investigated using infinite and finite dilution simulations to uncover their self-assembly properties, independently, as described in the following sections.

9.2.2 Infinite Dilution Simulations Investigating the Peptides' Conformational Properties

We performed infinite dilution simulations of the two selected peptides with sequence NH_3^+-KYRSGAITIGY-$CONH_2$ and NH_3^+-KYKGAIIGNIK-$CONH_2$ in aqueous solution, independently, using REMD simulations[133,134,135,136,137,138] in CHARMM[110]. The REMD simulations consisted of eight replicas with temperatures 283, 300, 318, 336, 356, 377, 403, and 432 K. The initial structures of the peptides corresponded to linear structures build in CHARMM[110]. The peptides were modeled using the CHARMM19 force-field[322], and the aqueous environment was modeled using the FACTS19[123] implicit solvent model with the surface tension coefficient set to 0.015 kcal mol^{-1} $Å^{-2}$ [23,24,192]. For all simulations, we used Langevin dynamics with a friction coefficient of 5 ps^{-1} introduced on all heavy atoms and simulation snapshots extracted in 100 ps intervals. The duration of each simulation per temperature was 150 ns for both peptides, independently, for an aggregate total simulation duration of 1.2 μs per peptide. Upon completion

of the infinite dilution simulations, we extracted the 15000 simulation snapshots of the 300 K trajectory. We subsequently performed a root mean squared deviation (RMSD)-based clustering analysis on the extracted snapshots using Wordom[141,142] and selected representative conformations from the six most populated clusters for each peptide resulting in six conformations per peptide. The clustering analysis was performed based on the backbone atoms of the peptides using a clustering radius of 2 Å and the quality-clustering algorithm. The six extracted conformations per peptide corresponded to the centers, or most representative structure, of each cluster and were used as initial structures in the finite dilution REMD simulations investigating the peptides' self-assembly properties described in the latter section.

9.2.3 Finite Dilution Simulations Investigating the Peptides' Self-Assembly Properties

We performed independent REMD simulations in CHARMM[110] for both designed peptides to investigate their self-assembly properties, analogously to refs[22,23,24,40,130,131,192]. The simulation system corresponded to six copies of the peptide KYRSGAITIGY in a 157 Å cubic periodic box and KYKGAIIGNIK in a 154 Å cubic periodic box, resulting in an approximately 3 mg/ml concentration for the two simulation systems. The initial conformation of the six peptide copies were obtained from the six conformations extracted from the corresponding aforementioned finite dilution REMD simulations. For each of the two simulations, all six copies of each peptide were initially placed in the center of the cubic periodic box, and then translated by +25 Å in the x-direction, -25 Å in the x-direction, +25 Å in the y-direction, -25 Å in the y-direction respectively, +25 Å in the z-direction, or -25 Å in the z-direction such that a peptide was placed at the center of each face of a 50 Å cubic box. The REMD simulations, for each peptide independently, consisted of sixteen replicas with temperatures 290, 295, 300, 310, 305, 315, 321, 327, 333, 339, 345, 352, 359, 366, 373 and 380 K. The simulations were performed using the CHARMM19 force-field[322]

and the FACTS[123] implicit solvent model with the surface tension coefficient set to 0.015 kcal mol^{-1} Å^{-2} [23,24,192] and simulation snapshots extracted in 10 ps intervals. For all simulations, we used Langevin dynamics with a friction coefficient of 5 ps^{-1} on all heavy atoms. The duration of each simulation per temperature was 1,000 ns for both peptides, independently, for an aggregate total simulation duration of 16 μs per peptide. Upon completion of the finite dilution simulations, we extracted the 100,000 simulation snapshots of the 300 K trajectory for each peptide independently.

9.2.4 Categorization of β-sheet Structures and Key β-sheet Interactions

We analyzed the formation of intermolecular β-sheet structures and categorized the intermolecular β-sheet patterns into antiparallel, parallel, or mixed 2-, 3-, 4-, 5-, 6- stranded β-sheet structures, similarly to refs[22,23,24,40,130,131,192] for designed peptides KYRSGAITIGY and KYKGAIIGNIK. The intermolecular β-sheet structures were identified using DSSP[185] and categorized using in-house FORTRAN programs. We calculated the (%) moving average number of structures containing antiparallel, parallel, or mixed β-sheet structures in each of the 2-,3-,4-,5-, 6- stranded peptides. For both designed peptides, the analysis showed that the peptides preferentially form antiparallel β-sheet structures over parallel β-sheet structures. Following this analysis, we focused our further analysis on 4-,5-, and 6- stranded antiparallel β-sheet structures, which we considered more complex compared to 2- and 3- stranded β-sheet[22,23,24,40,130,131,192], and thus can potentially correspond to patterns naturally occurring fibrils. Thus, we extracted the 4-, 5-, and 6-stranded antiparallel β-sheet structures from each of the simulation trajectories corresponding to the two designed peptides and calculated the (%) probability of a pair of residues belonging to separate adjacent peptides to be involved in a β-bridge conformation. From this analysis, we identified the predominant configuration (parallel or antiparallel) and key patterns of β-sheet interactions which indicate the key amyloidogenic regions of each designed peptide. The

specific analysis, in addition, also identifies if the designed residues are predominantly participating in the β-sheet interactions, which would hinder their availability for functionality.

9.2.5 Identification of Well-Aligned and Highly Ordered β-sheet Conformation Using Polar (P1) and Nematic (P2) Order Parameters

We assessed the extent of peptide alignment and relative orientation of the individual peptides within the 4-,5-, and 6-stranded antiparallel β-sheet structures using the P_1 and P_2 parameters. The analysis was performed in line with refs.[22,23,24,40,130,131,192], with the unit vector, $\vec{z}_i$, defined as the segment spanning from the Cα atom of residue 4 to the Cα atom of residue 9. According to the computed P_1 and P_2 parameters, among the 4-,5-, and 6-stranded antiparallel β-sheet structures, highly populated and highly ordered β-sheet structures were observed only in the 4-stranded antiparallel β-sheet structures for both peptide systems. This can be attributed to the limited number of peptides within the finite dilution self-assembly simulations. Thus, we focused our analysis on 4-stranded antiparallel β-sheet structures formed by the two designed peptides and constructed free energy landscapes for each peptide system using two dimensional probability $P(P_1,P_2)$ and *Eq. 9.1*.

$$G(P_1,P_2) = -k_B T \ln\left[P(P_1,P_2)\right] \qquad 9.1$$

From the free energy basins within the free energy landscapes, we extracted the highly ordered and well-aligned β-sheet structures for each peptide system.

9.2.6 Structural Analysis of the 4-strandend Highly Ordered and Well-Aligned Structures

We determined the solvent accessibility of the designed residues to assess their exposure and their potential functionality. The degree of solvent accessibility of residues at positions 1, 2, 3, and 11 within the extracted highly ordered and well-aligned 4-stranded antiparallel β-sheet structures for both designed peptides, KYRSGAITIGY and KYKGAIIGNIK, was determined using Wordom[141,142]. Similarly to refs.[24,192] , we calculated the solvent accessible surface area : total accessible surface area ratio of residues at positions 1, 2, 3, and 11 of the two central peptides in each highly ordered and well-aligned 4-stranded antiparallel β-sheet structures. As the degree of solvent accessibility of the two outer peptides should be artificially high due to the absence of interacting peptides on both sides, their values are not reported. The total accessible surface areas were defined as the maximum solvent accessible surface area (SASA), which is the SASA of the set of residues with all other atoms removed as in refs.[24,192].

9.3 Computational Results

9.3.1 Categorization of β-sheet Structures and Key β-sheet Interactions

We investigated the self-assembly properties of designed peptides KYRSGAITIGY and KYKGAIIGNIK using independent REMD simulations. Within the simulations, the peptides frequently formed β-sheet structures comprising 2-, 3-, 4-, 5-, and 6-stranded β-sheet structures (Figure 9.1). 2-stranded β-sheet structures were observed in both parallel and antiparallel conformations (Figure 9.1). 3-,4-,5-, and 6-stranded β-sheet structures were observed in parallel, antiparallel, and mixed conformations (Figure 9.1). As in refs.[22,23,24,40,130,131,192], structures comprising mixed conformations, in which the peptides are forming at least one pair of parallel and one pair of antiparallel β-sheet conformations, or complex conformations, in which one peptide is forming β-sheet interactions with more than two separate peptides, were not investigated as they are not expected to be representative of naturally occurring amyloid fibrils. In line with the

dominant configuration of the designable scaffolds that were used to produce the designed peptides sequences, both peptides showed a tendency to form antiparallel β-sheet structures over parallel β-sheet structures (Figure 9.1C to 9.1J). Thus, we extracted the antiparallel 4-,5-, and 6-stranded β-sheet structures for further analysis as they are more likely to represent naturally occurring fibrils than the less complex 2- and 3-stranded β-sheet structures.

Using these structures, we calculated the propensities of any two residues belonging to separate adjacent β-sheet bonded peptides to form intermolecular β-bridge interactions (Figure 9.2). In the designed peptide KYRSGAITIGY, the residues involved in β-bridge interactions were predominantly within the GAITIG motif (Figure 9.2A, 9.2C, 9.2E). In the designed peptide KYKGAIIGNIK, the residues involved in β-bridge interactions were predominantly within the GAIIGN motif (Figure 9.2B, 9.2D, 9.2F). Importantly, for both peptides, the designed residues are not involved in β-bridge interactions, indicating that the designed residues may be exposed and possess functional properties, including cell penetration and DNA binding.

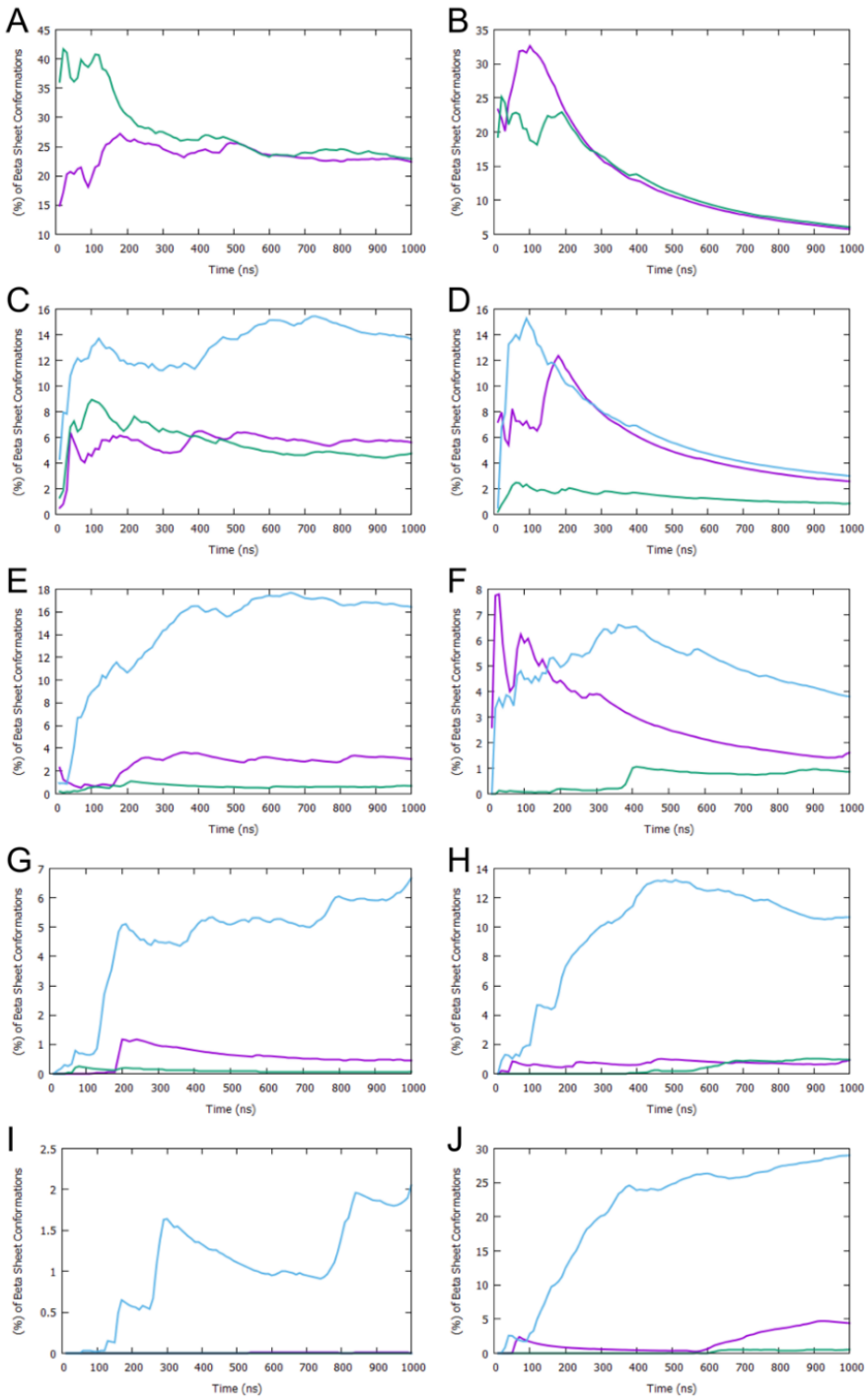


Figure 9.1 Moving averages of the fraction (%) of conformations with intermolecular β-sheets (Y-axis) in the REMD simulations at 300 K with respect to time (X-axis) for KYRSGAITIGY and KYKGAIIGNIK. Figures A, C, E, G, and I correspond to 2-, 3-, 4-, 5-, and 6-stranded β-sheets of peptide KYRSGAITIGY, respectively. Figures B, D, F, H, and J correspond to 2-, 3-, 4-, 5-, and 6-stranded β-sheets of peptide KYKGAIIGNIK, respectively. Purple color corresponds to antiparallel β-sheets, green color corresponds to parallel β-sheets, and blue color corresponds to mixed β-sheets.

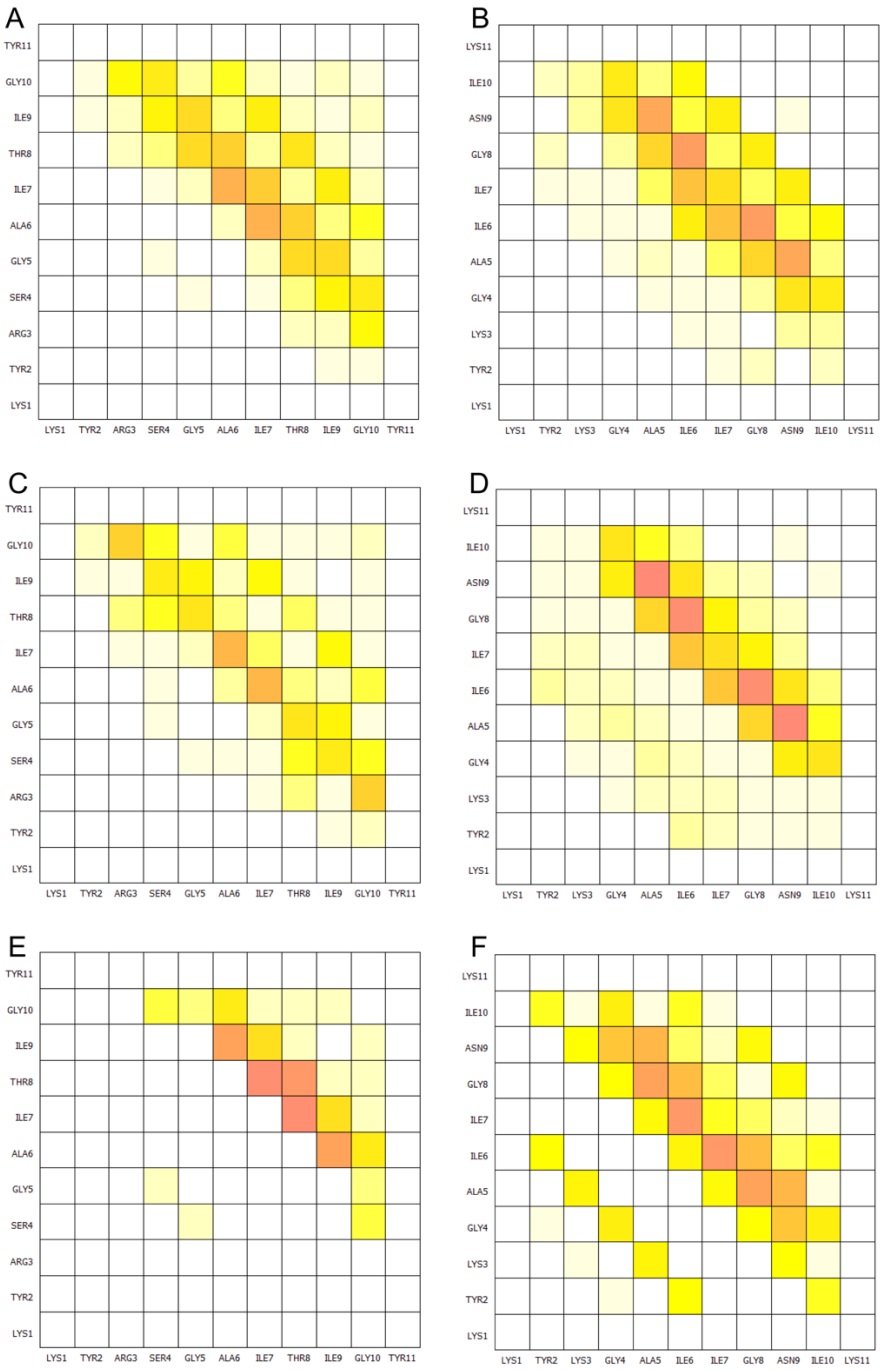


Figure 9.2 Density (%) maps of residue pairs forming intermolecular β-bridges for KYRSGAITIGY and KYKGAIIGNIK. The pairs of residues belong to nearest neighboring peptides participating in an isolated β-bridge or extended β-sheet conformation in the REMD simulations at 300 K. Figures A, C, and E correspond to 4-, 5-, and 6-stranded antiparallel configurations of peptide KYRSGAITIGY respectively. Figures B, D, and F correspond to 4-, 5-, and 6-stranded antiparallel configurations of peptide KYKGAIIGNIK respectively.

9.3.2 Identification of Well-aligned and Highly Ordered β-sheet Conformation Using Polar (P_1) and Nematic (P_2) Order Parameters

We determined the degree of alignment and order of the peptides within the 4-,5-, and 6-stranded antiparallel β-sheet structures based on P_1 and P_2 parameters. For both of the designed peptides, KYRSGAITIGY and KYKGAIIGNIK, similarly to our previous studies[23,24,192], highly ordered and well-aligned antiparallel β-sheet structures predominantly occurred within 4-stranded, rather than 5- or 6-stranded, antiparallel β-sheet structures, potentially due to the larger population of 4-stranded antiparallel β-sheet structures. Thus, we constructed free energy landscapes based on the P_1 and P_2 parameters corresponding to the 4-stranded antiparallel β-sheet structures (Figure 9.3A, 9.3B). From the free energy basin, indicated by black dotted lines in Figure 9.3A, 9.3B, we extracted the highly ordered and well-aligned 4-stranded antiparallel β-sheet structures, as these structures can potentially correspond to elementary β-sheet structural units of the naturally occurring fibrils. Representative conformations of the 4-stranded antiparallel β-sheet structures are shown in Figure 9.3C, 9.3D. The ensemble of highly ordered and well-aligned 4-stranded antiparallel β-sheet structures for both the designed were collected and analyzed below to determine the functionality of the designed residues.

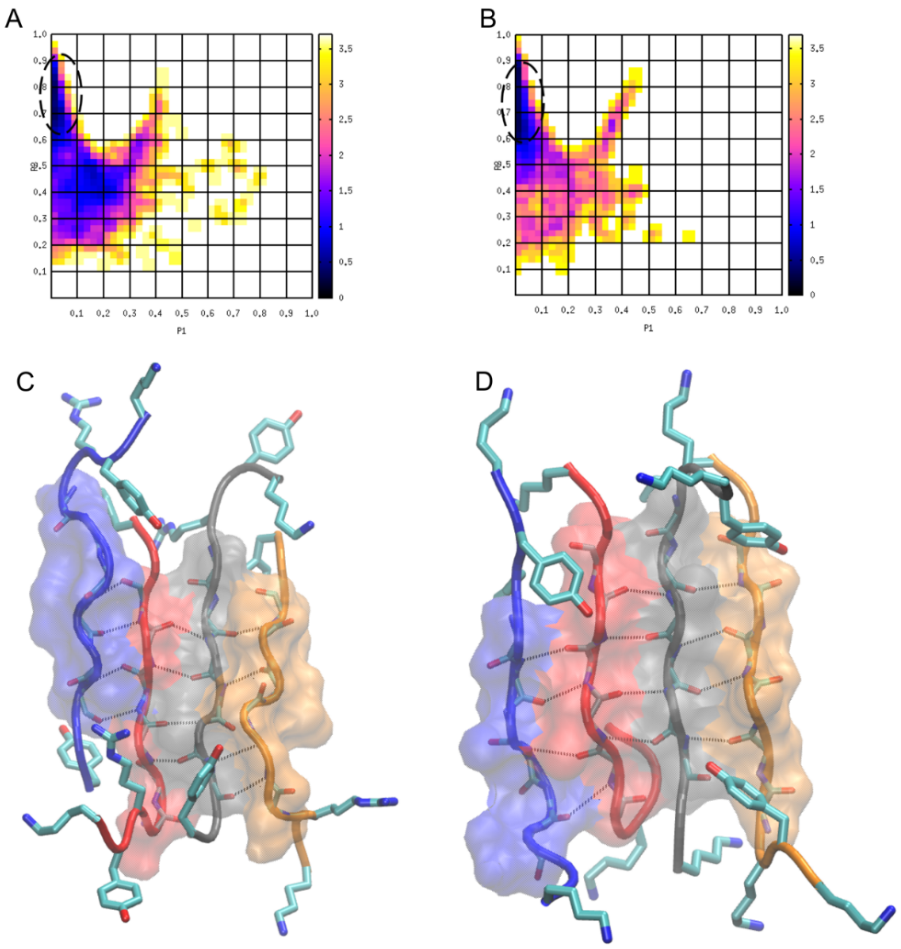


Figure 9.3 Free energy surfaces constructed from the 2-D probabilities of order parameters P_1 and P_2 and molecular graphics images of representative structures of KYRSGAITIGY and KYKGAIIGNIK extracted from the free energy minima. Upper panel: Free energy surfaces constructed from the 2-D probabilities of order parameters P_1 and P_2, calculated using the 4-stranded antiparallel β-sheets observed in the replica exchange MD simulations at 300 K. Surface in panel A, and B, correspond to 4-stranded antiparallel β-sheets formed by KYRSGAITIGY, and KYKGAIIGNIK, respectively. The global free energy minima in the plots are located in basins which are marked using black dashed lines, and within these we observe the presence of highly ordered and well-aligned β-sheet conformations. Bottom panel: Molecular graphic images of representative highly ordered and well-aligned conformations of peptides KYRSGAITIGY, and KYKGAIIGNIK in antiparallel arrangement. The peptides' backbone is shown in tube representation and the backbone atoms forming β-bridges are shown in thin licorice representation and are colored by name, and the β-bridge associated hydrogen bonds are shown using black dashed lines. The peptides are colored in blue, red, gray and orange from left to right. Residue moieties 4-9 in each of the two designed peptides form amyloid-zipper like patterns and are shown in transparent surface representation. Side chain atoms of residues at positions 1, 2, 3, and 11 are shown in thick licorice representation, and are colored by name.

9.3.3 Structural Analysis of the 4-Strandend Highly Ordered and Well-Aligned Structures

In the 4-stranded highly ordered and well-aligned antiparallel β-sheet structures of the designed peptides, the amyloid steric zipper comprises residues GAITIG for KYRSGAITIGY and residues GAIIGN for KYKGAIIGNIK, in line with our previous analysis (Figure 9.2). Apart from this, we observe that the designed residues of both peptides occasionally form specific interactions to neighboring residues. Within the structures of KYRSGAITIGY, the sidechain amide group of Lys1 occasionally forms hydrogen bonds to the backbone oxygen of Gly10 of an adjacent peptide (e.g., between peptides 1-2, 2-3, 3-4), and rarely forms hydrogen bonds to the backbone oxygen of Tyr11 of an adjacent peptide and to the sidechain hydroxyl group of Ser4 of a non-adjacent peptide (e.g., between peptides 1-3, 2-4). The side chain of Tyr2 occasionally forms hydrogen bonds to the sidechain hydroxyl group of Ser4 of a non-adjacent peptide, π-π interactions with Y11 of an adjacent peptide, and cation-π interactions with Arg3 of a non-adjacent peptide, and rarely forms hydrogen bonds to the backbone atoms of Gly10 and Tyr11. The side chain of Arg3 occasionally forms cation-π interactions with Tyr11 of an adjacent peptide and hydrogen bonds with the backbone oxygen or side chain hydroxyl group of Ty11 of an adjacent peptide. The side chain of Tyr 11 occasionally forms cation-π or hydrogen bonds to the side chain of Arg3 of an adjacent peptide and π-π interactions with Tyr2 of an adjacent peptide, and rarely forms hydrogen bonds to the backbone oxygen of Ser4 of an adjacent peptide. Within the structures of KYKGAIIGNIK, the sidechain amide group of Lys1 rarely forms cation-π or hydrogen bonds with the side chain of Tyr2 of a non-adjacent peptide. The side chain of Tyr2 rarely forms hydrogen bonds to the backbone atoms of Ile10 or Lys11 of an adjacent peptide, sidechain atoms of Asn9 of an adjacent peptide, and sidechain amide group of Lys1 of a non-adjacent peptide. The side chain amide group of Lys3 occasionally forms hydrogen bonds to the backbone oxygen of Lys1 of a

non-adjacent peptide. The side chain amide group of Lys11 rarely forms hydrogen bonds to the side chain atoms of Tyr2 of a non-adjacent peptide or Asn9 of an adjacent peptide.

Motivated by the occasional formation of interactions formed by the designed residues in both the peptides, we determined the degree of solvent accessibility of the designed residues within the highly ordered and well-aligned 4-stranded antiparallel β-sheet structures to assess their exposure and thus their potential functionality. The degree of solvent accessibility of the designed residues are presented in Table 9.1. For both of the designed peptides, KYRSGAITIGY and KYKGAIIGNIK, the degree of solvent accessibility of all four designed residues, with respect to the initial scaffolds, are above 0.40. According to our previous study on a peptide with sequence NH_3^+-RGDSGAITIGC-$CONH_2$, the solvent accessibility of rationally designed Cys was 0.34 ± 0.09, which was experimentally shown to be adequate for metal binding properties[24]. Thus, all of the designed residues of the two designed peptides of the current study should be considered adequately solvent exposed for DNA and cell penetrating functionality. Combined with their sufficient degree of exposure, the low instances of the designed residues interacting with other residues within the highly ordered and well-aligned 4-stranded antiparallel β-sheet structures supports that the designed residues could yield potentially cell penetrating and DNA binding functionality. Experiments evaluating the computationally designed materials functional properties are currently being performed in Dr. Mitraki's lab, and by additional experimentalists at the University of Crete.

Table 9.1 Degree of solvent accessibility of the four designed residues in the two designed peptides KYRSGAITIGY and KYKGAIIGNIK. The degree of solvent accessibility is calculated for the sidechains of the designed residues based on the definitions provided in the ref[24,192]. The designed residues of the peptide sequences are indicated in bold text.

Peptide	Residue Position 1	Residue Position 2	Residue Position 3	Residue Position 11
NH_3^+-**KYR**SGAITIG**Y**-$CONH_2$	0.58 ± 0.06	0.45 ± 0.10	0.49 ± 0.08	0.43 ± 0.11
NH_3^+-**KYK**GAIIGNI**K**-$CONH_2$	0.63 ± 0.06	0.42 ± 0.09	0.53 ± 0.06	0.53 ±0.07

9.4 Concluding Remarks

In this study, we rationally designed two peptides with sequences KYRSGAITIGY and KYKGAIIGNIK using a combination of data-driven, computational and rational approaches. The two peptides were subsequently investigated for their self-assembly properties. Within the simulations, we observed the frequent formation of intermolecular β-sheets for both the peptides indicating their self-assembly properties. We also observed that both the designed peptides formed antiparallel as their dominant configuration in line with the configuration of the designable scaffolds from which each of them was designed. The key β-sheet forming residue motifs of the peptides KYRSGAITIGY and KYKGAIIGNIK are predicted to be GAITIG and GAIIGN, respectively, indicating the potential functionality of the designed residues which are outside the predicted amyloid core. Subsequently, the designed residues were predicted to be sufficiently solvent exposed suggesting their potential functionality for cell-penetration and DNA binding properties. Experiments evaluating the computationally designed materials functional properties are currently being performed in Dr. Mitraki's lab, and by additional experimentalists at the University of Crete.

9.5 Innovation in Methods

In this study, we innovatively used a combination of rational and data-driven approaches to design two peptides with potential cell-penetration and DNA binding properties. In contrast to our previous study where materialphore models were used to computationally design peptides, here, we used sequence-based bioinformatics data to enable the performed design. This indicates the versatility of the computational protocol presented in chapter 7 to operate under different type of constraints.

9.6 New Findings and Potential Applications

In this study, we aimed to design amyloid materials that can act as nucleic acid delivering agents. We have observed that CPPs are one group of peptides which were frequently used to deliver nucleic acids to cells. We have also observed that CPPs have a majority of positively charges residues, thus we hypothesized that positively charged amyloid materials can act as DNA delivering agents with cell-penetration abilities. Using a combination of computational, and rational approaches we have designed two peptides with cell-penetration and DNA binding abilities. The two peptides were subsequently computational investigated for their (i) self-assembly properties, and (ii) the exposure of designed residues. Our computational results suggest that both the peptides could potentially have cell-penetration and DNA binding properties.

10. COMPUTATIONAL DESIGN OF FUNCTIONAL AMYLOID MATERIALS WITH CELL ADHESION AND CALCIUM/STRONTIUM BINDING PROPERTIES

10.1 Introduction

10.1.1 Overview and Significance

Increasing number of people each year suffer from diseases related to bone and joints, accidents and traumas that affect bones[323]. Bone tissue engineering is the development of viable alternates that can perform the functions of a bone when they are replaced[25]. Seeding porous scaffold which is shaped in the form of desired bone with cell and growth factors is one of the most popular strategies in bone tissue engineering[25]. The porous scaffolds are subsequently cultured and implanted to induce and direct the growth of the desired bone[25].

Commonly used tissue engineering scaffolds such as collagen, gelatin, hyaluronic acid, chitosan, alginate, and decellularized extracellular matrix (ECM) are typically derived from animal or plant sources. Native tissues are composed of hierarchical organization of biological structures from the nanometer to the centimeter scale and their complex biological functions closely link to their high order organization; therefore such scaffolds need to mimic the hierarchical structure of natural tissues in order to provide the necessary structural and biomechanical framework. Mimicking the ECM can therefore be the best strategy to develop advanced functional materials to control cellular behavior and to determine cell fate. Biological self-assembly offers a versatile bottom-up fabrication technology to construct such nanoscale materials imitating the complex hierarchical order of natural tissues. In particular, protein and peptide self-assembling scaffolds are gaining increasing interest compared to synthetic biodegradable polymers.[10]

Bone matrix proteins such as bone sialoprotein and osteopontin contain the sequence motif Arg-Gly-Asp (RGD) [324]. Therefore, RGD bearing self-assembling peptides could serve as ideal scaffolds to induce bone tissue growth and biomineralization, to induce osteogenic differentiation and can used for the functionalization of bone implants[194,195,196]. In addition, as discussed in chapters 3 and 4, RGD is the principal integrin-binding domain present within extracellular matrix (ECM) proteins such as fibronectin, vitronectin, fibrinogen, osteopontin, and bone sialoprotein; due to the capacity of the RGD sequence to bind to multiple integrin species, synthetic RGD peptides offer several advantages for biomaterial applications.

Moreover, bone is made from calcium phosphate in the form of the mineral hydroxyapatite and 99% of the human bodies' calcium is stored in bones[325]. Calcium functions as a stimulant for bone cell differentiation, proliferation of osteoblasts (cells that secret the matrix for bone formation), bone metabolism, and bone mineralization[323,325,326,327,328,329]. Specifically, calcium regulates the proliferation of osteoblasts by increasing the expression of insulin like growth factors (IGFs)[325,327]. Additionally, the recruitment, differentiation and survival of bone cells is modulated by seven transmembrane-spanning extracellular calcium sensing proteins[327]. For example, Tejeda-Montes et al.[330] combined cell attachment motifs such as RGDS or REDV with the calcium mineralizing sequences within the same elastin-like recombinamer membrane in order to test any synergistic effects towards bone regeneration applications, and reported on the enhanced mineralization, osteogenesis, and *in vivo* bone regeneration properties of a bioactive elastin-like recombinamer (ELR) membrane.

Strontium is generally also stored in the human body in the bones which replaces calcium in some of the hydroxyapatite crystal lattices[325]. It has been shown that low concentrations of strontium can stimulate the bone formation[325,331,332,333,334,335,336]. Specifically, strontium has the

ability to increase bone mass through the activation of osteoblast (bone forming) activity and suppression of osteoclast (bone resorbing) function. This also results in the improvement of bone mechanical properties in normal humans. The recognition of strontium effects on bone has led to its expected application in a variety of biomaterial scaffolds used in tissue engineering strategies aiming at bone repairing and regeneration. For example, Place et al.[337] proposed a bone tissue engineering approach in which arginine-glycineaspartic acid (RGD)-modified alginate hydrogels were crosslinked with bioactive strontium, and calcium. It has been shown that strontium released from the gel promotes the osteo-differentiation as shown by the increase of alkaline phosphatase activity, suggesting that the Sr-containing gel could represent a new strategy in bone tissue engineering.

Organic/inorganic composite scaffolds, particularly for bone tissue engineering applications, made of bioceramics or bioactive glasses and biodegradable polymers[338], often include metallic ions as part of the bioceramic or bioactive glass structural composition. These inorganic materials enable metal ion release during their degradation *in vitro* or *in vivo*[339,340]. In addition, a wide range of biomaterials have been developed for tissue-engineering applications, which include the RGD motif as a bioactive signal sequence motif that could be recognized by cells via integrins (cell transmembrane receptors).[10] Interaction between integrins and bioactive epitopes of ECM could activate signal transduction mechanisms, which can induce specific cellular functions including adhesion, migration, proliferation and differentiation.[10]

10.1.2 Goal

Taking into account the potential advanced properties of amyloid materials, including their potential biocompatibility and excellent mechanical properties, and the ability of computational design approaches to introduce changes at their sequence level to yield specific functionality (i.e.,

binding to specific ions/compounds) through our computation protocol presented in chapter 7, here we aimed to design amyloid materials which combine cell-adhesion and calcium/strontium binding properties.

10.1.3 Hypothesis

We hypothesized that our aim can be achieved through the use of our computational design protocol for the functionalization of amyloid materials, by using as inputs:

(1) The flexible structural templates of the amyloid designable scaffolds YATGAIIGNII (chapter 5 and ref[23]), and

(2) An ensemble of materialphore models extracted from a set of experimentally-resolved protein structures from the PDB entailing information on how amino acids bind to calcium and strontium ions.

Following knowledge gained in previous chapters 2, 3 showing the beneficial effect of introducing the RGD motif at the N-terminal domain for its exposure and functionality, we hypothesized that cell-adhesion properties can be achieved by introducing the well-known, and widely studied in the current thesis, cell-adhesion motif RGD at the amino acid positions 1, 2, 3 as in chapter 2, and 3 and refs[24,192] and additional studies.[178,193] Based on that, we hypothesized that functionalization and binding to calcium, and strontium ions could be achieved by the introduction of mutations at residue positions 9, 10 and 11, in combination with the presence of an aspartic acid at position 3, as a fixed mutation; actually the fixed mutation at position 3 can be justified by the fact that most materialphore models collected consist of amino acids binding to calcium and strontium are rich in aspartates.

10.1.4 Objectives Accomplished

The computational protocol described in chapter 7 has been implemented to computationally design amyloid materials with cell adhesion and calcium/strontium binding properties. The computational design was performed using the amyloid designable scaffolds formed by the peptide YATGAIIGNII (chapter 5 and ref[23]) with the amino acid positions 1, 2, and 3 mutated to Arg, Gly, and Asp for cell-adhesion properties. The designable positions 9, 10, and 11 were independently designed for calcium and strontium based on the materialphore models calculated for the two specific ions. Subsequently, following the procedure of the protocol, MD simulations and two-component computational validation were performed for a set of selected designed peptides to evaluate their binding ability to calcium and strontium. In what follows, we analytically describe the methods and results of computational designing amyloid material with cell adhesion and calcium/strontium binding properties.

10.2 Computational Methods

10.2.1 Selection of an Amyloid Designable Scaffold, Structural Analysis and Determination of its Mutable Positions

In chapter 5, we have investigated the amyloid self-assembly properties of the peptide with sequence YATGAIIGNII[23] by performing self-assembly REMD simulations according to the protocol by Tamamis and Archontis[130]. Within the simulations, we observed rich β-sheet content and thus we classified the observed β-sheet conformations into 2-, 3-, 4-, 5-, and 6-stranded antiparallel, parallel, and mixed β-sheet conformations. We observed that the peptide predominantly forms antiparallel β-sheets and GAIIG motif primarily forms the β-sheet core. We also observed that the residues at the termini were not participating in β-sheets and solvent exposed. Thus, we postulated that the specific peptide can be considered as a designable scaffold

and the terminal amino acids can be amenable for modification yielding the desired functionality. We collected the top 50 highly ordered and well-aligned conformations formed by the peptide produced with in the simulations; these structures were used as flexible structural templates in the subsequent design purposes.

10.2.2 Uncovering How Amino Acid Motifs Bind to Calcium and Strontium Ions According to Experimentally Resolved Protein Structures

For the purpose of the current chapter, we collected a small subset of the experimentally resolved structures of proteins binding to calcium and strontium ions. We collected sixteen experimentally resolved structures of proteins binding to calcium ions with the following PDB IDs: 3ws2, 4rhz, 4tx8, 4x3n, 4y5q, 4yu5, 4zg7, 4zpl, 4zps, 5aqe, 5ar2, 5b5s, 5bwg, 5cxf, and 5dpe; we collected fifteen experimentally resolved structures of proteins binding to strontium ions with the following PDB IDs: 1dd9, 1oo0, 1oq7, 1qrk, 1s8f, 1tjm, 1wc3, 2glq, 2woh, 3e4p, 3ws5, 4bw7, 4g25, 4iak, and 4kbn. From each experimentally resolved protein structure, we independently collected amino acid : calcium and strontium ion binding motifs, to construct materialphore models, similar to chapters 6, and 8 (which are defined as the set of amino acids whose side chain center of mass was within 6.5 Å of the resolved calcium and strontium ion). From each of the binding motif, we identified, calculated, and stored the identity of the amino acids as well as (i) distances between amino acid pairs, and (ii) distances between an amino acid and the resolved ion. Points (i) and (ii) along with the identities of the amino acid were defined as primary and secondary materialphore models in the chapters 6, 7, and 8. Both materialphore models combined can capture the geometric properties of the binding motifs.

10.2.3 Optimization-Based Design of Functional Amyloid Materials Combining Cell-adhesion and Calcium/Strontium Binding Properties

We used both the flexible structural templates of the designable scaffold YATGAIIGNII[23] collected in section *10.2.1* and primary materialphore models collected in the section *10.2.2* as inputs to computationally design amyloid material with both cell-adhesion and calcium/strontium binding properties following the computational protocol described in chapter 7. Through the optimization-based computational design model, we aimed to introduce mutations at the positions 1, 2, 3, 9, 10, and 11 of the designable scaffold **YAT**GAIIG**NII** (intended mutable positions in bold). From previous studies, we observed the cell-adhesion properties for a peptide could be achieved through the incorporation of RGD motif in a peptide[24,192] (presented in chapters 3, and 4). Thus, we fixed the mutable positions 1, 2, and 3 to arginine, glycine, and aspartic acid (RGD) on the designable scaffold through the incorporation of suitably selected constraints. To achieve the calcium/strontium binding properties, we mutated the designable positions 3 (fixed as aspartic acid), 9, 10, and 11 by minimizing the energy due to mutations with respect to the constraints of primary materialphore models. We independently solved two optimization-based design models, one for calcium binding, with primary materialphore models collected for calcium ions as input, and one for strontium binding, with primary materialphore models collected for strontium ions as input. From the solution sets, namely, (i) designed peptides with cell-adhesion and calcium binding, and (ii) designed peptides with cell-adhesion and strontium binding, we extracted the designed peptides which were common to both sets since we aimed to design peptide(s) with cell-adhesion and calcium/strontium binding. This resulted in seven peptides which were subjected to further investigation described as follows.

10.2.4 MD Simulation-based Modeling of the Designed Amino Acids on the Designable Scaffold and Computational Validation against Primary and Secondary Materialphore Models

Appropriate mutations were introduced through CHARMM[110] on one of the top 50 flexible structural template of the designable scaffold to create a starting conformation for each of the selected designed peptides' elementary β-sheet structural units. We performed a short 20 ns (at 300 K) simulation for each of designed peptide with water modeled implicitly using FACTS19[123] solvation model. GAIIG core for each of designed peptide was constrained in order to maintain the structural integrity of the β-sheets. Additional technical details for the simulations are detailed in the chapter 6 and ref[22]. Upon completion of the simulations, we extracted snapshots at 10 ps interval which resulted in 2000 snapshots for each of the designed peptide.

Subsequently, we performed the two-component computational validation on the collected simulation snapshots for each of the designed peptide as described in the chapter 7. In summary, in the first component, we evaluated the ability of the designed amino acids to mimic the slice of the primary materialphore model from which the specific design originated. The evaluation was done by comparing the distances between the designed amino acids' side chain center of mass with the corresponding distances in the slice of the primary materialphore model. In the second component of the computational validation, we evaluated the ability of the hypothetical center of mass of the calcium/strontium ion and designed amino acids to mimic the slice of the secondary materialphore model from which the specific design originated. The evaluation was done by the comparing the distance between the designed amino acid side chain center of mass and the hypothetical center of mass of the calcium/strontium ion with the corresponding distance in the slice of the secondary materialphore model. The second component of the computational validation was performed only on the designed pockets which passed the first component of the

computational validation. In addition, we also calculated the percentage of snapshots which passed the first component (designated as primary matches) and calculated the percentage of designed pockets that passed the second component given that they passed the first component (designated as secondary matches) as presented in chapter 7. Both the components of computational validation were performed independently for calcium and strontium binding properties for each of the designed peptides.

10.2.5 MD Simulations Investigating the Self-assembly Properties of the Designed Peptides and Final Computational Validation against Primary and Secondary Materialphore Models

The resulting designed peptides following the computational validation were subjected to self-assembly replica exchange MD (REMD) simulations, analogously to the refs[22,23,24,130,131,192]. In summary, the self-assembly simulations were performed using CHARMM[110] with FACTS19[123] implicit solvation model with a surface tension coefficient of 0.015 kcal mol^{-1} $Å^{-2}$. Langevin dynamics were employed with a 5.0 ps^{-1} friction coefficient on all non-hydrogen atoms of the designed peptides. Sixteen temperatures (290, 295, 300, 310, 305, 315, 321, 327, 333, 339, 345, 352, 359, 366, 373 and 380 K) were employed with a total simulation time of 16 μs and each replica exchange run's duration equal to 10 ps. Upon completion of the simulations, snapshots at 300 K were extracted at 10 ps intervals which resulted in 10,000 snapshots. The extracted snapshots were analyzed according to Tamamis and Archontis protocol[130] summarized in chapter 2.

10.3 Computational Results

10.3.1 Selection of an Amyloid Designable Scaffold, Structural Analysis and Determination of its Mutable Positions

We used the elementary β-sheet structural units formed by the amyloid designable scaffold YATGAIIGNII[23] (elucidated in chapter 5) as flexible structural templated for the computational design with only antiparallel β-sheets considered for design purposed in this chapter. Residue positions 1, 2, 3, 9, 10, and 11 were considered mutable positions on the top50 extracted highly ordered and well-aligned conformations were used as the flexible structural templates in the optimization-based design model (in what follows).

10.3.2 Uncovering How Amino Acid Motifs Bind to Calcium and Strontium Ions According to Experimentally Resolved Protein Structures

We collected sixteen and fifteen experimentally resolved protein structures for calcium and strontium ions, respectively. From the collected protein structures, we extracted materialphore models for both the ions, independently. Within a binding pocket of an ion interacting with amino acids in a specific PDB structures, primary materialphore models represent the all possible relative distances of pairs of interacting amino acids, and secondary materialphore models represent all distances between the amino acids and the ions. Amino acids and ions in the aforementioned definitions are geometrically described by their centers of mass, and an amino acid is considered to be within a compound's binding pocket if their distance is below 6.5 Å.

Our analysis resulted in the extraction of multiple materialphore models for both the ions. The total number of materialphore models extracted for each ion is equal to the sum of the number

of experimentally resolved proteins multiplied by the number of complexed ions with each protein which resulted in 49 and 43 materialphore models for calcium and strontium respectively.

In the present study, the number of designable residue positions is equal to two per peptide per binding pocket, and thus upon β-sheet formation, the total number of residue positions amenable for subsequent design in each binding pocket is four. Since the primary and secondary materialphore models can contain more than four amino acids interacting with the ions (as the number of amino acids interacting with the compound can be greater than four), we additionally considered "slices" of materialphore models to be used for subsequent design purposes. We followed a similar approach in chapters 6 and 8. For a given materialphore model, the slices represent combinations of four amino acids interacting with the compounds and are used as input information in the optimization-based design.

10.3.3 Optimization-Based Design of Functional Amyloid Materials Combining Cell-adhesion and Calcium/Strontium Binding Properties

We introduced the optimization-based design model detailed in the chapter 7. The optimization model aimed to introduce amino acids at the mutable, underlined residue positions of antiparallel flexible structural templates of the amyloid designable scaffold YATGAIIGNII. We independently solved two optimization-based computational design models for the design of amyloid-forming peptides combining, (i) cell-adhesion and calcium binding properties, and (ii) cell-adhesion and strontium binding properties. Positions 1, 2, 3 were mutated to RGD for cell-adhesion, thus positions 9, 10, and 11 were the three mutable positions. Within the design, we aimed that the substituted amino acids could be placed to reproduce the geometry of a slice of primary materialphore for both the ions, independently, at least in a portion λ of the fifty antiparallel flexible structural templates. As in chapter 6, both λ (25% in this case) and the fact

that we allowed a specific deviation for the matching between the distances of introduced amino acids in the flexible structural template and the corresponding distances in the primary materialphore model, enabled our design and accounted for the difference in distances that could originate due to the different size of the substituted amino acids compared to the original amino acids Thr3, Asn9, Ile10 and Ile11.

The optimization model was solved in such a way that primary materialphore model constraints were introduced first to eliminate number of substitutions and result in a total number of feasible number of solutions. Out of 20^3 possible solutions, design for calcium binding properties resulted in 31 designed peptide sequences and design for strontium binding properties resulted in 45 designed peptide sequences. Since we aimed to design peptides with cell-adhesion and both calcium/strontium binding properties, we extracted designed peptides which were common in both sets of solutions. This resulted in a total of seven peptides which were: RGDGAIIGDAD, RGDGAIIGDDA, RGDGAIIGDEN, RGDGAIIGADD, RGDGAIIGDNE, RGDGAIIGDTD, and RGDGAIIGTDD. The resulting seven peptides were investigated further below. For the selected limited number of materialphore models that were extracted overall the entire possible materialphore models available in the PDB, and for the solutions produced aiming at both calcium and strontium binding, we did not consider the energy minimization-based ranking provided by the objective function as a criterion to select the top solutions (or discard any peptides) due to low number of feasible solutions produced for further investigation.

10.2.4 MD Simulation-based Modeling of the Designed Amino Acids on the Designable Scaffold and Computational Validation against Primary and Secondary Materialphore Models

We performed short MD simulations at 300 K to investigate the ability of the designed peptides' elementary β-sheet structural unit to form binding pockets resembling the materialphore

models from which their design originated. Upon completion of the simulations, we performed two-component computational validation as described in chapter 7. The percentage values of primary and secondary component of the validation are presented in Table 10.1. The designed peptides can be considered promising for further evaluation (see below) with the exception of RGDGAIIGTDD for its strontium binding properties. Nevertheless, all peptides were investigated analytically in what follows.

Table 10.1 Designed peptide sequences for calcium/strontium binding and their corresponding first and second validation component percentage statistics based on short MD simulations. The validation statistics were calculated over snapshots extracted from the short MD simulations. The listed percentages for the first and second validation components correspond to the percentage of snapshots containing primary and secondary matches, within the snapshots containing primary matches, which were identified from the first validation component (see additional details in chapter 6 and ref[22]). Here, the designed pockets with four out of six distances falling within 1.5 Å and two out of six distances within 2.0 Å of the corresponding distances in primary materialphore models were considered primary matches and the designed pockets with all the four distances falling within 1.5 Å of the corresponding distances in the secondary materialphore models were considered secondary matches.

	Calcium		**Strontium**	
Peptide	First validation component (%)	Second validation component (%)	First validation component (%)	Second validation component (%)
RGDGAIIGDAD	100.00	51.31	99.75	61.25
RGDGAIIGDDA	100.00	68.88	100.00	62.53
RGDGAIIGDEN	94.65	35.21	75.60	42.91
RGDGAIIGADD	100.00	59.15	99.90	52.94
RGDGAIIGDNE	97.90	36.62	50.10	38.24
RGDGAIIGDTD	100.00	61.88	86.95	41.59
RGDGAIIGTDD	99.40	30.62	99.45	9.70

10.2.5 MD Simulations Investigating the Self-assembly Properties of the Designed Peptides and Final Computational Validation against Primary and Secondary Materialphore Models

The resulting seven peptides were investigated for their self-assembly properties using the self-assembly REMD simulations, analogously to the refs[22,23,24,130,131,192]. Upon completion of the simulations, we collected 100,000 snapshots which corresponded to 1 μs for each of the seven designed peptides. We classified the observed β-sheet conformations in the simulations into 2-, 3-, 4-, 5-, and 6-stranded antiparallel, parallel, mixed, and complex β-sheets. We observed that the designed peptides RGDGAIIGDTD, RGDGAIIGDAD, RGDGAIIGTDD, and RGDGAIIGADD showed an overall tendency to form antiparallel β-sheets (Figure 10.1) whereas the designed peptides RGDGAIIGDEN, RGDGAIIGDNE, and RGDGAIIGDDA showed an overall tendency to form parallel β-sheets (Figure 10.2). The three peptides which favored parallel configuration were discarded and not subjected to further investigation as the flexible structural templates on which the computational design was performed were antiparallel and the formation of the parallel β-sheets for the three peptides would disallow the formation of intended functional designed binding pockets. For the remaining four designed peptides, we analyzed the propensity of the peptides to form intermolecular β-sheet structure for the 4-, 5-, and 6-stranded antiparallel β-sheets (Figure 10.3). We observed that residues in the GAIIG core are the primary residues that are involved in β-sheet formation. We also observed that residues 3, and 9 in the designed peptides occasionally participated in β-sheet formation. We constructed free energy landscapes based on the P_1 and P_2 parameters corresponding to the 4-stranded antiparallel β-sheet structures (Figure 10.4). From the free energy basin, we extracted the highly ordered and well-aligned 4-stranded antiparallel β-sheet structures, as these structures can potentially correspond to elementary β-sheet structural units of the naturally occurring fibrils. For the designed peptide RGDGAIIGDAD, we

did not observe the formation of highly ordered and well-aligned β-sheet conformations. Thus, the specific peptide was discarded from further investigation. For the designed peptides RGDGAIIGDTD, RGDGAIIGTDD, and RGDGAIIGADD the collected highly ordered and well-aligned β-sheet conformation were subjected to two-component computational validation as described in *Methods*. The results of the computational validation are presented in the Table 10.2

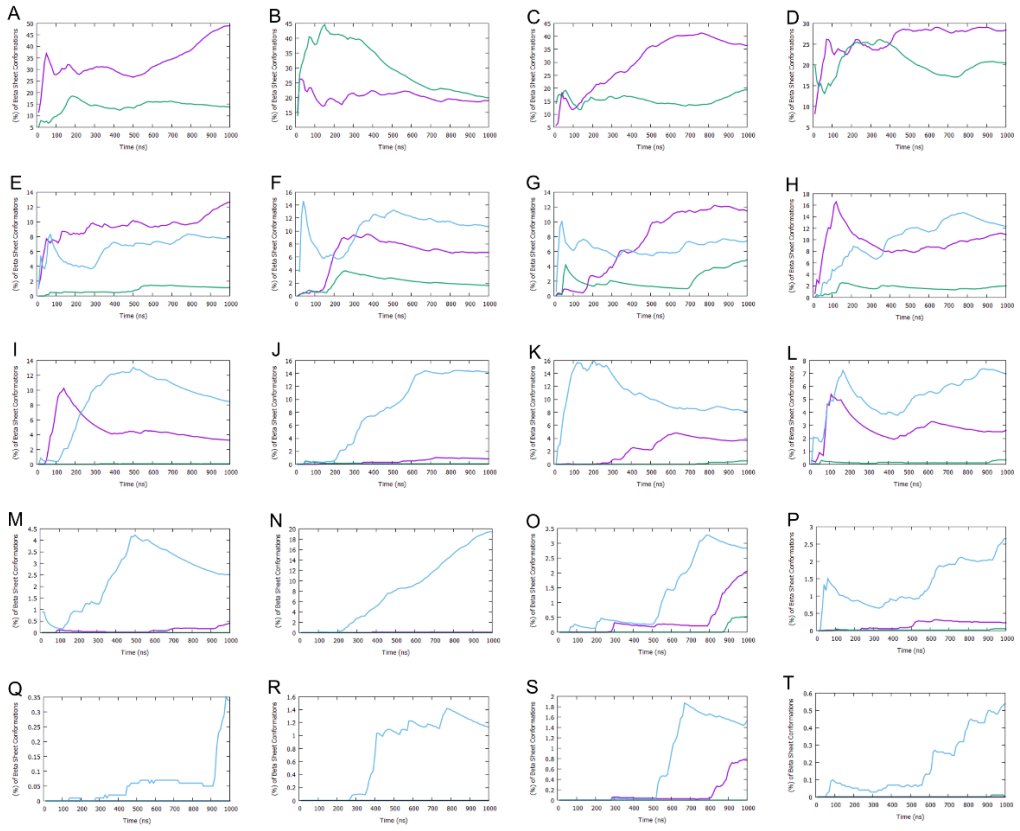


Figure 10.1 Moving averages of the fraction (%) of conformations with intermolecular β-sheets (Y-axis) in the REMD simulations at 300 K with respect to time (X-axis) for the peptides RGDGAIIGDTD, RGDGAIIGDAD, RGDGAIIGTDD, and RGDGAIIGADD. Figures A, E, I, M, and Q correspond to 2-, 3-, 4-, 5-, and 6-stranded β-sheets of peptide RGDGAIIGDTD, respectively. Figures B, F, J, N, and R correspond to 2-, 3-, 4-, 5-, and 6-stranded β-sheets of peptide RGDGAIIGDAD, respectively. Figures C, G, K, O, and S correspond to 2-, 3-, 4-, 5-, and 6-stranded β-sheets of peptide RGDGAIIGTDD, respectively. Figures D, H, L, P, and T correspond to 2-, 3-, 4-, 5-, and 6-stranded β-sheets of peptide RGDGAIIGADD, respectively. Purple color corresponds to antiparallel β-sheets, green color corresponds to parallel β-sheets, and blue color corresponds to mixed β-sheets.

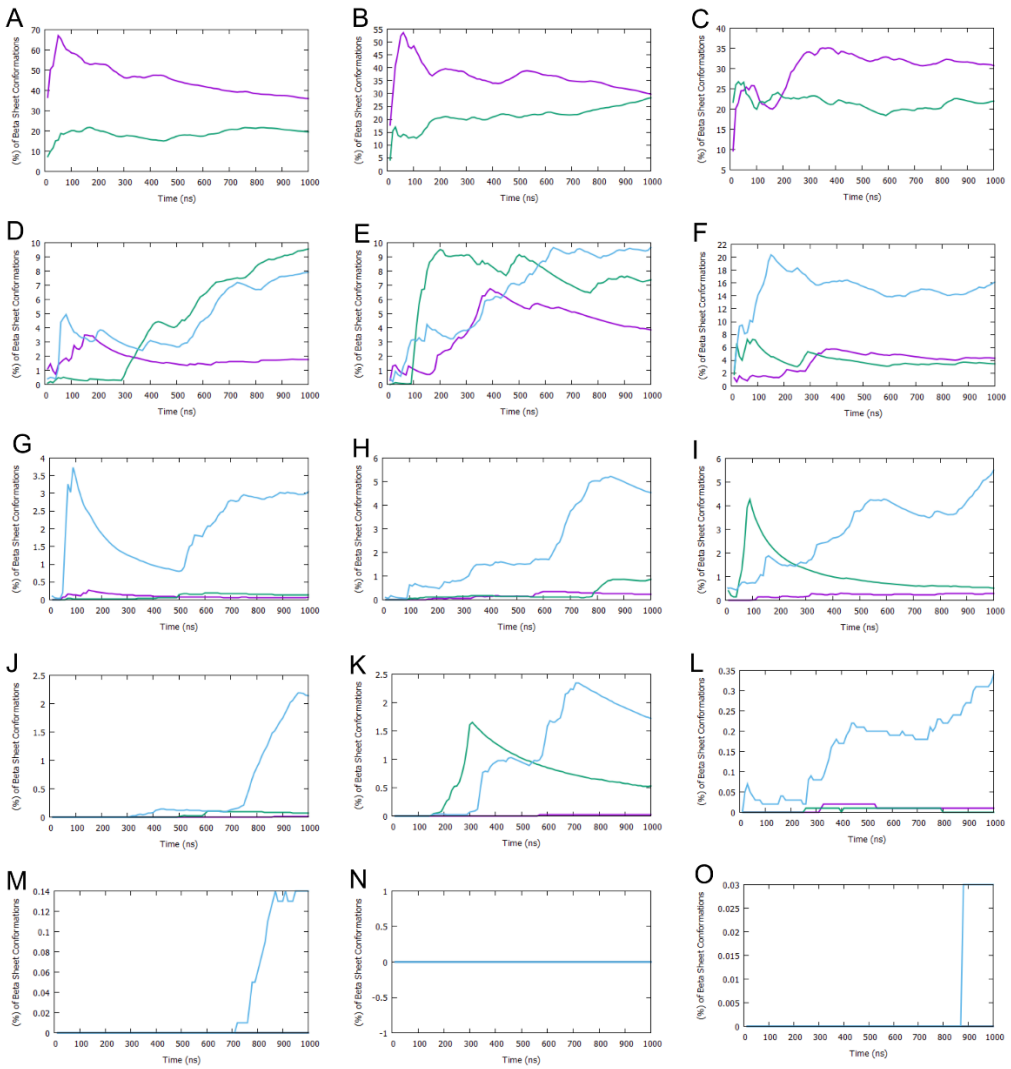


Figure 10.2 Moving averages of the fraction (%) of conformations with intermolecular β-sheets (Y-axis) in the REMD simulations at 300 K with respect to time (X-axis) for the peptides RGDGAIIGDEN, RGDGAIIGDNE, RGDGAIIGDDA. Figures A, D, G, J, and M correspond to 2-, 3-, 4-, 5-, and 6-stranded β-sheets of peptide RGDGAIIGDEN, respectively. Figures B, E, H, K, and N correspond to 2-, 3-, 4-, 5-, and 6-stranded β-sheets of peptide RGDGAIIGDNE, respectively. Figures C, F, I, L, and O correspond to 2-, 3-, 4-, 5-, and 6-stranded β-sheets of peptide RGDGAIIGDDA, respectively. Purple color corresponds to antiparallel β-sheets, green color corresponds to parallel β-sheets, and blue color corresponds to mixed β-sheets.

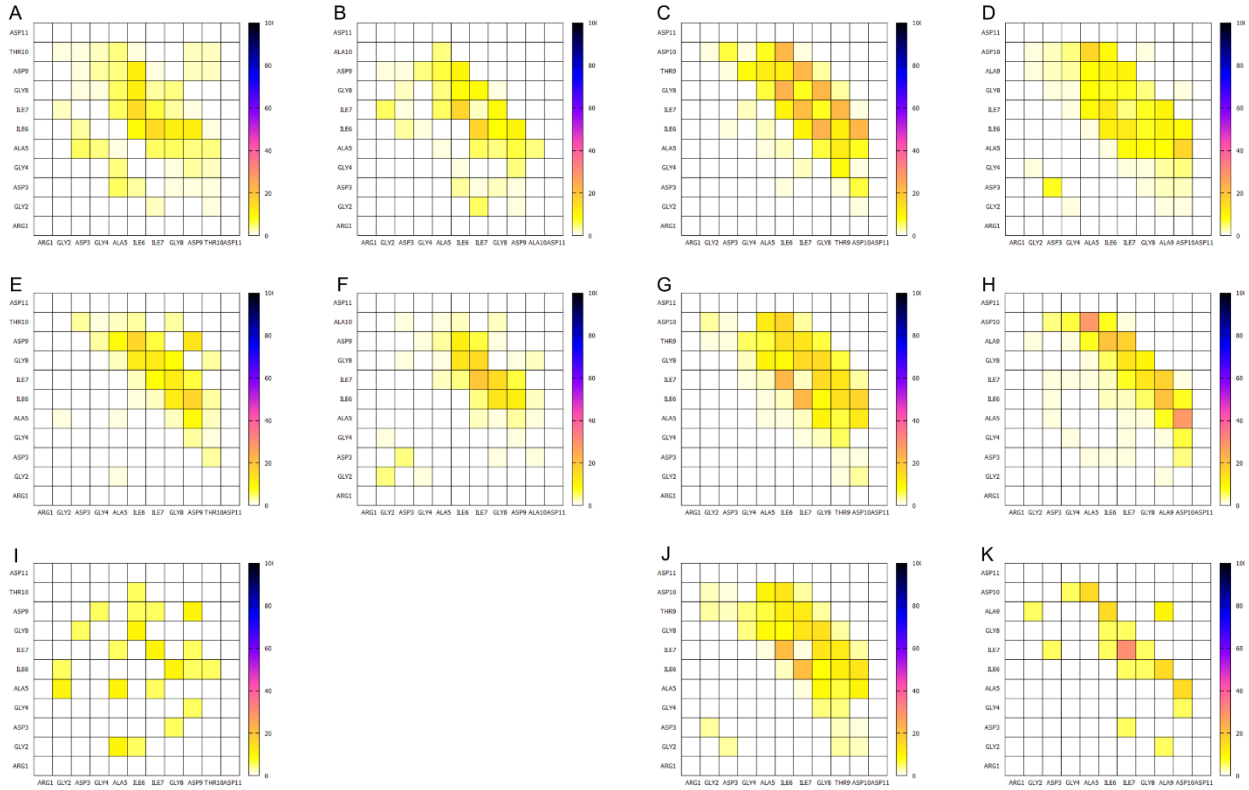


Figure 10.3 Density (%) maps of residue pairs forming intermolecular β-bridges for the peptides RGDGAIIGDTD, RGDGAIIGDAD, RGDGAIIGTDD, and RGDGAIIGADD. The pairs of residues belong to nearest neighboring peptides participating in an isolated β-bridge or extended β-sheet conformation in the REMD simulations at 300 K. Figures A, E, and I correspond to 4-, 5-, and 6-stranded antiparallel configurations of peptide RGDGAIIGDTD, respectively. Figures B, and F correspond to 4-, and 5-stranded antiparallel configurations of peptide RGDGAIIGDAD, respectively. Figures C, G, and J correspond to 4-, 5-, and 6-stranded antiparallel configurations of peptide RGDGAIIGTDD, respectively. Figures D, H, and K correspond to 4-, 5-, and 6-stranded antiparallel configurations of peptide RGDGAIIGADD, respectively.

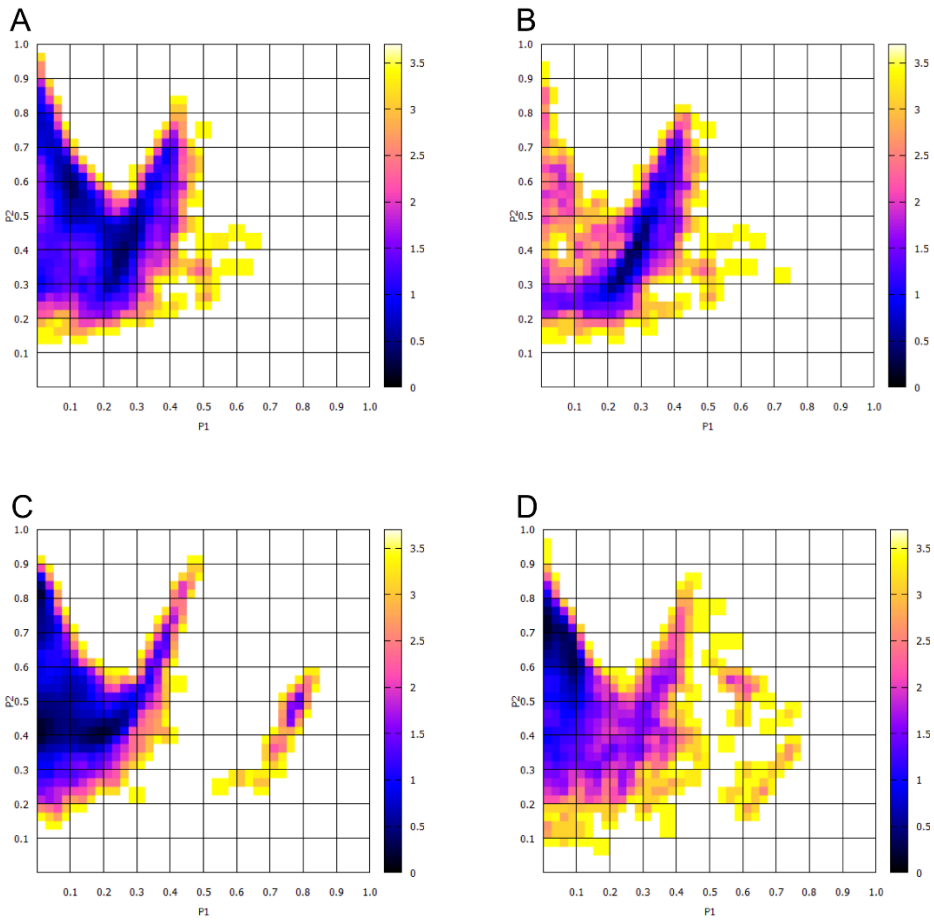


Figure 10.4 Free energy surfaces constructed from the 2-D probabilities of order parameters P_1 and P_2, calculated using the 4-stranded antiparallel β-sheets observed in the replica exchange MD simulations at 300 K for the peptides RGDGAIIGDTD, RGDGAIIGDAD, RGDGAIIGTDD, and RGDGAIIGADD. Surface in panel A, B, C, and D correspond to 4-stranded antiparallel β-sheets formed by RGDGAIIGDTD, RGDGAIIGDAD, RGDGAIIGTDD, and RGDGAIIGADD, respectively.

Table 10.2 Designed peptide sequences for calcium/strontium binding and their corresponding first and second validation component percentage statistics based on REMD simulations. The validation statistics were calculated over the highly ordered and well-aligned β-sheet conformations extracted from the self-assembly REMD simulations. The listed percentages for the first and second validation components correspond to the percentage of snapshots containing primary and secondary matches, within the snapshots containing primary matches, which were identified from the first validation component (see additional details in chapter 6 and ref[22]). Here, the designed pockets with four out of six distances falling within 1.5 Å and two out of six distances within 2.0 Å of the corresponding distances in primary materialphore models were considered primary matches and the designed pockets with all the four distances falling within 1.5 Å of the corresponding distances in the secondary materialphore models were considered secondary matches.

	Calcium		**Strontium**	
Peptide	First validation component (%)	Second validation component (%)	First validation component (%)	Second validation component (%)
RGDGAIIGTDD	97.00	26.55	86.20	10.78
RGDGAIIGDTD	80.40	57.57	38.80	35.42
RGDGAIIGADD	96.60	77.77	100.00	83.01

10.4 Concluding Remarks

In this study, we aimed to computationally design peptides with potential cell-adhesion, calcium, and strontium binding properties. The computational design was achieved by implementing the generalized computational protocol presented in chapter 7 by using the amyloid designable scaffold YATGAIIGNII. The amino acid positions 1, 2, and 3 were mutated to Arg, Gly, and Asp for cell-adhesion properties and the positions 9, 10, and 11 were independently designed for calcium and strontium based on the materialphore models calculated for the two specific ions. Subsequently, MD simulations and two-component computational validation were performed for the designed peptides to evaluate their binding ability to calcium and strontium. Out of the seven designed peptides, the self-assembly simulations performed, and subsequent two-

component computational validation performed indicated that the three peptides with sequences NH_3^+-RGDGAIIGTDD-$CONH_2$, NH_3^+-RGDGAIIGDTD-$CONH_2$, and NH_3^+-RGDGAIIGADD-$CONH_2$ could be considered for further computational and experimental investigation. According to the second component of validation (on the ability of an ion to be introduced in the pocket of the designed amino acid based on the materialphore model it originated from), the latter two peptides could presumably better incorporate both calcium and strontium binding properties.

10.5 Innovation in Methods

In chapter 8, we showed that the computational protocol presented in chapter 7 can be solved for multiple problems resulting in the design of multifunctional materials in terms of binding properties. In chapter 9, we used sequence-based bioinformatics data to enable the use of the computational protocol in tandem with rational approaches to suggest the design of two peptides with potential cell-penetration and DNA binding properties. Here, we combined sequence-based constraints in conjunction with the solution of two problems, one for each ion under investigation. We showed the capacity of the computational protocol to be combinedly used in designing functional amyloid materials combining multi-functional properties, including cell-attachment introduced through a sequenced-based constraint (placement of RGD at first three positions), in tandem with binding to calcium and strontium ions, introduced through materialphore model-based constraints at positions 3, 9, 10 and 11, with position 3 constituting a "fixed" mutation as part of both aimed functionalities.

10.6 New Findings and Potential Applications

In this study, we aimed to design novel potential tissue engineering scaffolds specifically for bone tissue engineering applications. From the previous studies, we have observed that bone

tissue engineering scaffolds have cell-adhesion properties and release metallic ions such as calcium, and strontium. The specific metallic ions are play a huge role in bone growth on the scaffolds. We aimed to use amyloid materials as scaffolds, as they could combine potential advanced properties such as potential biocompatibility, excellent mechanical properties, and the ability to introduce changes at their sequence level to yield specific functionality. We suggested the design of two peptides using our computational protocol, potentially incorporating cell-adhesion, calcium binding, and strontium binding properties. These peptides have not been experimentally validated yet.

11. SUMMARY AND FUTURE SCOPE

11.1 Summary

Protein self-assembly is defined as the self-organization of the individual protein molecules, usually in a less ordered state, into high ordered three-dimensional protein architectures. The aggregation of protein into amyloid fibrils is a key example of protein self-assembly, where individual proteins self-assemble into β-sheets perpendicular to the fibril axis. Even though amyloid self-assembly was discovered in the context of diseases, naturally occurring sequences from amyloid peptides were used as elementary blocks for the formation of amyloid materials. Functional amyloid materials have been rationally designed to bind ions, compounds, and surfaces (cell attachment) which resulted in several applications including the fields of tissue engineering, separations and drug delivery.

Computational tools are increasing becoming powerful tools in predicting the structure of amyloid which enables the design of functional amyloid materials. Here, we developed new computational tools, and used them in tandem with existing computational tools and in combination with rational approaches to design novel functional materials. We have successfully designed a bifunctional self-assembling peptide RGDSGAITIGC with cell-adhesion and metal-binding properties. [24] The specific peptide could potentially be exploited to fabricate novel biomaterials with promising tissue engineering applications.[24] Subsequently, we developed an in-house computational design strategy to energetically stabilize and potentially functionalize an amyloid peptide fibril.[192] We stabilized the fibrils formed by the peptide RGDSGAITIGC by introducing mutations at position 11 which resulted in the designed peptide RGDAGAITIGY.[192] In addition to stabilizing the fibril, the introduction of tyrosine at position 11 was computationally predicted to possess cross-linking properties.[192]

Subsequently, we aimed to computationally design amyloid materials on demand with the capacity to bind ions, compounds or possess cell-attachment properties. To achieve this, we identified self-assembling peptides with a β-sheet core and additional residue at the termini.[23] The elementary β-sheet structural units formed by those peptides were referred by us as amyloid designable scaffolds.[23] For example, we have elucidated that peptide sequences like YATGAIIGNII,[23] and AGKGAIIGFIK[22] can act as amyloid designable scaffolds as GAIIG can act as amyloidogenic core, and additional (e.g. 2-3) residues at both termini are outside the amyloid β-sheet, as glycine residues can act as β-turn promoters halting β-sheet elongation.

Subsequently, we developed a computational protocol for the design of amyloid materials binding to an ion of interest.[22] The protocol introduces mutations at the terminal non-β-sheet forming residues of an amyloid designable scaffold through an optimization-based computational model which mimics how amino acids bind to ions in materialphore models.[22] Materialphore models represent how amino acids bind to ions according to the experimentally resolved structures containing the specific ion. The computational model simultaneously minimizes the energy due to the mutations aimed at stabilizing the bound conformation of the designed amino acids, as well as to avoid destabilizing the self-assembly properties of the designed peptides.[22] The optimum designs were then computationally validated using a series of simulations and structural analysis to select the top designed peptides, which are predicted to form fibrils with ion binding properties.[22] The computational developed was first applied for the design of amyloid materials with cesium ion deposition and capture properties.[22] The designed amyloid materials could potentially be useful in the development of technologies to capture radioactive cesium ions from acidic wastewater conditions or as biocompatible materials removing cesium from blood.[22]

The developed protocol was subsequently formulated toward a generalized protocol to design amyloid materials that can bind to an ion or compound of interest. Additionally, the computational protocol was also generalized in such in a way that it can be used in combination with rational design approaches or intuition, such as introducing the motif RGD to combine cell adhesion properties. The resulting computational protocol was implemented to design and suggest functional amyloid materials binding to AD drugs, including donepezil and tacrine. Such potentially successful and experimentally validated amyloid materials designed could act as novel drug delivery carriers which can enhance the half-life and effectivity of current AD drugs and potentially the stabilization of cognition (enhancement of current AD treatments). The resulting computational protocol was also used in combination with data-driven approach (enabled by sequence-based bioinformatics data) to design amyloid materials with cell-penetration and DNA binding properties. Such potentially successful and experimentally validated could act as efficient nucleic acid delivery agents to the cell nuclei by effectively penetrating the cell membrane. Lastly, the resulting computational protocol was used in tandem with rational approaches to design functional amyloid materials combining cell-adhesion, calcium binding, and strontium binding properties. Such potentially successful and experimentally validated amyloid materials could act as bone tissue engineering scaffolds by releasing the specific metallic ions which play an important role in bone growth on the scaffolds.

11.2 Future Scope

This study laid the foundations for the computational design of functional amyloid materials, specifically for applications of binding to an ion/compound of interest. In what follows a list of potential improvements currently under consideration in the lab is provided:

1. Amyloid designable scaffolds: In this dissertation, we defined and elucidated the structure of amyloid designable scaffolds, representing short amyloid peptides with the capacity to self-assemble into amyloid-like self-assembled structures with exposed residues at the termini not participating in β-sheets. The elementary structural β-sheet unit of such scaffolds are used as input the computational functionalization-based design according to which functional amino acids are mutated on the scaffolds' non-β-sheet forming residue positions. One significant limitation is associated with the limited number of amyloid designable scaffolds (and their corresponding elementary β-sheet structures) available, which are currently limited to the amyloid designable scaffolds formed by the peptides RGDSGAITIGC[24], YATGAIIGNII[23], and AGKGAIIGFIK[22]. Thus, the discovery of a series of amyloid designable scaffolds with diverse geometries particularly on the non-β-sheet forming positions is a significant future direction which can significantly enhance the functionalization of materials and increase success rates in design. Such amyloid designable scaffolds can be inspired by similar approaches reported in our published studies.[22,23]
2. Materialphore models: In this dissertation, we defined materialphore models to represent amino acid motifs and their corresponding geometries, extracted from experimentally resolved protein structures, denoting how amino acids bind to a specific ion/compound of interest. Such models were used as constraints in the computational functionalization-based design, enabling the production of combinations of designed amino acids on the amyloid materials with the potential to bind an ion/compound of interest. Two related significant limitations of the currently produced materialphore models is that (i) all materialphore models extracted are considered equally important in the functionalization

based design without any prioritization on the most important and (potentially naturally) conserved amino acid motifs which are key for the recognition of an ion/compounds, and (ii) all amino acids within a materialphore model are considered equally important in the functionalization-based design without any prioritization of the most important amino acids within a model. Thus, the use of support vector machines is a direction currently considered in the lab to resolve this, and select the most important amino acids to be substituted during design. Such approaches could be used in the future to significantly expand the materialphore models which can be considered, as they would reduce the space by prioritizing the solutions; this could be helpful when the number of available materialphore models that can be extracted for specific ions/compounds is significantly large (e.g., calcium and strontium ions presented in chapter 10, where only a limited subset of all possible materialphore models was considered).

3. The computational functionalization-based design is handled through the substitution of amino acids, by considering their energy in a coarse-grained representation and their ability to form the right pocket in a sufficiently small number of pockets (considered to be important for functionalization). Future considerations currently in Dr. Tamamis' lab aim to improve the accuracy of energy representation by including an all-atom representation of the designed amino acids and the ion/compound, by taking into account additionally the actual interaction energy between the ion/compound with the designed amino acids. These will alleviate any issues with the limitations of the current considerations outlined above.
4. The validation of the designed peptides is currently achieved in two stages. While the first stage is not computationally demanding, it may not provide an accurate representation of the self-assembled structures, and therefore, the second stage, at which the self-assembly

properties of the designed peptides are investigated computationally is needed. While the second stage is more accurate, its computational cost is prohibitive for including it in a potential server where users in the future could potentially use the capabilities of the computational design to validate the designed peptides. Future considerations currently in Dr. Tamamis' lab aim at alleviating this by using other types of methods that do not necessarily need the use of computationally costly REMD simulations. In addition, the use of sequence-based prediction algorithms, including the ones outlined in chapter 2, could be used to improve the amyloid-predicting ability of the newly designed peptides, as an extra validation stage.

REFERENCES

[1] Whitesides, G. M.; Boncheva, M. Beyond molecules: self-assembly of mesoscopic and macroscopic components. *Proc. Natl. Acad. Sci. U.S.A.* **2002,** *99* (8), 4769-74.

[2] Desiraju, G. R. Crystal Engineering: The Design of Organic Solids. *Elsevier, New York*. **1998**.

[3] Evans, D. F.; Wennerstrom, H. The Colloidal Domain: Where Physics, Chemistry, Biology, and Technology Meet. *Wiley, New York*. **1999**.

[4] Thomas, E. L. The ABCs of Self-Assembly. *Science*. **1999,** *286* (5443), 1307.

[5] Gazit, E.; Mitraki, A., *Plenty of Room for Biology at the Bottom*. IMPERIAL COLLEGE PRESS: 2012; p 216.

[6] Khalid, S.; Bond, P. J.; Holyoake, J.; Hawtin, R. W.; Sansom, M. S. DNA and lipid bilayers: self-assembly and insertion. *J. R. Soc. Interface.* **2008,** *5 (3)*, S241-50.

[7] van Meer, G.; Voelker, D. R.; Feigenson, G. W., Membrane lipids: where they are and how they behave. *Nat. Rev. Mol. Cell. Biol.* **2008,** *9* (2), 112-24.

[8] Sleytr, U. B.; Schuster, B.; Egelseer, E. M.; Pum, D. S-layers: principles and applications. *FEMS Microbiol. Rev.* **2014,** *38* (5), 823-64.

[9] Sleytr, U. B.; Glauert, A. M. Ultrastructure of the cell walls of two closely related clostridia that possess different regular arrays of surface subunits. *J. Bacteriol.* **1976,** *126* (2), 869-82.

[10] Loo, Y.; Goktas, M.; Tekinay, A.B.; Guler, M.O.; Hauser, C.A.; Mitraki, A. Self-Assembled Proteins and Peptides as Scaffolds for Tissue Regeneration. *Adv. Healthcare Mater.* **2015**, *4*, 2557-2586.

[11] Sutherland, T. D.; Young, J. H.; Weisman, S.; Hayashi, C. Y.; Merritt, D. J. Insect silk: one name, many materials. *Annu. Rev. Entomol.* **2010,** *55*, 171-88.

[12] Kubik, S., High-performance fibers from spider silk. *Angew. Chem. Int. Ed. Engl.* **2002,** *41* (15), 2721-3.

[13] Rambaran, R.N.; Serpell, L.C. Amyloid fibrils Abnormal Protein Assembly. *Prion.* **2008**, *2(3)*, 112–117.

[14] Iadanza, M. G.; Jackson, M. P.; Hewitt, E. W.; Ranson, N. A.; Radford, S. E. A new era for understanding amyloid structures and disease. *Nat. Rev. Mol. Cell. Biol.* **2018,** *19* (12), 755-773.

[15] Sawaya, M.R.; Sambashivan, S.; Nelson, R.; Ivanova, M.I.; Sievers, S.A.; Apostol M.I.; Thompson, M.J.; Balbirnie, M.; Wiltzius, J.J.; McFarlane, H.T.; Madsen A.Ø.; Riekel, C.; Eisenberg, D. Atomic structures of amyloid cross-beta spines reveal varied steric zippers. *Nature*.

2007, *447(7143)*, 453-457.

[16] Knowles, T. P.; Vendruscolo, M.; Dobson, C. M. The amyloid state and its association with protein misfolding diseases. *Nat. Rev. Mol. Cell. Biol.* **2014,** *15* (6), 384-96.

[17] Smith, J. F.; Knowles, T. P.; Dobson, C. M.; Macphee, C. E.; Welland, M. E. Characterization of the nanoscale properties of individual amyloid fibrils. *Proc. Natl. Acad. Sci. U.S.A.* **2006,** *103* (43), 15806-11.

[18] Jackson, M. P.; Hewitt, E. W. Why are Functional Amyloids Non-Toxic in Humans? *Biomolecules* **2017,** *7* (4), 71.

[19] Tao, K.; Wang, J.; Zhou, P.; Wang, C.; Xu, H.; Zhao, X.; Lu, J.R. Self-assembly of short aβ(16-22) peptides: effect of terminal capping and the role of electrostatic interaction. *Langmuir.* **2011**, *27(6)*, 2723-2730.

[20] Tenidis, K.; Waldner, M.; Bernhagen, J.; Fischle, W.; Bergmann, M.; Weber, M.; Merkle, M.L.; Voelter, W.; Brunner, H.; Kapurniotu, A. Identification of a penta- and hexapeptide of islet amyloid polypeptide (IAPP) with amyloidogenic and cytotoxic properties. *J. Mol. Biol.* **2000**, *295*, 1055–1071.

[21] Görbitz C.H. Nanotube formation by hydrophobic dipeptides. *Chemistry.* **2001**, *7(23)*, 5153-5159.

[22] Jonnalagadda, S.V.R.; Kokotidou, C.; Orr, A.A.; Fotopoulou, E.; Henderson, K.J.; Choi, C.H.; Lim, W.T.; Choi, S.J.; Jeong, H.K.; Mitraki, A.; Tamamis, P. Computational Design of Functional Amyloid Materials with Cesium Binding, Deposition, and Capture Properties. *J. Phys. Chem. B* **2018**, *122(30)*, 7555-7568.

[23] Kokotidou, C.; Jonnalagadda, S. V. R.; Orr, A. A.; Seoane-Blanco, M.; Apostolidou, C. P.; van Raaij, M. J.; Kotzabasaki, M.; Chatzoudis, A.; Jakubowski, J. M.; Mossou, E.; Forsyth, V.T.; Mitchell, E.P.; Bowler, M.W.; Llamas-Saiz, A.L.; Tamamis, P.; Mitraki, A. A novel amyloid designable scaffold and potential inhibitor inspired by GAIIG of amyloid beta and the HIV-1 V3 loop. *FEBS Lett*. **2018**, *592*, 1777-1788.

[24] Deidda, G.; Jonnalagadda, S.V.R.; Spies, J.W.; Ranella, A.; Mossou, E.; Forsyth V.T.; Mitchell, E.P.; Bowler, M.W.; Tamamis, P.; Mitraki, A. Self-Assembled Amyloid Peptides with Arg-Gly-Asp (RGD) Motifs as Scaffolds for Tissue Engineering. *ACS Biomater. Sci. Eng.* **2017**, *3*, 1404-1416.

[25] Terzaki, K.; Kalloudi, E.; Mossou, E.; Mitchell, E.P.; Forsyth V.T.; Rosseeva, E.; Simon, P.; Vamvakaki, M.; Chatzinkolaidou, M.; Mitraki, A.; Farsari, M. Mineralized Self-Assembled Peptides on 3D Laser-Made Scaffolds: A New Route toward 'Scaffold on Scaffold' Hard Tissue Engineering. *Biofabrication* **2013***, 5*, 045002.

[26] Kasotakis, E.; Mossou, E.; Adler-Abramovich, L.; Mitchell, E.P.; Forsyth, V.T.; Gazit, E.;

Mitraki, A. Design of Metal-Binding Sites onto Self-Assembled Peptide Fibrils. *Biopolymers* **2009**, *92*, 164−172.

[27] Kasotakis, E.; Mitraki, A. Designed Self-Assembling Peptides as Templates for the Synthesis of Metal Nanoparticles. *Methods Mol. Biol.* **2013***, 996*, 195-202.

[28] Kasotakis, E.; Mitraki, A. Silica Biotemplating by Self-Assembling Peptides via Serine Residues Activated by the Peptide Amino Terminal Group. *Biopolymers* **2012**, *98*, 501−509.

[29] Kasotakis, E.; Kostopoulou, A.; Spuch-Calvar, M.; Androulidaki, M.; Pelekanos, N.; Kanaras, A.G.; Lappas, A.; Mitraki, A. Assembly of Quantum Dots on Peptide Nanostructures and Their Spectroscopic Properties. *Applied Physics* **2014**, *116*, 977-985.

[30] Adler-Abramovich, L.; Gazit, E. Controlled Patterning of Peptide Nanotubes and Nanospheres Using Inkjet Printing Technology. *J. Pept. Sci.* **2008**, *14*, 217–223.

[31] Reches, M.; Gazit, M. Casting Metal Nanowires within Discrete Self-Assembled Peptide Nanotubes. *Science* **2003**, *300*, 625–627.

[32] Li, D.; Furukawa, H.; Deng, H.; Liu, C.; Yaghi, O.M.; Eisenburg, D.S. Designed Amyloid Fibers as Materials for Selective Carbon Dioxide Capture. *Proc. Natl. Acad. Sci. USA* **2014**, *111*, 191-196.

[33] Li, D. Jones, E.M.; Sawaya, M.R.; Furukawa, H.; Luo, F.; Ivanova, M.; Sievers, S.A.; Wang, W.; Yaghi, O.M.; Liu, C.; Eisenberg, D.S. Structure-Based Design of Functional Amyloid Materials. *J. Am. Chem. Soc.* **2014***, 136*, 18044–18051.

[34] Hauser, C.A.; Deng, R.; Mishra, A.; Loo, Y.; Khoe, U.; Zhuang, F.; Cheong, D.W.; Accardo, A.; Sullivan, M.B.; Riekel, C.; Yang, J.Y.; Hauser, U.A.; Natural Tri- to Hexapeptides Self-Assemble in Water to Amyloid Beta-Type Fiber Aggregates by Unexpected Alpha-Helical Intermediate Structures. *Proc. Natl. Acad. Sci. USA* **2011**, *108*, 1361−1366.

[35] Pellarin, R.; Schuetz, P.; Guarnera, E.; Caflisch, A. Amyloid Fibril Polymorphism is Under Kinetic Control. *J Am. Chem. Soc.* **2010**, *132*, 14960-14970.

[36] Knowles, T.P.; Mezzenga, R. Amyloid Fibrils as Building Blocks for Natural and Artificial Functional Materials. *Adv. Mater.* **2016***, 28*, 6546-6561.

[37] Reches, M.; Gazit, E. Formation of Closed-Cage Nanostructures by Self-Assembly of Aromatic Dipeptides. *Nano Lett.* **2004**, *4*, 581-585.

[38] Kasai, S.; Ohga, Y.; Mochizuki, M.; Nishi, N.; Kadoya, Y.; Nomizu, M. Multifunctional peptide fibrils for biomedical materials. *Biopolymers.* **2004**, *76(1)*, 27-33.

[39] Li, C.; Adamcik, J.; Mezzenga, R. Biodegradable nanocomposites of amyloid fibrils and graphene with shape-memory and enzyme-sensing properties. *Nat. Nanotechnol.* **2012**, *7(7)*, 421-

427.

[40] Tamamis, P.; Terzaki, K.; Kassinopoulos, M.; Mastrogiannis, L,; Mossou, E.; Forsyth, V.T.; Mitchell, E.P.; Mitraki, A.; Archontis, G. Self-Assembly of an Aspartate-Rich Sequence from the Adenovirus Fiber Shaft: Insights from Molecular Dynamics Simulations and Experiments. *J. Phys. Chem. B* **2014**, *118*, 1765-1774.

[41] Bolisetty, S.; Reinhold, N.; Zeder, C.; Orozco, M.N.; Mezzenga, R. Efficient Purification of Arsenic-Contaminated Water Using Amyloid-Carbon Hybrid Membranes. *Chem. Commun. (Camb).* **2017**, *253*, 5714-5717.

[42] Bolisetty, S.; Mezzenga, R. Amyloid-Carbon Hybrid Membranes for Universal Water Purification. *Nat. Nanotechnol.* **2016**, *11*, 365-371.

[43] Silva, R.F.; Araújo, D.R.; Silva, E.R.; Ando, R.A.; Alves, W.A. L-diphenylalanine microtubes as a potential drug-delivery system: characterization, release kinetics, and cytotoxicity. *Langmuir.* **2013**, *29(32)*, 10205-10212.

[44] Marchesan, S.; Qu, Y.; Waddington, L.J.; Easton, C.D.; Glattauer, V.; Lithgow, T.J.; McLean, K.M.; Forsythe J.S.; Hartley, P.G. Self-assembly of ciprofloxacin and a tripeptide into an antimicrobial nanostructured hydrogel. *Biomaterials.* **2013**, *34(14)*, 3678-3687.

[45] Shen, Y.; Posavec, L.; Bolisetty, S.; Hilty, F. M.; Nystrom, G.; Kohlbrecher, J.; Hilbe, M.; Rossi, A.; Baumgartner, J.; Zimmermann, M. B.; Mezzenga, R. Amyloid fibril systems reduce, stabilize and deliver bioavailable nanosized iron. *Nat. Nanotechnol.* **2017,** *12* (7), 642-647.

[46] Westermark, P. Aspects on Human Amyloid Forms and their Fibril Polypeptides. *FEBS J.* **2005**, *272(23)*, 5942-5949.

[47] Biancalana, M.; Koide, S. Molecular mechanism of Thioflavin-T binding to amyloid fibrils. *Biochim. Biophys. Acta.* **2010,** *1804* (7), 1405-12.

[48] Xue, C.; Lin, T. Y.; Chang, D.; Guo, Z., Thioflavin T as an amyloid dye: fibril quantification, optimal concentration and effect on aggregation. *R Soc Open Sci* **2017,** *4* (1), 160696.

[49] Vadukul, D. M.; Al-Hilaly, Y. K.; Serpell, L. C. Methods for Structural Analysis of Amyloid Fibrils in Misfolding Diseases. *Methods. Mol. Biol.* **2019,** *1873*, 109-122.

[50] Micsonai, A.; Wien, F.; Kernya, L.; Lee, Y. H.; Goto, Y.; Refregiers, M.; Kardos, J. Accurate secondary structure prediction and fold recognition for circular dichroism spectroscopy. *Proc. Natl. Acad. Sci. U.S.A.* **2015,** *112* (24), E3095-103.

[51] Papanikolopoulou, K.; Schoehn, G.; Forge, V.; Forsyth, V. T.; Riekel, C.; Hernandez, J. F.; Ruigrok, R. W.; Mitraki, A. Amyloid fibril formation from sequences of a natural beta-structured fibrous protein, the adenovirus fiber. *J. Biol. Chem.* **2005,** *280* (4), 2481-90.

[52] Riek, R.; Eisenberg, D. S. The activities of amyloids from a structural perspective. *Nature.* **2016,** *539* (7628), 227-235.

[53] Rodriguez, J. A.; Ivanova, M. I.; Sawaya, M. R.; Cascio, D.; Reyes, F. E.; Shi, D.; Sangwan, S.; Guenther, E. L.; Johnson, L. M.; Zhang, M.; Jiang, L.; Arbing, M. A.; Nannenga, B. L.; Hattne, J.; Whitelegge, J.; Brewster, A. S.; Messerschmidt, M.; Boutet, S.; Sauter, N. K.; Gonen, T.; Eisenberg, D. S. Structure of the toxic core of alpha-synuclein from invisible crystals. *Nature.* **2015,** *525* (7570), 486-90.

[54] Soriaga, A. B.; Sangwan, S.; Macdonald, R.; Sawaya, M. R.; Eisenberg, D. Crystal Structures of IAPP Amyloidogenic Segments Reveal a Novel Packing Motif of Out-of-Register Beta Sheets. *J. Phys. Chem. B* **2016,** *120* (26), 5810-6.

[55] Tycko, R. Physical and structural basis for polymorphism in amyloid fibrils. *Protein. Sci.* **2014,** *23* (11), 1528-39.

[56] Li, Y.; Zhao, C.; Luo, F.; Liu, Z.; Gui, X.; Luo, Z.; Zhang, X.; Li, D.; Liu, C.; Li, X. Amyloid fibril structure of alpha-synuclein determined by cryo-electron microscopy. *Cell. Res.* **2018,** *28* (9), 897-903.

[57] Gremer, L.; Scholzel, D.; Schenk, C.; Reinartz, E.; Labahn, J.; Ravelli, R. B. G.; Tusche, M.; Lopez-Iglesias, C.; Hoyer, W.; Heise, H.; Willbold, D.; Schroder, G. F. Fibril structure of amyloid-beta(1-42) by cryo-electron microscopy. *Science* **2017,** *358* (6359), 116-119.

[58] Fitzpatrick, A. W. P.; Falcon, B.; He, S.; Murzin, A. G.; Murshudov, G.; Garringer, H. J.; Crowther, R. A.; Ghetti, B.; Goedert, M.; Scheres, S. H. W. Cryo-EM structures of tau filaments from Alzheimer's disease. *Nature* **2017,** *547* (7662), 185-190.

[59] Guerrero-Ferreira, R.; Taylor, N. M.; Mona, D.; Ringler, P.; Lauer, M. E.; Riek, R.; Britschgi, M.; Stahlberg, H. Cryo-EM structure of alpha-synuclein fibrils. *Elife* **2018,** *7*.

[60] Rousseau, F.; Schymkowitz, J.; Serrano, L., Protein aggregation and amyloidosis: confusion of the kinds? *Curr. Opin. Struct. Biol.* **2006,** *16* (1), 118-26.

[61] Fernandez-Escamilla, A. M.; Rousseau, F.; Schymkowitz, J.; Serrano, L. Prediction of sequence-dependent and mutational effects on the aggregation of peptides and proteins. *Nat. Biotechnol.* **2004,** *22* (10), 1302-6.

[62] Linding, R.; Schymkowitz, J.; Rousseau, F.; Diella, F.; Serrano, L. A comparative study of the relationship between protein structure and beta-aggregation in globular and intrinsically disordered proteins. *J. Mol. Biol.* **2004,** *342* (1), 345-53.

[63] Conchillo-Sole, O.; de Groot, N. S.; Aviles, F. X.; Vendrell, J.; Daura, X.; Ventura, S. AGGRESCAN: a server for the prediction and evaluation of "hot spots" of aggregation in polypeptides. *BMC Bioinformatics* **2007,** *8*, 65.

[64] Maurer-Stroh, S.; Debulpaep, M.; Kuemmerer, N.; Lopez de la Paz, M.; Martins, I. C.; Reumers, J.; Morris, K. L.; Copland, A.; Serpell, L.; Serrano, L.; Schymkowitz, J. W.; Rousseau, F. Exploring the sequence determinants of amyloid structure using position-specific scoring matrices. *Nat. Methods* **2010,** *7* (3), 237-42.

[65] Lopez de la Paz, M.; Serrano, L. Sequence determinants of amyloid fibril formation. *Proc. Natl. Acad. Sci. U.S.A.* **2004,** *101* (1), 87-92.

[66] Galzitskaya, O. V.; Garbuzynskiy, S. O.; Lobanov, M. Y. Prediction of amyloidogenic and disordered regions in protein chains. *PLoS Comput. Biol.* **2006,** *2* (12), e177.

[67] Garbuzynskiy, S. O.; Lobanov, M. Y.; Galzitskaya, O. V. FoldAmyloid: a method of prediction of amyloidogenic regions from protein sequence. *Bioinformatics* **2010,** *26* (3), 326-32.

[68] Zhang, Z.; Chen, H.; Lai, L. Identification of amyloid fibril-forming segments based on structure and residue-based statistical potential. *Bioinformatics* **2007,** *23* (17), 2218-25.

[69] Zibaee, S.; Makin, O. S.; Goedert, M.; Serpell, L. C. A simple algorithm locates beta-strands in the amyloid fibril core of alpha-synuclein, Abeta, and tau using the amino acid sequence alone. *Protein Sci.* **2007,** *16* (5), 906-18.

[70] Tian, J.; Wu, N.; Guo, J.; Fan, Y. Prediction of amyloid fibril-forming segments based on a support vector machine. *BMC Bioinformatics* **2009,** *10 Suppl 1*, S45.

[71] Hamodrakas, S. J.; Liappa, C.; Iconomidou, V. A. Consensus prediction of amyloidogenic determinants in amyloid fibril-forming proteins. *Int. J. Biol. Macromol.* **2007,** *41* (3), 295-300.

[72] O'Donnell, C. W.; Waldispuhl, J.; Lis, M.; Halfmann, R.; Devadas, S.; Lindquist, S.; Berger, B. A method for probing the mutational landscape of amyloid structure. *Bioinformatics* **2011,** *27* (13), i34-42.

[73] Kim, C.; Choi, J.; Lee, S. J.; Welsh, W. J.; Yoon, S. NetCSSP: web application for predicting chameleon sequences and amyloid fibril formation. *Nucleic Acids Res.* **2009,** *37* (Web Server issue), W469-73.

[74] Yoon, S.; Welsh, W. J.; Jung, H.; Yoo, Y. D. CSSP2: an improved method for predicting contact-dependent secondary structure propensity. *Comput. Biol. Chem.* **2007,** *31* (5-6), 373-7.

[75] Yoon, S.; Welsh, W. J. Rapid assessment of contact-dependent secondary structure propensity: relevance to amyloidogenic sequences. *Proteins* **2005,** *60* (1), 110-7.

[76] Yoon, S.; Welsh, W. J. Detecting hidden sequence propensity for amyloid fibril formation. *Protein Sci.* **2004,** *13* (8), 2149-60.

[77] Tsolis, A. C.; Papandreou, N. C.; Iconomidou, V. A.; Hamodrakas, S. J. A consensus method for

the prediction of 'aggregation-prone' peptides in globular proteins. *PLoS One* **2013,** *8* (1), e54175.

[78] Trovato, A.; Seno, F.; Tosatto, S. C. The PASTA server for protein aggregation prediction. *Protein Eng. Des. Sel.* **2007,** *20* (10), 521-3.

[79] Bryan, A. W., Jr.; Menke, M.; Cowen, L. J.; Lindquist, S. L.; Berger, B. BETASCAN: probable beta-amyloids identified by pairwise probabilistic analysis. *PLoS Comput. Biol.* **2009,** *5* (3), e1000333.

[80] Morriss-Andrews, A.; Shea, J. E. Computational studies of protein aggregation: methods and applications. *Annu. Rev. Phys. Chem.* **2015,** *66*, 643-66.

[81] Carballo-Pacheco, M.; Strodel, B. Advances in the Simulation of Protein Aggregation at the Atomistic Scale. *J. Phys. Chem. B* **2016,** *120* (12), 2991-9.

[82] Simunovic, M.; Mim, C.; Marlovits, T. C.; Resch, G.; Unger, V. M.; Voth, G. A. Protein-mediated transformation of lipid vesicles into tubular networks. *Biophys. J.* **2013,** *105* (3), 711-9.

[83] Simunovic, M.; Srivastava, A.; Voth, G. A. Linear aggregation of proteins on the membrane as a prelude to membrane remodeling. *Proc. Natl. Acad. Sci. U. S. A.* **2013,** *110* (51), 20396-401.

[84] Zhang, J.; Muthukumar, M. Simulations of nucleation and elongation of amyloid fibrils. *J. Chem. Phys.* **2009,** *130* (3), 035102.

[85] Auer, S.; Meersman, F.; Dobson, C. M.; Vendruscolo, M. A generic mechanism of emergence of amyloid protofilaments from disordered oligomeric aggregates. *PLoS Comput. Biol.* **2008,** *4* (11), e1000222.

[86] Friedman, R.; Caflisch, A. Surfactant effects on amyloid aggregation kinetics. *J. Mol. Biol.* **2011,** *414* (2), 303-12.

[87] Magno, A.; Pellarin, R.; Caflisch, A., Mechanisms and kinetics of amyloid aggregation investigated by a phenomenological coarse-grained model. *Computational Modeling of Biological Systems: From Molecules to Pathways*, ed. NV Dokholyan, **2012**,191–214. New York: Wiley

[88] Pellarin, R.; Caflisch, A. Interpreting the aggregation kinetics of amyloid peptides. *J. Mol. Biol.* **2006,** *360* (4), 882-92.

[89] Pellarin, R.; Guarnera, E.; Caflisch, A. Pathways and intermediates of amyloid fibril formation. *J. Mol. Biol.* **2007,** *374* (4), 917-24.

[90] Pellarin, R.; Schuetz, P.; Guarnera, E.; Caflisch, A. Amyloid fibril polymorphism is under kinetic control. *J. Am. Chem. Soc.* **2010,** *132* (42), 14960-70.

[91] Friedman, R.; Pellarin, R.; Caflisch, A. Amyloid aggregation on lipid bilayers and its impact on

membrane permeability. *J. Mol. Biol.* **2009,** *387* (2), 407-15.

[92] Bellesia, G.; Shea, J. E. Self-assembly of beta-sheet forming peptides into chiral fibrillar aggregates. *J. Chem. Phys.* **2007,** *126* (24), 245104.

[93] Morriss-Andrews, A.; Bellesia, G.; Shea, J. E. beta-sheet propensity controls the kinetic pathways and morphologies of seeded peptide aggregation. *J. Chem. Phys.* **2012,** *137* (14), 145104.

[94] Bellesia, G.; Shea, J. E. Effect of beta-sheet propensity on peptide aggregation. *J. Chem. Phys.* **2009,** *130* (14), 145103.

[95] Bellesia, G.; Shea, J. E. Diversity of kinetic pathways in amyloid fibril formation. *J. Chem. Phys.* **2009,** *131* (11), 111102.

[96] Morriss-Andrews, A.; Shea, J. E. Kinetic pathways to peptide aggregation on surfaces: the effects of beta-sheet propensity and surface attraction. *J. Chem. Phys.* **2012,** *136* (6), 065103.

[97] Cheon, M.; Chang, I.; Hall, C. K. Extending the PRIME model for protein aggregation to all 20 amino acids. *Proteins* **2010,** *78* (14), 2950-60.

[98] Phelps, E. M.; Hall, C. K. Structural transitions and oligomerization along polyalanine fibril formation pathways from computer simulations. *Proteins* **2012,** *80* (6), 1582-97.

[99] Cheon, M.; Chang, I.; Hall, C. K. Influence of temperature on formation of perfect tau fragment fibrils using PRIME20/DMD simulations. *Protein Sci.* **2012,** *21* (10), 1514-27.

[100] Wagoner, V. A.; Cheon, M.; Chang, I.; Hall, C. K. Fibrillization propensity for short designed hexapeptides predicted by computer simulation. *J. Mol. Biol.* **2012,** *416* (4), 598-609.

[101] Wagoner, V. A.; Cheon, M.; Chang, I.; Hall, C. K. Computer simulation study of amyloid fibril formation by palindromic sequences in prion peptides. *Proteins* **2011,** *79* (7), 2132-45.

[102] Cheon, M.; Chang, I.; Hall, C. K. Spontaneous formation of twisted Abeta(16-22) fibrils in large-scale molecular-dynamics simulations. *Biophys. J.* **2011,** *101* (10), 2493-501.

[103] Wagoner, V. A.; Cheon, M.; Chang, I.; Hall, C. K. Impact of sequence on the molecular assembly of short amyloid peptides. *Proteins* **2014,** *82* (7), 1469-83.

[104] Mehta, A. K.; Lu, K.; Childers, W. S.; Liang, Y.; Dublin, S. N.; Dong, J.; Snyder, J. P.; Pingali, S. V.; Thiyagarajan, P.; Lynn, D. G. Facial symmetry in protein self-assembly. *J. Am. Chem. Soc.* **2008,** *130* (30), 9829-35.

[105] Balbach, J. J.; Ishii, Y.; Antzutkin, O. N.; Leapman, R. D.; Rizzo, N. W.; Dyda, F.; Reed, J.; Tycko, R. Amyloid fibril formation by A beta 16-22, a seven-residue fragment of the Alzheimer's beta-amyloid peptide, and structural characterization by solid state NMR. *Biochemistry* **2000,** *39*

(45), 13748-59.

[106] Tjernberg, L. O.; Naslund, J.; Lindqvist, F.; Johansson, J.; Karlstrom, A. R.; Thyberg, J.; Terenius, L.; Nordstedt, C. Arrest of beta-amyloid fibril formation by a pentapeptide ligand. *J. Biol. Chem.* **1996,** *271* (15), 8545-8.

[107] Spill, Y. G.; Pasquali, S.; Derreumaux, P. Impact of Thermostats on Folding and Aggregation Properties of Peptides Using the Optimized Potential for Efficient Structure Prediction Coarse-Grained Model. *J. Chem. Theory Comput.* **2011,** *7* (5), 1502-10.

[108] Chebaro, Y.; Jiang, P.; Zang, T.; Mu, Y.; Nguyen, P. H.; Mousseau, N.; Derreumaux, P. Structures of Aβ17–42 Trimers in Isolation and with Five Small-Molecule Drugs Using a Hierarchical Computational Procedure. *J. Phys. Chem. B* **2012,** *116* (29), 8412-8422.

[109] Sterpone, F.; Melchionna, S.; Tuffery, P.; Pasquali, S.; Mousseau, N.; Cragnolini, T.; Chebaro, Y.; St-Pierre, J. F.; Kalimeri, M.; Barducci, A.; Laurin, Y.; Tek, A.; Baaden, M.; Nguyen, P. H.; Derreumaux, P. The OPEP protein model: from single molecules, amyloid formation, crowding and hydrodynamics to DNA/RNA systems. *Chem. Soc. Rev.* **2014,** *43* (13), 4871-93.

[110] Brooks, B. R.; Brooks, C. L., 3rd; Mackerell, A. D., Jr.; 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. CHARMM: the biomolecular simulation program. *J. Comput. Chem.* **2009,** *30* (10), 1545-614.

[111] Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M., CHARMM: A program for macromolecular energy, minimization, and dynamics calculations. *J. Comput. Chem.* **1983,** *4* (2), 187-217.

[112] MacKerell, A. D.; Bashford, D.; Bellott, M.; Dunbrack, R. L.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph-McCarthy, D.; Kuchnir, L.; Kuczera, K.; Lau, F. T.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D. T.; Prodhom, B.; Reiher, W. E.; Roux, B.; Schlenkrich, M.; Smith, J. C.; Stote, R.; Straub, J.; Watanabe, M.; Wiorkiewicz-Kuczera, J.; Yin, D.; Karplus, M. All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J. Phys. Chem. B* **1998,** *102* (18), 3586-616.

[113] Mackerell, A. D., Jr.; Feig, M.; Brooks, C. L., 3rd. Extending the treatment of backbone energetics in protein force fields: limitations of gas-phase quantum mechanics in reproducing protein conformational distributions in molecular dynamics simulations. *J. Comput. Chem.* **2004,** *25* (11), 1400-15.

[114] Best, R. B.; Zhu, X.; Shim, J.; Lopes, P. E.; Mittal, J.; Feig, M.; Mackerell, A. D., Jr. Optimization of the additive CHARMM all-atom protein force field targeting improved sampling of the backbone phi, psi and side-chain chi(1) and chi(2) dihedral angles. *J. Chem. Theory Comput.*

2012, *8* (9), 3257-3273.

[115] Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules. *J. Am. Chem. Soc.* **1995,** *117* (19), 5179-5197.

[116] Cheatham, T. E., 3rd; Cieplak, P.; Kollman, P. A. A modified version of the Cornell et al. force field with improved sugar pucker phases and helical repeat. *J. Biomol. Struct. Dyn.* **1999,** *16* (4), 845-62.

[117] Jorgensen, W. L.; Tirado-Rives, J. The OPLS [optimized potentials for liquid simulations] potential functions for proteins, energy minimizations for crystals of cyclic peptides and crambin. *J. Am. Chem. Soc.* **1988,** *110* (6), 1657-66.

[118] Roux, B.; Simonson, T. Implicit solvent models. *Biophys. Chem.* **1999,** *78* (1-2), 1-20.

[119] Feig, M.; Brooks, C. L., 3rd. Recent advances in the development and application of implicit solvent models in biomolecule simulations. *Curr. Opin. Struct. Biol.* **2004,** *14* (2), 217-24.

[120] Warshel, A.; Russell, S. T. Calculations of electrostatic interactions in biological systems and in solutions. *Q. Rev. Biophys.* **1984,** *17* (3), 283-422.

[121] Warshel, A.; Papazyan, A. Electrostatic effects in macromolecules: fundamental concepts and practical modeling. *Curr. Opin. Struct. Biol.* **1998,** *8* (2), 211-7.

[122] Onufriev, A. V.; Case, D. A. Generalized Born Implicit Solvent Models for Biomolecules. *Annu. Rev. Biophys.* **2019,** *48*, 275-296.

[123] Haberthur, U.; Caflisch, A. FACTS: Fast analytical continuum treatment of solvation. *J. Comput. Chem.* **2008,** *29* (5), 701-15.

[124] Lazaridis, T.; Karplus, M. Effective energy functions for protein structure prediction. *Curr. Opin. Struct. Biol.* **2000,** *10* (2), 139-45.

[125] Hassan, S. A.; Guarnieri, F.; Mehler, E. L. A General Treatment of Solvent Effects Based on Screened Coulomb Potentials. *J. Phys. Chem. B* **2000,** *104* (27), 6478-6489.

[126] Ferrara, P.; Apostolakis, J.; Caflisch, A. Evaluation of a fast implicit solvent model for molecular dynamics simulations. *Proteins* **2002,** *46* (1), 24-33.

[127] Im, W.; Lee, M. S.; Brooks, C. L., 3rd. Generalized born model with a simple smoothing function. *J. Comput. Chem.* **2003,** *24* (14), 1691-702.

[128] Im, W.; Feig, M.; Brooks, C. L., 3rd. An implicit membrane generalized born theory for the study of structure, stability, and interactions of membrane proteins. *Biophys. J.* **2003,** *85* (5), 2900-

18.

[129] Chocholousova, J.; Feig, M. Balancing an accurate representation of the molecular surface in generalized born formalisms with integrator stability in molecular dynamics simulations. *J. Comput. Chem.* **2006,** *27* (6), 719-29.

[130] Tamamis, P.; Kasotakis, E.; Archontis, G.; Mitraki, A., Combination of theoretical and experimental approaches for the design and study of fibril-forming peptides. *Methods Mol. Biol.* **2014,** *1216*, 53-70.

[131] Tamamis, P.; Kasotakis, E.; Mitraki, A.; Archontis, G. Amyloid-like self-assembly of peptide sequences from the adenovirus fiber shaft: insights from molecular dynamics simulations. *J. Phys. Chem. B* **2009**, *113* (47), 15639-15647.

[132] Tamamis, P.; Archontis, G. Amyloid-like Self-Assembly of a Dodecapeptide Sequence from the Adenovirus Fiber Shaft: Perspectives from Molecular Dynamics Simulations. *J. Non-Cryst. Solids* **2011**, *357* (2), 717-722.

[133] Sugita, Y.; Okamoto, Y. Replica-exchange molecular dynamics method for protein folding. *Chem. Phys. Lett.* **1999,** *314* (1), 141-151.

[134] Swendsen, R. H.; Wang, J.-S. Replica Monte Carlo Simulation of Spin-Glasses. *Phys. Rev. Lett.* **1986,** *57* (21), 2607-2609.

[135] Hukushima, K.; Nemoto, K. Exchange Monte Carlo Method and Application to Spin Glass Simulations. *J. Phys. Soc. Jpn.* **1995,** *65*, 1604-1608.

[136] Hansmann, U. H. E. Parallel tempering algorithm for conformational studies of biological molecules. *Chem. Phys. Lett.* **1997,** *281* (1), 140-150.

[137] Sanbonmatsu, K. Y.; García, A. E. Structure of Met-enkephalin in explicit aqueous solution using replica exchange molecular dynamics. *Proteins: Struct., Funct., Bioinf.* **2002,** *46* (2), 225-234.

[138] Nymeyer, H.; Gnanakaran, S.; García, A. E., Atomic Simulations of Protein Folding, Using the Replica Exchange Algorithm. In *Methods in Enzymology*, Academic Press: 2004; Vol. 383, pp 119-149.

[139] Rosenman, D. J.; Connors, C. R.; Chen, W.; Wang, C.; Garcia, A. E. Abeta monomers transiently sample oligomer and fibril-like configurations: ensemble characterization using a combined MD/NMR approach. *J. Mol. Biol.* **2013,** *425* (18), 3338-59.

[140] Wu, C.; Murray, M. M.; Bernstein, S. L.; Condron, M. M.; Bitan, G.; Shea, J. E.; Bowers, M. T. The structure of Abeta42 C-terminal fragments probed by a combined experimental and theoretical study. *J. Mol. Biol.* **2009,** *387* (2), 492-501.

[141] Seeber, M.; Cecchini, M.; Rao, F.; Settanni, G.; Caflisch, A. Wordom: a program for efficient analysis of molecular dynamics simulations. *Bioinformatics* **2007,** *23* (19), 2625-7.

[142] Seeber, M.; Felline, A.; Raimondi, F.; Muff, S.; Friedman, R.; Rao, F.; Caflisch, A.; Fanelli, F. Wordom: a user-friendly program for the analysis of molecular structures, trajectories, and free energy surfaces. *J. Comput. Chem.* **2011,** *32* (6), 1183-94.

[143] Coluzza, I. Computational protein design: a review. *J. Phys. Condens. Matter* **2017,** *29* (14), 143001.

[144] Samish, I.; MacDermaid, C. M.; Perez-Aguilar, J. M.; Saven, J. G. Theoretical and computational protein design. *Annu. Rev. Phys. Chem.* **2011,** *62*, 129-49.

[145] Li, Z.; Yang, Y.; Zhan, J.; Dai, L.; Zhou, Y. Energy functions in de novo protein design: current challenges and future prospects. *Annu. Rev. Biophys.* **2013,** *42*, 315-35.

[146] Goldenzweig, A.; Fleishman, S. J. Principles of Protein Stability and Their Application in Computational Design. *Annu. Rev. Biochem.* **2018,** *87*, 105-129.

[147] Butterfoss, G. L.; Kuhlman, B. Computer-based design of novel protein structures. *Annu. Rev. Biophys. Biomol. Struct.* **2006,** *35*, 49-65.

[148] Huang, P. S.; Boyken, S. E.; Baker, D. The coming of age of de novo protein design. *Nature* **2016,** *537* (7620), 320-7.

[149] Baker, D. What has de novo protein design taught us about protein folding and biophysics? *Protein Sci.* **2019,** *28* (4), 678-683.

[150] Bryson, J. W.; Betz, S. F.; Lu, H. S.; Suich, D. J.; Zhou, H. X.; O'Neil, K. T.; DeGrado, W. F. Protein design: a hierarchic approach. *Science* **1995,** *270* (5238), 935-41.

[151] Kuhlman, B.; Dantas, G.; Ireton, G. C.; Varani, G.; Stoddard, B. L.; Baker, D. Design of a novel globular protein fold with atomic-level accuracy. *Science* **2003,** *302* (5649), 1364-8.

[152] Fung, H. K.; Welsh, W. J.; Floudas, C. A. Computational De Novo Peptide and Protein Design: Rigid Templates versus Flexible Templates. *Ind. Eng. Chem. Res.* **2008,** *47* (4), 993-1001.

[153] Ponder, J. W.; Richards, F. M. Tertiary templates for proteins. Use of packing criteria in the enumeration of allowed sequences for different structural classes. *J. Mol. Biol.* **1987,** *193* (4), 775-91.

[154] Dunbrack, R. L., Jr.; Cohen, F. E. Bayesian statistical analysis of protein side-chain rotamer preferences. *Protein Sci.* **1997,** *6* (8), 1661-81.

[155] Dahiyat, B. I.; Mayo, S. L. De novo protein design: fully automated sequence selection. *Science*

1997, *278* (5335), 82-7.

[156] Gordon, D. B.; Hom, G. K.; Mayo, S. L.; Pierce, N. A. Exact rotamer optimization for protein design. *J. Comput. Chem.* **2003,** *24* (2), 232-43.

[157] Mooers, B. H.; Datta, D.; Baase, W. A.; Zollars, E. S.; Mayo, S. L.; Matthews, B. W. Repacking the Core of T4 lysozyme by automated design. *J. Mol. Biol.* **2003,** *332* (3), 741-56.

[158] Ross, S. A.; Sarisky, C. A.; Su, A.; Mayo, S. L. Designed protein G core variants fold to native-like structures: sequence selection by ORBIT tolerates variation in backbone specification. *Protein Sci.* **2001,** *10* (2), 450-4.

[159] Harbury, P. B.; Tidor, B.; Kim, P. S. Repacking protein cores with backbone freedom: structure prediction for coiled coils. *Proc. Natl. Acad. Sci. U. S. A.* **1995,** *92* (18), 8408-12.

[160] Klepeis, J. L.; Floudas, C. A.; Morikis, D.; Tsokos, C. G.; Argyropoulos, E.; Spruce, L.; Lambris, J. D. Integrated computational and experimental approach for lead optimization and design of compstatin variants with improved activity. *J. Am. Chem. Soc.* **2003,** *125* (28), 8422-3.

[161] Klepeis, J. L.; Floudas, C. A.; Morikis, D.; Tsokos, C. G.; Lambris, J. D. Design of Peptide Analogues with Improved Activity Using a Novel de Novo Protein Design Approach. *Ind. Eng. Chem. Res.* **2004,** *43* (14), 3817-3826.

[162] Morikis, D.; Soulika, A. M.; Mallik, B.; Klepeis, J. L.; Floudas, C. A.; Lambris, J. D. Improvement of the anti-C3 activity of compstatin using rational and combinatorial approaches. *Biochem. Soc. Trans.* **2004,** *32* (Pt 1), 28-32.

[163] Fung, H. K.; Floudas, C. A.; Taylor, M. S.; Zhang, L.; Morikis, D. Toward full-sequence de novo protein design with flexible templates for human beta-defensin-2. *Biophys. J.* **2008,** *94* (2), 584-99.

[164] Jacak, R.; Leaver-Fay, A.; Kuhlman, B. Computational protein design with explicit consideration of surface hydrophobic patches. *Proteins* **2012,** *80* (3), 825-38.

[165] Kuhlman, B.; Baker, D. Native protein sequences are close to optimal for their structures. *Proc. Natl. Acad. Sci. U. S. A.* **2000,** *97* (19), 10383-8.

[166] Rohl, C. A.; Strauss, C. E.; Misura, K. M.; Baker, D. Protein structure prediction using Rosetta. *Methods Enzymol.* **2004,** *383*, 66-93.

[167] Pokala, N.; Handel, T. M. Energy functions for protein design: adjustment with protein-protein complex affinities, models for the unfolded state, and negative design of solubility and specificity. *J. Mol. Biol.* **2005,** *347* (1), 203-27.

[168] Liang, S.; Grishin, N. V. Effective scoring function for protein sequence design. *Proteins* **2004,**

54 (2), 271-81.

[169] Kaltofen, S.; Li, C.; Huang, P. S.; Serpell, L. C.; Barth, A.; Andre, I. Computational de novo design of a self-assembling peptide with predefined structure. *J. Mol. Biol.* **2015,** *427* (2), 550-62.

[170] King, N. P.; Sheffler, W.; Sawaya, M. R.; Vollmar, B. S.; Sumida, J. P.; Andre, I.; Gonen, T.; Yeates, T. O.; Baker, D. Computational design of self-assembling protein nanomaterials with atomic level accuracy. *Science* **2012,** *336* (6085), 1171-4.

[171] Chen, Z.; Johnson, M. C.; Chen, J.; Bick, M. J.; Boyken, S. E.; Lin, B.; De Yoreo, J. J.; Kollman, J. M.; Baker, D.; DiMaio, F. Self-Assembling 2D Arrays with de Novo Protein Building Blocks. *J. Am. Chem. Soc.* **2019**.

[172] Fallas, J. A.; Ueda, G.; Sheffler, W.; Nguyen, V.; McNamara, D. E.; Sankaran, B.; Pereira, J. H.; Parmeggiani, F.; Brunette, T. J.; Cascio, D.; Yeates, T. R.; Zwart, P.; Baker, D. Computational design of self-assembling cyclic protein homo-oligomers. *Nat. Chem.* **2017,** *9* (4), 353-360.

[173] King, N. P.; Bale, J. B.; Sheffler, W.; McNamara, D. E.; Gonen, S.; Gonen, T.; Yeates, T. O.; Baker, D. Accurate design of co-assembling multi-component protein nanomaterials. *Nature* **2014,** *510* (7503), 103-8.

[174] Gonen, S.; DiMaio, F.; Gonen, T.; Baker, D. Design of ordered two-dimensional arrays mediated by noncovalent protein-protein interfaces. *Science* **2015,** *348* (6241), 1365-8.

[175] Smadbeck, J.; Chan, K. H.; Khoury, G. A.; Xue, B.; Robinson, R. C.; Hauser, C. A.; Floudas, C. A. De novo design and experimental characterization of ultrashort self-associating peptides. *PLoS Comput. Biol.* **2014,** *10* (7), e1003718.

[176] Clarke, N. D.; Yuan, S. M. Metal search: a computer program that helps design tetrahedral metal-binding sites. *Proteins* **1995,** *23* (2), 256-63.

[177] Hellinga, H. W.; Caradonna, J. P.; Richards, F. M. Construction of new ligand binding sites in proteins of known structure. II. Grafting of a buried transition metal binding site into Escherichia coli thioredoxin. *J. Mol. Biol.* **1991,** *222* (3), 787-803.

[178] Ruoslahti, E. RGD and other recognition sequences for integrins. *Annu. Rev. Cell. Dev. Biol.* **1996**, *12*, 697-715.

[179] Orbach, R.; Adler-Abramovich, L.; Zigerson, S.; Mironi-Harpaz, I.; Seliktar, D.; Gazit, E. Self-assembled Fmoc-peptides as a platform for the formation of nanostructures and hydrogels. *Biomacromolecules* **2009,** *10* (9), 2646-51.

[180] Horii, A.; Wang, X.; Gelain, F.; Zhang, S. Biological designer self-assembling peptide nanofiber scaffolds significantly enhance osteoblast proliferation, differentiation and 3-D migration. *PLoS One* **2007,** *2* (2), e190.

[181] Seow, W. Y.; Hauser, C. A. Tunable mechanical properties of ultrasmall peptide hydrogels by crosslinking and functionalization to achieve the 3D distribution of cells. *Adv. Healthc. Mater.* **2013,** *2* (9), 1219-23.

[182] Wohlrab, S.; Muller, S.; Schmidt, A.; Neubauer, S.; Kessler, H.; Leal-Egana, A.; Scheibel, T. Cell adhesion and proliferation on RGD-modified recombinant spider silk proteins. *Biomaterials* **2012,** *33* (28), 6650-9.

[183] Spieß, K.; Wohlrab, S.; Scheibel, T., Structural characterization and functionalization of engineered spider silk films. *Soft Matter* **2010,** *6* (17), 4168-4174.

[184] Tamamis, P.; Adler-Abramovich, L.; Reches, M.; Marshall, K.; Sikorski, P.; Serpell, L.; Gazit, E.; Archontis, G. Self-assembly of phenylalanine oligopeptides: insights from experiments and simulations. *Biophys. J.* **2009,** *96* (12), 5020-9.

[185] Kabsch, W.; Sander, C. Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* **1983**, *22* (12), 2577-2637.

[186] Cecchini, M.; Rao, F.; Seeber, M.; Caflisch, A. Replica exchange molecular dynamics simulations of amyloid peptide aggregation. *J. Chem. Phys.* **2004,** *121* (21), 10748-56.

[187] Allen, M. P.; Tildesley, D. J. *Computer Simulation of Liquids*; Oxford Science: Oxford, U.K., 1987.

[188] Samanta, U.; Bahadur, R. P.; Chakrabarti, P. Quantifying the accessible surface area of protein residues in their local environment. *Protein Eng.* **2002**, 15 (8), 659-667.

[189] Main, A. L.; Harvey, T. S.; Baron, M.; Boyd, J.; Campbell, I. D. The three-dimensional structure of the tenth type III module of fibronectin: an insight into RGD-mediated interactions. *Cell.* **1992**, *71* (4), 671-678.

[190] Mosmann, T. Rapid colorimetric assay for cellular growth and survival: application to proliferation and cytotoxicity assays. *J. Immunol. Methods* **1983**, *65* (1-2), 55-63.

[191] Guskova, O.; Savchenko, V.; König, U.; Uhlmann, P.; Sommer, J.U. How do immobilised cell-adhesive Arg–Gly–Asp-containing peptides behave at the PAA brush surface? *Mol. Sim.* **2018**, *44(16)*, 1325-1337.

[192] Jonnalagadda, S.V.R.; Ornithopoulou, E.; Orr, A.A.; Mossou, E.; Forsyth, V.T.; Mitchell, E.P.; Bowler, M.W.; Mitraki, A.; Tamamis, P. Computational Design of Amyloid Self-Assembling Peptides Bearing Aromatic Residues and the Cell Adhesive Motif Arg-Gly-Asp. *Mol. Syst. Des. Eng.* **2017***, 2*, 321-335.

[193] Bellis, S.L. Advantages of RGD peptides for directing cell association with biomaterials. *Biomaterials* **2011**, *32(18)*, 4205-4210.

[194] Sargeant, T.D.; Aparicio, C.; Goldberger, J.E.; Cui, H.; Stupp, S.I. Mineralization of peptide amphiphile nanofibers and its effect on the differentiation of human mesenchymal stem cells. *Acta Biomater.* **2012**, *8(7)*, 2456-2465.

[195] Hosseinkhani, H.; Hosseinkhani, M.; Tian, F.; Kobayashi, H.; Tabata, Y. Osteogenic differentiation of mesenchymal stem cells in self-assembled peptide-amphiphile nanofibers. *Biomaterials* **2006**, *27(22)*, 4079-4086.

[196] Sargeant, T.D.; Rao, M.S.; Koh, C.Y.; Stupp, S.I. Covalent functionalization of NiTi surfaces with bioactive peptide amphiphile nanofibers. *Biomaterials* **2008**, *29(8)*, 1085-1098.

[197] Min, K.I.; Yun, G.; Jang, Y.; Kim, K.R.; Ko, Y.H.; Jang, H.S.; Lee, Y.S.; Kim, K.; Kim, D.P. Covalent Self-Assembly and One-Step Photocrosslinking of Tyrosine-Rich Oligopeptides to Form Diverse Nanostructures. *Angew Chem. Int. Ed. Engl.* **2016**, *55(24)*, 6925-6928.

[198] Ding, Y.; Li, Y.; Qin, M.; Cao, Y.; Wang, W. Photo-cross-linking approach to engineering small tyrosine-containing peptide hydrogels with enhanced mechanical stability. *Langmuir* **2013**, *29(43)*, 13299-13306.

[199] Partlow, B. P.; Applegate, M. B.; Omenetto, F. G.; Kaplan, D. L. Dityrosine Cross-Linking in Designing Biomaterials. *ACS Biomater. Sci. Eng.* **2016,** *2* (12), 2108-2121.

[200] Huang, Y. F.; Lu, S. C.; Huang, Y. C.; Jan, J. S. Cross-linked, self-fluorescent gold nanoparticle/polypeptide nanocapsules comprising dityrosine for protein encapsulation and label-free imaging. *Small.* **2014,** *10* (10), 1939-44.

[201] Smadbeck, J.; Peterson, M.B.; Khoury, G.A.; Taylor, M.S.; Floudas, C.A. Protein WISDOM: a workbench for in silico de novo design of biomolecules. *J. Vis. Exp.* **2013**, *(77).*

[202] Tamamis, P.; López de Victoria, A.; Gorham, R.D. Jr.; Bellows-Peterson M.L.; Pierou, P.; Floudas, C.A.; Morikis, D.; Archontis, G. Molecular dynamics in drug design: new generations of compstatin analogs. *Chem. Biol. Drug Des.* **2012**, *79(5)*, 703-718.

[203] Fung, H.K.; Taylor M.S.; Floudas, C.A. Novel Formulations for the Sequence Selection Problem in De Novo Protein Design with Flexible Templates. *Optim. Methods Softw.* **2007**, *22*, 51-71.

[204] Fung, H.K.; Floudas, C.A.; Taylor, M.S.; Zhang, L.; Morikis, D. Toward Full-Sequence De Novo Protein Design with Flexible Templates for Human Beta-Defensin-2. *Biophys. J.* **2008**, *94(2)*, 584–599.

[205] Pons, C.; Talavera, D.; de la Cruz, X.; Orozco, M.; Fernandez-Recio, J. Scoring by Intermolecular Pairwise Propensities of Exposed Residues (SIPPER): A New Efficient Potential for Protein-Protein Docking. *J. Chem. Inf. Model* **2011**, *51*, 370–377.

[206] Rajgaria, R.; McAllister, S.R.; Floudas, C.A. A novel high resolution Calpha--Calpha distance

dependent force field based on a high quality decoy set. *Proteins* **2006**, *65(3)*, 726-741.

[207] Rajgaria, R.; McAllister, S.R.; Floudas, C.A. Distance Dependent Centroid to Centroid Force Fields using High Resolution Decoys. *Proteins* **2008**, *70*, 950-970.

[208] Tobi, D.; Elber, R. Distance-dependent, pair potential for protein folding: results from linear optimization. *Proteins.* **2000**, *41,* 40-46.

[209] Laage, D.; Stirnemann, G.; Hynes, J.T.; Why Water Reorientation Slows without Iceberg Formation around Hydrophobic Solutes. *J. Phys. Chem. B* **2009**, *113*, 2428-2435.

[210] Du, Q.; Freysz, E.; Shen, Y.R. Surface Vibrational Spectroscopic Studies of Hydrogen Bonding and Hydrophobicity. *Science* **1994**, *264*, 826-828.

[211] Despa, F.; Berry, R.S. The Origin of Long-Range Attraction between Hydrophobes in Water. *Biophys J.* **2007**, *92(2)*, 373-378.

[212] Spolar, R.S.; Hat, J.; Record, M.T. Hydrophobic Effect in Protein Folding and Other Noncovalent Processes Involving Proteins. *Proc. Natl. Acad. Sci. U. S. A.* **1989**, *86*, 8382-8385.

[213] Swanson, J.M.J.; Henchman, R.H.; McCammon, J.A. Revisiting free energy calculations: a theoretical connection to MM/PBSA and direct calculation of the association free energy. *Biophysics. J.* **2004**, *86(1)*, 67–74.

[214] Pascual-Ahuir, J.L.; Silla, E.; Tuñon, I. GEPOL: An improved description of molecular surfaces. III. A new algorithm for the computation of a solvent-excluding surface *J. Comput. Chem.* **1994**, *15(10***)**, 1127 – 1138.

[215] Gazit, E. A possible role for pi-stacking in the self-assembly of amyloid fibrils. *FASEB J.* **2002,** *16* (1), 77-83.

[216] Gazit, E. Mechanisms of amyloid fibril self-assembly and inhibition. Model short peptides as a key research tool. *FEBS J.* **2005,** *272* (23), 5971-8.

[217] Gazit, E. Self assembly of short aromatic peptides into amyloid fibrils and related nanostructures. *Prion* **2007,** *1* (1), 32-5.

[218] Irwin, J.J.; and Shoichet, B.K. ZINC – A Free Database of Commercially Available Compounds for Virtual Screening. *J. Chem. Inf. Model.* **2005**, *45(1)*, 177-182.

[219] Cecchini, M.; Rao, F.; Seeber, M.; Caflisch, A. Replica exchange molecular dynamics simulations of amyloid peptide aggregation. *J. Chem. Phys.* **2004,** *121* (21), 10748-56.

[220] Allen, M. P.; Tildesley, D. J. *Computer Simulation of Liquids*; Oxford Science: Oxford, U.K., 1987.

[221] Wesson, L.; Eisenberg, D. Atomic solvation parameters applied to molecular dynamics of proteins in solution. *Protein Sci.* **1992,** *1* (2), 227-35.

[222] Ke, P. C.; Sani, M. A.; Ding, F.; Kakinen, A.; Javed, I.; Separovic, F.; Davis, T. P.; Mezzenga, R. Implications of peptide assemblies in amyloid diseases. *Chem. Soc. Rev.* **2017,** *46* (21), 6492-6531.

[223] Wei, G.; Su, Z.; Reynolds, N. P.; Arosio, P.; Hamley, I. W.; Gazit, E.; Mezzenga, R. Self-assembling peptide and protein amyloids: from structure to tailored function in nanotechnology. *Chem. Soc. Rev.* **2017,** *46* (15), 4661-4708.

[224] Dharmadana, D.; Reynolds, N. P.; Conn, C. E.; Valery, C. Molecular interactions of amyloid nanofibrils with biological aggregation modifiers: implications for cytotoxicity mechanisms and biomaterial design. *Interface Focus* **2017,** *7* (4), 20160160.

[225] Iconomidou, V. A.; Vriend, G.; Hamodrakas, S. J. Amyloids protect the silkmoth oocyte and embryo. *FEBS Lett.* **2000,** *479* (3), 141-5.

[226] Mitraki, A. Protein aggregation from inclusion bodies to amyloid and biomaterials. *Adv Protein Chem. Struct. Biol.* **2010,** *79*, 89-125.

[227] Paci, E.; Gsponer, J.; Salvatella, X.; Vendruscolo, M. Molecular dynamics studies of the process of amyloid aggregation of peptide fragments of transthyretin. *J. Mol. Biol.* **2004,** *340* (3), 555-69.

[228] Cecchini, M.; Rao, F.; Seeber, M.; Caflisch, A. Replica exchange molecular dynamics simulations of amyloid peptide aggregation. *J. Chem. Phys.* **2004,** *121* (21), 10748-56.

[229] Colletier, J.P.; Laganowsky, A.; Landau, M.; Zhao, M.; Soriaga, A.B.; Goldschmidt, L.; Flot, D.; Cascio, D.; Sawaya, M.R.; Eisenberg, D.S. Molecular basis for amyloid-β polymorphism. *Proc. Natl. Acad. Sci. U. S. A.* **2011**, *108(41)*, 16938-16943.

[230] van Raaij, M.J.; Mitraki, A. Beta-structured viral fibres: assembly, structure and implications for materials design. *Curr. Opin. Solid State Mater. Sci.* **2004**, *8*, 151–156.

[231] Mitraki, A.; Papanikolopoulou, K.; Van Raaij, M.J. Natural triple beta-stranded fibrous folds. *Adv. Protein Chem.* **2006**, ***73***, 97-124.

[232] Lakshmanan, A.; Cheong, D. W.; Accardo, A.; Di Fabrizio, E.; Riekel, C.; Hauser, C. A. Aliphatic peptides show similar self-assembly to amyloid core sequences, challenging the importance of aromatic interactions in amyloidosis. *Proc. Natl. Acad. Sci. U. S. A.* **2013,** *110* (2), 519-24.

[233] Lopez De La Paz, M.; Goldie, K.; Zurdo, J.; Lacroix, E.; Dobson, C. M.; Hoenger, A.; Serrano, L. De novo designed peptide-based amyloid fibrils. *Proc. Natl. Acad. Sci. U. S. A.* **2002,** *99* (25), 16052-7.

[234] Colvin, M.T.; Silvers, R.; Ni, Q.Z.; Can, T.V.; Sergeyev, I.; Rosay, M.; Donovan, K.J.; Michael, B.; Wall, J.; Linse, S.; et al. Atomic Resolution Structure of Monomorphic Aβ42 Amyloid Fibrils. *J. Am. Chem. Soc.* **2016**, *138*, 9663-9674.

[235] Gremer, L.; Schölzel, D.; Schenk, C.; Reinartz, E.; Labahn, J.; Ravelli, R.B.; Tusche, M.; Lopez-Iglesias, C.; Hoyer, W.; Heise, H,; et al. Fibril Structure of Amyloid-β(1-42) by Cryo-Electron Microscopy. *Science* **2017**, *358*, 116-119.

[236] Wälti, M.A.; Ravotti, F.; Arai, H.; Glabe, C.D.; Wall, J.S.; Böckmann, A.; Güntert, P.; Meier, B.H.; Riek, R. Atomic-Resolution Structure of a Disease-Relevant Aβ(1-42) Amyloid Fibril. *Proc. Natl. Acad. Sci. U. S. A.* **2016**, *113*, E4976-E4984.

[237] Tamamis, P.; Floudas, C.A. Molecular Recognition of CXCR4 by a Dual Tropic HIV-1 gp120 V3 Loop. *Biophys. J.* **2013**, *105(6)*, 1502-1514.

[238] Tamamis, P.; and Floudas, C.A. Molecular Recognition of CCR5 by an HIV-1 gp120 V3 Loop. *PLoS ONE*, **2014,** *9(4)*, e95767.

[239] Luckey, M.; Hernandez, J.; Arlaud, G.; Forsyth, V. T.; Ruigrok, R. W.; Mitraki, A. A peptide from the adenovirus fiber shaft forms amyloid-type fibrils. *FEBS Lett.* **2000,** *468* (1), 23-7.

[240] Bowler, M. W.; Nurizzo, D.; Barrett, R.; Beteva, A.; Bodin, M.; Caserotto, H.; Delageniere, S.; Dobias, F.; Flot, D.; Giraud, T.; Guichard, N.; Guijarro, M.; Lentini, M.; Leonard, G. A.; McSweeney, S.; Oskarsson, M.; Schmidt, W.; Snigirev, A.; von Stetten, D.; Surr, J.; Svensson, O.; Theveneau, P.; Mueller-Dieckmann, C. MASSIF-1: a beamline dedicated to the fully automatic characterization and data collection from crystals of biological macromolecules. *J. Synchrotron Radiat.* **2015,** *22* (6), 1540-7.

[241] Troyer, R.M.; Collins, K.R.; Abraha, A.; Fraundorf, E.; Moore, D.M.; Krizan, R.W.; Toossi, Z.; Colebunders, R.L.; Jensen, M.A.; Mullins, J.I.; Vanham, G.; Arts, E.J. Changes in human immunodeficiency virus type 1 fitness and genetic diversity during disease progression. *J. Virol.* **2005**, *79(14)*, 9006-9018.

[242] Szyk, A.; Deaconescu, A.M.; Spector, J.; Goodman, B.; Valenstein, M.L.; Ziolkowska, N.E.; Kormendi, V.; Grigorieff, N.; Roll-Mecak, A. Molecular Basis for Age-Dependent Microtubule Acetylation by Tubulin Acetyltransferase. *Cell* **2014**, *157*, 1405-1415.

[243] Wagner, T.; Bellinzoni, M.; Wehenkel, A.; O'Hare, H.M.; Alzari, P.M. Functional Plasticity and Allosteric Regulation of α-Ketoglutarate Decarboxylase in Central Mycobacterial Metabolism. *Chem. Biol.* **2011**, *18(8)*, 1011-1020.

[244] Poonsiri, T.; Wright, G.S.A.; Diamond, M.S.; Turtle, L.; Solomon T.; Antonyuk, S.V. Structural Study of the C-Terminal Domain of Nonstructural Protein 1 from Japanese Encephalitis Virus. *J. Virol.* **2018**, *92*, e01868- 01917.

[245] Howell, D. W.; Tsai, S. P.; Churion, K.; Patterson, J.; Abbey, C.; Atkinson, J. T.; Porterpan, D.;

You, Y. H.; Meissner, K. E.; Bayless, K. J.; Bondos, S. E. Identification of multiple dityrosine bonds in materials composed of the Drosophila protein Ultrabithorax. *Adv. Funct. Mater.* **2015,** *25* (37), 5988-5998.

[246] Yang, D.; Sarina, S.; Zhu, H.; Liu, H.; Zheng. Z.; Xie, M.; Smith, S.V.; Komarneni, S. Capture of Radioactive Cesium and Iodide Ions from Water by Using Titanate Nanofibers and Nanotubes. *Angew. Chem. Int. Ed. Engl.* **2011**, *50*, 10594-10598.

[247] Qian, J.; Xu. J.; Kuang, L.; Hua, D. Cesium Removal from Human Blood by Poly (ethylene glycol)-Decorated Prussian Blue Magnetic Nanoparticles. *ChemPlusChem* **2017**, *82*, 888–895.

[248] Sun, J.; Yang, D.; Sun, C.; Liu, L.; Yang, S.; Alec Jia, Y.; Cai, R.; Yao, X. Potassium Niobate Nanolamina: A Promising Adsorbent for Entrapment of Radioactive Cations from Water. *Sci. Rep.* **2014**, *4*, 7313.

[249] Yang, D.; Liu, H.; Zheng, Z.; Sarina, S.; Zhu, H. Titanate-Based Adsorbents for Radioactive Ions Entrapment from Water. *Nanoscale* **2013**, *5*, 2232-2242.

[250] Sarina, S.; Bo, A.; Liu, D.; Liu, H.; Yang, D.; Zhou, C.; Maes, N.; Komarneni, S.; Zhu, H. Separate or Simultaneous Removal of Radioactive Cations and Anions from Water by Layered Sodium Vanadate-Based Sorbents. *Chem. Mater.* **2014**, *26*, 4788–4795.

[251] Lee, H.; Kim, H.S.; Jeong, H.K.; Park, M.; Chung, D.Y.; Lee, K.Y.; Lee, E.H.; Lim, W.T.; Selective Removal of Radioactive Cesium from Nuclear Waste by Zeolites: On the Origin of Cesium Selectivity Revealed by Systematic Crystallographic Studies, *J. Phys. Chem. C* **2017**, *121*, 10594-10608.

[252] Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. The Protein Data Bank. *Nucleic Acids Res.* **2000**, *28*, 235-242.

[253] Tiwari, M.K.; Singh, R.; Singh, R.K.; Kim, I.N.; Lee, J.K. Computational Approaches for Rational Design of Proteins with Novel Functionalities. *Comput. Struct. Biotechnol. J.* **2012**, *2*, e201209002.

[254] Pfoh, R.; Li, A.; Chakrabarti, N.; Payandeh, J.; Pomès, R.; Pai, E. F. Structural Asymmetry in the Magnesium Channel CorA Points to Sequential Allosteric Regulation. *Proc. Natl. Acad. Sci. U. S. A.* **2012**, *109*, 18809-18814.

[255] Best, R. B.; Zhu, X.; Shim, J.; Lopes, P. E.; Mittal, J.; Feig, M.; Mackerell, Jr, A. D. Optimization of the Additive CHARMM All-Atom Protein Force Field Targeting Improved Sampling of the Backbone φ, ψ and Side-Chain χ(1) and χ(2) Dihedral Angles. *J. Chem. Theory Comput.* **2012**, *8*, 3257-3273.

[256] Lamoureux, G.; Roux, B. Modeling Induced Polarization with Classical Drude Oscillators: Theory and Molecular Dynamics Simulation Algorithm. *J. Chem. Phys.* **2003**, *119*, 3025–3039.

[257] Lamoureux, G.; Harder, E.; Vorobyov, I.V.; Roux, B.; MacKerell, Jr, A.D. A Polarizable Model of Water for Molecular Dynamics Simulations of Biomolecules. *Chem. Phys. Lett.* **2006**, *418*, 245–249.

[258] Harder, E.; Anisimov, V.M.; Vorobyov, I.V.; Lopes, P.E.M.; Noskov, S.Y.; MacKerell, Jr, A.D.; Roux, B. Atomic Level Anisotropy in the Electrostatic Modeling of Lone Pairs for a Polarizable Force Field Based on the Classical Drude Oscillator. *J. Chem. Theory Comput.* **2006**, *2*, 1587–1597.

[259] Daisy, P.; Singh, S.K.; Vijayalakshmi, P.; Selvaraj, C.; Rajalakshmi, M.; Suveena, S. A Database for the Predicted Pharmacophoric Features of Medicinal Compounds. *Bioinformation* **2011**, *6*, 167-168.

[260] Kim, K.H.; Kim, N.D.; Seong, B.L. Pharmacophore-Based Virtual Screening: A Review of Recent Applications. *Expert Opin. Drug Discov.* **2010***, 5*, 205-222.

[261] Koes, D.R.; Camacho, C.J. ZINCPharmer: Pharmacophore Search of the ZINC Database. *Nucleic Acids Res.* **2012**, *40*, W409-W414.

[262] Koes, D.R.; Dömling, A.; Camacho, C.J. AnchorQuery: Rapid Online Virtual Screening for Small-Molecule Protein-Protein Interaction Inhibitors. *Protein Sci.* **2018**, *27*, 229-232.

[263] Armentrout, P.B.; Yang, B.; Rodgers, M.T. Metal Cation Dependence of Interactions with Amino Acids: Bond Dissociation Energies of Rb(+) and Cs(+) to the Acidic Amino Acids and their Amide Derivatives. *J. Phys. Chem. B* **2014**, *118*, 4300-4314.

[264] Pfoh, R.; Li, A.; Chakrabarti, N.; Payandeh, J.; Pomès, R.; Pai, E.F. Structural Asymmetry in the Magnesium Channel CorA Points to Sequential Allosteric Regulation. *Proc. Natl. Acad. Sci. U. S. A.* **2012**, *109*, 18809-18814.

[265] Baconguis, I.; Bohlen, C.J.; Goehring, A.; Julius, D.; Gouaux, E. X-ray Structure of Acid-Sensing Ion Channel 1-Snake Toxin Complex Reveals Open State of a Na(+)-Selective Channel. *Cell* **2014**, *156*, 717-729.

[266] Farenc, C.; Spinelli, S.; Vinogradov, E.; Tremblay, D.; Blangy, S.; Sadovskaya, I.; Moineau, S.; Cambillau, C. Molecular Insights on the Recognition of a Lactococcus Lactis Cell Wall Pellicle by the Phage 1358 Receptor Binding Protein. *J. Virol.* **2014**, *88*, 7005-7015.

[267] Petersen, F.N.; Jensen, M.Ø.; Nielsen, C.H. Interfacial Tryptophan Residues: A Role for the Cation-Pi Effect? *Biophys. J.* **2005**, *89*, 3985-3996.

[268] Gallivan, J.P.; Dougherty, D.A. A Computational Study of Cation-pi Interactions vs Salt Bridges in Aqueous Media: Implications for Protein Engineering. *J. Am. Chem. Soc.* **2000**, *122*, 870-874.

[269] Daza, E.A.; Misra, S.K.; Schwartz-Duval, A.S.; Ohoka, A.; Miller, C.; Pan, D. Nano-Cesium for Anti-Cancer Properties: An Investigation into Cesium Induced Metabolic Interference. *ACS Appl. Mater. Interfaces* **2016**, *8*, 26600-26612

[270] Clarke, N.D.; Yuan, S.M. Metal search: a computer program that helps design tetrahedral metal-binding sites. *Proteins.* **1995**, *23(2)*, 256-263.

[271] Dahiyat, B.I.; Mayo, S.L. Protein design automation. *Protein Sci.* **1996**, *5(5)*, 895-903.

[272] Richter, F.; Leaver-Fay, A.; Khare, S.D.; Bjelic, S.; Baker, D. De novo enzyme design using Rosetta3. *PLoS One.* **2011**, *6(5)*:e19230.

[273] Burns, A., Lliffe, S. Alzheimer's disease. *Br. Med. J.* **2009**, *338*, b158.

[274] Alzheimer's Disease Facts and Figures. *Alzheimer's Association* **2015**, *11*, 332-84.

[275] What Are the Signs of Alzheimer's Disease. *National Institute on Aging* **2017**.

[276] Mufson, E. J.; Counts, S. E.; Perez, S. E.; Ginsberg, S. D. Cholinergic System During the Progression of Alzheimer's Disease: Therapeutic Implications. *Expert Rev. Neurother.* **2008**, *8*, 1703–1718.

[277] Ferreira-Viera, T. H.; Guimaraes, I. M.; Silvia, F. R.; Ribeiro, F. M. Alzheimer's Disease: Targeting the Cholinergic System. *Curr. Neuropharmacol.* **2016**, *14*, 101-115.

[278] Hasselmo, M. E. The Role of Acetylcholine in Learning and Memory. *Curr. Opin. Neurobiol.* **2006**, *16*, 710–715.

[279] Garcia-Ayllon, M.; Small, D. H.; Azila, J.; Saez-Valero, J. Revisiting the Role of Acetylcholinesterase in Alzheimer's Disease: Cross-talk with P-tau and ß-amyloid. *Front. Mole. Neurosci.* **2011**, *4*, 1-7.

[280] McGleenon B. M., Dynan K. B., Passmore A. P. Acetylcholinesterase Inhibitors in Alzheimer's Disease. *Br. J. Clin. Pharmacol.* **2001**, *48*, 471-480.

[281] Hansen, R. A.; Gartlehner, G.; Webb, A. P.; Morgan, L. C.; Moore, C. G.; Jonas, D. E. Efficacy and Safety of Donepezil, Galantamine, and Rivastigmine for the Treatment of Alzheimer's Disease: A Systematic Review and Meta-analysis. *Clin. Interv. Aging*. **2008**, *3*, 211–225.

[282] Casey, D.; Antimisiaris, D.; O'Brien, J. Drugs for Alzheimer's disease: Are they effective? *P&T*. **2010**, *35*, 208–211.

[283] Korabecny, J.; Spilovska, K.; Mezeiova, E.; Benek, O.; Juza, R.; Kaping, D.; Soukup, O. A. Efficacy and Safety of Donepezil, Galantamine, Rivastigmine, and Memantine for the Treatment of Alzheimer's Disease: a Systematic Review and Meta-Analysis. *J. Alzheimers Dis.* **2014***, 41,*

615-631.

[284] Korabecny, J.; Spilovska, K.; Mezeiova, E.; Benek, O.; Juza, R.; Kaping, D.; Soukup, O.A. Systematic Review on Donepezil-based Derivatives as Potential Cholinesterase Inhibitors for Alzheimer's Disease. *Curr. Med. Chem*. **2018**, doi: 10.2174/0929867325666180517094023.

[285] Hu, B.; Shen, Y.; Adamcik, J.; Fischer, P.; Schneider, M.; Loessner, M.J.; Mezzenga, R. Polyphenol-Binding Amyloid Fibrils Self-Assemble into Reversible Hydrogels with Antibacterial Activity. *ACS Nano.* **2018**, *12*(4), 3385-3396.

[286] Shen, Y.; Posavec, L.; Bolisetty, S.; Hilty, F. M.; Nyström, G.; Kohlbrecher, J.; Hilbe, M.; Rossi, A.; Baumgartner, J.; Zimmermann, M. B.; Mezzenga, R. Amyloid Fibril Systems Reduce, Stabilize and Deliver Bioavailable Nanosized Iron. *Nat. Nanotechnol*. **2017**, *12(7)*, 642-647.

[287] Duraj-Thatte, A.M.; Praveschotinunt, P.; Nash, T. R.; Ward, F. R.; Joshi, N. S. Modulating Bacterial and Gut Mucosal Interactions with Engineered Biofilm Matrix Proteins. *Sci. Rep*. **2018**, *8*, 3475.

[288] Jackson, M. P.; Hewitt, E. W. Why Are Functional Amyloids Non-Toxic in Humans? *Biomolecules* **2017**, *7*, 71.

[289] Gauthier, S.; Molinuevo, J.L. Benefits of Combined Cholinesterase Inhibitor and Memantine Treatment in Moderate-Severe Alzheimer's Disease. *Alzheimers Dement.* **2013**, *9(3)*, 326-331.

[290] Piovesan, D.; Minervini, G.; Tosatto, S. C. The RING 2.0 web server for high quality residue interaction networks. *Nucleic Acids Res.* **2016**, *44(W1)*, W367-74.

[291] Trott, O.; Olson, A. J. AutoDock Vina: Improving the Speed and Accuracy of Docking With a New Scoring Function, Efficient Optimization and Multithreading. *J. Comput. Chem.* **2010**, *31(2)*, 455-461.

[292] Cheung, J.; Rudolph, M.J.; Burshteyn, F.; Cassidy, M.S.; Gary, E.N.; Love, J.; Franklin, M.C.; Height, J.J. Structures of human acetylcholinesterase in complex with pharmacologically important ligands. *J. Med. Chem.* **2012**, *55*(22), 10282-10286.

[293] Ulens, C.; Spurny, R.; Thompson, A.J.; Alqazzaz, M.; Debaveye, S.; Han, L.; Price, K.; Villalgordo, J.M.; Tresadern, G.; Lynch, J.W.; Lummis, S.C. The prokaryote ligand-gated ion channel ELIC captured in a pore blocker-bound conformation by the Alzheimer's disease drug memantine. *Structure* **2014**. *22*(10), 1399-1407.

[294] Kryger, G.; Silman, I.; Sussman, J.L. Structure of acetylcholinesterase complexed with E2020 (Aricept): implications for the design of new anti-Alzheimer drugs. *Structure* **1999**, *7*(3), 297-307.

[295] Nachon, F.; Carletti, E.; Ronco, C.; Trovaslet, M.; Nicolet, Y.; Jean, L.; Renard, P.Y. Crystal structures of human cholinesterases in complex with huprine W and tacrine: elements of specificity for anti-Alzheimer's drugs targeting acetyl- and butyryl-cholinesterase. *Biochem. J.* **2013**, *453*(3),

393-399.

[296] Horton, J.R.; Sawada, K.; Nishibori, M.; Cheng, X. Structural Basis for Inhibition of Histamine N-Methyltransferase by Diverse Drugs. *J. Mol. Biol.* **2005**, *353*(2), 334–344.

[297] Bencharit, S.; Morton, C.L.; Hyatt, J.L.; Kuhn, P.; Danks, M.K.; Potter, P.M.; Redinbo, M.R. Crystal structure of human carboxylesterase 1 complexed with the Alzheimer's drug tacrine: from binding promiscuity to selective inhibition. *Chem. Biol.* **2003**, *10*(4), 341-349.

[298] Harel M, Schalk I, Ehret-Sabatier L, Bouet F, Goeldner M, Hirth C, Axelsen PH, Silman I, Sussman JL. Quaternary Ligand Binding to Aromatic Residues in the Active-Site Gorge of Acetylcholinesterase. *Proc. Natl. Acad. Sci. U. S. A.* **1993***, 90*(19), 9031-9035.

[299] Hansen, S.B.; Taylor, P. Galanthamine and non-competitive inhibitor binding to ACh-binding protein: evidence for a binding site on non α-subunit interfaces of heteromeric neuronal nicotinic receptors. *J. Mol. Biol.* **2007**, *369*(4), 895–901.

[300] Greenblatt, H.M.; Guillou, C.; Guénard, D.; Argaman, A.; Botti, S.; Badet, B.; Thal, C.; Silman, I.; Sussman, J.L. The complex of a bivalent derivative of galanthamine with torpedo acetylcholinesterase displays drastic deformation of the active-site gorge: implications for structure-based drug design. *J. Am. Chem. Soc.* **2004**, *126*(47), 15405-15411.

[301] Greenblatt, H.M.; Kryger, G.; Lewis, T.; Silman, I.; Sussman, J.L. Structure of acetylcholinesterase complexed with (-)-galantamine at 2.3 A resolution. *FEBS Lett.* **1999**, *463*(3), 321-326.

[302] Bartolucci, C.; Perola, E.; Pilger, C.; Fels, G.; Lamba, D. Three-dimensional structure of a complex of galanthamine (Nivalin) with acetylcholinesterase from Torpedo californica: implications for the design of new anti-Alzheimer drugs. *Proteins* **2001**, *42*(2), 182-191.

[303] Kwok, A.; Hart, S. L. Comparative structural and functional studies of nanoparticle formulations for DNA and siRNA delivery. *Nanomedicine* **2011,** *7* (2), 210-9.

[304] Kwok, A.; McCarthy, D.; Hart, S. L.; Tagalakis, A. D. Systematic Comparisons of Formulations of Linear Oligolysine Peptides with siRNA and Plasmid DNA. *Chem. Biol. Drug Des.* **2016,** *87* (5), 747-63.

[305] Vaughan, E. E.; Dean, D. A. Intracellular trafficking of plasmids during transfection is mediated by microtubules. *Mol. Ther.* **2006,** *13* (2), 422-8.

[306] Lehto, T.; Ezzat, K.; Wood, M. J. A.; El Andaloussi, S. Peptides for nucleic acid delivery. *Adv. Drug Deliv. Rev.* **2016,** *106* (Pt A), 172-182.

[307] Suhorutsenko, J.; Oskolkov, N.; Arukuusk, P.; Kurrikoff, K.; Eriste, E.; Copolovici, D. M.; Langel, U. Cell-penetrating peptides, PepFects, show no evidence of toxicity and immunogenicity

in vitro and in vivo. *Bioconjug. Chem.* **2011,** *22* (11), 2255-62.

[308] Guidotti, G.; Brambilla, L.; Rossi, D. Cell-Penetrating Peptides: From Basic Research to Clinics. *Trends Pharmacol. Sci.* **2017,** *38* (4), 406-424.

[309] Raucher, D.; Ryu, J. S. Cell-penetrating peptides: strategies for anticancer treatment. *Trends Mol. Med.* **2015,** *21* (9), 560-70.

[310] Bechara, C.; Sagan, S. Cell-penetrating peptides: 20 years later, where do we stand? *FEBS Lett.* **2013,** *587* (12), 1693-702.

[311] Wang, F.; Wang, Y.; Zhang, X.; Zhang, W.; Guo, S.; Jin, F. Recent progress of cell-penetrating peptides as new carriers for intracellular cargo delivery. *J. Controlled Release* **2014,** *174*, 126-36.

[312] Ignatovich, I. A.; Dizhe, E. B.; Pavlotskaya, A. V.; Akifiev, B. N.; Burov, S. V.; Orlov, S. V.; Perevozchikov, A. P. Complexes of plasmid DNA with basic domain 47-57 of the HIV-1 Tat protein are transferred to mammalian cells by endocytosis-mediated pathways. *J. Biol. Chem.* **2003,** *278* (43), 42625-36.

[313] Veiman, K. L.; Mager, I.; Ezzat, K.; Margus, H.; Lehto, T.; Langel, K.; Kurrikoff, K.; Arukuusk, P.; Suhorutsenko, J.; Padari, K.; Pooga, M.; Lehto, T.; Langel, U. PepFect14 peptide vector for efficient gene delivery in cell cultures. *Mol. Pharm.* **2013,** *10* (1), 199-210.

[314] Freimann, K.; Arukuusk, P.; Kurrikoff, K.; Vasconcelos, L. D. F.; Veiman, K. L.; Uusna, J.; Margus, H.; Garcia-Sosa, A. T.; Pooga, M.; Langel, U. Optimization of in vivo DNA delivery with NickFect peptide vectors. *J. Controlled Release* **2016,** *241*, 135-143.

[315] Freimann, K.; Arukuusk, P.; Kurrikoff, K.; Parnaste, L.; Raid, R.; Piirsoo, A.; Pooga, M.; Langel, U. Formulation of Stable and Homogeneous Cell-Penetrating Peptide NF55 Nanoparticles for Efficient Gene Delivery In Vivo. *Mol. Ther. Nucleic Acids* **2018,** *10*, 28-35.

[316] Dai, B.; Li, D.; Xi, W.; Luo, F.; Zhang, X.; Zou, M.; Cao, M.; Hu, J.; Wang, W.; Wei, G.; Zhang, Y.; Liu, C. Tunable assembly of amyloid-forming peptides into nanosheets as a retrovirus carrier. *Proc. Natl. Acad. Sci. U. S. A* **2015,** *112* (10), 2996-3001.

[317] Yolamanova, M.; Meier, C.; Shaytan, A. K.; Vas, V.; Bertoncini, C. W.; Arnold, F.; Zirafi, O.; Usmani, S. M.; Muller, J. A.; Sauter, D.; Goffinet, C.; Palesch, D.; Walther, P.; Roan, N. R.; Geiger, H.; Lunov, O.; Simmet, T.; Bohne, J.; Schrezenmeier, H.; Schwarz, K.; Standker, L.; Forssmann, W. G.; Salvatella, X.; Khalatur, P. G.; Khokhlov, A. R.; Knowles, T. P.; Weil, T.; Kirchhoff, F.; Munch, J. Peptide nanofibrils boost retroviral gene transfer and provide a rapid means for concentrating viruses. *Nat. Nanotechnol.* **2013,** *8* (2), 130-6.

[318] Munch, J.; Rucker, E.; Standker, L.; Adermann, K.; Goffinet, C.; Schindler, M.; Wildum, S.; Chinnadurai, R.; Rajan, D.; Specht, A.; Gimenez-Gallego, G.; Sanchez, P. C.; Fowler, D. M.; Koulov, A.; Kelly, J. W.; Mothes, W.; Grivel, J. C.; Margolis, L.; Keppler, O. T.; Forssmann, W. G.; Kirchhoff, F. Semen-derived amyloid fibrils drastically enhance HIV infection. *Cell* **2007,** *131*

(6), 1059-71.

[319] Madani, F.; Lindberg, S.; Langel, U.; Futaki, S.; Gräslund, A. Mechanisms of cellular uptake of cell-penetrating peptides. *J. Biophys.* **2011,** *2011*, 414729-414729.

[320] Wu, J.; Shancheng, Y.; Lihua, T.; Dong, H. In *Computational analysis of propensities of amino acids and nucleotides usage at protein-nucleic acid interfaces*, International Conference on Information Science and Technology, 26-28 March 2011; 2011; pp 1342-1349.

[321] Agrawal, P.; Bhalla, S.; Usmani, S. S.; Singh, S.; Chaudhary, K.; Raghava, G. P. S.; Gautam, A. CPPsite 2.0: a repository of experimentally validated cell-penetrating peptides. *Nucleic Acids Res.* **2016,** *44* (D1), D1098-D1103.

[322] Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. CHARMM: A program for macromolecular energy, minimization, and dynamics calculations. *J. Comput. Chem.* **1983,** *4* (2), 187-217.

[323] Zhou, H.; Wei, J.; Wu, X.; Shi, J.; Liu, C.; Jia, J.; Dai, C.; Gan, Q. The bio-functional role of calcium in mesoporous silica xerogels on the responses of osteoblasts in vitro. *J. Mater. Sci. Mater. Med.* **2010,** *21* (7), 2175-85.

[324] Pytela, R.; Pierschbacher, M.D.; Argraves, S.; Suzuki, S.; Ruoslahti, E. Arginine-glycine-aspartic acid adhesion receptors. *Methods Enzymol.* **1987**, *144*, 475-489.

[325] Mourino, V.; Cattalini, J. P.; Boccaccini, A. R. Metallic ions as therapeutic agents in tissue engineering scaffolds: an overview of their biological applications and strategies for new developments. *J. R. Soc. Interface.* **2012,** *9* (68), 401-19.

[326] Maeno, S.; Niki, Y.; Matsumoto, H.; Morioka, H.; Yatabe, T.; Funayama, A.; Toyama, Y.; Taguchi, T.; Tanaka, J. The effect of calcium ion concentration on osteoblast viability, proliferation and differentiation in monolayer and 3D culture. *Biomaterials* **2005,** *26* (23), 4847-55.

[327] Marie, P. J. The calcium-sensing receptor in bone cells: a potential therapeutic target in osteoporosis. *Bone* **2010,** *46* (3), 571-6.

[328] Valerio, P.; Pereira, M. M.; Goes, A. M.; Leite, M. F. Effects of extracellular calcium concentration on the glutamate release by bioactive glass (BG60S) preincubated osteoblasts. *Biomed. Mater.* **2009,** *4* (4), 045011.

[329] Hinoi, E.; Takarada, T.; Yoneda, Y. Glutamate signaling system in bone. *J. Pharmacol. Sci.* **2004,** *94* (3), 215-20.

[330] Tejeda-Montes, E.; Klymov, A.; Nejadnik, M.R.; Alonso, M.; Rodriguez-Cabello, J.C.; Walboomers, X.F.; Mata, A. Mineralization and bone regeneration using a bioactive elastin-like

recombinamer membrane. *Biomaterials* **2014**, 35(29), 8339-8347.

[331] Boivin, G.; Deloffre, P.; Perrat, B.; Panczer, G.; Boudeulle, M.; Mauras, Y.; Allain, P.; Tsouderos, Y.; Meunier, P. J., Strontium distribution and interactions with bone mineral in monkey iliac bone after strontium salt (S 12911) administration. *J. Bone Miner. Res.* **1996,** *11* (9), 1302-11.

[332] Dahl, S. G.; Allain, P.; Marie, P. J.; Mauras, Y.; Boivin, G.; Ammann, P.; Tsouderos, Y.; Delmas, P. D.; Christiansen, C. Incorporation and distribution of strontium in bone. *Bone* **2001,** *28* (4), 446-53.

[333] Grynpas, M. D.; Hamilton, E.; Cheung, R.; Tsouderos, Y.; Deloffre, P.; Hott, M.; Marie, P. J. Strontium increases vertebral bone volume in rats at a low dose that does not induce detectable mineralization defect. *Bone* **1996,** *18* (3), 253-9.

[334] Marie, P. J.; Ammann, P.; Boivin, G.; Rey, C. Mechanisms of action and therapeutic potential of strontium in bone. *Calcif. Tissue Int.* **2001,** *69* (3), 121-9.

[335] Zhu, L. L.; Zaidi, S.; Peng, Y.; Zhou, H.; Moonga, B. S.; Blesius, A.; Dupin-Roger, I.; Zaidi, M.; Sun, L. Induction of a program gene expression during osteoblast differentiation with strontium ranelate. *Biochem. Biophys. Res. Commun.* **2007,** *355* (2), 307-11.

[336] Tan, S.; Zhang, B.; Zhu, X.; Ao, P.; Guo, H.; Yi, W.; Zhou, G.Q. Deregulation of bone forming cells in bone diseases and anabolic effects of strontium-containing agents and biomaterials. *Biomed Res. Int.* **2014**, *2014*:814057.

[337] Place, E. S.; Rojo, L.; Gentleman, E.; Sardinha, J. P.; Stevens, M. M. Strontium- and zinc-alginate hydrogels for bone tissue engineering. *Tissue Eng. Part A.* **2011**, *17(21-22)*, 2713-2722.

[338] Rezwan, K.; Chen, Q. Z.; Blaker, J. J.; Boccaccini, A. R. Biodegradable and bioactive porous polymer/inorganic composite scaffolds for bone tissue engineering. *Biomaterials* **2006,** *27* (18), 3413-31.

[339] Habibovic, P.; Barralet, J. E. Bioinorganics and biomaterials: bone repair. *Acta Biomater.* **2011,** *7* (8), 3013-26.

[340] Hoppe, A.; Guldal, N. S.; Boccaccini, A. R. A review of the biological response to ionic dissolution products from bioactive glasses and glass-ceramics. *Biomaterials* **2011,** *32* (11), 2757-74.