Genetically engineered repetitive polypeptides: design, synthesis, characterization, and applications

Natalya I. Topilina

University at Albany, State University of New York, nt992267@albany.edu

The University at Albany community has made this article openly available. Please share how this access benefits you.

Follow this and additional works at: https://scholarsarchive.library.albany.edu/legacy-etd

Part of the Analytical Chemistry Commons, Biochemistry Commons, and the Polymer Chemistry Commons

Recommended Citation

https://scholarsarchive.library.albany.edu/legacy-etd/123

This Dissertation is brought to you for free and open access by the The Graduate School at Scholars Archive. It has been accepted for inclusion in Legacy Theses & Dissertations (2009 - 2024) by an authorized administrator of Scholars Archive. Please see Terms of Use. For more information, please contact scholarsarchive@albany.edu.
Genetically Engineered Repetitive Polypeptides: 
Design, Synthesis, Characterization, and Applications

by

Natalya I. Topilina

A Dissertation
Submitted to the University at Albany, State University of New York
in Partial Fulfillment of
the Requirements for the Degree of
Doctor of Philosophy

College of Arts & Sciences
Department of Chemistry
2009
Genetically Engineered Repetitive Polypeptides:
Design, Synthesis, Characterization, and Applications

by

Natalya I. Topilina

COPYRIGHT 2009
ABSTRACT

This work is dedicated to the investigation of protein folding with the emphasis on the aggregation and amyloid fibrils formation. The focus of the research is to elucidate the relationship between polypeptide chain sequence and folding/aggregation properties. For this purpose, i) libraries of the de novo macromolecular repetitive β-sheet forming polypeptides were designed, and ii) the strategy for their rapid creation was developed. The strategy enables retention of a basic polypeptide core, yet allowing variations in the repetitive sequence, degree of polymerization and distribution of specific amino acids. The polypeptides were expressed in *E. coli*, purified, and characterized by various biophysical methods, including circular dichroism (CD), deep UV resonance Raman spectroscopy (DUVRR), atomic force microscopy (AFM) and transmission electron microscopy (TEM). The polypeptides form β-sheet structure along with simultaneous coacervation and self-assembly into amyloid-like fibrils. The repetitive nature eliminates sequence specificity in polypeptide interactions and enables their utilization as macromolecular models of large amyloidogenic proteins. These models were employed for structural characterization of amyloid fibrils and elucidation of the influence of polypeptide chain properties such as charge distribution, number of intramolecular β-strands and hydrophobicity on folding and aggregation.
ACKNOWLEDGEMENTS

I would like to take this opportunity to thank people who helped me while I was undertaking my graduate studies.

I would like to give my most sincere and deepest thanks to my mentor and advisor, Prof. John T. Welch. I am truly indebted for him for his patience and integrity, warm encouragement and generosity.

I am very grateful to Prof. Igor K. Lednev for the opportunity to collaborate with his research group on the characterization of polypeptides by deep UV resonance Raman spectroscopy.

I would certainly like to acknowledge my doctoral committee members, Prof. Georges Belfort, Prof. Carla Theimer, Prof. Igor K. Lednev, and Prof. Jayanti Pande.

To the past and present members of the Welch’s and Lednev’s groups, it was a very beneficial experience, and I am thankful for all your assistance throughout my graduate studies.

And a very special thank you is extended to my family for their love, support, and encouragement.
TABLE OF CONTENTS

Abstract ..................................................................................................................................................iii

Acknowledgements ...............................................................................................................................iv

Table of Contents ....................................................................................................................................v

List of Figures ..........................................................................................................................................viii

List of Tables ..........................................................................................................................................xii

CHAPTER 1 .............................................................................................................................................. 1

Design and Preparation of β-sheet Forming Repetitive and Block-copolymerized Polypeptides by Genetic Engineering .................................................................................................................. 1

1.1 Introduction ....................................................................................................................................... 1

1.1.1 Design of genetically engineered protein-based polymers ......................................................... 2

1.1.2 Applications of genetically engineered protein-based polymers ................................................ 5

1.1.3 Cloning methods for genes coding repetitive polypeptides ......................................................... 6

1.2 Results and Discussion ..................................................................................................................... 11

1.2.1 Design of polypeptide repetitive unit ......................................................................................... 11

1.2.2 Construction of libraries of repetitive polypeptides .................................................................... 13

1.2.2 Expression, purification and post translational chemical modification ........................................ 19

1.2.2.1 First generation libraries: aggregation during the expression ................................................. 19

1.2.2.2 Second generation libraries: introduction of amphiphilic character .................................... 20

1.3 Conclusions ...................................................................................................................................... 22

1.4 Experimental Section ........................................................................................................................ 23

1.4.1 Materials ..................................................................................................................................... 23

1.4.2 Design of synthetic DNA for monomers and adaptive sequences .............................................. 24

1.4.3 Preparation of recipient vectors .................................................................................................. 24

1.4.4 Construction of the first generation repetitive polypeptide libraries ............................................ 25

1.4.4.1 Concatenation of 1 (8F) and 2 (16YF): .................................................................................. 25

1.4.4.2 Expression vectors construction ............................................................................................... 26

1.4.5 Construction of the second generation repetitive polypeptide libraries ...................................... 26

1.4.5.1 Assembly of 3 (8Y), 4 (8E), 5 (8H) and 6 (8K) for construction of repeating units 24a (16YE), 25a (16HK), 16YK ........................................................................................................ 26

1.4.5.2 Assembly of 24a (16YE), 25a (16HK) for construction repeating units 26a (32YEHK), 26b (32HKYE) .............................................................................................................................. 27
1.4.5.3 Concatenation of \(26a\) (32YEHK), \(26b\) (32HKYE) and construction of the expression vectors for low molecular weight polypeptides ....................... 27
1.4.5.4 Concatenation of \(26a\) (32YEHK) and construction of expression vectors for high molecular weight polypeptides .................................................. 27
1.4.5.5 Construction of (YE)n, (HK)n and their block-copolymerization and concatenation ........................................................................................................ 28
1.4.6 Expression and purification of repetitive polypeptides ................................................................. 29
1.5 References .................................................................................................................. 31
CHAPTER 2 ..................................................................................................................... 38
Characterization of the Genetically Engineered Repetitive Polypeptide YEHK21 .... 38
2.1 Introduction ................................................................................................................ 38
2.1.1 \(\beta\)-Sheet folding, aggregation and amyloid diseases ............................................ 38
2.1.2 Characterization methods of amyloid systems .................................................... 40
2.2 Results and Discussion ............................................................................................... 43
2.2.1 Peptide isolation and functionalization ............................................................... 43
2.2.2 Characterization of YEHK21 by far-UV CD and deep UV resonance Raman spectroscopies................................................................. 45
2.2.3 Characterization of YEHK21 by AFM ............................................................... 47
2.2.4 Characterization of YEHK21 by TEM ............................................................... 51
2.2.5 Fibril formation ................................................................................................... 52
2.3 Conclusions ................................................................................................................ 55
2.4 Experimental Section ................................................................................................. 56
2.4.1 Fluorescence, UV absorption and CD spectroscopy ........................................... 56
2.4.2 Resonance Raman measurements ........................................................................ 56
2.4.3 Atomic force microscopy (AFM) measurements ................................................ 56
2.4.4 Transmission electron microscopy (TEM) measurements .................................. 57
2.5 References .................................................................................................................. 58
CHAPTER 3 ..................................................................................................................... 67
Applications of Genetically Designed Repetitive Polypeptides............................... 67
3.1 Introduction ................................................................................................................ 67
3.1.1 Experimental shortcut to the protein amyloidoisis .............................................. 67
3.1.2 Macromolecular systems derived from genetically engineered repetitive polypeptides................................................................. 67
3.1.3 Amyloidosis of intrinsically disordered proteins and peptides ....................... 68
3.1.4 Role of electrostatic interactions on the aggregation of peptides and proteins.. 69
3.1.5 \(\beta\)-Sheet folding by multi-stranded peptide models ........................................ 70
3.2 Results and Discussion

3.2.1 YE8 as a model for amyloidogenic intrinsically unfolded proteins
3.2.1.1 Design of YE8
3.2.1.2 Preparation of the initial YE8 sample
3.2.1.3 pH dependence of folding and aggregation
3.2.1.4 The effect of ionic strength on folding and aggregation
3.2.1.5 Concentration dependence and molecular template effect
3.2.1.6 YE8 as a model for a completely amyloidogenic protein
3.2.1.7 The mechanism of the YE8 polypeptide β-sheet folding and fibrillation

3.2.2 Charge Distribution and Amyloid Fibril Formation
3.2.2.1 Preparation of the YEHK21 samples
3.2.2.2 The folding of YEHK21
3.2.2.3 AFM characterization of YEHK21
3.2.2.4 Folding and aggregation of YEHK21 and YE8 in highly charged states
3.2.2.5 Role of electrostatic interactions and charge distribution in YEHK21 and YE8
3.2.2.6 Implications to the deposition of intrinsically unfolded proteins

3.2.3 Effect of a β-strand number
3.2.3.1 Design of the polypeptide family
3.2.3.2 Sample preparation
3.2.3.3 Folding
3.2.3.4 Fibril formation
3.2.3.5 Gelatinous phase formation
3.2.3.6 Polymorphism in YEHKn aggregation
3.2.3.7 Role of hydrophobic interactions. Model of the type II fibril
3.2.3.8 Amyloid fibril “sticky ends”

3.2.4 Structural characterization of amyloid-like fibrils

3.3 Conclusions

3.4 References
LIST OF FIGURES

Figure 1. Schematic representation of the organization of amorphous and crystalline domains in silk fibers. Silk-based design of peptide sequence ((AG)₃EG)n is encircled. Adapted from ref. 21.

Figure 2. (A) Schematic illustration of the design of a combinatorial library of de novo proteins containing six β-strands (arrows) punctuated by turns. (B) Designed binary sequence pattern. (C) Amino acid sequences (single letter code) of 17 de novo proteins from the combinatorial library. Adapted from ref. 26.

Figure 3. Generation of double-stranded DNA cassette encoding sequence of monomer. (A) by hybridization; (B) by hybridization of partially complementary sequences followed by dNTP filling-in and restriction enzyme digestion; (C) a hybrid formed by partially complementary sequences can be used as a PCR template. Adapted from ref. 55.

Figure 4. Illustration of in vivo amplification of double-stranded DNA, concatenation of monomers and ligation of DNA oligomer into plasmid. Adapted from ref. 55.

Figure 5. The molecular biology steps of RDL. (A) A synthetic monomer gene is inserted into a cloning vector. (B) The gene is designed to contain recognition sites for two different restriction endonucleases, RE1 and RE2, at each of the coding sequence. (C) An insert is prepared by digestion of the vector with both RE1 and RE2 and subsequently ligated into the vector that has been linearized by digestion with only RE1. (D) The product contains two head-to-tail repeats of the original gene, with the RE1 and RE2 sites maintained only at the ends of the gene. (E) Additional rounds of RDL proceed identically, using products from previous rounds as starting materials. Adapted from ref. 57.

Figure 6. A) General design of GAGAGAGE homopolypeptide, B) Peptide fragments coded by smallest DNA building blocks (strand GAGAGA + turn GX).

Figure 7. A) Peptide fragments encoded by the smallest DNA building blocks and comprised of GAGAGA strand and GX turn. B) Design of repetitive units YKYE, YEHK, YEYE, HKHK.

Figure 8. Synthetic DNA sequences for minimal coding and adaptive units.

Figure 9. Schematic representation of the construction of the repetitive and block-copolymerized polypeptide libraries.

Figure 10. Construction of repetitive DNA sequences and insertion of DNA multimers into the expression vector.

Figure 11. Oligomerization with adaptive sequences, agarose gel electrophoresis, and enrichment of longer adapted oligomers. Lane 1: 1000 bp molecular marker ladder, lane 2: high 200 bp molecular marker ladder, lane 3: C2-32YEHKn-H6 oligomers. Enclosed triangular area was purified and used for cloning.
Figure 12. The extent of DNA oligomerization at two subsequent concatenation steps.\textsuperscript{62}

Figure 13. Western-blot analysis of expression of A2-16YF4 (23) and A2-16YF4 (23). Lanes 1-4 (A2-16YF4 (23)) and lanes 6-9 (A2-16YF4 (23)): before induction, 30 min, 1 h and 2 h after induction, lane 5: molecular weight marker (29 kD protein shows the signal).\textsuperscript{62}

Figure 14. 15% SDS PAGE of purified short repetitive polypeptides H6-YEHKn-H6 (27a-h) and H6-HKYEn-H6 (29a-d). Lanes 1-8: 27a-h, lane 9: molecular weight marker (from the top, 66, 45, 36, 29, 24, 20, 14.2 kD), lanes 10-13: 29a-d, respectively.\textsuperscript{62}

Figure 15. 8% SDS PAGE of intact and carbamylated repetitive polypeptides H6-(HK8YE8)n-H6 (n=1-7, 39a-g). Lanes 1-7: 39a-g, lanes 8: molecular weight marker (from the top, 205, 116, 97, 84, 66, 55, 45, 36 kD, lane 9-15: carbamylated 39a’-g’, respectively.\textsuperscript{62}

Figure 16. Structural model of repetitive YEHK unit. The β-sheet is formed by Ala and Gly residues (shown backbone only). The turns are functionalized by Tyr, Glu, His and Lys residues forming the arrays at the edges of β-sheet.\textsuperscript{43}

Figure 17. 8% SDS PAGE gel electrophoretic analysis of induction, purification, and modification of YEHK21. lane 1, before induction; 2, 4 h induction; 3, whole cell lysate; 4, flow through 1; 5, flow through 2; 6, 10 mM; 7, 20 mM; 8, 300 mM; 9, 500 mM imidazole 8 M urea PBS; 11, 300 mM fraction heated at 98 ºC for 1 min, 12, 2 min; 13, 3 min, 14, 4 min, 15, 5 min, 16, 7 min, 17, 10 min, 18, 20 min; 19, 60 min, 20, 180 min. Lanes 10,21, molecular weight marker (SIGMAMARKER Wide Molecular Range).\textsuperscript{70}

Figure 18. Far-UV CD spectra of the solution YEHK21 polypeptide after dialysis.\textsuperscript{45}

Figure 19. 195-nm excitation RR spectra of YEHK21 backbone obtained at 22 ºC (blue curve) and 90 ºC (green curve) and poly-L-lysine (PLL) β-sheet (red curve) and PLL random coil (black curve) Spectra recorded by Dr. Vladimir V. Ermolenkov.\textsuperscript{70}

Figure 20. Tapping mode-AFM topograph of YEHK21 fibrils on HOPG.\textsuperscript{48}

Figure 21. Probability density plot of fibril thickness derived from AFM topographic data of YEHK21 fibrils. 79 samplings were employed to generate the distribution profile.\textsuperscript{49}

Figure 22. YEHK repeat units with g-turns forming antiparallel β-sheet structure. Width (turn-strand turn) of the computed model, 3.4-3.8 nm; thickness, 0.72 nm.\textsuperscript{49}

Figure 23. The implication of the turn motif on amphiphilicity of the polypeptide surface. Left: Formation of surfaces with equal hydrophobicity employing β-turn motive. Right: Formation of surfaces with different hydrophobicity employing γ-turn motive.\textsuperscript{50}

Figure 24. TEM micrographs of YEHK21 fibrils on carbon-coated Cu grids. The measurements were recorded by Dr. Kathleen Dunn.\textsuperscript{70}
Figure 25. Probability density plot of fibril width derived from TEM of YEHK21 fibrils. 161 samplings were employed to generate the distribution profile. 70

Figure 26. AFM topography of assembled domains of YEHK21 on HOPG. 53

Figure 27. In a tentative structural model for the most common fibrils observed both the dimensions of those fibrils and amphiphilic nature of proposed anti-parallel β-sheet can be accommodated by the interaction of four polypeptide molecules. 54

Figure 28. The β-hairpin motif, in which two antiparallel strands are connected by a short loop. The most common loop size is two residues (n = 1). R represents amino acid side chains. Dashed lines represent hydrogen bonds. 45

Figure 29. Peptide β-sheet folding models. Schematic representation illustrating cooperativity in model β-sheets propagated (top) along a β-hairpin peptide from the turn (parallel with the β-strands) and (bottom) perpendicular to the strand direction as the docking of a third strand enhances the stability of the existing hairpin template. 46

Figure 30. Structural model of YE peptide construct. The β-sheet is formed by Ala and Gly residues (shown backbone only). The turns are functionalized by Glu and Tyr residues forming arrays at the edges of β-sheet. 73

Figure 31. DUVRR spectra of YE8 at pH 2 (left panel) and pH 3.5 (right panel) incubated at room temperature and recorded in 0, 2, 3, 13, 27 and 56 days after the preparation. Excitation with quasi-CW radiation at 197 nm. For comparison, all Raman spectra were normalized to the intensity of the strongest tyrosine Raman band at 1618 cm⁻¹. Spectra recorded by Dr. Vladimir V. Ermolenkov. 75

Figure 32. Characterization of YE8 samples incubated for two month at different pH employing A) Far-UV CD, B) fluorescence and C) UV-vis spectroscopes. 76

Figure 33. Comparison of YE8 polypeptide aggregation at different conditions. TM-AFM topographs on HOPG and TEM micrographs on carbon-coated Cu grids of YE8 samples incubated at different pH: A) and D) at pH 2, B) and E) at pH 3.5, C) and F) at pH 5 respectively. 78

Figure 34. DUVRR spectra of 44 µM YE8 incubated at room temperature for 1, 7, 14, 21, 35 and 50 days after the sample preparation (solid blue curves), results of their fitting with a linear combination of basis spectra (dotted red curves) and their differences (solid black curves). Folded YE8, unfolded YE8 and tyrosine spectra were used as basis spectra. The experimental spectra were recorded at 197 nm with quasi-CW excitation. 80

Figure 35. Kinetics of YE8 folding and the relative proportions of the YE8 polypeptide in the folded and unfolded forms obtained for the samples with different concentrations of 9, 12, 15, 25, 35 and 44 µM as well as for the samples subjected to stirring or seeding with folded YE8. 81

Figure 36. Proposed mechanism of YE8 aggregation driven β-sheet folding. 83

Figure 37. The 197-nm excited Raman spectra of YEHK21 dialyzed against water at pH 2.0, pH 3.5, pH 5.0, pH 6.5 and pH 10 (blue), and spectra of gelatinous
fractions obtained by sample centrifugation and supernatant decantation (red). Spectra recorded by Vitali Sikirzhitsky.

**Figure 38.** TM-AFM topographs of YEHK21 aggregates at A) pH 5 and B) pH 6.5.

**Figure 39.** The 197-nm excited Raman spectra of YEHK21 incubated in solution with pH 2 (red), its gelatinous (blue) and aqueous (green) fractions. Spectra recorded by Vitali Sikirzhitsky.

**Figure 40.** The 197-nm excited Raman spectra of YE8 incubated in solution with pH 6.5 (red), its gelatinous (blue) and aqueous (green) fractions. Spectra recorded by Dr. Vladimir V. Ermolenkov.

**Figure 41.** TM-AFM topographs of YEHK21 fibrils at pH 2 (A) and YE8 at pH 6.5 (B).

**Figure 42.** Schematic representation of charge distribution in YEHK21 and YE8 at different pH. The square illustrates unit containing one strand and turn; the ovals stand for unfolded hexahistidinyl tracks. The red, blue and white colors represent the positively charged, negatively charged or neutral units.

**Figure 43.** Schematic representation of the YEHK7, YEHK14 and YEHK21 polypeptides.

**Figure 44.** Resonance Raman spectra of S0 solutions of YEHK7 (blue curve), YEHK14 (red), YEHK21 (black). Spectra recorded by Dr. Vladimir V. Ermolenkov.

**Figure 45.** TM-AFM topographs of YEHK7 (A, D), YEHK14 (B, E) and YEHK21 (C, F) fibrils.

**Figure 46.** Resonance Raman spectra of YEHKn polypeptides. Initial solution after dialysis (green), its aqueous fraction (F1, red) and solution with rendered aggregated polypeptides (F2, blue). Spectra recorded by Dr. Vladimir V. Ermolenkov.

**Figure 47.** AFM topographs of the YEHK7 (A), YEHK14 (B) and YEHK21 (C) aggregates rendered from gelatinous phase.

**Figure 48.** The models of the polypeptide chain organization. One color represents one polypeptide chain. A) Fibrils with pronounced beaded morphology, B) very smooth and linear fibrils, C) AFM topograph of the YEHK21 fibrils.

**Figure 49.** Proposed anti-parallel β-sheet formed by four polypeptide molecules.

**Figure 50.** A proposed model of a monomer addition.

**Figure 51.** Hypothetical scheme of the YEHK fibril cross-β core.

**Figure 52.** TEM micrographs of YEHK fibrils on carbon coated Cu grids.

**Figure 53.** DUVRR spectra of YEHK fibrils. Spectra recorded by Dr. Vladimir V. Ermolenkov.
Figure 54. The Raman signatures of a β-sheet and turns obtained from the differential spectra of 32YEHK7, 40YEHK7 and 48YEHK7 polypeptides. Spectra obtained by Vitali Sikirzhitsky.

LIST OF TABLES

Table 1. Predicted charges of YE8 polypeptide, the β-sheet core and different amino acids at different pH values.\textsuperscript{a} .......................................................................................................................... 74

Table 2. Predicted charges of modified YEHK21 and YE8 at different pH.\textsuperscript{a} ................. 89

Table 3. The average width of amyloid fibrils formed by polypeptides with different repetitive units. ........................................................................................................... 104
CHAPTER 1

Design and Preparation of β-sheet Forming Repetitive and Block-copolymerized Polypeptides by Genetic Engineering

1.1 Introduction

Genetic engineering of polypeptide chains and utilization of living cells as a factory for the artificial protein production have become powerful tools in modern science and biotechnology. Genetically engineered artificial proteins have been widely utilized for the development of new protein-based materials and as model systems for the investigation of protein folding mechanisms, molecular recognition, and self-assembly. Genetically engineered protein-based polymers can be prepared with precise control of the sequence, molecular weight, and stereochemistry.\textsuperscript{1-5} Employing native forces, biopolymers can self-assemble into monolayers, hydrogels, or crystalline materials with appealing mechanical and biophysical properties. Programmed incorporation of functional domains at specific location in a biopolymer facilitates the development of environmentally responsive materials, with biorecognition and controlled crosslinking abilities. A unique aspect of genetically engineered polymeric materials is that the preparation of the bio-template, the coding DNA sequence, is required only once. Subsequent biosynthetic processes obviate the need for difficult or complex synthetic manipulations.

In our laboratory we have been interesting in the design, synthesis and characterization of water soluble, repetitive, β-sheet forming polypeptides. In particular, the focus of our research is the investigation of the effect of polypeptide architecture (repetitive sequence, degree of polymerization and copolymerization) on the physical
properties of polypeptides. Thus, the design of polypeptide sequence and the development of cloning methods for the rapid construction of gene libraries of genes coding repetitive polypeptides are crucial steps of this project.

1.1.1 Design of genetically engineered protein-based polymers

*Elastin-like polymers (ELPs)* have been designed employing the repetitive domain of the extracellular matrix protein elastin. The majority of studies were completed with the VPGXG pentapeptide motif where X can be any amino acid but proline.\(^6,7\) Such polymers can be soluble in aqueous solutions below the invert transition temperature \(T_t\) while a sharp phase transition leads to the rapid aggregation of polypeptides above this temperature. The design of *silk-like polymers (SLPs)* is based on silk proteins produced by spiders and other insects. These proteins contain repetitive sequences of crystalline and amorphous domains.\(^8\) The design of the majority of SLPs is based on *B. mori* fibroin with the repetitive sequence \([GAGAGSGAAG][SGAGAG]_Y\] and spidroin 1 and spidroin 2 from *N. clavipes* \([GGAGQGGYGGGLGSQAGRGGLGQGGAG]\)\(^9\) and \([GPGGYGGPGQGYPGYPVPQPSGPGS],^{9-11}\) respectively. Most SLPs are poorly soluble in aqueous media and not well characterized. *Silk-elastin-like block copolymers* consist of blocks of silk-like and elastin-like polymers with general representation \(\{(SLP)_a(ELP)_b\}_c\). In contrast to SLPs, silk-elastin-like block copolymers are soluble and have been well characterized. The biophysical properties of silk-elastin-like block copolymers are predetermined by the length of the SLP and ELP regions within the block copolymer.\(^12,13\) Increasing the number of the ELP blocks facilitates solubility of a
copolymer while the increasing of the number of the SLP domains results in the formation of a gelatinous phase.\textsuperscript{14}

\textit{β-Sheet forming polymers are} a family of polymers that can adopt an anti-parallel β-sheet structure and subsequently self-assemble in very regular aggregates. Several groups have reported on the preparation of high molecular weight polypeptides\textsuperscript{3,15-29} employing two principal approaches to the polypeptide design. The first approach is based on the oligomerization of the predefined repeating sequences adapted from β-sheet folded proteins. Following this method Tirrell has utilized the poly(alanine-glycine) (poly(AG)) motif from silk fibroin and created a family of β-sheet forming homopolymers.\textsuperscript{15,16,18,19,30-33} The design of the polypeptide sequence ((AG)\textsubscript{3}EG)\textsubscript{n}\textsuperscript{16,17} employing crystalline domains of silk fibers is shown on Figure 1.

\begin{figure}[h]
\centering
\includegraphics[width=0.5\textwidth]{schematic.png}
\caption{Schematic representation of the organization of amorphous and crystalline domains in silk fibers. Silk-based design of peptide sequence ((AG)\textsubscript{3}EG)\textsubscript{n} is encircled. Adapted from ref. 21.}
\end{figure}

Several polypeptide series including [(AG)mEG]\textsubscript{n}\textsuperscript{17} (m, n =3-6) and [(AG)\textsubscript{3}XG]\textsubscript{n} families\textsuperscript{34,15} (X = N, F, S, V, or Y) were designed and applied to study the effect of the polymer repetitive unit on a polypeptide crystal structure. X-ray diffraction of samples
obtained by crystallization of the polymers from aqueous formic acid solution revealed that they form a highly ordered, lamellar structure where β-sheet strands are connected through the γ turns\textsuperscript{16,35} and organized in an anti-parallel fashion.

A second approach to the design of β-sheet forming polypeptides relies on binary patterning of hydrophobic and hydrophilic amino acids to create β-strand sequences. The alternation of glutamic acid (E) or lysine (K) and alanine (A) residues were employed for the design of the β-sheet folded AEAEAKAK sequence.\textsuperscript{36} More general elaboration of this approach was developed by Hecht\textsuperscript{23,25,26,28} who created a combinatorial library of β-sheet forming polymers by binary patterning of random hydrophobic and hydrophilic amino acids in β-strand sequences while consistently employing a specified amino acid sequence for the turn induction at the desired position (Figure 2).

**Figure 2.** (A) Schematic illustration of the design of a combinatorial library of de novo proteins containing six β-strands (arrows) punctuated by turns. (B) Designed binary sequence pattern. (C) Amino acid sequences (single letter code) of 17 de novo proteins from the combinatorial library. Adapted from ref. 26.
1.1.2 Applications of genetically engineered protein-based polymers

Rational design of polypeptide-based polymers may be desirable for physical and biomedical applications. The possibility of the controlled release of biologically active agents, for drug or gene delivery was demonstrated with several genetically engineered biopolymers. These polypeptides, soluble in aqueous media, can easily be injected. The formation of a viscoelastic gelatinous phase under physiological conditions occurs in a time dependent and stimuli responsive manner. Biopolymers may also be employed in materials design for tissue culture and tissue engineering. Protein engineering enables the production of materials where the programmable viscoelastic properties of protein polymers can be coupled with cell recognition sites. Such biomaterials can be used for the replacement of extracellular matrix proteins, providing the necessary structure, attachment sites, and support for cell culture and tissue regeneration. Fusion of an elastomeric protein-based domain with a desired protein was employed for the purification of proteins by simple centrifugation. The inverse temperature transitional properties of these elastomeric protein-based polymers were central to their separation. Rationally designed polypeptide-based polymers may fold via established rubrics forming specific secondary structures. The construction of β-sheet forming polypeptides for the investigation of β-sheet folding, aggregation, and amyloid formation is another application. The self-assembly of folded molecular building blocks into functional protein-based nanomaterials is of a great interest in modern nanotechnology where precise dimensional control and selective functionalization are important.
1.1.3 Cloning methods for genes coding repetitive polypeptides

The most common strategy for the bio-synthesis of repetitive polypeptides requires the chemical synthesis of the DNA sequences that are utilized for the construction of double-stranded DNA cassettes, subsequent enzymatic oligomerization of these cassettes, construction of an appropriate expression vector, and expression of polypeptide in a suitable cellular system. Double-stranded DNA cassettes can be constructed following a few different approaches (Figure 3).

![Diagram of cloning methods](Figure 3)

**Figure 3.** Generation of double-stranded DNA cassette encoding sequence of monomer. (A) by hybridization; (B) by hybridization of partially complementary sequences followed by dNTP filling-in and restriction enzyme digestion; (C) a hybrid formed by partially complementary sequences can be used as a PCR template. Adapted from ref. 55.

In the hybridization approach, (Figure 3A), a DNA cassette is created by the duplex DNA encoding for the desired amino acid sequence and flanked with appropriate restriction sites at both termini. Two other approaches toward the DNA cassette
assembly employ hybridizations of two synthetic DNA sequences, followed by polymerase filling\textsuperscript{54} (Figure 3B) or the polymerase chain reaction\textsuperscript{6} (PCR) (Figure 3C) with subsequent enzymatic digestion. The DNA manipulations require a substantial amount of the double-stranded DNA cassettes, so amplification of DNA is commonly effected in \textit{in vivo} systems.\textsuperscript{14,53} Figure 4 illustrates the steps involved in the amplification procedure.

\textbf{Figure 4.} Illustration of \textit{in vivo} amplification of double-stranded DNA, concatenation of monomers and ligation of DNA oligomer into plasmid. Adapted from ref. 55.

First, double-stranded DNA is inserted into linear plasmid DNA with compatible ends. The plasmid with the monomer sequence is then inserted into \textit{E. coli}. Upon
positive clone selection, the necessary double-stranded DNA cassette is recovered by the enzymatic digestion of the plasmids harvested from cells. The most difficult step in the production of genetically engineered repetitive polypeptides is the development of rapid, precise and flexible ways to assemble the DNA cassettes. Two different approaches were suggested: concatenation or “ex plasmid” gene assembly and iteration or “in plasmid” gene assembly.

**Concatenation or “ex plasmid” gene assembly** has been used to create linear head-to-tail self-ligated DNA polymers.\(^\text{14}\) It is a reliable method for the one-step creation of libraries of homopolymers comprised of a variable number of DNA monomers. However the concatenation has several disadvantages: i) the degree of polymerization is not well-controlled, ii) linear head-to-tail self-ligation is not possible for the controlled synthesis of block-copolymers or for variation in codon usage in highly repetitive systems, iii) oligomer cyclization competes with the synthesis of linear ligation products. The cyclization of the growing oligomer can be obviated by the method developed by Urry\(^\text{43}\) where special pendent DNA sequences are incorporated in the ligation reaction mixture to serve as “chain terminators“ that promote formation of linear polymers with desired molecular weight. Unfortunately, the yield of high molecular weight concatemers is often low. Additional concatenation steps are usually required for the production of high molecular weight biopolymers.

During **iterative or “in plasmid” gene assembly** the repetitive blocks are sequentially assembled within a plasmid by the repeated insertion of a monomer resulting in a controllable increase of the DNA molecular weight.\(^\text{54,56}\) The example of gene assembly utilizing the “in plasmid” approach is shown on Figure 5. This outlines
the application of Recursive Directional Ligation (RDL) method\textsuperscript{57} for the construction of elastine-like polypeptides with precise molecular weight.

**Figure 5.** The molecular biology steps of RDL. (A) A synthetic monomer gene is inserted into a cloning vector. (B) The gene is designed to contain recognition sites for two different restriction endonucleases, RE1 and RE2, at each of the coding sequence. (C) An insert is prepared by digestion of the vector with both RE1 and RE2 and subsequently ligated into the vector that has been linearized by digestion with only RE1. (D) The product contains two head-to-tail repeats of the original gene, with the RE1 and RE2 sites maintained only at the ends of the gene. (E) Additional rounds of RDL proceed identically, using products from previous rounds as starting materials. Adapted from ref. 57.

A combination of both approaches to repetitive gene assembly was employed in the preparation of elastin-like block-copolymers.\textsuperscript{37} Following the encoding of hydrophilic and hydrophobic blocks in separate DNA cassettes, the blocks were concatenated independently and then joined using “in plasmid” gene assembly. When hydrophilic and hydrophobic blocks were introduced to the plasmid via sequential
ligation, the block-copolymers formed had the desired head-to-tail structure. Although this approach guarantees the correct fusion order in the produced block-copolymers it requires extensive cloning.

Most of the latest work on the repetitive sequence construction employs a seamless cloning method,\textsuperscript{36,58} utilizing the type IIs restriction endonucleases which effect DNA cleavage at specified distance from the recognition site. This approach has several advantages: i) the uni-directional ligation and insertion into plasmid is possible, ii) the insertion of extraneous sequence at the junction between the desired monomer sequences is avoided, and iii) the monomer sequence is liberated from restriction site sequence constraints.
1.2 Results and Discussion

1.2.1 Design of polypeptide repetitive unit

The first approach discussed in Introduction, the utilization of sequences adapted from β-sheet folded proteins, was chosen for the design of the β-sheet forming polypeptides. However, in contrast to previously reported work, several sequences were employed for the construction of each polypeptide resulting in the synthesis of various polymers and block-copolymers. The amino acid sequences encoded by monomeric DNA building blocks were adapted from the work of Tirrell,\textsuperscript{15-17,19,33,59-61} where the GAGAGA sequence was utilized for the β-strand forming motif and pendant amino acids X were located at the turns (Figure 6A).

![Figure 6](image)

**Figure 6.** A) General design of GAGAGAGE homopolypeptide, B) Peptide fragments coded by smallest DNA building blocks (strand GAGAGA + turn GX).

The β-sheet forming polypeptides were based on the hierarchical assembly of a small number of monomeric building blocks coding (GA)\textsubscript{3}GX sequences, comprised of a single β-strand GAGAGA and turn functionalized by specific amino acid residue (Figure 6B). This design strategy affords flexible variation of repetitive unit sequence while
permitting precise control over any particular turn or a strand position within the repetitive polypeptide.

**Design of the first generation polypeptide libraries.** Our initial experiments reproduced literature precedent\textsuperscript{33} in construction and expression of repetitive homopolymer with phenylalanine (F) residue in turn position (Figure 6A, $X_1^1=X_2^2=X_3^3=X_4^4=F$) and slightly asymmetrically modified heteropolymer with phenylalanine (F) and tyrosine (Y) residues in the turn positions. (Figure 6A, $X_1=X_3=Y$, $X_2=X_4=F$). It was determined that the hydrophobicity of these peptides imparted a profound tendency to aggregate.

**Design of the second generation polypeptide libraries.** To improve the solubility of the polypeptides more hydrophilic amino acids were incorporated at the turn sites to introduce amphiphilicity. The amphiphilic polypeptide design enabled the introduction of separate hydrophilic and hydrophobic edges to the $\beta$-sheet to enhance water solubility. Starting with the four building blocks (8Y, 8E, 8H, 8K where each coding block consists of a $\beta$-strand forming GAGAGA sequence and a turn sequence GX, Figure 6B) the heterogeneous unit YEHK was assembled. Negatively charged glutamate (E) and protonated lysine (K) were introduced to generate a hydrophilic edge, where stabilizing salt bridges could form. The aromatic tyrosine (Y) and histidine (H) residues were postulated to form an intramolecular $\beta$-stacked hydrophobic array.

Assembly of the DNA coding repetitive unit YEHK from separate building blocks coding 8Y, 8E, 8H and 8K (Figure 7a) enabled the examination of the impact of the systematic displacement of the turn residues, e.g., creation of YKYE where histidinyl (H) turn residues of the YEHK repeat are replaced with tyrosines (Y).
Figure 7. A) Peptide fragments encoded by the smallest DNA building blocks and comprised of GAGAGA strand and GX turn. B) Design of repetitive units YKYE, YEHK, YEYE, HKHK.62

Deletion of the lysine residues from YKYE forms the simple di-strand repeat 16YE which results in an unbalanced negative charge in the resultant polypeptides. By analogy, the HK peptide presents histidine (H) residues along one edge and array of protonated lysines (K) on the other (Figure 7b).

1.2.2 Construction of libraries of repetitive polypeptides

Each β-sheet forming polypeptide was derived from the assembly of a small number of minimal DNA cassettes encoding the elementary constructs for a single strand and turn. Figure 8 shows the synthetic DNA sequences employed for the construction of DNA cassettes encoding the minimal amino acid sequence (cassettes 1-6) and adaptive DNA cassettes utilized for DNA manipulation during libraries construction (cassettes 7-14).
Figure 8. Synthetic DNA sequences for minimal coding and adaptive units.62
The overall procedure for the combination of genetic cassettes encoding various combinations of individual β-sheet strands and turns is described in Figure 9.

**Figure 9.** Schematic representation of the construction of the repetitive and block-copolymerized polypeptide libraries.\(^{62}\)

**Building block assembly and oligomerization.** Oligonucleotide units were assembled based on unidirectional, head-to-tail oligomerization\(^{14,30,57}\) with adaptive DNA sequences for incorporation at the appropriate cloning sites (Figure 10).\(^{14,30,43}\) Each of the minimal
DNA coding units was joined at BanI sites allowing unidirectional concatenation and reduction of concatenated units to minimal coding units. Inclusion of the adaptive DNA sequences, utilized for intermediate cloning steps and for construction of the expression vectors, allows cloning at virtually any restriction site with common, conventional, commercially available cloning vectors.

**Figure 10.** Construction of repetitive DNA sequences and insertion of DNA multimers into the expression vector.\(^{62}\)

*The adaptive DNA for cloning.* A1 adaptive DNA (15) contains the recognition sites for the type IIs restriction endonuclease, _BsaI_, where cloning at both the 3’ and 5’ terminii leads to facile recovery of assembled DNA units, facilitating further recursive, seamless oligomerization and block copolymerization.\(^{36,43,58,63,64}\) Self-ligation of the adaptive sequences during the concatenation was suppressed by selective phosphorylation of 5’-hydroxyl groups of synthetic adaptive DNA strands that constitute the termini of the oligomerized repetitive DNA units (7b, 8a, 9b, 10a, 12b, 13a, and 14b). Oligomerized
DNA containing the adaptive sequences was purified by agarose gel electrophoresis. The bands containing the desired adapted oligomers were enriched as shown in Figure 11 and then were cloned into recipient vectors.

**Figure 11.** Oligomerization with adaptive sequences, agarose gel electrophoresis, and enrichment of longer adapted oligomers. Lane 1: 1000 bp molecular marker ladder, lane 2: high 200 bp molecular marker ladder, lane 3: C2-32YEHKn-H6 oligomers. Enclosed triangular area was purified and used for cloning.\(^62\)

**Adaptive DNA for expression.** Adaptive DNA for expression was utilized at the last concatenation step to terminate the oligomerization and to enable the insertion of the construct into the desired expression vector. Simultaneously, the appropriate fusion amino acid sequences at both the head and tail of the repetitive and block-copolymerized polypeptides were introduced thereby enabling rapid shotgun library construction with various terminal amino acids (Figure 9b). Adaptive DNA sequence \(16\) was designed to facilitate the preparation of previously published polypeptides\(^{17,19,33,60}\) by cloning into the pET-3a or pET-9a vectors. Resulting polypeptides have the repetitive sequence fused with a 4.9 kDa domain that could be easily removed enabling recovery of the purified repetitive polypeptide sequence. Adaptive sequence \(17\) was designed to introduce a hexahistidinyl track not only to facilitate polypeptide purification but also to terminate
the β-sheet construct with a specific terminal sequence. Sequence 18 allowed
differentiation of the C- and N- terminus with specific sequences, e.g., a hexahistidinyl
track at the C-terminus and a Cys-Cys repeat at N-terminus.

**Extent of Oligomerization.** In contrast with previously described methods, the sequential oligomerization of the DNA product recovered from the first concatenation step was effected in the presence of the appropriate adaptive DNA sequence prior to insertion into an expression vector (Figure 9c and Figure 12). This approach permits a second level of concatenation. In a single step, a library of repetitive polypeptides can be formed bearing the appropriate C- and N-terminal fusion sequences where there is a gradual increase in molecular weight e.g., 1X, 2X, 3X, 4X, 5X (where X is the selected concatemer from previous oligomerization). This approach affords a diversity of oligomers not accessible by arithmetic synthesis based on ligation of two blocks e.g., 1X, 2X, 4X, 8X, 16X series.

![Figure 12](image)

**Figure 12.** The extent of DNA oligomerization at two subsequent concatenation steps.

**Block-copolymerization.** The designed strategy can be utilized for synthesis of polypeptide block-copolymers. The constructs prepared via this step-wise concatenation strategy have distinct domains comprised of the simple motifs (Figure 9d).
1.2.2 Expression, purification and post translational chemical modification

RecA deficient hosts were used for both cloning and expression, since DNA repetitive sequences frequently result in rapid recombination. Rare codon supplemented expression hosts were utilized to maximize the yields of target polypeptides. In most cases, since the product repetitive polypeptides are harmful to the bacterial hosts, i) 1% of glucose was added as a supplement to the agarose plates and preculture media to suppress expression leakage and ii) polypeptide expression was induced at relatively high OD (>1).

1.2.2.1 First generation libraries: aggregation during the expression

During the course of expression of 22 (A2-8F13) and 23 (A2-16YF12) polypeptides the extremely high tendency of these molecules to aggregate was observed. Figure 13 illustrates the time course of polypeptide expression revealed by Western Blotting analysis. Within 2 hours of expression, the majority of the expression products aggregated and remained in the stacking gel wells during the electrophoresis (at the top of the figure).

![Figure 13. Western-blot analysis of expression of A2-16YF4 (23) and A2-16YF4 (23). Lanes 1-4 (A2-16YF4 (23)) and lanes 6-9 (A2-16YF4 (23)): before induction, 30 min, 1 h and 2 h after induction, lane 5: molecular weight marker (29 kD protein shows the signal).]

62
Once aggregation occurred, 22 (A2-8F13) and 23 (A2-16YF12) polypeptides could not be dissociated using denaturing reagents suitable for use with Ni-NTA column chromatography such as urea or guanidine hydrochloride.

1.2.2.2 Second generation libraries: introduction of amphiphilic character

Repetitive and block-copolymerized polypeptides 27, 28, 32 (YEHKn), 29, 30 (HKYEn), 36 (16YEn), 37 (16HKn), 38 ((YE8HK8)n), 39 ((HK8YE8)n), 40, 41 (KYEYn), and 42, 43 (YKYEn) were constructed with various N- and C-terminal sequences by combination of 8Y, 8E, 8H, and 8K coding DNA units. In contrast with 22 (A2-8F13 ) and 23 (A2-16YF12), the second generation polypeptides were the hydrophilic nature and are easily expressed. These peptides were monomeric in 8 M urea or 3 to 6 M guanidine hydrochloride during the purification and remained soluble after dialysis against water.

As an illustration of polypeptide expression and purification, gel electrophoretic analysis of repetitive polypeptides 27a-h (H6-YEHKn-H6) and 29a-d (H6-HKYEn-H6) is shown on Figure 14. Expression and post-translational chemical modification of the repetitive block-copolymers 39a-g (H6-(HK8YE8)n-H6) are depicted on Figure 15. The average yield of purified repetitive polypeptide was about 20-50 mg per 1 L of cell culture. After Ni-NTA purification selected fractions of purified polypeptides were dialyzed against double distilled water and used for characterization.
**Figure 14.** 15% SDS PAGE of purified short repetitive polypeptides H6-YEHKn-H6 (27a-h) and H6-HKYEn-H6 (29a-d). Lanes 1-8: 27a-h, lane 9: molecular weight marker (from the top, 66, 45, 36, 29, 24, 20, 14.2 kD), lanes 10-13: 29a-d, respectively.~\(^{62}\)

**Figure 15.** 8% SDS PAGE of intact and carbamylated repetitive polypeptides H6-(HK8YE8)n-H6 (n=1-7, 39a-g). Lanes 1-7: 39a-g, lanes 8: molecular weight marker (from the top, 205, 116, 97, 84, 66, 55, 45, 36 kD, lane 9-15: carbamylated 39a’-g’, respectively.~\(^{62}\)
1.3 Conclusions

Repetitive polypeptide libraries were designed to fold into an anti-parallel β-sheet structure and to remain soluble in aqueous solution. The sequence of each polypeptide was derived from the assembly of a small number of elementary constructs, each coding for a single β-sheet strand GAGAGA and a turn GX. Such design affords precise control over polypeptide sequence while permitting flexible variation of the polymer architecture.

The strategy for the rapid creation of repetitive polypeptide libraries was developed based on the hierarchical assembly of DNA cassettes. In this approach concatenation or block-copolymerized products were utilized to produce new more complex DNA constructs for additional DNA manipulations. By straightforward extension, combination of various DNA cassettes allows easy construction of various complex yet still oligomeric DNA sequences. The developed strategy retains the basic polypeptide core, yet allows variations at precise positions of repetitive polypeptide affecting such important polypeptide characteristics such as: sequence and degree of polymerization and distribution of charged, hydrophobic and hydrophilic amino acids. Repetitive and block-polymerized polypeptides ranging from 3 to 130 kDa were successfully expressed in E. coli and purified employing Ni\textsuperscript{2+}-NTA affine chromatography with yield at the 20-50 mg/L level.

Part of this work has been published:

1.4 Experimental Section

1.4.1 Materials

Restriction endonucleases and modifying enzymes were purchased from New England Biolabs (Beverly, MA) unless otherwise mentioned. Inoue ultra-competent and electrocompetent cells of XL1-Blue (Stratagene, La Jolla, CA) and DH5αF’ (Invitrogen, Carlsbad, CA), were prepared, transformed, and selected on LB agar plates containing ampicillin (100 µg/mL) with blue-white selection for pUC18 clones or kanamycin sulfate (50 µg/mL) for pET-28 clones according to literature methods. Competent cells of BLR(DE3) and BLR(DE3)pLysS were purchased from Novagen, Inc. (Madison, WI). BLR(DE3)pLysSRARE was prepared from BLR(DE3) transformed by pLysSRARE isolated from Rosetta(DE3)pLysS (Novagen, Inc., Madison, WI) and the standard competent cells were prepared according to literature methods (glycerol was added at 15% to prevent the autolysis upon thawing). Plasmids and DNA fragments separated by agarose gel electrophoresis were purified using QIAprep Spin Miniprep Kit and QIAquick Gel Extraction Kit (Qiagen Inc., Valencia, CA), respectively.

Repetitive polypeptides are represented first by the N-terminal sequence then the repeat unit and finally the C-terminal sequence. The number preceding the repeat sequence refers to the total number of amino acid residues in a single monomer unit and the suffix value indicates the number of repeats. For example, 32YEHK7 and 16YF12 represent 7 repeats of 32-amino acid unit \((GA)_3GY(GA)_3GE(GA)_3GH(GA)_3GK\) and 12 repeats of 16-amino acid unit \((GA)_3GY(GA)_3GF\), respectively. To simplify, the same compound number was used for both the constructed expression vectors and the repetitive polypeptide of that vector product unless specified.
1.4.2 Design of synthetic DNA for monomers and adaptive sequences

Synthetic DNA for coding units and adaptive DNA are listed in Figure 8. Each DNA coding pair 1 (8F), 2 (16YF), 3 (8Y), 4 (8E), 5 (8H), and 6 (8K), respectively, was mixed at 20 mM and the 5’ ends were phosphorylated for 30 min to 1 h at 37 ºC using T4 polynucleotide kinase (New England Biolabs, Beverly, MA) and then annealed with simultaneous deactivation of T4 polynucleotide kinase. The 5’ end of each of the oligonucleotides units 7b, 8a, 9b, 10a, 11b, 12b, 13a, and 14b was phosphorylated for 30 min to 1 h at 37 ºC, respectively, and then the kinase was deactivated at 80 ºC for 20 min. To each solution was added the respective complimentary oligonucleotides 7a, 8b, 9a, 10b, 11a, 12a, 13b, and 14a, at a final concentration of 20 mM. The mixtures were then annealed to give 7 (A1-1), 8 (A1-2), 9 (A2-1), 10 (A2-2), 11 (A3-1-H6), 12 (NdeI-A3-1-H6), 13 (A3-2-H6), and 14 (A3-1-C2), respectively. The resulting ‘internally-phosphorylated’ adaptive sequences were appropriately mixed to give 15 (A1, 7+8), 16 (A2, 9+10), 17 (H6-H6, 11+13), 17’ (NdeI-H6-H6, 12+13), and 18 (C2-H6, 14+13), respectively, at the final concentration of 1 mM (See phosphorylated sites represented as ‘p’ in Figure 8).

1.4.3 Preparation of recipient vectors

pUC18 (10 µg, Bayou Biolabs, Harahan, LA) or pET-28a (1 µg, Novagen, Inc., Madison, WI) was digested by BamHI and EcoRI and the desired fragments were separated by agarose gel electrophoresis, purified using QIAquick Gel Extraction Kit and eluted with TE (pH 8.0) 100 µL to give 19a (pUC18/BamHI-EcoRI) and 40 µL to give 19b (pET-28a/BamHI-EcoRI). One of pET-28a-c vectors (1 µg, Novagen, Inc., Madison, WI) was digested with NcoI and BamHI and then purified in the same manner to give 19c
(pET-28/Ncol-BamHI). pET-9a (1 µg, Novagen, Inc., Madison, WI) was digested with NdeI and BamHI to give 19d (pET-9a/NdeI-BamHI).

1.4.4 Construction of the first generation repetitive polypeptide libraries

1.4.4.1 Concatenation of 1 (8F) and 2 (16YF):

The oligonucleotide solutions 1 (8F) and 2 (16F) (1~2 µL each) were mixed with the adaptive DNA solution 15 in a ratio of 10~100:1 (0.2~2 µL), PEG8000 (final concentration at 4%), and H2O (up to 9 µL), respectively, and annealed at 45 °C for 5 min. Then T4 DNA ligase (1 µL) and the supplied ligation buffer (1 µL) were added, and then system allowed to incubate at 4 °C overnight or longer. The resulting population of oligomers bearing the adaptive sequence at both terminii was subjected to electrophoresis on agarose gel. Oligomerized mixtures with the desired molecular weight were excised from the gel, purified, and precipitated with EtOH in a 0.6 mL microcentrifuge tube. To the each tube was added vector solution 19a (pUC18/BamHI-EcoRI) (2 µL), PEG8000 (final concentration at 4%), T4 DNA kinase (0.25 µL), the supplied kination buffer (0.5 µL), and H2O (up to 5 µL), then the mixture was incubated at 37 °C for 30 min. T4 DNA ligase (0.25 µL) was added at room temperature and the mixture was incubated at 4 °C overnight or longer. The ligation mixture was used to transform XL1-Blue or DH5αF’ cells. Plasmids of selected colonies were isolated and after BamHI digestion, were examined for the appropriate fragment sizes. Plasmids harboring 13 repeats of 8F (20, 8F13) and 4 repeats of 16YF (21, 16YF4) were selected and used for further oligomerization.
1.4.4.2 Expression vectors construction

20 (8F13) and 21 (16YF4) were digested by BsaI, respectively, and fragments of the oligomerized units were isolated, purified, and recursively oligomerized with adaptive sequence 16 as described. The desired oligomer was isolated, purified, ligated with recipient vector 19b (pET-28a/BamHI-EcoRI), and the resultant ligation mixtures used to transform E. coli. Plasmids of selected colonies were isolated and, after NcoI-BamHI digestion, the fragment sizes determined. Plasmids harboring 13 repeats of 8F (22, pET-28a/A2-8F13) and 12 repeats of 16YF (23, pET-28a/A2-16YF12) were identified. A schematic representation of cloning is depicted in Figure 10.

1.4.5 Construction of the second generation repetitive polypeptide libraries

During the development of the second generation polypeptide libraries following repetitive polypeptide families were constructed: 27 (H6-(32YEHK)n-H6), 28 (C2-(32YEHK)n-H6), 29 (H6-(32HKYE)n-H6), 30 (C2-(32HKYE)n-H6), 32 (A2-32YEHKn), 36 (H6-16YEn-H6), 37b (H6-(YE8HK8)n-H6), 38 (H6-(HK8YE8)n-H6), 39 (H6-32KYEYn-H6), 40 (C2-32KYEYn-H6) and 41 (C2-32YKYEn-H6).

1.4.5.1 Assembly of 3 (8Y), 4 (8E), 5 (8H) and 6 (8K) for construction of repeating units 24a (16YE), 25a (16HK), 16YK

Initially, 3 (8Y) and 4 (8E) were copolymerized with adaptive sequence 15 and cloned to 19a using the method described for construction of expression vectors 8F and 16YF. Plasmids harboring dimerized 24a (16YE), 24b (16YY), 24c (16EY), and 24d (16EE) were identified. Similarly, plasmids harboring 25a (16HK) were prepared from 5 (8H) and 6 (8K) along with 25b (16HH), 25c (16KH), and 25d (16KK).
1.4.5.2 Assembly of 24a (16YE), 25a (16HK) for construction repeating units 26a (32YEHK), 26b (32HKYE)

Equal numbers of plasmids harboring 24a (16YE) and 25a (16HK) were digested by BsaI. The resulting mixture of 16YE and 16HK units was purified and ligated with adaptive sequence 15. The desired fragments of adapted dimers were isolated, purified, and cloned to 19a (pUC18/BamHI-EcoRI). Plasmids harboring 26a (32YEHK) were identified along with 26b (16YE2 or 32YEYE), 26c (32HKYE), and 26d (16HK2 or 32HKHK).

1.4.5.3 Concatenation of 26a (32YEHK), 26b (32HKYE) and construction of the expression vectors for low molecular weight polypeptides

26a (32YEHK) and 26c (32HKYE) were digested by BsaI and the resulting 32YEHK and 32HKYE fragments were purified, oligomerized with adapters 17 and 18, respectively, and cloned to 19c (pET-28/NcoI-BamHI), to give a library of expression vectors harboring 27a-h (H6-32YEHKn-H6, n=1 to 8), 28a-h (C2-32YEHKn-H6, n=1 to 8), 29a-d (H6-32HKYEn-H6, n=1 to 4), or 30a-h (C2-32HKYEn-H6, n=1 to 6).

1.4.5.4 Concatenation of 26a (32YEHK) and construction of expression vectors for high molecular weight polypeptides

32YEHK fragments were further oligomerized with adaptive sequence 15, cloned to 19a (pUC18/BamHI-EcoRI), and a plasmid harboring 7 repeats of 32YEHK (31, 32YEHK7) was identified along with other various repeats units. 31 (32YEHK7) was digested by BsaI and the resulting 32YEHK7 fragment was purified, oligomerized with 16, 17 and 18, respectively, and then cloned to 19c(pET-28/NcoI-BamHI) to give a library of expression vectors harboring (32a-c, pET-28a/A2-32YEHKn, n=14,21,28).
(27g.h.i, pET-28/H6-32YEHKn-H6, n=7,14,21) and (28g.h-k, pET-28/C2-32YEHKn-H6, n=7,14,21,28,35).

1.4.5.5 Construction of (YE)n, (HK)n and their block-copolymerization and concatenation

16YEn, 16Kn, (16YE-16HK)n and (16HK-16YE)n repetitive coding sequences were similarly prepared from 26b (32YEYE) and 26d (32HKHK). First, a library of 33a (16YE4), 33b (16HK4), 33c (16YE2-16HK2), and 33d (16HK2-16YE2), a collection of 34a (16YE8), 34b (16HK8), 34c (16YE4-16HK4), and 34d (16HK4-16YE4) from 33a (16YE4) and 33b (16HK4), and then a library of 35a (16YE16), 35b (16HK16), 35c (16YE8-16HK8), and 35d (16HK8-16YE8) from 34a (16YE8) and 34b (16HK8) was constructed. Finally, 34a (16YE8), 34b (16HK8), and 35d (16HK8-16YE8) were oligomerized with 17, respectively, and then cloned into 19c (pET-28/NcoI-BamHI). 35c (16YE8-16HK8) was oligomerized with 17’ and then cloned to 19d (pET-9a/NdeI-BamHI). The resulting libraries of plasmids harboring repetitive coding sequences with H6-H6 adaptive sequence (36a-d, pET-28/H6-16YEn-H6, n=8,16,24,32. 37a-d, pET-28/H6-16HKEn-H6, n=8,16,24. 38a-b, pET-9a/H6-(YE8HK8)n-H6, n=1,2, 39a-g pET-28/H6-(HK8YE8)n-H6, n=1-7) were obtained. KYEY (40a-d, pET-28/H6-32KYEYn-H6, n=8,16,24,32. 41a-g, pET-28/C2-32KYEYn-H6, n=8,16,24) and YKYE series (42a-d, pET-28/H6-32YKYEYn-H6, n=8,16. 43a-c, pET-28/C2-32YKYEYn-H6, n=8,16,24) were prepared from 16KY and 16YK obtained from 3(8Y) and 6(8K), and 24a (16YE) and 24c (16EY), respectively.
1.4.6 Expression and purification of repetitive polypeptides

Expression vectors harboring coding sequences were used to transform expression hosts, BLR(DE3)pLys or BLR(DE3)pLysSRARE (42 °C for 90 sec then 4 volumes of SOC medium, 37 °C for 1 h). The transformants were selected on LB agar plates containing chloramphenicol (34 µg/mL), 1% glucose, and 50 µg/mL of kanamycin sulfate without tetracycline. Selected colonies were inoculated into 2xYT broth (2 to 5 mL for preliminary expression on a 5 to 50 mL scale and in 200 mL for 4 L scale expression) containing kanamycin (50 µg/mL), chloramphenicol (34 µg/mL), and 1% glucose. The overnight precultures were used to inoculate the desired quantities of expression culture (2xYT broth containing kanamycin (10 µg/mL) and chloramphenicol (34 µg/mL). Expression of polypeptides was initiated between 1.0 to 2.0 OD by the addition of IPTG (final concentration 1 mM) and induced for 3 to 4 h with aeration. The expression and purification of the polypeptides were analyzed by SDS-PAGE. The polyhistidine-tagged polypeptides were detected by western-blot using a SuperSignal West HisProbe kit. Polypeptides expressed from plasmids harboring 8F and 16YF with A2 adaptive sequences were purified and cleaved according to the literature protocol\textsuperscript{15,16,59} with three additional washes with 1% Triton-X100. Small and large scale purification of other polyhistidine-tagged repetitive polypeptides, under denaturing conditions in 8 M urea, was possible using Ni-NTA Spin Column and Ni-NTA Superflow (Qiagen Inc., Valencia, CA), respectively. Solutions of 10, 20, 40, 100, 300, and 500 mM imidazole in 8 M urea and 1x PBS (phosphate buffered saline, 0.1 M phosphate buffer (pH 7.4), 0.5 M NaCl at final concentration) were prepared for separation of the polyhistidine-tagged repetitive polypeptides. The required quantities of
2 M imidazole (pH 7.4), urea, and 8x stock solution of PBS were combined and then filtered through a 0.45 µm filter. For large-scale purification, cells pelleted by centrifugation at 3500 G for 30 min at 4 °C were resuspended in H₂O (15 mL per 1 L culture) and frozen. The frozen cells were lysed by freeze-thaw sonication method for 20 min in the presence of benzonase (2 µL per 1 L culture, Novagen, Inc., Madison, WI) and PMSF (phenylmethansulfonyl fluoride, 2 mM). β-Mercaptoethanol (5 µL/1 L culture, 0.2 µL/mL lysate) was added in the cases of C2-H6 polypeptides. The lysate was allowed to incubate for 30 min at rt. For purification of carbamylated polypeptides, urea (11 g per 1 L culture, 8 M at final concentration) was added to the lysate and the solution was heated in a boiling water bath for 4 h with occasional mixing. For purification of un-modified peptides, the lysate was added to guanidine hydrochloride (6 M at final concentration) and then incubated on boiling water bath for 1 h with occasional mixing. In both cases, the resultant solution was centrifuged at 15000 G for 30 min at 20 °C, and then the supernatant was diluted twice with 10 mM imidazole in 8 M urea in PBS or H₂O with 10 mM imidazole and recentrifuged at 25000 G for 1 h at 20 °C. The supernatant was applied to Ni-NTA column and the H6-H6 polypeptides adsorbed on the column. The column was eluted with 40 mM imidazole in 8 M urea in PBS and then with 300 mM imidazole in 8 M urea PBS while A2 and C2-H6 modified polypeptides were washed with 10 to 20 mM imidazole 8 M urea in PBS and eluted with 100 mM imidazole 8 M urea in PBS. Carbamylation could also be performed after purification by incubation of the column eluent containing 8 M urea on boiling water bath for 1 to 2 h. In each case, the eluent was dialyzed against H₂O for at least 2 days (dialysis membrane MW cutoff = 3,500 Da).”
1.5 References


(5) Ratner, B. D.; Bryant, S. J. Biomaterials: Where we have been and where we are going. *Annu. Rev. Biomed. Eng.* **2004**, *6*, 41-75.


(32) Panitch, A.; Matsuki, K.; Cantor, E. J.; Cooper, S. J.; Atkins, E. D. T.; Fournier, M. J.; Mason, T. L.; Tirrell, D. A. Poly(L-alanylglucose): multigram-scale biosynthesis,


CHAPTER 2
Characterization of the Genetically Engineered Repetitive Polypeptide YEHK21

2.1 Introduction

The relationship between protein folding, misfolding and aggregation, and protein sequence and structure is one of the important questions in modern biology and molecular medicine. Identification of factors affecting polypeptide chain behavior develops a fundamental understanding of protein biophysical properties and facilitates the development of pharmaceutics for curing protein related diseases, rational design of proteins with specific structure and function, and design of modern protein-based materials.

2.1.1 β-Sheet folding, aggregation and amyloid diseases

Protein folding is the physical process by which a polypeptide chain assumes a characteristic and functional three-dimensional structure.\(^1\) For the majority of proteins folding into the correct native structure is essential for biological function. A number of proteins can fold and unfold reversibly in the absence of environmental factors.\(^2\) These observations led to a hypothesis that the information necessary for the correct folding of a polypeptide chain is coded within the amino acid sequence.\(^2\) It has been widely accepted that protein folding is the result of a series of hierarchical steps where structural complexity increases with each step\(^3\) and is driven by local interactions programmed by the amino acid sequence. Protein folding is controlled by conserved (native) contacts among strongly interacting amino acids, the same contacts that stabilize local elementary
structures (LES). Latent secondary structures such as α-helixes and β-sheets\textsuperscript{4} develop on the minimization of exposure of the nonpolar residues to water.\textsuperscript{5}

**β-Sheet Formation.** With the recognition of the importance of secondary structures on the induction of tertiary structure,\textsuperscript{6-8} β-sheet systems emerged as context-independent models for probing secondary structural propensities, to include the nature and magnitude of weak stabilizing interactions and cooperativity.\textsuperscript{9} The parameters that affect the preferential folding into an α-helix or a β-sheet\textsuperscript{10} and influence the rate of folding (α-helix, 500 ns and β-sheet, 10 ms) have been recently reviewed.\textsuperscript{11-14}

**β-Sheet tendency to self-assemble and amyloid diseases.** In contrast to the existence of α-helixes as stable elements, β-strands requires the stabilization of hydrogen bonding between neighboring strands. As a consequence, the β-strands at the edges of any β-sheet have only partially satisfied hydrogen bonding requirements and are, therefore, capable of additional interactions.\textsuperscript{15} Furthermore, upon folding, the surfaces of a β-sheet functionalized by amino acid side chains are typically hydrophobic resulting in a tendency to undergo face-to-face aggregation in aqueous media.\textsuperscript{15}

Of profound importance in molecular biology and medicine is the involvement of β-sheet forming peptides and proteins in various neurodegenerative disordered and systemic amyloidosis.\textsuperscript{16} The formation of β-sheet-rich, proteinaceous fibrillar aggregates\textsuperscript{17} principally composed of a protein or peptide specific to a particular disease,\textsuperscript{18-21} is associated with nearly 20 protein deposition diseases. Despite differences in protein structure and function in the soluble form, the morphology of fibrillar aggregates and the general steps of protein deposition are surprisingly similar.\textsuperscript{22} The mechanism of amyloid fibril formation can be characterized by “nucleated growth”, a lag
phase followed by a rapid exponential growth.\textsuperscript{17, 23-28} Amyloid formation by peptides and proteins has been suggested to be a generic phenomenon. The observation that proteins unrelated to any protein deposition diseases can form fibrillar aggregates with morphological, structural, and tinctorial properties characteristic of amyloidogenic fibrils \textit{in vitro} \textsuperscript{28,33-41} validates this hypothesis. Not restricted to specific amino acid sequences,\textsuperscript{8,29-33,42} protein folding into intermolecular $\beta$-sheet structure represent an alternate folding path governed by intermolecular interactions. The process by which proteins and peptides can deposit into amyloid aggregates is under intensive investigation.

\subsection*{2.1.2 Characterization methods of amyloid systems}

The structural characterization of the amyloid fibrils and the elucidation of the protein aggregation mechanism are crucial for understanding of the factors governing the transformation of polypeptide chains during amyloid diseases, prospecting approaches to cure amyloid diseases\textsuperscript{43,44} and develop new protein-based materials.\textsuperscript{45,46}

\textbf{Structural characterization of amyloid fibrils.} The understanding of the molecular basis of protein and peptide fibrillogenesis requires a detailed knowledge about the structural organization of polypeptide chains within amyloid aggregates. However, the structural characterization of amyloid fibrils is significantly limited due to the amorphous nature of fibrillar aggregates. As a result, the classical tools of structural biology such as single-crystal X-ray diffraction and solution NMR have profound limitations in characterization of amyloid systems.\textsuperscript{47,48} Several types of imaging techniques including transmission electron microscopy (TEM) and atomic force microscopy (AFM) were utilized in characterization of fibril morphology and topology.\textsuperscript{18}
According to these studies, amyloid assemblages appear as long non-branching fibrils typically consisted of several protofilaments of 2 to 5 nm in diameter. The protofilaments can either twist together forming rope-like fibrils or have lateral association forming long ribbons-like fibrils. The characterization of amyloid fibrils by X-ray diffraction showed that polypeptide chains are organized into β-strands which run perpendicular to the fibril axes and form intermolecular cross-β cores. The recently reported applications of solid-state NMR, electron paramagnetic resonance (EPR) spectroscopy, X-ray and neutron scattering, and studies employing the biochemical methods including limited proteolysis, H/D exchange coupled with mass, solution NMR and Raman spectroscopies revealed valuable insights into the high-resolution structural organization of amyloid fibrils.

Investigation of the amyloid fibrils formation mechanism. “Nucleated growth” has been proposed as a mechanism for aggregation of amyloidogenic proteins and peptides. In contrast to the “nucleated growth” associated with crystallization, the amyloid fibril formation is complicated by the secondary structure rearrangement during aggregation and formation of various polymorphic assemblages as the intermediates and the end products of the aggregation. The complete elucidation of the amyloid formation mechanism involves the characterization of the polypeptide chain conformation and aggregation state during the protein amyloidogenesis, and the understanding of a relationship between protein sequence and folding, misfolding, and aggregation properties. Considering the amorphous nature of protein assemblages, the most complex problem in the characterization of the amyloid formation mechanism is the limitation of most of experimental methods to follow all steps of protein passage from...
the monomeric functional state to intermolecular β-sheet enriched fibrillar aggregates. In addition, simultaneous investigation of structural and dynamic aspects of protein transitions is particular problematic and requires the usage of several complementary experimental methods. Various biophysical methods including ThT fluorescence, circular dichroism (CD), vibrational circular dichroism (VCD), Raman and infrared (IR) spectroscopies have been applied in the effort of characterization of the secondary structure evolution during protein aggregation. Unfortunately, methods utilized for the quantitative studies, such as CD, IR and NMR spectrosopies can only be successfully utilized at initial steps of fibrillogenesis. In this prospect, deep UV resonance Raman (DUVRR) spectroscopy is an effective tool for structural characterization during all stages of protein fibrillation providing the quantitative analysis of the secondary structure evolution. Aggregation of proteins and peptides can be studied by the dynamic light scattering (DLS), analytical centrifugation, mass spectroscopy, yet these techniques have little or no information on the structural organization of the aggregated species and should be used in combinations with imaging methods.

Chapter II describes the characterization of folding and aggregation of the YEHK21 polypeptide specifically designed to be a model of large amyloidogenic proteins. The obtained experimental data were also utilized for the development of a tentative model of the polypeptide chain organization inside the predominant type of amyloid fibrils formed by YEHK21.
2.2 Results and Discussion

The β-sheet-forming repetitive polypeptide H6YEHK21H6 (referred latter as YEHK21) was designed with selected amino acids at the turn positions in order to facilitate control of coacervation and formation targeted supramolecular structures. The defined volumes and uniquely modified edges of the molecular subunits are controlled by the selection of turn-inducing amino acids. The design and cloning of YEHK21 construct consisting of repetitive polypeptides with 32 amino acid repeats, (GA)3GY(GA)3GE(GA)3GH(GA)3GK or 32YEHK (Figure 16), was described in details in Chapter 1.

![Figure 16](image)

**Figure 16.** Structural model of repetitive YEHK unit. The β-sheet is formed by Ala and Gly residues (shown backbone only). The turns are functionalized by Tyr, Glu, His and Lys residues forming the arrays at the edges of β-sheet.

2.2.1 Peptide isolation and functionalization

On purification the desired β-sheet polypeptide (YEHK21) was isolated. (Figure 17) The negatively charged glutamic acid (E) and positively charged lysine (K) residues originally were introduced at adjacent turn sites to facilitate anti-parallel β-sheet assembly by salt bridge formation. (See Figure 16) It was found however that strong electrostatic interactions between the glutamic acid (E) and positively charged lysine (K)
residues facilitate the polypeptide precipitation into amorphous aggregates during dialysis. Carbamylation of the lysines residues prior to or after affinity chromatography develops the considerable uncompensated charge in YEHK21 (discussed in Chapter III) and facilitates peptide folding and aggregation into amyloid–like fibrils. The dialysis of carbamylated YEHK21 polypeptides usually results in the formation of the gelatinous phase composed by YEHK21 aggregated into intermolecular $\beta$-sheet containing fibrils. The expression, purification, and modification of YEHK21 polypeptide is shown on Figure 17.

![Figure 17](image)

**Figure 17.** 8% SDS PAGE gel electrophoretic analysis of induction, purification, and modification of YEHK21. lane 1, before induction; 2, 4 h induction; 3, whole cell lysate; 4, flow through 1; 5, flow through 2; 6, 10 mM; 7, 20 mM; 8, 300 mM; 9, 500 mM imidazole 8 M urea PBS; 11, 300 mM fraction heated at 98 °C for 1min, 12, 2 min; 13, 3 min, 14, 4 min, 15, 5 min, 16, 7min, 17, 10 min, 18, 20 min; 19, 60 min, 20, 180 min. Lanes 10,21, molecular weight marker (SIGMAMARKER Wide Molecular Range).70

It was found that polypeptide constructs containing homocitrulline residues exhibited a more pronounced propensity to undergo intermolecular aggregation to form fibrils than material prepared where the conversion was avoided. Therefore, our studies have focused on these homocitrulline containing peptides.
2.2.2 Characterization of YEHK21 by far-UV CD and deep UV resonance Raman spectroscopies

Folding of repetitive polypeptides were studied employing the far-UV circular dichroism (CD) and deep-UV resonance Raman measurements.

Far-UV CD spectroscopy. Far-UV CD spectroscopy is a conventional technique for the determination protein secondary structure and for characterization of protein structural rearrangements.69-72

Figure 18 illustrates the far-UV CD spectra of the YEHK21 polypeptide after dialysis.

![Far-UV CD spectra](image)

**Figure 18.** Far-UV CD spectra of the solution YEHK21 polypeptide after dialysis.

The far-UV CD spectrum YEHK21 consists of a trough at ~207 nm and a peak at ~197-200 nm. The overall shape of the spectra resembles qualitatively the far-UV CD signature of the β-sheet conformation of polypeptides and proteins. Despite strong indication of the formation of a β-sheet structure in the solution of YEHK21, traditional deconvolution programs do not provide good quantitative analysis of polypeptide secondary structure presumably because globular protein databases were used in the comparison of CD spectra.73
Deep UV Raman spectra. The Raman scattering phenomenon is based on short-range interactions. Resonance Raman scattering from amide chromophores makes the major contribution to Raman spectra of proteins under deep UV excitation.\textsuperscript{74-78} The amide chromophore Raman signature is very sensitive to the polypeptide backbone conformation and provides direct quantitative information about the secondary structure of proteins.\textsuperscript{74-77,79,80} Tyrosine contribution was subtracted. Red and black curves represent the RR spectra of poly-L-lysine (PLL) $\beta$-sheet and PLL random coil, respectively. Illumination of the polypeptide with 195 nm light, near the amide chromophore $\pi-\pi^*$ electronic transition (absorption at $\sim$ 190 nm), resonantly enhances Raman scattering from the polypeptide backbone. Consequently, the Raman spectrum, dominated by amide bands, represents a vibrational signature of secondary structural elements.\textsuperscript{76,81} Figure 19 shows resonance Raman (RR) spectra of YEHK21 (blue and green lines) at 22 °C and 90 °C.

![Raman spectra](image)

**Figure 19.** 195-nm excitation RR spectra of YEHK21 backbone obtained at 22 °C (blue curve) and 90 °C (green curve) and poly-L-lysine (PLL) $\beta$-sheet (red curve) and PLL random coil (black curve) Spectra recorded by Dr. Vladimir V. Ermolenkov.\textsuperscript{70}
A RR spectrum of tyrosine was subtracted from both these spectra to eliminate the contribution of the aromatic amino acid side chain. The \( \beta \)-sheet conformation dominates the YEHK21 secondary structure at room temperature as revealed by comparison of its RR spectrum with that of poly-L-lysine (red lines in Figure 19) in a predominantly \( \beta \)-sheet conformation (poly-L-lysine \( \beta \)-sheet assemblage was obtained by incubation at 52 °C in a pH 11.3 solution for 3 h). At 90 °C, the YEHK21 RR spectrum closely resembles that of a poly-L-lysine random coil (black lines in Figure 19 recorded at room temperature and pH 4.0) indicating the temperature-induced melting of the YEHK21 \( \beta \)-sheet structure. Discrepancies between YEHK21 and poly-L-lysine spectra in the \( C_\alpha \)-H bending mode region are attributed to differences in the amino acid residue sequences. The observed temperature-induced structural changes in the YEHK21 were found to be reproducibly reversible.

### 2.2.3 Characterization of YEHK21 by AFM

The spectroscopic study of polypeptide folding was followed by an investigation of the controlled self-assembly of the polypeptide on a defined surface. For that purpose, purified YEHK21 in doubly-distilled water (DDW) was employed for solvent-based deposition on highly oriented pyrolytic graphite (HOPG) or polycrystalline Ni surfaces to investigate \( \beta \)-sheet self-organization. Deposition from the liquid phase was followed by DDW rinsing and dry-N\(_2\) drying protocols. Topographic investigations utilized tapping-mode atomic force microscopy (TM-AFM). Figure 20 displays a TM-AFM topographic image of the HOPG surface following YEHK21 deposition. In contrast, the pre-deposition HOPG surface exhibited an atomically flat surface marked by the absence of any discernible features.
The most notable feature in Figure 20 is a presence of highly linear fibrillar YEHK21 structures. These structures are stable under ambient conditions and exhibit no conformational change after extended storage times. The fibril lengths range from tens to thousands of nanometers. The YEHK21 fibril width is uniform with an average value of 15±2 nm as determined prior to deconvolution. Figure 21 displays a probability density plot of fibril thickness derived from an analysis of 79 fibril samplings derived from multiple deposition experiments. Observed fibril thicknesses as determined by TM-AFM did not exhibit a continuum of thicknesses as might be expected from lamellar stacking.\(^82\) Fibril thickness varies in discrete increments of approximately 0.8±0.2 nm. Figure 21 clearly shows a substantially higher probability for fibrils with a thickness of 1.7±0.2 nm, a thickness not inconsistent with a loose bilayer or ‘ribbon’ configuration.\(^83\) (Figure 22)
Figure 21. Probability density plot of fibril thickness derived from AFM topographic data of YEHK21 fibrils. 79 samplings were employed to generate the distribution profile.

Figure 22. YEHK repeat units with g-turns forming antiparallel β-sheet structure. Width (turn-strand turn) of the computed model, 3.4-3.8 nm; thickness, 0.72 nm.

The observed fibril thickness increments are consistent with thickness estimates from the computationally derived YEHK21 polypeptide β-sheet structure. We speculate
that the intrasheet interactions and the disposition of the hexahistidinyl tracts at the YEHK21 termini strongly favor a bilayer configuration. Nonetheless we are cognizant of the complexity of turn formation and the implication of turn motif on amphiphilicity (Figure 23).

![Figure 23](image)

**Figure 23.** The implication of the turn motif on amphiphilicity of the polypeptide surface. Left: Formation of surfaces with equal hydrophobicity employing β-turn motive. Right: Formation of surfaces with different hydrophobicity employing γ-turn motive.

Differing turn motifs can be assumed by β-hairpins as reflected in the number of amino acids per turn and the number of hydrogen bonds formed between the distal strands with a consequence that the faces of β-sheet may be differentiated by the orientation of the methyl groups of constituent strands toward a single face of the sheet. It has been previously shown that closely related constructs poly[(AG)_3EG], poly[(AG)_3YG] and poly[(AG)_3KG] form an amphphilic β-sheet structures via the intermediacy of β-turns to redirect the β-strand subunits. If a β-turn containing structure is assumed for YEHK21 as shown in Figure 22, the amphiphilic character of the resultant β-sheet would be consistent with the observed bilayer formation. The non-polar alanine methyl bearing faces would be shielded from the solvent, while the glycyl derived moieties would be exposed to the aqueous environment.
2.2.4 Characterization of YEHK21 by TEM

To confirm this structure and further investigate YEHK21 fibril morphology transmission electron microscope (TEM) studies were carried out. YEHK21 was deposited on carbon-coated support grids. Deposition protocols were similar to those for HOPG substrates. A uranyl acetate stain was utilized on post-dried samples to enhance edge contrast of the fibril assemblies in the TEM. Figure 24 displays resulting TEM micrographs.

![TEM micrographs of YEHK21 fibrils on carbon-coated Cu grids. The measurements were recorded by Dr. Kathleen Dunn.](image)

**Figure 24.** TEM micrographs of YEHK21 fibrils on carbon-coated Cu grids. The measurements were recorded by Dr. Kathleen Dunn.

The highly linear assemblies in the left image of Figure 24 agree qualitatively with the TM-AFM topographs of Figure 20. The lower image of Figure 24 displays a higher magnification micrograph of the fibril assemblies. On analysis of the TEM data, the average width of the fibrillar structures was found to be about 6.5±1 nm as shown in Figure 25. An average measured fibril width of 6.5±1 nm was predominant for all TEM data studied. This is in approximate agreement with the width of a structure comprised of two β-sheets side-by-side where the hydrophobic turn groups (Tyr and His) of the two
sheets pack closely. (See width estimates in Figure 22.) As expected, TM-AFM data yielded a much larger average fibril width of 15±2 nm due to tip-convolution effects. Calibration of the tip-convolution of the DI Nanoscope III AFM utilized Au nanospheres with a nominal diameter of 1.4 nm, chosen for the similarity of the sphere diameter to the fibril thickness. This calibration revealed a tip convolution-induced feature broadening of 8±1 nm and therefore resulted in a very good agreement between the TEM and TM-AFM fibril-width measurements.

Figure 25. Probability density plot of fibril width derived from TEM of YEHK21 fibrils. 161 samplings were employed to generate the distribution profile.

2.2.5 Fibril formation

A higher-resolution TM-AFM micrograph of crossed fibrils is shown in Figure 26. Note the apparent increase in the fibril thickness at the intersections of the fibrils. A similar crossing phenomenon is also evidenced in the TEM image (Figure 24, red circle) where the distinct edges of two overlapping fibrils are observed. Both these image data sets imply that the fibrils form in solution prior to deposition on the substrate. This
contrasts with the surface-templated assembly of low molecular weight polypeptide β-sheets on HOPG that have been reported recently.\textsuperscript{85-87}

\begin{figure}[h]
\centering
\includegraphics[width=0.5\textwidth]{image}
\caption{AFM topography of assembled domains of YEHK21 on HOPG.}
\end{figure}

The surface disposition of the fibrillar structures presented herein more closely resemble observations by Marini et al.,\textsuperscript{88} although the polypeptide β-sheets reported in that work consisted of only eight residues, nearly two orders of magnitude fewer than the polypeptides considered here. It is important to note that β-sheets frequently exhibit a right handed twist. An increase in the number of strands with a concomitant increase in number of H-bonds should improve sheet stability, however the twist induces dissymmetry in the interaction of side chains from neighboring strands.\textsuperscript{89} Cooperative β-sheet formation has been reported in the formation of helical ribbons based upon the self assembly of the octapeptide FKFEGKF\textsuperscript{90} where the KE and F domains lead to self assembly of the individual molecular strands into a ribbon with an extended β-sheet geometry. Over longer times, the ribbons tend to form super helices around the initial strand.\textsuperscript{88} It has also been reported that low molecular weight peptides assemble into elongated β-sheets containing tapes which then dimerize to form helical ribbons and
subsequently aggregate to form fibrils by forming twisted lamellae\textsuperscript{83} In this work, helical assemblies were not commonly observed for YEHK21 in contrast to either of the above materials or the products described in ref. 88

Figure 27. In a tentative structural model for the most common fibrils observed both the dimensions of those fibrils and amphiphilic nature of proposed anti-parallel $\beta$-sheet can be accommodated by the interaction of four polypeptide molecules.

A possible assembling of four YEHK21 molecules which can accommodate the observed dimensions of the fibrils is depicted in Figure 27. The amphiphilic character induced by the involvement of the $\beta$-turn in reversing strand direction is consistent with the facile solution formation of bilayer thick fibrils. By design, the $\beta$-sheet forming YEHK21 molecule has relatively hydrophobic and hydrophilic edges dominated by the turn groups, as a consequence, in the proposed model the more hydrophobic surfaces and turn groups of the molecules are shielded from the aqueous environment. Further experiments to validate this hypothesis are underway.
2.3 Conclusions

A de novo, genetically engineered 687 residue polypeptide expressed in *E. coli* has been found to form β-sheet structure and self-assemble into highly rectilinear, fibrillar structures in aqueous solution. Atomic force microscopy (AFM) and transmission electron microscopy (TEM) definitively established the tendency of the fibrils to predominantly display an apparently planar bilayer or ribbon assemblage.

Part of this work has been published:


2.4 Experimental Section

2.4.1 Fluorescence, UV absorption and CD spectroscopy

Fluorescence spectra were measured in a 1-cm rectangular quartz cell with a magnetic stirrer using a Jobin Yvon Fluoromax-3 spectrofluorometer. Typically, 275-nm excitation, with 2-nm excitation and 4-nm emission slits, 1-nm data interval and 0.5-s integration time were used for fluorescence measurements. UV absorption spectra were measured in 1-mm quartz cell using Hewlett-Packard HP 8452 diode array spectrophotometer. Far-UV CD spectra were measured in 0.05-cm temperature-controlled quartz cell using Jasco J-720 spectropolarimeter.

2.4.2 Resonance Raman measurements

Deep-UV Raman instrument is described in. The fourth harmonic (195 nm, ~2 mW) of the Indigo S laser system (Coherent) is used to generate Raman scattering from a 0.6-mm diameter thermostatically controlled sample solution stream. A custom-built subtractive double spectrograph equipped with a Roper Scientific Spec-10:400B CCD camera (liquid nitrogen cooled) is utilized for recording Raman spectra. The spectral resolution of the system is 4 cm\(^{-1}\). Spectra are analysed using GRAMS/AI (7.01) software.

2.4.3 Atomic force microscopy (AFM) measurements

Sample preparation: After dialysis, aqueous solutions of the polypeptide were used at a pH of 7.0 for further studies. A concentration of 0.12 mg/mL was determined using UV-visible spectroscopy. 50 µL of the aqueous solution was dropped on a 12 mm by 12 mm sample of highly oriented pyrolytic graphite (HOPG) (Veeco Instruments). The
polypeptide solution so deposited on HOPG was incubated at room temperature for 30 minutes and then removed by a micropipette.

**AFM imaging conditions.** AFM imaging was performed in resonant tapping mode under ambient conditions with a MultiMode microscope (Digital Instruments) using a Nanoscope IIIa control system. A J scanner was used with a lateral range of ~ 125 µm. Images were collected using Nanoscope III software version 4.42r8 in height and phase mode simultaneously. The standard silicon TESP cantilevers (nominal spring constant, 40 N/m, resonance frequency about 300 kHz, and tip radius less than 10 nm purchased from BudgetSensors®) were used for imaging. The force was minimized during the imaging by choosing a setpoint corresponding to more than 90% of the free oscillation amplitude of about 12 nm. Typical scan frequency was 1 Hz and images were collected at resolution of 512 x 512 points. The tip convolution of about 8 nm for features 1.4 nm in height was calculated by imaging the 1.4 nm diameter gold nanoparticles generally used for TEM calibration. Offline Nanoscope image analysis software (version 5.12r3) was used for image analysis. The images were flattened and no further image processing was performed.

**2.4.4 Transmission electron microscopy (TEM) measurements**

Specimens were prepared for transmission electron microscope observation by placing a droplet of the peptide-containing solution on a standard carbon-coated support grid. After 30 minutes solution was removed and the specimen then stained using uranyl acetate to accentuate the morphology of the structures. Bright field TEM images were recorded on film in a JEOL 200CX operating at 80 kV.
2.5 References


(71) Chen, E.; Gensch, T.; Gross, A. B.; Hendriks, J.; Hellingwerf, K. J.; Kliger, D. S. Dynamics of protein and chromophore structural changes in the photocycle of


(82) Parkhe, A. D.; Cooper, S. J.; Atkins, E. D.; Fournier, M. J.; Mason, T. L.; Tirrell, D. A. Effect of local sequence inversions on the crystalline antiparallel beta-sheet


CHAPTER 3

Applications of Genetically Designed Repetitive Polypeptides

3.1 Introduction

3.1.1 Experimental shortcut to the protein amyloidosis

Peptide models derived from the amyloidogenic proteins and peptides have been widely used to simplify investigations of the mechanism of amyloid formation.\(^1\)\(^,\)\(^2\) Small peptide models were successfully employed in probing fibrillar structure \(^1\)\(^-\)\(^4\) thus facilitated atomic-level insights into the \(\beta\)-strand assembly of fibrils.\(^5\)\(^-\)\(^7\) A number of chimeric proteins and \textit{de novo} designed peptide models were also successfully used to study the mechanism of amyloid formation.\(^7\) The advantage of using the \textit{de novo} designed models is the possibility to test only a specific aspect of the complex overall process and thereby directly access the effects of protein sequence on behavior. Another advantage of work with model peptides is the ability to generate under physiological conditions intermediates in the amyloid formation process. In stark contrast, natural proteins normally require non-native conditions such as high concentration, lowered pH, or increased ionic strength to facilitate protein misfolding.\(^8\)\(^-\)\(^12\)

3.1.2 Macromolecular systems derived from genetically engineered repetitive polypeptides

Our study of protein folding, focusing on aggregation and amyloid formation is based on the development of the large, fibril-forming polypeptides, and identification of the relationship between the amino acid sequence and folding/aggregation properties. The polypeptides are good experimental models for amyloidogenic protein as the
substances i) represent a macromolecular model, ii) reduce sequence specificity enabling the study of a generic polypeptide chain, iii) form intermolecular β-sheets structure and aggregate into amyloid-like fibrils, and iv) respond to environmental stimuli.

Chapter 3 describes the utilization of the polypeptides as models of protein amyloidogenesis; among them are 1) amyloidogenesis of intrinsically unfolded proteins, 2) influence of electrostatic interactions, 3) the role of β-strand numbers, and 4) structural characterization of the cross-beta core of amyloid fibrils.

3.1.3 Amyloidoisis of intrinsically disordered proteins and peptides

The number of human amyloid diseases including Alzheimer’s, Parkinson’s, Huntington’s, and Creutzfeld-Jacob’s diseases entail the deposition of natively disordered proteins. Intrinsically disordered proteins and peptides do not have rigid tertiary structures under native conditions rather existing in a dynamic ensemble of conformations due to the combination of high net charge and low hydrophobicity of polypeptide chain. Despite an unstructured nature, intrinsically disordered proteins and peptides can suffer induced folding under a variety of environmental stimuli carrying out various biological functions. Recently, significant progress has been made in the understanding the etiology of amyloid diseases associated with intrinsically disordered proteins. The formation of intermediates on induced folding was proposed as a requirement for the initiation of amyloidogenesis. However a fundamental understanding of the forces that direct the induced folding and importantly the role of resultant intermolecular interactions of the partially folded intermediates has not yet been achieved.
3.1.4 Role of electrostatic interactions on the aggregation of peptides and proteins

Electrostatic interactions play an important role on the aggregation of peptides and proteins into amyloid-like fibrillar aggregates. The number and position of charged amino acid residues define the conformational stability of a polypeptide chain and the nature of electrostatic interactions.

Small peptides. Charge complementarity within highly charged molecules has been shown to drive the assembly of well-defined amyloid-like fibrils and other nanosstructures.\(^7,17-19\) On the contrary, repulsive electrostatic interactions impede the oligomerization of small amyloidogenic peptides,\(^20\) and have been widely used in the design of pharmaceutical agents to retard amyloid fibril formation.\(^20,21\)

Proteins. It is commonly accepted that the net charge of protein is inversely related to the tendency to aggregate,\(^22,23\) Charge neutralization can promote protein aggregation. The development of uncompensated charge at the edges of a β-sheet protein prevents aggregation.\(^24,25\) High net charge appears to be a key factor in the maintenance of intrinsically disordered proteins in monomeric and dynamic state.\(^13\) Despite the evidence that reduction of protein net charge facilitates aggregation,\(^22,23\) there are numerous examples of amyloid fibril formation from charged peptides and proteins \textit{in vivo} and \textit{in vitro}.\(^26\) Many human amyloid diseases involve the deposition of natively disordered proteins\(^14,15\) that are highly charged under physiological conditions and yet retain the ability to form amyloid fibrils. Aggregation of many amyloid proteins \textit{in vitro} occurs at acidic pH when there is significant charge.\(^13,27-31\) Moreover, the importance of repulsive electrostatic interactions on amyloid fibril formation has been recognized.\(^32\) Uncompensated charge has been shown to facilitate the formation of
ordered amyloid-like assemblages, while the screening or compensation of charge led to rapid collapse into amorphous aggregates.\textsuperscript{32} The deposition of charged proteins and peptides to form fibrillar assemblages with a pronounced polyelectrolyte nature\textsuperscript{33,34} was illustrated by the amyloid driven aggregation and precipitation of soluble proteins.\textsuperscript{34} Amyloid peptides, proteins and corresponding aggregates may interact with various electrolytes and polyelectrolytes. These substances not only bind amyloid fibrils\textsuperscript{35-38} but also promote \textit{in vitro} fibrilogenesis\textsuperscript{37,39} affecting the nucleation\textsuperscript{40} and seeds\textsuperscript{41} and fibrils\textsuperscript{42,43} stabilization.

\textbf{3.1.5 β-Sheet folding by multi-stranded peptide models}

Non-aggregative β-\textit{hairpin} models, consisting of two anti-parallel β-strands connected by a β-turn (Figure 28),\textsuperscript{44-46} were successfully utilized to study the correlations between peptide sequences and folding ability in the absence of any tertiary interactions. The nature of amino acid residues at the turn position,\textsuperscript{47-51} the intrinsic propensity of amino acids to form or participate in β-sheet conformations,\textsuperscript{52} interstrand side chain interactions including both hydrophobic interactions\textsuperscript{44,53,54} and salt bridge formation\textsuperscript{46,55} were defined as important factors affecting β-hairpin folding.

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{figure28}
\caption{The β-hairpin motif, in which two antiparallel strands are connected by a short loop. The most common loop size is two residues (n = 1). R represents amino acid side chains. Dashed lines represent hydrogen bonds.\textsuperscript{45}}
\end{figure}
Several multistrand β-sheet models were employed in order to investigate cooperative interactions perpendicular to the axis of β strand propagation (Figure 29). The evaluation of cooperative interactions on H-bonding in a family of three- and four-stranded models revealed only weakly incremental cooperative effects on the stability of the constructs upon addition of each extra β-strand.56

**Figure 29.** Peptide β-sheet folding models. Schematic representation illustrating cooperativity in model β-sheets propagated (top) along a β-hairpin peptide from the turn (parallel with the β-strands) and (bottom) perpendicular to the strand direction as the docking of a third strand enhances the stability of the existing hairpin template.46
3.2 Results and Discussion

3.2.1 YE8 as a model for amyloidogenic intrinsically unfolded proteins

The YE8 polypeptide was designed to form a finite but large set of weakly interacting β-strands to enable examination of the role of intramolecular/intermolecular nucleation in the absence of tertiary contacts and establish the importance of intramolecular and intermolecular β-sheet propagation. The charged amino acids were introduced into turn position to control polypeptide folding and coacervation. The study of the YE8 polypeptide behavior under varying environmental conditions can mimic common steps in intrinsically disordered protein aggregation and reveal generic aspects of the mechanism of protein deposition.

3.2.1.1 Design of YE8

The 11 kDa repetitive polypeptide YE8 consists of 8 repeats of the 16 amino acid monomer, (GA)$_3$GY(GA)$_3$GE or YE (Figure 30). It has a primary sequence GH$_6$[(GA)$_3$GY(GA)$_3$GE]$_8$GAH$_6$ where both N- and C- termini bear hexahistidinyl tracts. YE8 was proposed to form 16 anti-parallel β-strands GAGAGA capable of weak interstrand interactions via hydrogen bonding and weak hydrophobic attraction. YE8 has an array of negatively charged glutamic acids (E) on one hydrophilic edge of β-sheet and uncharged tyrosine residues (Y) on the other. The repetitive YE track is flanked by the C- and N-terminal hexahistidinyl sequences. This design permits precise control of the charge distribution along the β-sheet edge via variation of the pH or ionic strength of the medium.
3.2.1.2 Preparation of the initial YE8 sample

After expression and purification, a solution containing the YE8 polypeptide was dialyzed against doubly distilled H₂O at 4 °C. The dialyzed solution was centrifuged at 15000 g for 45 min. The aggregates were then separated by decantation of the supernatant. Analysis by Raman and CD spectroscopy of the YE8-containing supernatant revealed that the repetitive polypeptide exists as a random coil (Data not shown). This solution was used as a stock solution for all other experiments.

3.2.1.3 pH dependence of folding and aggregation

A pH range between 2 and 5 was chosen so that the protonation or deprotonation of glutamic acid residues could be modulated while the histidine groups remained effectively unchanged and positively charged (Table 1).
Table 1. Predicted charges of YE8 polypeptide, the β-sheet core and different amino acids at different pH values.$^a$

<table>
<thead>
<tr>
<th>Peptide/aa</th>
<th>pH 2</th>
<th>pH 3.5</th>
<th>pH 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>H6(YE)8H6</td>
<td>12.9</td>
<td>11.4</td>
<td>5.2</td>
</tr>
<tr>
<td>(YE)8</td>
<td>0</td>
<td>-0.9</td>
<td>-6.4</td>
</tr>
<tr>
<td>8E</td>
<td>0</td>
<td>-0.9</td>
<td>-6.4</td>
</tr>
<tr>
<td>8Y</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>2·H6</td>
<td>12</td>
<td>12</td>
<td>11.6</td>
</tr>
</tbody>
</table>

$^a$ The prediction was based on the pKa values for isolated amino acid residues using Protein Calculator v3.3.

The initial YE8 solution was diluted to 44 µM with simultaneous adjustment of the acidity to pH 2, pH 3.5 and pH 5. The samples were incubated at room temperature with small aliquots withdrawn for measurements. Folding and aggregation of YE8 under the various conditions were followed by Raman, CD and AFM. Polypeptide solutions incubated at pH 2 and pH 3.5 did not show any visual changes over the incubation period, while the sample incubated at pH 5 exhibited significant precipitate formation by the second and third days of incubation.

Figure 31 shows the RR spectra of YE8 incubated for various length of time at room temperature at pH 2 and pH 3. The measured Raman spectra were dependent on both the sample age and on the solution pH. The shape and relative intensities of the amide bands indicated that all YE8 Raman spectra acquired shortly after modulation of the solution pH were dominated by the RR spectroscopic signature of an unordered polypeptide. After 56 days incubation at pH 5, there was no substantial change in the spectrum of YE8 (data not shown). Very different behavior was found for solutions at pH 2 and pH 3.5, the Raman spectra exhibited substantial changes in all amide bands (Figure 31).
**Figure 31.** DUVRR spectra of YE8 at pH 2 (left panel) and pH 3.5 (right panel) incubated at room temperature and recorded in 0, 2, 3, 13, 27 and 56 days after the preparation. Excitation with quasi-CW radiation at 197 nm. For comparison, all Raman spectra were normalized to the intensity of the strongest tyrosine Raman band at 1618 cm\(^{-1}\). Spectra recorded by Dr. Vladimir V. Ermolenkov.

In particular, the amide I band narrowed and intensified with incubation time were indicative of β-sheet formation. It is noteworthy, that β-sheet formation occurred more quickly at pH 3.5 than at pH 2 as assessed by changes in the amide I intensity.

YE8 samples incubated for two months at various pH conditions were also characterized by UV-vis, Far-UV CD fluorescence spectroscopy. The far-UV CD spectra of both YE8 samples incubated at low pH, had a strong trough at ~215 nm and a peak at ~193 nm (Figure 32A).
Figure 32. Characterization of YE8 samples incubated for two month at different pH employing A) Far-UV CD, B) fluorescence and C) UV-vis spectroscopes.
While the overall shape of these curves qualitatively resembles the far-UV CD signature of the \( \beta \)-sheet conformation in polypeptides and proteins, traditional deconvolution programs for quantitative analysis of YE8 secondary structure are not applicable.\(^{53}\) The far-UV CD spectra of YE8 incubated at pH 5 is not shown due to high levels of scattering within this sample. The fluorescence spectra of YE8 polypeptide solution are shown on Figure 32B. It is well known that fluorescence of tyrosine residues is very sensitive to the local environment and is often employed as a natural biomarker.\(^{57}\) The fluorescence spectra of YE8 polypeptides under all experimental conditions resemble the spectrum of an aqueous tyrosine tripeptide (H-Gly-Tyr-Gly-OH) solution suggesting that the tyrosine side chain is significantly exposed to water in all samples including samples in a fully folded conformation. Although no aggregates were visually detectable in YE8 samples at pH 3.5 and pH 2 during incubation, significant scattering was detected by UV-vis spectroscopy (Figure 32C). To separate possible aggregates, YE8 solutions were centrifuged at 15000 G for 45 min at 4 °C. After centrifugation, small portions of a gelatinous phase were separated by decantation. Irrespective of the YE8 incubation conditions or progress of polypeptide folding, folded YE8 was found only in the gelatinous phase. The peptide in the supernatant had a random coil conformation.

The morphology of the YE8 aggregates formed at various pH was studied by atomic force microscopy (AFM) and transmission electron microscopy (TEM). Consistent with \( \beta \)-sheet assemblage at pH 3.5 and pH2, YE8 forms linear, non-branching amyloid-like fibrillar aggregates (Figure 33). On incubation at pH 5, only amorphous aggregates of YE8 were detected.
Figure 33. Comparison of YE8 polypeptide aggregation at different conditions. TM-AFM topographs on HOPG and TEM micrographs on carbon-coated Cu grids of YE8 samples incubated at different pH: A) and D) at pH 2, B) and E) at pH 3.5, C) and F) at pH 5 respectively.

3.2.1.4 The effect of ionic strength on folding and aggregation

The presence of sodium perchlorate influenced polypeptide folding of YE8 solutions at different pH values. In contrast to the previously described experiments, YE8 precipitation began after two-to-three days of incubation without evidence for β-sheet formation. Even after 30 days, no β-sheet formation was detectable.

3.2.1.5 Concentration dependence and molecular template effect

Based on the results above, pH 3.5 was chosen as the preferred condition for YE8 folding and aggregation into amyloid-like fibrils. YE8 solutions (pH 3.5, 44 - 9 µM)
were incubated at room temperature. The absence of reliable basis DUVRR spectra for YE8 β-sheet and β-turns made it impossible to retrieve complete quantitative information about structural rearrangements of the polypeptide during the folding process. Instead, three spectra were utilized, including the YE8 backbone Raman spectra obtained for the initial disordered and the final completely folded polypeptide, and the Raman spectrum of tyrosine, to fit Raman spectra measured at various stages of fibrillation as described above.

These basis spectra were utilized in the fitting of the YE8 spectra acquired at various incubation times and with various polypeptide concentrations as well as for samples subjected to stirring and seeding. All spectra were fit satisfactorily. For example, Figure 34 illustrates the fitting of 44-µM YE8 obtained at selected incubation times. In addition to experimental spectra and best linear combinations of three basis spectra, the residual spectra are also shown following incubation for 1, 7, 14, 21, 35 and 50 days.

The very good quality fit permits several conclusions to be drawn. Firstly, these results support the hypothesis that the tyrosine contribution to the Raman spectrum of YE8 did not change on polypeptide folding. Second, the two conformations tentatively described as folded and unfolded are dominant in various proportions at different stages of β-sheet folding process. The β-sheet/turn conformation occurred directly from the disordered polypeptide with the PPII conformation without the involvement of any intermediate states during the folding process. The two-state behavior found for YE8 folding did not necessarily result from an all-or-none transition. As discussed below, the YE8 aggregation plays an important role in folding.
Figure 34. DUVRR spectra of 44 μM YE8 incubated at room temperature for 1, 7, 14, 21, 35 and 50 days after the sample preparation (solid blue curves), results of their fitting with a linear combination of basis spectra (dotted red curves) and their differences (solid black curves). Folded YE8, unfolded YE8 and tyrosine spectra were used as basis spectra. The experimental spectra were recorded at 197 nm with quasi-CW excitation.
The fitting results of the Raman spectra of YE8 obtained at various stages of β-sheet folding process allowed the determination of the relative proportions of the YE8 polypeptide in the folded and unfolded forms. The corresponding kinetic curves are shown in Figure 35.

**Figure 35.** Kinetics of YE8 folding and the relative proportions of the YE8 polypeptide in the folded and unfolded forms obtained for the samples with different concentrations of 9, 12, 15, 25, 35 and 44 µM as well as for the samples subjected to stirring or seeding with folded YE8.

Variation of concentration had little effect on the spectra that remained indicative of β-sheet formation. However, the kinetics of YE8 structural rearrangements were strongly concentration dependent, e.g., the amide I Raman band intensity was dependent upon both concentration and incubation time. The kinetic curves obtained, typical for a fibrillation process, had a lag period prior to β-sheet formation that was strongly dependent on YE8 concentration. The shortest lag time, ~7 days was found for the solutions with concentrations of 44 and 35 µM of YE8.
Both shear and the presence of molecular templates accelerate the fibrillation of
global proteins in vitro. For template studies, 10 µL of 44 µM YE8 solution previously
incubated for 60 days and subsequently sonicated was added to 3.99 mL of a freshly
prepared 44-µM YE8 solution (0.25 % templating material was added). In good
agreement with published results on the fibrillation of amyloidogenic proteins, no lag
phase was evident. (Figure 35).

Except for the absence of the lag phase, the RR spectra and the characteristic
fibrillation times were similar for 44 µM YE8 samples incubated with and without
template. Agitation by magnetic stirring of a 44 µM solution of YE8 led to a dramatic
acceleration in β-sheet formation. With continuous stirring, the changes in the RR
spectrum of YE8 were complete in five days. The final RR spectrum, in particular the
intensity of the amide I band, was nearly identical for the stirred sample and samples
incubated 50 days without stirring.

3.2.1.6 YE8 as a model for a completely amyloidogenic protein

Consistent with intrinsically unfolded protein sequences, YE8 lacks order-
promoting amino acids such as hydrophobic residues (Val, Ile, Leu) and most of aromatic
species (Trp, Phe). These amino acids initiate protein folding via hydrophobic collapse
to form hydrophobic core of the folded protein. On the other hand, YE8 is rich in
charged (Glu and His) residues that seldom support the formation of ordered ensembles.
The large net charge makes possible the flexibility and solubility of the YE8 chain.
Modulation of pH affects the protonation or deprotonation of the glutamic acid and
histidine residues enabling control over YE8 solubility, folding, aggregation and
amyloid-like fibril formation. YE8 has all the major characteristic of an amyloidogenic
protein providing an excellent model system for detailed investigations of the mechanism of amyloid fibril formation. The identical GAGAGA strands obviate sequence specificity of local interactions and allow investigation of the effects of \textit{intra- or intermolecular} assistance on initiation and propagation of $\beta$-sheet folding and amyloid fibril formation. The repetitive nature of YE8 polypeptide facilitates future investigations of \textit{intra- or intermolecular} interactions with homologs of the (YE8)$_n$ polypeptide family based on the parent YE8 construct. The influence of systematic variation of the number of identical weakly interacting $\beta$-stands on the folding and amyloidogenic nature of polypeptides can be directly studied.

3.2.1.7 The mechanism of the YE8 polypeptide $\beta$-sheet folding and fibrillation

A nucleated growth mechanism is proposed for YE8 folding (Figure 36) as YE8 secondary structure evolution exhibits a lag phase that can be reduced by introduction of traces of amyloid fibrils or seeds.

![Figure 36](image)

\textbf{Figure 36.} Proposed mechanism of YE8 aggregation driven $\beta$-sheet folding.

The lag phase for YE8 folding increases significantly on polypeptide dilution (Figure 35) revealing a pronounced concentration dependence. For the most dilute sample (9 $\mu$M) the lag time exceeds 30 days clearly signifying that absence of substantial \textit{intramolecular} assistance significantly retards folding. The acceleration of $\beta$-sheet formation at higher concentration is indicative of the role of aggregation in folding.
Moreover, the observation that YE8 β-sheets were detected only within the fibrillar assemblies supports the importance of the intermolecular interactions. Concentration dependence was not observed on incubation of the two highest concentration solutions suggestive of polypeptide supersaturation\(^{60}\) where the rate of aggregation no longer influences folding. Under these conditions only the evolution of secondary structure defines the rate of the initiation and propagation of YE8 folding. Two different processes, such as polypeptide aggregation and maturation of the secondary structure to form a functional β-sheet nucleus take place during the nucleation stage. Polypeptide aggregation necessarily precedes and facilitates the initiation of folding. Identification and elucidation of the role of these two processes on polypeptide amyloidogenesis is difficult. The non-specific/specific aggregation of polypeptides appears to be strongly concentration dependent while β-sheet formation and aggregate maturation to the amyloidogenic nucleus is likely concentration independent.\(^{60}\) At high polypeptide concentrations, aggregation occurs much more rapidly than secondary structure conversion. Initially, the majority of pre-amyloidogenic aggregates form with very little β-sheet content. On the contrary, in the low concentration regime, secondary structure conversion could be faster than/or comparable to polypeptide aggregation. Polypeptide aggregates can constitute a significant portion of the amyloidogenic nuclei.

### 3.2.2 Charge Distribution and Amyloid Fibril Formation

Considering the polyelectrolyte nature of the majority of amyloid fibrils and especially the ability of highly charged intrinsically disordered proteins to form amyloid deposits \textit{in vivo}, we propose that the ability of highly charged polypeptide chain to form amyloid aggregates is predetermined by the position of charges within the ordered
assemblages (amyloid fibril). To test this hypothesis and investigate the influence of electrostatic interactions, we utilized the two model polypeptides YE8 (Figure 30, p.73) and YEHK21 (Figure 16, p. 43).

3.2.2.1 Preparation of the YEHK21 samples

Upon a dialysis against doubly distilled H$_2$O the modified YEHK21 exists in a gelatinous phase composed of folded polypeptides assembled into amyloid-like fibrils. To study the folding and aggregation of YEHK21 at varying conditions, an 8M urea solution of unfolded polypeptide solution (solution after polypeptide purification) was dialyzed at 4 ºC against aqueous acid or base at pH 10.0, pH 6.5, pH 5.0, pH 3.5 and pH 2.0.

3.2.2.2 The folding of YEHK21

Upon the dialysis, the YEHK21 samples were characterized by deep UV resonance Raman spectroscopy (Figure 37). The analysis of Raman spectra suggests that the YEHK21 existed in a β-sheet conformation at pH 6.5 and was unfolded at pH 2 and pH10. Intermediate states with β-sheet and random coil components are observed at pH 3.5 and pH 5 (Figure 37, blue spectra). Importantly, all folded polypeptides were identified within the gelatinous phase obtained by centrifugation and supernatant decantation (Figure 37, red spectra). All supernatants contain YEHK21 in unfolded state. The detailed discussion of the Raman spectra of YEHK21 can be found elsewhere.$^{61}$
**Figure 37.** The 197-nm excited Raman spectra of YEHK21 dialyzed against water at pH 2.0, pH 3.5, pH 5.0, pH 6.5 and pH 10 (blue), and spectra of gelatinous fractions obtained by sample centrifugation and supernatant decantation (red). Spectra recorded by Vitali Sikirzhitsky.
3.2.2.3 AFM characterization of YEHK21

AFM characterization of YEHK21 deposited from solutions (pH 6.5 and pH 5) on highly oriented pyrolytic graphite (HOPG) revealed the formation of non-branching amyloid-like fibrils by folded polypeptides (Figure 38).

![Figure 38. TM-AFM topographs of YEHK21 aggregates at A) pH 5 and B) pH 6.5.](image)

3.2.2.4 Folding and aggregation of YEHK21 and YE8 in highly charged states

Samples containing YEHK21 at pH 2 and YE8 at pH 6.5 had no detectible β-sheet structure and did not form a gelatinous phase immediately after dialysis. After long incubation at room temperature, 3 months for YEHK21 and 1 month for YE8, Raman spectroscopy revealed low concentrations of the polypeptides with a β-sheet conformation (Figure 39 and Figure 40, blue spectra). Centrifugation of both samples revealed the existence of gelatinous and aqueous fractions in polypeptide solutions that were separately analyzed. The predominant presence of the β-sheet conformation can be clearly identified by the presence of narrow and intense Amide I band in the samples containing aggregated fractions of YEHK21 and YE8 (blue spectra). The spectra of supernatant solutions are characteristic of disordered polypeptides (green spectra).
Figure 39. The 197-nm excited Raman spectra of YEHK21 incubated in solution with pH 2 (red), its gelatinous (blue) and aqueous (green) fractions. Spectra recorded by Vitali Sikirzhitsky.

Figure 40. The 197-nm excited Raman spectra of YE8 incubated in solution with pH 6.5 (red), its gelatinous (blue) and aqueous (green) fractions. Spectra recorded by Dr. Vladimir V. Ermolenkov.

The AFM studies of the morphology of the polypeptide aggregates in the gelatinous phase revealed the formation of well-defined amyloid-like fibrils (Figure 41).
Figure 41. TM-AFM topographs of YEHK21 fibrils at pH 2 (A) and YE8 at pH 6.5 (B).

3.2.2.5 Role of electrostatic interactions and charge distribution in YEHK21 and YE8

By the design, the edges of YEHK21 are functionalized by charged lysine (K), histidine (H), and glutamic (E) acid residues (Figure 16). The modified YEHK21 represents a simplification of the parent YEHK construct where the charged lysine (K) residues were converted into neutral homocitrulene. As a further simplification of the system, the YE8 polypeptide has repetitive domain with only one type of charged residues: negatively charged glutamic acids (E). Variation of pH or ionic strength control electrostatic interactions which are in turn regulate solubility, folding and aggregation of polypeptides (Table 2).

Table 2. Predicted charges of modified YEHK21 and YE8 at different pH.\(^a\)

<table>
<thead>
<tr>
<th>pH</th>
<th>H(_6)YE(_6)H(_6)</th>
<th>YE(_8)</th>
<th>8E</th>
<th>8Y</th>
<th>2H(_6)</th>
<th>H(<em>6)YEHK(</em>{21})H(_6)</th>
<th>YEHK(_{21})</th>
<th>21E</th>
<th>21K</th>
<th>21H</th>
<th>21Y</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>12.9</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>12</td>
<td>32.9</td>
<td>20.9</td>
<td>-0.1</td>
<td>21</td>
<td>21</td>
<td>0</td>
</tr>
<tr>
<td>3.5</td>
<td>11.4</td>
<td>-0.9</td>
<td>-0.9</td>
<td>0</td>
<td>12</td>
<td>30.6</td>
<td>18.6</td>
<td>-2.4</td>
<td>21</td>
<td>21</td>
<td>0</td>
</tr>
<tr>
<td>5</td>
<td>5.2</td>
<td>-6.4</td>
<td>-6.4</td>
<td>0</td>
<td>11.6</td>
<td>15.2</td>
<td>3.6</td>
<td>-16.8</td>
<td>21</td>
<td>20.4</td>
<td>0</td>
</tr>
<tr>
<td>6.5</td>
<td>-2</td>
<td>-8</td>
<td>-8</td>
<td>0</td>
<td>6</td>
<td>-4.4</td>
<td>-10.4</td>
<td>-20.9</td>
<td>21</td>
<td>10.5</td>
<td>0</td>
</tr>
<tr>
<td>10</td>
<td>-12</td>
<td>-12</td>
<td>-8</td>
<td>-4</td>
<td>0</td>
<td>-31.5</td>
<td>-31.5</td>
<td>-21</td>
<td>10.5</td>
<td>0</td>
<td>-10.5</td>
</tr>
</tbody>
</table>

\(^a\) The prediction was based on the pKa values of the isolated amino acid residues using Protein Calculator (v3.3).

YE8 and YEHK21 have two functionally and sequentially different domains: the repetitive tracks (YEHK)\(_{21}\) and (YE)\(_8\) designed to form β-sheet structure and build up
the core of amyloid fibril and terminal domains H6 representing the unfolded part of proteins which are not involved in fibril formation. The charge distribution (Table 2) within different domains is schematically depicted on Figure 42.

**Figure 42.** Schematic representation of charge distribution in YEHK21 and YE8 at different pH. The square illustrates unit containing one strand and turn; the ovals stand for unfolded hexahistidinyl tracks. The red, blue and white colors represent the positively charged, negatively charged or neutral units.

At low pH, both polypeptides have high positive charge, while with the increase in pH, the positive charge uniformly diminishes by the diminishing of the number of positively charged amino acid residues and by the charge compensation by the negatively charged residues. The polypeptides become negatively charged at neutral pH and uniformly negatively charged at high pH. Depending on the solution pH and the repetitive sequence, the YEHK21 and YE8 polypeptides can have different charge distribution in the β-sheet forming repetitive tracks and terminal unfolded domains. They can have
combinations of positive and negative charges or only one type of charges. Importantly, the uniformly charged residues can be located i) only in repetitive β-sheet forming tracks (Figure 42, 5 and 11), ii) only in terminal sequences (Figure 42, 7), or iii) in both parts (Figure 42, 1).

Charge distribution within the β-sheet forming domain defines the electrostatic control of polypeptide folding and coacervation. Several important insights into the influence of electrostatic interactions within amyloid β-sheet core have been made. Firstly, a decrease in the net charge and subsequent decrease in electrostatic repulsion within the β-sheet forming domain results in aggregation of YE8 into well defined fibrils (Figure 42, 7 and 8), an observation in accord with previously reported studies. Secondly, reduction of the local electrostatic repulsion rather than a decrease in net charge also can enforce the amyloidogenic character. At pH 6.5, YEHK21 (Figure 42, 4) has significantly higher charge at the β-sheet edge than that in YE8 (Figure 42, 10) but easily aggregates at these conditions. Presumably aggregation can be achieved by the localization of negative charges at every other turn in YEHK21 in comparison with the adjunct charges located in every turn in the YE8 construct. Our hypothesis suggests that the correct arrangement of electrostatic interactions within polypeptides, rather than the reduction or elimination of these interactions leads to the amyloid fibril formation. Thus, it is important to consider amyloid fibrils as polyelectrolytes. The ratio of the polypeptide charge to length rather than solely the polypeptide net charge should be considered in polypeptide aggregation and coacervation. 3) Strong attractive electrostatic interactions in large macromolecular polypeptides commonly result in fast collapse and aggregation into amorphous assemblages. In contrast, small peptides form
ordered aggregates under these conditions. This can be illustrated by the fact that YEHK21 self-assemble into amorphous aggregates (Figure 42, 6) while YEHK21 where the charge is uncompensated (Figure 42, 4) forms ordered fibrils. Incubation of highly charged YEHK21 (Figure 42, 4) and YE8 (Figure 42, 7-10) in the presence of electrolytes led to the formation of unordered aggregates.

Charge distribution within the unfolded domain does not influence polypeptide behavior. Analysis of pH dependence of the YE8 aggregation revealed that a polypeptide having significant (Figure 42, 9) or even fully uncompensated (Figure 42, 8) positive charge on the terminal domains easily forms amyloid fibrils.

Electrostatic interactions between β-sheet forming and unfolded domains can affect polypeptide aggregation. Thus, strong attraction between highly charged β-sheet forming and unfolded domains in polypeptide YE8 (Figure 42, 9) leads to the polypeptide precipitation. Interestingly, small attractive electrostatic interactions between β-sheet forming domain and terminal unfolded parts can facilitate polypeptide aggregation. At pH 3.5, YE8 aggregates (Figure 42, 8) significantly faster then at pH 2 (Figure 42, 7) clearly showing the assistance of electrostatic interactions. However, if the net charge on the β-sheet forming domain is too high, electrostatic interactions with unfolded domains can not compensate the repulsion and polypeptide remains unfolded and monomeric (Figure 42, 10).

Amyloid fibril formation by the highly charged polypeptide chains was observed in both model constructs. The YE8 (Figure 42, 10) and YEHK21 (Figure 42, 1 and 5) have very significant charges in the β-sheet forming amino acid tracks polypeptides under specified conditions. However, both polypeptides form β-sheets and aggregate into
amyloid fibrils, although the process is rather slow. This result supports the capability of highly charged polypeptide chains to fold and aggregate into ordered structures. Moreover it illustrates the possibility of the cross-beta core of amyloid fibril to be comprised of the extremely charged polypeptide chains.

3.2.2.6 Implications to the deposition of intrinsically unfolded proteins

Our study of model polypeptides showed that the electrostatic interactions control the folding and aggregation state of polypeptide chain. Importantly, regardless of the molecular weight aggregation-driven only folding was observed for both polypeptides at all conditions reported. The revealed intermolecular assistance in the model polypeptide folding is highly relevant to the deposition of intrinsically disordered proteins and of a great importance. A proposed mechanism for intrinsically disordered protein deposition \(^ {13}\) includes partial refolding as a required initial step. *In vitro*, the induced folding is usually promoted by the global environmental changes such as variations in pH, ionic strength or temperature. \(^ {13,21}\) In *in vivo* environment, intermolecular interactions cause local environmental changes and can affect the development of induced folding and further fibrillation. \(^ {13}\)

The described models for the first time show that highly charged model polypeptides can fold into β-sheet structure and deposit as amyloid-like fibrils *in vitro* in the absence of any environmental factors such as polyelectrolytes, salts, etc. In the process, self-sufficiency of the polypeptide to promote folding and further self-assembly is revealed.
3.2.3 Effect of a β-strand number

3.2.3.1 Design of the polypeptide family

Three repetitive polypeptides that possess identical strands GAGAGA and differ only in the degree of polymerization of the 32 amino acid unit 32YEHK, [(GA)3GY(GA)3GE(GA)3GH(GA)3GK] have been constructed. The resultant polypeptides YEHK7 (18 kDa), YEHK14 (34 kDa) and YEHK21 (51 kDa) consist of seven, fourteen and twenty one repeats of the 32YEHK unit and are proposed to form up to 28, 56 and 84 intramolecular antiparallel β-strands repeats (Figure 43).

**Figure 43.** Schematic representation of the YEHK7, YEHK14 and YEHK21 polypeptides

3.2.3.2 Sample preparation

The polypeptides YEHK7, YEHK14, and YEHK21 were expressed and purified and carbamoylated as described in Chapter 1. Upon the dialysis against an aqueous pH 6.5 solution, carbamoylated YEHK21 was completely folded and existed in a gelatinous phase formed of amyloid-like fibrils (Chapter 2). Anticipating that the lower molecular
weight polypeptides have a diminished tendency to fold and aggregate under these conditions. 8M urea solutions containing the unfolded YEHK7, YEHK14, and YEHK21 polypeptides, with concentration normalized to 0.3 mg/mL, were dialyzed against double distilled water at pH 6.5. This allowed us to follow the progress of polypeptide folding and aggregation using sequentially smaller polypeptides. Upon dialysis, as mentioned above, the highest molecular weight polypeptide YEHK21 forms a gelatinous phase, while YEHK14 is only partially aggregated, and YEHK7 remains largely soluble.

3.2.3.3 Folding

The progression of model polypeptide folding was investigated by deep UV resonance Raman spectroscopy. The Analysis of Raman spectra of YEHK7, YEHK14 and YEHK21 (Figure 44) reveals significant difference in β-sheet contents of polypeptide samples.

![Resonance Raman spectra of S0 solutions of YEHK7 (blue curve), YEHK14 (red), YEHK21 (black). Spectra recorded by Dr. Vladimir V. Ermolenkov.](image)

**Figure 44.** Resonance Raman spectra of S0 solutions of YEHK7 (blue curve), YEHK14 (red), YEHK21 (black). Spectra recorded by Dr. Vladimir V. Ermolenkov.
The dominant contribution to the Raman spectrum of the YEHK21 polypeptide is derived from the β-sheet signature. The shortest YEHK7 polypeptide presents principally spectral features associated with a predominantly random coil conformation. The spectrum of YEHK14 has features from both a random coil and a β-sheet.

3.2.3.4 Fibril formation

Tapping-mode atomic force microscopy (TM-AFM) revealed that all polypeptides aggregate into non-branching fibrils (Figure 45).

![Figure 45](image)

**Figure 45.** TM-AFM topographs of YEHK7 (A, D), YEHK14 (B, E) and YEHK21 (C, F) fibrils.

The density of fibrils varied significantly depending on the polypeptide length. YEHK7 had only a few linear fibrils while the longer YEHK14 and YEHK21 polypeptides formed significant numbers of fibrils. The fibril density of the longest YEHK21 polypeptide was so high that the polypeptide solution had to be diluted three fold for
investigation of the fibril morphology (Figure 45 C,F). In addition to the difference in fibril density, the morphology of aggregates was different. Most of the YEHK7 assemblages represent unordered branched aggregates and only small amount of linear unbranched fibrils was detected. The majority of these fibrils were comprised of beaded assemblies of 2-5 nm in height which form curled chains (Figure 45D). The YEHK14 and YEHK21 polypeptides (Figure 45 E,F) deposited only as linear fibrils having two distinct morphologies. The first of these fibrils (referred to as type I fibrils) are similar to the beaded-chain assemblages observed in the YEHK7 sample. A second morphology (type II fibrils) is thin and smooth aggregates lacking irregularities or beaded features.

3.2.3.5 Gelatinous phase formation

The polypeptide solutions were centrifuged, and polypeptides existing in gelatinous (G) and aqueous fractions (F1) were separated in an effort to characterize the folded polypeptides and to define the relationship between polypeptide folding and aggregation. Upon centrifugation, a gelatinous phase formed with all samples however the volume of this phase decreases dramatically within the polypeptide series. The gelatinous phase (G) was washed\(^1\) with an equal volume of doubly distilled water to prepare second suspended fraction (F2). The initial aqueous fractions from centrifugation (F1) contained polypeptides in an unfolded state (Figure 46, red curves). However, the aqueous fractions F2, in the case of YEHK7 was enriched in $\beta$-sheet. However with YEHK14 and YEHK21, the polypeptides were completely folded (Figure 46, blue curves). Thus, irrespective of molecular weight and the progression of folding (Figure 46, green curves), the folded polypeptides congregated within the gelatinous phase suggesting the importance of intermolecular interactions on folding to form a $\beta$-
Moreover, the folded polypeptides can be washed from the gelatinous phase into an aqueous suspension while preserving the folded state.

**Figure 46.** Resonance Raman spectra of YEHKn polypeptides. Initial solution after dialysis (green), its aqueous fraction (F1, red) and solution with rendered aggregated polypeptides (F2, blue). Spectra recorded by Dr. Vladimir V. Ermolenkov.
The AFM characterization of the F2 fractions revealed that polypeptides in the gelatinous phase exist in fibrillar form as well (Figure 47).

Figure 47. AFM topographs of the YEHK7 (A), YEHK14 (B) and YEHK21 (C) aggregates rendered from gelatinous phase.

Type I fibrils were observed in all samples, (Figure 47) albeit the fibrils appeared to be substantially shorter than those described in Figure 3. Interestingly, thin and smooth type II fibrils were present in all three samples, being dominant in YEHK21 samples. The increase in abundance of the type II fibrils within the F2 samples in comparison with F1 samples may be a result of the greater stability of type II fibrils. Polypeptide samples can be enriched in type II fibrils simply by extraction from gelatinous phase. In contrast to the initial solution of YEHK7 (Figure 45A), unordered branched assemblages were not observed in the β-sheet enriched samples of YEHK7 (Figure 47A). This observation suggests that aggregation of unfolded polypeptides may be responsible for the formation of the beaded aggregates.
3.2.3.6 Polymorphism in YEHKn aggregation

Prospective models of the two polymorphs of the folded YEHKn aggregates, the thin and smooth Type I fibrils and the beaded Type II chains, are shown in Figure 48.

**Figure 48.** The models of the polypeptide chain organization. One color represents one polypeptide chain. A) Fibrils with pronounced beaded morphology, B) very smooth and linear fibrils, C) AFM topograph of the YEHK21 fibrils.

The type I fibrils have been proposed to have a less developed cross-β fibrillar core and constructed of partially unfolded polypeptide chains. In contrast, the type II fibrils have very well organized cross-β core formed by completely folded polypeptides that result in a very smooth linear fibril morphology. In the proposed models, the cross-β cores of fibrils consists of the tracks of continuous β-sheet structures formed by a single polypeptide chain. The repetitive nature of the polypeptides eliminates the sequence specificity of local interactions. The rate of intramolecular β-sheet propagation can be faster than intermolecular assisted folding, consequently β-sheet formation is faster for the longer polypeptides. Formation of the type II fibrils would likely be affected more by the polypeptide molecular weight on consideration of the longer intramolecular β-
sheet tracks involved in fibril formation. Indeed, the type I fibrils are formed by all YEHKn polypeptides, while the type II fibrils appear only on aggregation of the longer polypeptides YEHK14 (minor) and YEHK21 (major).

3.2.3.7 Role of hydrophobic interactions. Model of the type II fibril

The tentative model of the type II fibrils derived from the AFM and TEM characterization (Chapter II) has been proposed as an assemblage of four YEHK21 molecules (Figure 49).

![Figure 49](image)

**Figure 49.** Proposed anti-parallel $\beta$-sheet formed by four polypeptide molecules.

The model defines the importance of the hydrophobic interactions on the polypeptide chain aggregation. On folding the YEHK polypeptides $\beta$-sheets have the hydrophobic and hydrophilic edges (Figure 16, p. 43) with the amphiphilic surfaces functionalized either by alanine or glycine. The fibrillar assemblage is likely stabilized by significant shielding of the hydrophobic surfaces from the aqueous environment. The proposed model is illustrative of how a polypeptide chain with generally low hydrophobicity folds into $\beta$-sheet structure and in the process generates surfaces with the increased
hydrophobic character capable of promoting aggregation. The latter ability may be related to the amyloidosis of intrinsically disordered proteins.

3.2.3.8 Amyloid fibril “sticky ends”

Taking into account the involvement of four polypeptide molecules into the fibrillar assemblage (Figure 49) and observed very smooth and uniform morphology of fibrils, we propose that the fibril growth occurs by the addition of an unfolded monomer to the fibril ends rather than by a lateral assembly of protofibrils or other aggregated species. Considering the low probability of the perfect alignment of all four polypeptide molecules, the overhanging domain of the folded polypeptide at the end of fibril forms a “sticky end” that facilitates the aggregation and folding of an incoming molecule (Figure 50).

Figure 50. A proposed model of a monomer addition.

The “sticky ends” with a protein specific pattern of hydrophobicity and charge distribution may be reproduced by incoming protein preserving the existing molecular organization and may be relevant to the existence of cross-species barrier and polymorphism in amyloid fibrils.63, 64

3.2.4 Structural characterization of amyloid-like fibrils

The structural elucidating of the cross-β core of large amyloid fibrils is a challenging problem in modern structural biology. A set of de novo designed
polypeptides was genetically engineered to provide amyloid-like fibrils with similar morphology and yet different strand lengths. Three *de novo* designed polypeptides consisting of seven repeats with 32, 40 and 48 amino acid units, \((\text{GA})_n\text{GY(GA)}_n\text{GE(GA)}_n\text{GH(GA)}_n\text{GK} \quad (n = 3, \ 4 \ \text{or} \ 5: \ 32\text{YEHK7}, \ 40\text{YEHK7} \ \text{and} \ 48\text{YEHK7}, \ \text{respectively})\) have identical turn architecture and differ only in the strand length deriving from the numbers of GA repeats in each strand, \((\text{GA})_3, \ (\text{GA})_4 \ \text{and} \ (\text{GA})_5 \) (Figure 51).

**Figure 51.** Hypothetical scheme of the YEHK fibril cross-\(\beta\) core.

TEM images of the re-dispersed form of polypeptides from the gelatinous phase show the formation of typical fibrillar structures (Figure 52).
Analysis of the TEM data revealed that the average fibril width of the polypeptides is consistent with the cross-β core structure shown in Figure 51.

Table 3. The average width of amyloid fibrils formed by polypeptides with different repetitive units.

<table>
<thead>
<tr>
<th></th>
<th>32YEHK (32YEHK21)</th>
<th>40YEHK (40YEHK7)</th>
<th>48YEHK7 (48YEHK7)</th>
</tr>
</thead>
<tbody>
<tr>
<td>6.5±1 nm</td>
<td>7.8 ±1 nm</td>
<td>10 ±1 nm</td>
<td></td>
</tr>
</tbody>
</table>

The solutions of fibrils were analyzed by deep UV resonance Raman spectroscopy (Figure 53). The spectra of all samples were also indicative of the β-sheet structure conformation. The Raman signature of core β-strands was the same for all three types of fibrils strongly suggesting similar structure.
Differential ultraviolet Raman spectroscopy of repetitive polypeptides allowed the separation of the spectroscopic signatures of the highly ordered \( \beta \)-sheet strands forming the cross-\( \beta \) core of amyloid fibrils and, for the first time, the Raman signature of fibril turns (Figure 54).

**Figure 53.** DUVRR spectra of YEHK fibrils. Spectra recorded by Dr. Vladimir V. Ermolenkov.

**Figure 54.** The Raman signatures of a \( \beta \)-sheet and turns obtained from the differential spectra of 32YEHK7, 40YEHK7 and 48YEHK7 polypeptides. Spectra obtained by Vitali Sikirzhytsky.
The first two examples of cross-β core Raman signatures, those obtained for YEHK polypeptides and lysozyme, indicate the very ordered, crystalline–like structure of the anti-parallel β-sheet. The data obtained on the Ramachandran ψ angle differs in the YEHK and lysozyme fibrillar β-sheets. To the best of our knowledge this is the first direct evidence that the structure of the fibril cross-β core is a sequence dependent. Thus we can speculate that the controlled tuning of the polypeptide sequence by genetic engineering is a great tool for studies of the relationship between the sequence and the cross-β sheet structure.
3.3 Conclusions

The macromolecular polypeptide-based models have a repetitive nature that eliminates sequence specificity in inter- and intramolecular interactions yet vary in molecular weight (number of \(\beta\)-strands), turn and strand architecture, charge location, and hydrophobicity. Structural characterization of amyloid fibrils revealed the importance of non-specific aggregation in the fibrillation of intrinsically unfolded proteins, illustrated the influence of electrostatic interactions on intrinsically unfolded protein aggregation, and exposed the importance of the charge distribution rather than net charge only. We have proposed a model of amyloid fibril containing “sticky ends” which have a protein specific pattern of hydrophobicity and charge distribution. This sequence specific scaffold can be reproduced by every incoming protein and may be relevant to the existence of cross-species barrier and polymorphism in protein amyloidosis.

Part of this work has been published:


3.4 References


(27) Chiti, F.; Bucciantini, M.; Capanni, C.; Taddei, N.; Dobson, C. M.; Stefani, M. Solution conditions can promote formation of either amyloid protofilaments or mature fibrils from the HypF N-terminal domain. *Protein Sci* **2001**, *10*, 2541-2547.


(54) Wouters, M. A.; Curmi, P. M. G. An analysis of side chain interactions and pair correlations within antiparallel beta-sheets: the differences between backbone hydrogen-