

BIOSYNTHESIS AND GENETIC SELECTION OF OLIGOPEPTIDES
WITH POTENTIALLY THERAPEUTIC EFFECTS AGAINST ALZHEIMER'S DISEASE:
INHIBITORS OF AMYLOID- β AGGREGATION



Ilias Matis

DOCTORAL THESIS

Supervising Professor: F. Kolisis

National Technical University of Athens
School of Chemical Engineering, Biotechnology Laboratory

2018

Approval of the doctoral thesis from the School of Chemical Engineering of the National Technical University of Athens does not constitute acceptance of the author's opinions.

Η έγκριση της διδακτορικής διατριβής από την Ανώτατη Σχολή Χημικών Μηχανικών του Εθνικού Μετσόβιου Πολυτεχνείου δεν υποδηλώνει αποδοχή των γνώμων του συγγραφέα.

(N. 5343/1392, Άρθρο 202)

Supervising Committee (Τριμελής Συμβουλευτική Επιτροπή)

Professor Emeritus,
National Technical University of Athens (Supervisor)

Fragiskos Kollis

Research Associate Professor,
National Hellenic Research Foundation

Georgios Skretas

Associate Professor,
University of Athens

Dimitris G. Hatzinikolaou

Examination Committee (Επταμελής Εξεταστική Επιτροπή)

Professor Emeritus,
National Technical University of Athens (Supervisor)

Fragiskos Kollis

Research Associate Professor,
National Hellenic Research Foundation

Georgios Skretas

Associate Professor,
University of Athens

Dimitris G. Hatzinikolaou,

Professor,
National Technical University of Athens

Gerasimos Lyberatos

Assistant Professor,
National Technical University of Athens

Evangelos Topakas

Professor,
University of Athens

Spiros Efthimiopoulos

Professor,
Universitat Autònoma de Barcelona

Salvador Ventura

Contents

Acknowledgements:.....	7
Περίληψη.....	9
Abstract.....	12
Publications.....	14
1. Introduction	15
1.1. Protein folding	15
1.2. Protein misfolding alters the free energy landscape.....	17
1.3. Protein aggregation	18
1.3.1. Aggregation and amyloid formation.....	20
1.3.2. Protein misfolding, aggregation and related diseases.....	22
1.4. Alzheimer’s disease: a brief history.....	23
1.5. Association of Alzheimer’s disease with the amyloid-β peptide	25
1.6. The Amyloid Precursor Protein (APP)	26
1.7. The amyloid cascade hypothesis	27
1.8. Re-evaluation of the amyloid cascade hypothesis.....	28
1.8.1. Findings concerning APP processing.....	29
1.8.2. Findings in support of the amyloid cascade hypothesis	30
1.9. AD pathogenesis	33
1.9.1. Toxicity-inducing Aβ variants.....	33
1.9.2. Toxicity related to oligomeric and aggregated Aβ species	33
1.9.3. Alternative causes of Aβ-induced toxicity	37
1.10. Aggregation kinetics.....	37
1.11. Structural data for Aβ oligomers and aggregates.....	39
1.12. Aβ-related therapeutic strategies against AD.....	40
1.12.1. BACE1 and BACE2 inhibitors	41
1.12.2. γ-Secretase inhibitors	42
1.12.3. Inhibition of Aβ aggregation	43
1.12.4. Enhancing Aβ clearance	43
1.12.5. Immunotherapeutic strategies against Aβ aggregation	44
1.12.6. Preclinical treatment.....	45
1.12.7. Drug discovery pipeline for AD	45
1.12.8. Aβ-targeting small molecules	47
1.13. Microbial screens for the detection of Aβ aggregation inhibitors.....	50

1.13.1. Dye-based bacterial screens for aggregation inhibitors	50
1.14. Increasing the diversity of chemical libraries in microbial screens	55
1.15. Synthetic biology in the discovery of cyclic peptide pharmaceuticals	56
1.15.1. General introduction to peptide pharmaceuticals	56
1.15.2. Biosynthesis of cyclic peptides.....	57
1.16. Thesis outline	63
Chapter 2. Screening of a biosynthetic cyclic peptide library and identification of bioactive hits with inhibitory effects against the aggregation of A β	65
2.1. <i>In vivo</i> generation of a diverse library of cyclic peptides.....	65
2.1.1. A SICLOPPS-generated cyclic peptide library	65
2.2. Identification of cyclic peptide inhibitors of A β aggregation.....	69
2.2.1. FACS-enabled screening of the combined library of cyclic peptides.....	73
2.2.2. Isolation and evaluation of individual cyclic peptide-producing clones from the sorted pool	75
2.2.3. Individual peptide clones increase the soluble fraction of A β ₄₂ -EGFP.....	78
2.3. Selection of two cyclic peptides for further evaluation.....	82
2.3.1. The cyclo-SASPT and cyclo-TAFDR effect is A β -specific.....	82
2.3.2. The cyclo-SASPT and cyclo-TAFDR effect on A β isoforms/variants of A β	84
2.4. Remarks	85
Chapter 3. Effect of the selected peptides on A β aggregation and neurotoxicity	86
3.1. <i>In vitro</i> evaluation of the effect of the selected peptides on A β aggregation.....	86
3.1.1. Evaluation of the effect of the selected peptides on A β aggregation using circular dichroism spectroscopy and thioflavin T staining.....	86
3.1.2. Evaluation of the effect of the selected peptides on A β aggregation using transmission electron microscopy.....	89
3.1.3. Evaluation of the effect of the selected peptides on A β aggregation using dynamic light scattering.	89
3.2. Evaluation of the effect of the selected peptides on A β -induced neurotoxicity <i>in vitro</i>	93
3.2.1. Evaluation of the effect of the selected peptides on A β -induced neurotoxicity in primary mouse hippocampal neurons	93
3.2.2. Evaluation of the effect of the selected peptides on A β -induced neurotoxicity in neuronal cell lines.....	95
3.2.3. Evaluation of the effect of the selected peptides on the neurotoxicity caused by naturally secreted A β oligomers	96
3.2.4. Evaluation of the effect of the selected peptides on the binding of A β ₄₂ aggregates to the neuronal surface	97
3.3. Evaluation of the effect of the selected peptides on A β aggregation and neurotoxicity <i>in vivo</i>	98

3.4. Remarks	102
Chapter 4. Structure-activity analysis of the selected A β -targeting cyclic peptides	104
4.1. Identification of residues critical for bioactivity in A β C5-34 and A β C5-116 by site-directed mutagenesis	104
4.2. High-throughput sequencing of the sorted peptide library	107
4.2.1. Identification of valid cyclic peptide-encoding DNA sequences	108
4.2.2. Cyclic-peptide nomenclature	110
4.2.3. Initial observations on NGS results	110
4.2.4. Sequence read numbers and amino acid composition analyses	113
4.2.5. Identification of distinct pentapeptide motifs within the sorted peptide library	116
4.2.6. Identification of distinct hexapeptide motifs within the sorted peptide library	158
4.2.7. Investigation of the tetrapeptide pool of sequences	166
4.3. Remarks on structure-activity analysis	168
Chapter 5. Discussion: remarks and perspectives	170
5.1. Perspective and future work.....	172
5.1.1. Cyclic peptides as anti-aggregation agents.....	172
5.1.2. Further study and development of A β C5-34 and A β C5-116	173
5.1.3. Evaluation of additional selected cyclic peptides	175
5.1.4. Generation of secondary peptide libraries	176
5.2. Conclusion.....	176
Chapter 6. Methods and materials	178
6.1. Chemicals, reagents, kits and media.....	178
6.2. Bacterial cell transformation	178
6.3. Library construction	179
6.4. Expression vector construction.....	181
6.5. Cyclic peptide library screening.	183
6.6. Isolation of individual bacterial clones from the sorted peptide library	184
6.7. Bacterial cell fluorescence	185
6.8. Protein electrophoresis and western blot analysis.....	186
6.9. Preparation of synthetic A β ₄₀ and A β ₄₂ stocks and solutions	187
6.10. Recombinant production and purification of A β ₄₂	187
6.11. Circular dichroism measurements	189
6.12. Thioflavin T staining of A β fibril formations.....	189
6.13. Transmission electron microscopy	189
6.14. Dynamic light scattering	190

6.15. Neuronal cell cultures	190
6.16. Cell viability measurements.....	192
6.17. Immunocytochemistry	192
6.18. <i>In vivo</i> assays in <i>C. elegans</i>	193
6.18.1. Treatment of <i>C. elegans</i> nematodes with selected cyclic peptides.....	194
6.18.2. Analysis of cyclic-peptide-treated <i>C. elegans</i> nematodes	195
6.19. High-throughput sequencing analysis.....	196
6.20. Statistical analyses	197
Appendix	198
Supplementary Tables	198
Supplementary Figures	216
References	225

Acknowledgements:

The presented doctoral thesis was performed in the lab of Enzyme and Synthetic Biotechnology at the Institute of Biology, Medicinal Chemistry & Biotechnology of the National Hellenic Research Foundation, led by Res. Assoc. Prof. Georgios Skretas. This work was funded by the project entitled “BIOSYNTHESIS AND GENETIC SELECTION OF CYCLIC PEPTIDES WITH POTENTIAL THERAPEUTIC EFFECTS AGAINST ALZHEIMER'S DISEASE: INHIBITORS OF A β AGGREGATION” (Acronym: CYCLIPAD) financed by the Hellenic Ministry of Education, Research and Religious Affairs and the National Strategic Reference Framework (“ΘΑΛΗΣ-CYCLIPAD”-Cod: 657 Mis.380042).

Firstly, I would like to thank the members of my supervising committee, Prof. Emeritus Fragiskos Kolisis, Res. Assoc. Prof. Georgios Skretas and Assoc. Prof. Dimitris G. Hatzinikolaou for trusting me with this demanding task, for their guidance, the extended conversations and their counsel. I would especially like to thank Res. Assoc. Prof. Georgios Skretas for his daily supervision, patience and persistence, from the beginning until the end of this work. I would also like to extend my acknowledgments to Prof. Gerasimos Lyberatos (NTUA), Assis. Prof. Evangelos Topakas (NTUA), Prof. Spiros Efthimiopoulos (UOA) and Prof. Salvador Ventura (UOB) for kindly accepting to participate in the examination committee.

I gratefully acknowledge Res. Assoc. Prof. V. Papadimitriou, and Res. Prof. A. Xenakis for allowing me to perform dynamic light scattering experiments, as well as Res. Assoc. Prof. E. Chrysinia for the use of protein purification facilities. I also would like to express my sincere gratitude to Barbara Mavroiedi (NCSR, “Demokritos”, Lab of Res. Prof. M. Pelekanou), Nikole Papaevgeniou (NHRF, Lab of Res. Assoc. Prof. N. Chondrogianni), Res. Assoc. Prof. K. Vekrellis (Biomedical Research Foundation of the Academy of Athens, BRFAA), Zacharoula Linardaki and Alexandra Stavropoulou (Lab of Prof. S. Efthimiopoulos, UOA). Their *in vitro* and *in vivo* work was critical in validating the findings of this work. Their contribution is explicitly noted in the corresponding chapter. Finally yet importantly, I would like to thank the members of the Enzyme & Synthetic Biotechnology lab, Daphne Delivoria who

constructed and characterized the library that was utilized herein, Stefania Panoutsou, Dimitra Gialama, Dimitra Zarafeta, Evgenia Megalou, Panagiota Stathopoulou, Lina Bellou and Michael Fasseas, for their valuable assistance and daily contributions. They made this experience a memorable one.

Importantly, I am particularly thankful to my family and partner, Eleftheria. Their continuous support and patience significantly facilitated this undertaking.

Περίληψη

Η νόσος Alzheimer αποτελεί την πλέον ολέθρια νευροεκφυλιστική νόσο και την πιο διαδεδομένη μορφή άνοιας. Μέχρις στιγμής, υποψήφιες ενώσεις για την αντιμετώπιση της νόσου δεν έχουν επιδείξει ικανότητα αναχαίτησης της προόδου της νόσου, με αποτέλεσμα την αυξημένη αναγκαιότητα εύρεσης νέων θεραπευτικών ενώσεων. Η παθογένεια της νόσου Alzheimer έχει συσχετιστεί με το αμυλοειδές-β (Αβ), ένα πεπτίδιο 37-43 αμινοξέων το οποίο εμφανίζει εγγενή τάση σχηματισμού ολιγομερών και αδιάλυτων συσσωματωμάτων με νευροτοξικές ιδιότητες. Κατά συνέπεια, η ανακάλυψη μικρομοριακών βιολογικών αναστολέων της συσσωμάτωσης του Αβ αποτελεί μία ελκυστική θεραπευτική προσέγγιση για την αντιμετώπιση της νόσου, η οποία έχει αποτελέσει βασική επιδίωξη της φαρμακευτικής βιομηχανίας.

Η συγκεκριμένη διδακτορική διατριβή περιγράφει το σχεδιασμό και ανάπτυξη ενός βακτηριακού συστήματος το οποίο δύναται να λειτουργήσει ως πλατφόρμα ανακάλυψης βιοδραστικών ενώσεων, με αποτρεπτική δράση ως προς τη συσσωμάτωση του Αβ, καθώς και τη συνεπαγόμενη νευροτοξικότητά του. Για αυτό το σκοπό, βιβλιοθήκες κυκλικών πεπτιδίων παράγονται βιοσυνθετικά σε κύτταρα *E. coli* ενώ παράλληλα υποβάλλονται σε διαλογή με κριτήριο την ικανότητά τους να παρεμποδίζουν τη συσσωμάτωση του Αβ, χρησιμοποιώντας μια γενετική δοκιμασία υψηλού ρυθμού απόδοσης. Η εν λόγω γενετική δοκιμασία παρέχει τη δυνατότητα εντοπισμού βιοδραστικών κυκλικών πεπτιδίων τα οποία αυξάνουν το επίπεδο φθορισμού της πρωτεϊνικής χίμαιρας του Αβ με την πράσινη φθορίζουσα πρωτεΐνη (Green Fluorescence Protein) GFP. Η παραγωγή της χίμαιρας και των βιβλιοθηκών των κυκλικών πεπτιδίων πραγματοποιούνται παράλληλα σε τροποποιημένα βακτήρια, ενώ η διαλογή των μεμονωμένων φθορίζοντων κλώνων επιτελείται με τη χρήση κυτταρομετρίας ροής (fluorescence-activated cell sorting, FACS).

Το τροποποιημένο βακτηριακό σύστημα που χρησιμοποιήθηκε, επέτρεψε τη βιοσύνθεση και τη γρήγορη διαλογή 10 εκατομμυρίων κυκλικών πεπτιδίων, καθώς και την επιλογή εκατοντάδων βιοδραστικών πεπτιδίων με εν δυνάμει ρυθμιστικό ρόλο έναντι της συσσωμάτωσης του Αβ. Από

αυτά, δύο πεπτίδια που έφεραν τις αλληλουχίες cyclo-SASPT και cyclo-TAFDR και ονομάστηκαν AβC5-34 και AβC5-116 αντίστοιχα, παρήχθησαν μέσω σύνθεσης και χρησιμοποιήθηκαν για τη μελέτη της δράσης τους πάνω στη συσσωμάτωση του Αβ, μέσω βιοχημικών, βιοφυσικών και βιολογικών δοκιμασιών, *in vitro*. Οι δοκιμασίες αυτές φανέρωσαν ότι τα συγκεκριμένα κυκλικά πεπτίδια παρεμβάλλονται στη φυσιολογική συσσωμάτωση του Αβ, οδηγώντας στο σχηματισμό μη τυπικών συσσωματώσεων οι οποίες εμφανίζουν περιορισμένη νευροτοξικότητα. Επιπλέον, η προστατευτική δράση των AβC5-34 και AβC5-116 έναντι της συσσωμάτωσης και κυτταροτοξικότητας του Αβ αξιολογήθηκαν *in vivo* σε διαγονιδιακά στελέχη *Caenorhabditis elegans* που παράγουν Αβ. Οι δοκιμασίες αυτές έδειξαν πως τα πεπτίδια AβC5-34 και AβC5-116 παρουσιάζουν προστατευτικές ιδιότητες έναντι της συσσωμάτωσης και της κυτταροτοξικότητας του Αβ. Συνεπώς, τα πεπτίδια AβC5-34 και AβC5-116 παρουσιάζουν φαρμακολογικό ενδιαφέρον ενώ η περαιτέρω αξιολόγηση και βελτιστοποίηση τους θα μπορούσε να οδηγήσει σε ενώσεις με θεραπευτικές ιδιότητες έναντι της νόσου Alzheimer. Τέλος, ο συνδυασμός στοχευμένης μεταλλαξιγένεσης και ανάλυσης DNA νέας γενιάς, επέτρεψε την ανακάλυψη των σχέσεων δομής και δραστηριότητας για τον πληθυσμό των επιλεγμένων κυκλικών πεπτιδίων, καθώς και την ταυτοποίηση διακριτών οικογενειών βιοδραστικών πεπτιδίων.

Η γενετικά τροποποιημένη βακτηριακή πλατφόρμα που παρουσιάζεται στη συγκεκριμένη διδακτορική διατριβή εμφανίζει ορισμένα σημαντικά πλεονεκτήματα. Πρώτον, επιτρέπει εύκολη και αποτελεσματική διαλογή μοριακών βιβλιοθηκών. Δεύτερον, η ενσωμάτωση μιας μεθόδου διαλογής που στηρίζεται στο FACS, επέτρεψε γρήγορη διαλογή με αυξημένη απόδοση σε βιοδραστικά μόρια, συγκριτικά με τις συνηθισμένες τεχνικές διαλογής σε τρυβλία. Τρίτον, η προσέγγιση αυτή παρακάμπτει την ανάγκη για γνώση δομικών χαρακτηριστικών ή για την προετοιμασία συγκεκριμένων συσσωματωμάτων της Αβ. Τέλος, το βακτηριακό σύστημα που χρησιμοποιήθηκε έχει σχεδιαστεί με στόχο την παρακολούθηση της αναδίπλωσης και της συσσωμάτωσης της πρωτεΐνης-στόχου. Αυτό είχε ως αποτέλεσμα η εργασία που παρουσιάζεται σε αυτή τη διατριβή να λειτουργήσει ως βάση για τη χρήση αυτού του συστήματος για άλλες πρωτεΐνες-στόχους με προβληματική

αναδίπλωση, οι οποίες έχουν συσχετιστεί με τις ανάλογες νόσους. Τρέχον αλλά και δημοσιευμένο ερευνητικό έργο από το εργαστήριο Ενζυμικής & Συνθετικής Βιοτεχνολογίας του Εθνικού Ιδρύματος Ερευνών, έχει αποδείξει πως η συγκεκριμένη πλατφόρμα μπορεί να χρησιμοποιηθεί στην ανακάλυψη κυκλικών πεπτιδίων με βιοδραστικότητα έναντι ποικίλων πρωτεϊνικών στόχων, όπως τη μεταλλαγμένη δισμουτάση του υπεροξειδίου (SOD1), τη μεταλλαγμένη p53 και την πολυ-Q χαντινγκτίνη (polyQ-Htt).

Abstract

Alzheimer's disease (AD) is the most devastating neurodegenerative disease and the most common form of dementia. So far, AD drug candidates have been deemed unable to halt disease progression and, thus, effective therapeutics are in enormous demand. The pathogenesis of AD has been strongly linked to amyloid β ($A\beta$), a peptide composed of 37-43 amino acids with the tendency to form neurotoxic oligomeric and aggregated structures. As a result, the discovery of small-molecule and biological inhibitors of $A\beta$ aggregation presents an attractive therapeutic approach for the treatment of AD, which has been pursued aggressively by the pharmaceutical industry.

The work presented herein describes the development of engineered bacterial cells with the ability to function as a discovery platform for chemical rescuers of $A\beta$ misfolding and neurotoxic aggregation, with potentially therapeutic properties against AD. In this system, combinatorial libraries of cyclic peptides with extended chemical diversities are biosynthesized in *Escherichia coli* cells and simultaneously screened for their ability to inhibit the aggregation of $A\beta$, using an ultrahigh-throughput genetic assay. This assay detects bioactive cyclic peptides that enhance the fluorescence of a chimeric $A\beta$ fusion with the green fluorescent protein (GFP). Both the fusion and the cyclic oligopeptide libraries under investigation are co-expressed in the engineered bacteria, and selection of individual, fluorescent bacterial clones is performed via fluorescence-activated cell sorting (FACS).

By using this engineered system, a library of more than 10 million different short cyclic peptides was biosynthesized and rapidly screened, and hundreds of bioactive peptides with putative modulatory effects on $A\beta$ aggregation were identified. From the isolated hits, two cyclic pentapeptides with sequences cyclo-SASPT and cyclo-TAFDR, termed $A\beta$ C5-34 and $A\beta$ C5-116 respectively, were acquired in isolated form, following chemical synthesis, and their effects on $A\beta$ aggregation were investigated in a variety of biochemical, biophysical and biological assays *in vitro*. These analyses revealed that the selected cyclic peptides interfered with the normal course of $A\beta$ aggregation, leading to the formation of atypical $A\beta$ aggregates that exhibited reduced neurotoxicity compared to that of typical $A\beta$

aggregates. The protective effects of A β C5-34 and A β C5-116 against the aggregation and cytotoxicity of A β were also evaluated *in vivo*, using established transgenic *Caenorhabditis elegans* models of AD, expressing human A β . These assays revealed that A β C5-34 and A β C5-116 exerted a protective effect against the aggregation and cytotoxicity of A β . Thus, A β C5-34 and A β C5-116 constitute attractive lead molecules, which upon further evaluation can aid in the development of effective anti-amyloid therapies against AD. Finally, a combination of site-directed mutagenesis and next-generation sequencing analysis of the hits derived from screening, allowed the discovery of structure-activity relationships within the selected peptide population and revealed the existence of distinct families of bioactive cyclic peptides.

The herein engineered bacterial platform exhibits a number of important advantages. First, it enables the easy and efficient screening of molecular libraries with enhanced chemical diversities. Second, the incorporation of a FACS-based screening assay allowed for rapid selection with enhanced hit-yields, compared to conventional plate-based screening assays. Third, it bypasses the need for any prior structural knowledge or preparation of specific aggregated states of A β . Finally, since the utilized genetic system is designed to monitor protein misfolding and aggregation, the present work serves as a proof-of-concept for other misfolding-prone protein targets associated with conformational diseases. Indeed, published and ongoing work at the Laboratory of Enzyme & Synthetic Biotechnology of the National Hellenic Research Foundation has demonstrated that the same approach can be successfully applied in the discovery of cyclic peptide rescuers for a variety of disease-associated proteins, such as the mutant Cu/Zn superoxide dismutase, mutant p53 and polyglutaminated huntingtin.

Publications

1. Matis, I. *et al.* An integrated bacterial system for the discovery of chemical rescuers of disease-associated protein misfolding. *Nat. Biomed. Eng.* **1**, 838-852 (2017).
2. Kostelidou, K., Matis, I. & Skretas, G. Microbial genetic screens for monitoring protein misfolding associated with neurodegeneration: tools for identifying disease-relevant genes and for screening synthetic and natural compound libraries for the discovery of potential therapeutics. *Curr. Pharm. Des.* (E-pub ahead of press) (2018).

Manuscripts in preparation at the time of writing:

1. Delivoria, D.C., Chia S., Habchi, J., Perni, M., Papaevgeniou, N., Matis, I., Chondrogianni, N., Dobson, C.M., Vendruscolo, M., Skretas, G. 2018. Bacterial biosynthesis of a 200-million-member macrocycle library and direct functional screening for the discovery of rescuers of amyloid β misfolding and aggregation.
2. Delivoria, D.C., Papaevgeniou, N., Matis, I., Chondrogianni, Skretas, G. 2018. A bacterial discovery platform uncovers different families of peptide macrocycles against A β aggregation and neurotoxicity.

1. Introduction

1.1. Protein folding

Protein folding is a seminal factor in protein function and therefore, unsurprisingly, the mechanisms that govern protein folding have been the object of extensive research for decades. The discovery of reversible protein folding implied that the native conformation of a protein is guided by its inherent tendency to achieve the lowest free energy¹. Consequently, it was suggested that the amino acid sequence was responsible for the final polypeptide conformation. This suggestion, although true to an extent, omits the effect of numerous factors on protein folding.

During the past five decades, a number of efforts to explain protein folding have been made. The models that were proposed were developed in an effort to reach a compromise between the vast conformational variation of an amino acid sequence with the efficiency and swiftness of protein folding. The random acquisition of conformations towards the global free-energy minimum for a protein would require unrealistic amounts of time even for a relatively small protein (Levinthal's paradox). However, protein folding is completed within milliseconds or seconds². The first models for protein folding posited that in order to achieve folding in short periods, it was necessary for the conformational space to be restricted. The nucleation-growth model^{3,4}, the diffusion-collision model⁵ and the jigsaw-puzzle model⁶ were used to explain the above paradox, however efforts were eventually shifted towards an approach that uses potential energy biases to explain the transition to the native state⁷. This approach posits that the guiding force of protein folding is the free-energy landscape, composed of the potential energy and the configurational entropy². Simultaneously, this approach introduces the principle of minimal frustration, which suggests that the protein sequence has evolved into one where the unfavourable amino acid interactions are reduced during the folding process. Combined with the thermodynamic stability provided by the native interactions, this principle allows folding to occur in biologically relevant times. The free energy landscape can be illustrated in

the popular form of a folding funnel. Folding funnels correlate the potential energy to the configurational entropy for every protein conformation within the structure ensemble. As more native interactions form, they reduce the potential energy as well as the configurational entropy, and folding moves towards the native state^{2, 8-10}. (**Figure 1.1**).

The shape of the free energy landscape of a particular protein is indicative of its folding pathway. Proteins that show a smooth transition between the unfolded and the native states, also produce smooth folding funnels¹¹. However, most protein folding pathways require the formation of intermediates, giving rise to more rugged landscapes that contain multiple valleys, resultant of local energy minima^{9, 10}. Energetically favourable conformations that arise as products of misfolding can also contribute to the free energy landscape of certain proteins.

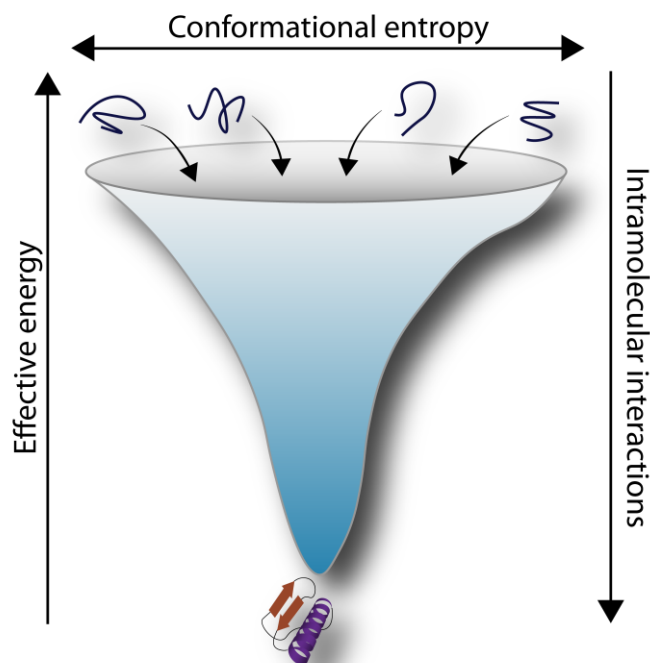


Figure 1.1. Free energy landscape. A vast number of unfolded protein conformations (upper funnel) transition to the native state (lower funnel), by following the principle of minimal frustration. This representation is idealized. Actual proteins rarely transition as smoothly to their native state. Figure adapted and redrawn from Karplus, 2011².

While the native state is usually located at the lowest end of the free energy landscape, a protein's ability to remain at this low energy conformation varies largely. Many proteins tend to alternate between close energy minima, simultaneously changing their conformation. While the principle of minimal frustration appears to describe the transition from the unfolded to the native state, local deviations from this principle (frustrations) offer the flexibility needed for various protein tasks. Ligand binding, allosteric activation, formation of heterodimeric complexes and other tasks, rely on energetic frustrations that enable the protein to participate in such tasks^{12, 13}.

1.2. Protein misfolding alters the free energy landscape

However, changes in parameters such as the pH, temperature, and mutations, can alter the free energy landscape of a protein, leading to unfolded or misfolded conformations¹⁴. The energy minima that occur following these changes can sometimes be thermodynamically favourable^{15, 16}. In some occasions the new energy minima involve protein conformation that will make the monomeric protein more prone to interact with other monomers or even other proteins. This interaction can then lead to further oligomerization of the protein and in the creation of aggregated species of various sizes **(Figure 1.2)**.

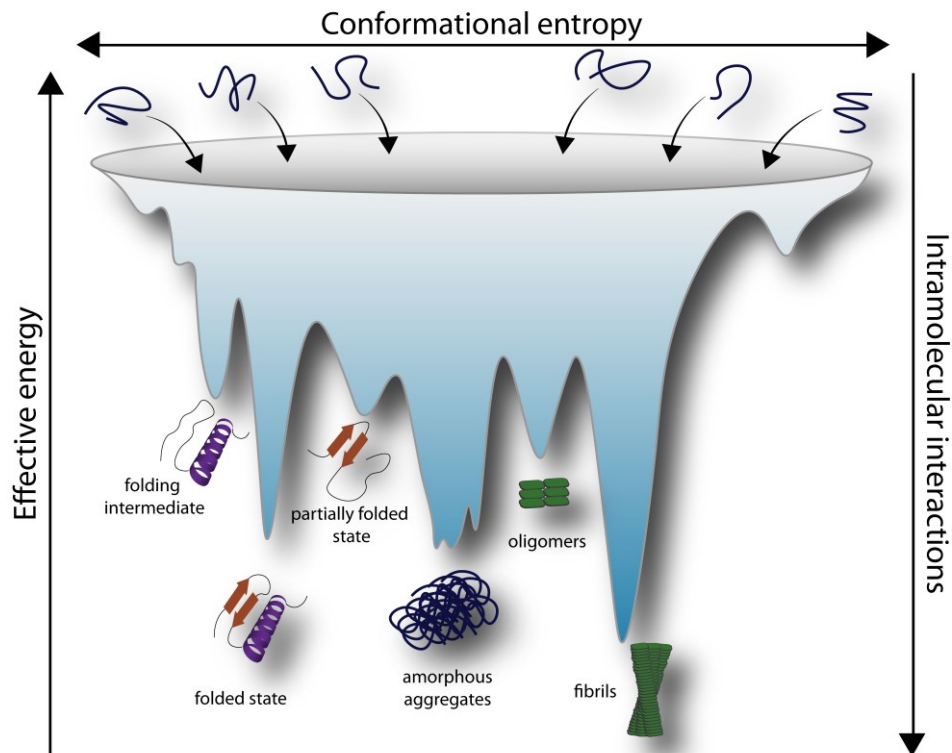


Figure 1.2. Free energy landscape of a misfolded protein. Various factors can cause the transition of an otherwise well folded protein to a misfolded or partially folded state. This transition can cause the formation of a variety of aggregated protein species, some even exhibiting enhanced thermodynamic stability.

1.3. Protein aggregation

Protein aggregation occurs in cases where the protein could not assume the native conformation or transitions from it, to a conformation that exposes hydrophobic segments of the sequence. In these cases, intermolecular interactions with other misfolded proteins will occur, causing the accumulation of the misfolded species in aggregated cores. The mode of aggregation and the resulting aggregates vary largely, owing to the diversity of proteins that aggregate. Most aggregates are formed through non-covalent interactions, while some can involve covalent bonds in the aggregation process. Aggregates also display variability in the conformational order underlying each species. Functional aggregates, such as the ones formed by elastomeric proteins, exhibit hydrophobic areas with a high

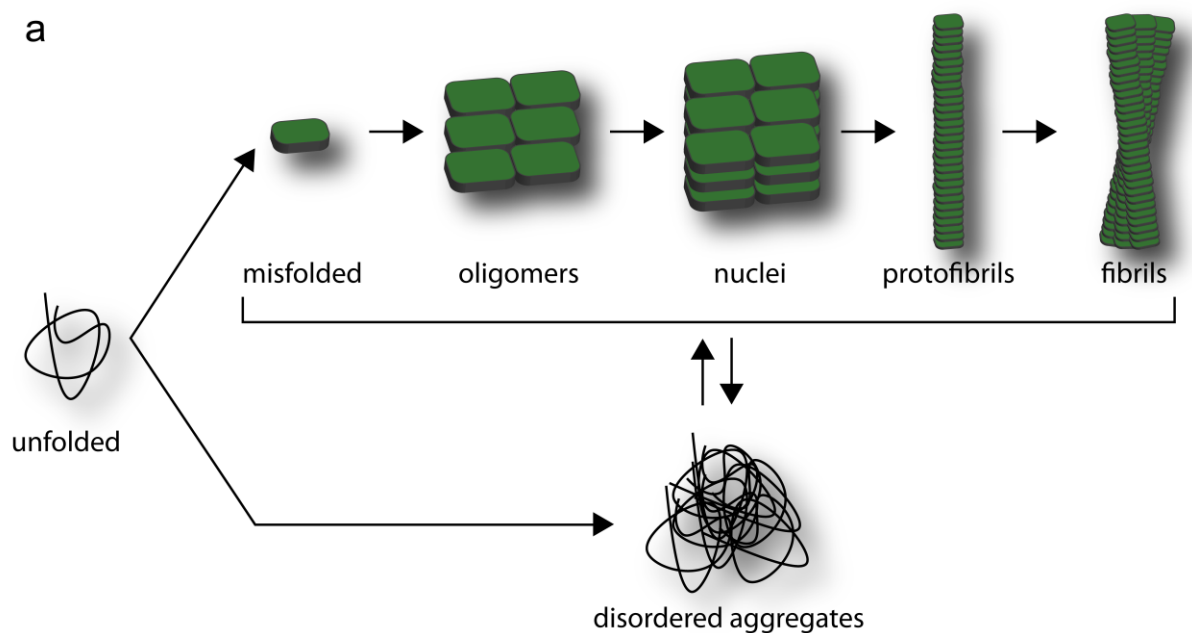
content in Pro and Gly, required for physiological function¹⁷. Aggregates that are the result of intrinsically disordered proteins and unwanted modifications to the environment or to the protein itself, usually have no apparent physiological function and can lead to cellular dysfunction. Proteins like amyloid- β (A β) and α -synuclein (asyn) are known to form a large variety of organized oligomers and aggregates known as amyloid fibrils. Depending on the conditions surrounding the aggregation process, the same proteins can also form amorphous aggregates. The species created by these amyloid forming proteins compose a network of species that are in equilibrium with each other¹⁴. However, larger species such as fibrils also exhibit limited solubility and therefore lead to deposition of proteins.

Structural studies on amyloid species are highly important as they could provide insight into the mechanisms of aggregate formation. A number of techniques have been utilized for the characterization of amyloid formation, some of which will be mentioned in the following paragraphs. Recently, the fibril structure of the 42-residue isoform of the amyloid- β peptide (A β_{42}) was solved¹⁸⁻²⁰. This Alzheimer's disease-linked peptide is an example of a difficult to characterize amyloid forming protein as it is mostly disordered as a monomer and tends to form a large variety of oligomeric and aggregated species^{14, 21, 22}.

Destabilization of the native state can lead to exposed hydrophobic areas that interact with one another and eventually aggregate²³. It has been suggested that that the transition to the native state takes place in a cooperative manner that protects against interactions that might favor aggregation²⁴. Watters *et al.* have used a small artificial protein to highlight the energy properties of evolved proteins. The authors showed that the artificial protein went through numerous transitions prior to achieving an equilibrium of stable conformations²⁵. This finding is in accordance with the assumption that natural proteins have indeed evolved to fold in a cooperative manner that protects them from acquiring disordered conformations which can lead to aggregation²⁴.

1.3.1. Aggregation and amyloid formation

Despite the lack of homology in their amino acid sequences, proteins that form amyloid structures exhibit similarities in their aggregation mechanism. The initial step of this process entails the formation of nuclei that promote amyloid formation. This rate-limiting initial step is not favoured from a free energy point of view. Once the nuclei or nucleation seeds have been formed however, they accelerate amyloid formation leading to structures like protofibrils and fibrils. The amyloid structures formed in this aggregation pathway are ordered and fibrils in particular are highly stable species, composed of multiple cross- β sheets forming an axis of around 10 nm in diameter²⁶⁻²⁹ (Figure 1.3a).



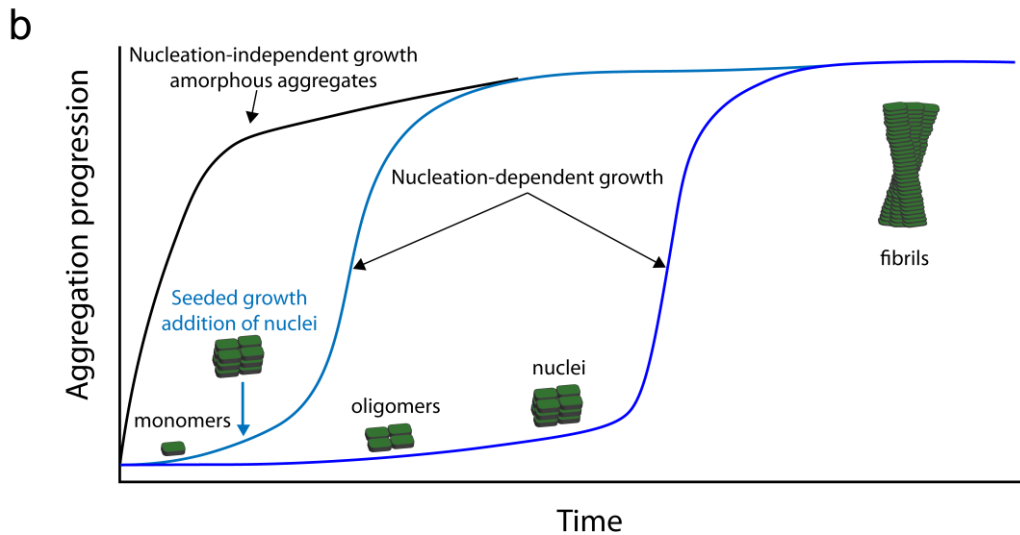


Figure 1.3. Formation of amyloid and amorphous aggregates. a) Simplified depiction of the aggregation process that leads to either amyloid formations or amorphous aggregates. Proteins are capable of transitioning between aggregated species of varying stability although the whole complicated dynamic equilibrium moves towards the formation of larger insoluble aggregates and amyloid deposition. b) Nucleation-dependent and nucleation-independent aggregation pathways progress under different kinetics. In nucleation-dependent aggregation, the initial step that leads to formation of aggregated nuclei consists the lag-phase. This phase can be accelerated through the addition of already formed nuclei. Nucleation-independent aggregation proceeds rapidly yielding much less defined structures.

Amyloid fibrils bind Congo red and show green, yellow or orange birefringence when the appropriately stained deposits are visualized by polarization microscopy²⁶. Nucleation-dependent amyloid growth has been associated with the toxic effect imposed by amyloid forming proteins^{30,31}, while seeding with aggregated nuclei is known to accelerate amyloid formation³⁰. On the contrary, nucleation-independent growth leads to the formation of disordered, non-amyloid-like aggregates³², that can also lead to toxicity³³. The two pathways are not completely isolated from each other, however, they lead to formations with different characteristics and behaviours^{32,34}.

1.3.2. Protein misfolding, aggregation and related diseases

A variety of human diseases with distinct pathologies, such as Alzheimer’s disease (AD), Parkinson’s disease (PD), type II diabetes, cystic fibrosis, prion diseases and many more, compose the group of “protein-misfolding diseases” (PMDs)¹⁴. The diseases that form this diverse group can be familial, acquired or transmissible. Pathogenesis is initiated either through the misfolding or inherent disorder of specific proteins, which can cause the loss of the protein’s physiological function, or cause them to produce toxicity-inducing aggregates such as amyloids¹⁴. Despite the fact that amyloid formations are not always directly linked to disease pathogenesis, they are central to a subset of diseases called amyloidoses²⁶. Protein misfolding and inherent protein disorder are important components of neurodegenerative diseases (ND) such as Alzheimer’s disease, amyotrophic lateral sclerosis (ALS) and Parkinson’s disease. Alzheimer’s disease is the most prominent among NDs (**Table 1.1**) and presents a brilliant example of the complicated nature of protein disorder and aggregation, as will be discussed below.

Table 1.1. Examples of misfolding-prone proteins associated with neurodegenerative diseases.

Disease	Misfolding-prone protein	Sequence length
Alzheimer’s disease	Amyloid β peptide (A β)	40, 42
	Tau	352-441 (6 isoforms)
Parkinson’s disease	α -synuclein (asyn)	140
Huntington’s disease	polyQ expansions of Huntingtin (Htt)	3144
Amyotrophic lateral sclerosis	Cu/Zn superoxide dismutase 1 (SOD1)	153
	Fused in sarcoma/translocated in liposarcoma (FUS/TLS)	526
	TAR DNA-binding (TDP-43)	414
	Dipeptide repeat expansions of <i>C9orf72</i>	481
Spongiform encephalopathies	Prion protein (PrP)	253
Frontotemporal dementia with Parkinsonism	Tau	352-441
Familial British dementia	ABri	23
Familial Danish dementia	ADan	23
Spinocerebellar ataxias	polyQ expansions of ataxins (ATXNs)	816 (ataxin-1)

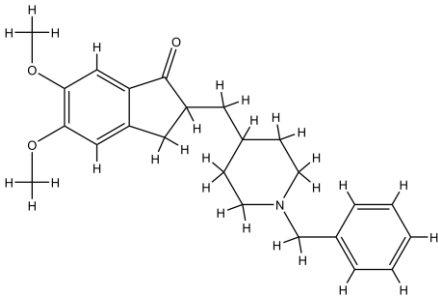
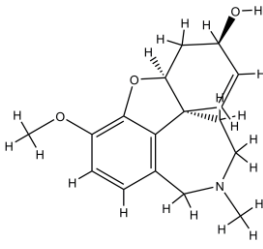
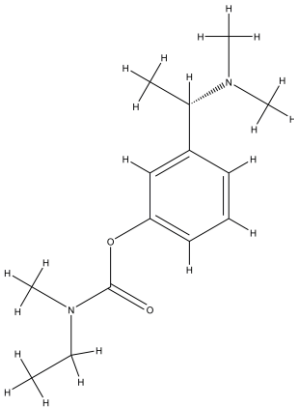
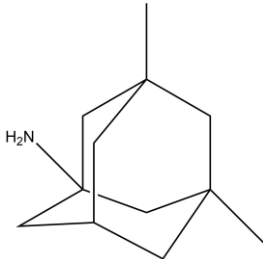
PMDs and NDs impose an enormous combined socio-economic impact. Age-related neurodegenerative diseases such as AD and PD are predicted to affect an increasing percent of the ageing population, exerting added pressure to health systems³⁵⁻³⁷.

1.4. Alzheimer's disease: a brief history.

Alzheimer's disease takes its name from Alois Alzheimer, a neuropathologist or a clinical psychiatrist as he self-identified. While working in Frankfurt, Alois Alzheimer was responsible for the clinical assessment and investigation of a female patient named Auguste D. Following the patient's death, Alzheimer, who was then working in Munich, was given brain samples belonging to the patient. This circumstance led to the discovery of important histopathological alterations that would eventually become known as amyloid plaques and neurofibrillary tangles. These alterations combined with the clinical assessment and evaluation of the patient during the course of her disease, led Alzheimer to present his findings in a psychiatry conference in Tübingen in 1906³⁸. His report was the first mention of the clinical and histological hallmarks of the disease that bears his name. Today, Alzheimer's disease (AD) is the most common form of dementia, affecting approximately 47 million people worldwide³⁹. Rising lifespans have led to an increase in documented AD occurrences which are expected to increase further in the near future⁴⁰. Regarding AD treatment, existing procedures mostly address the cognitive, psychological and behavioural symptoms of the disease. Pharmacologic treatments against the disease do not slow or stop the disease. Instead, the drugs approved by the U.S. Food and Drug Administration (FDA) offer temporary symptomatic treatment. FDA approved drugs are presented in

Table 1.2.

Table 1.2. FDA approved drugs against Alzheimer’s disease.

Compound	Structure	Proprietary name	Type
Donepezil ⁴¹	 <p>The structure of Donepezil consists of a central piperidine ring. At the 2-position of the piperidine ring, there is a benzamide group. The benzamide group has a benzene ring with a methoxy group (-OCH₃) at the 4-position and a methyl group (-CH₃) at the 3-position. At the 4-position of the piperidine ring, there is a methyl group (-CH₃).</p>	Aricept	Acetyl cholinesterase inhibitor
Galantamine ^{42, 43}	 <p>The structure of Galantamine is a tropane alkaloid. It features a tropane bicyclic core (8-azabicyclo[3.2.1]octane). At the 3-position of the tropane ring, there is a methyl group (-CH₃). At the 10-position, there is a propyl group (-CH₂CH₂CH₃). At the 11-position, there is a 2-methoxyphenyl group (-CH₂-C₆H₄-OCH₃).</p>	Razadyne	Acetyl cholinesterase inhibitor
Rivastigmine ^{44, 45}	 <p>The structure of Rivastigmine is a tropane alkaloid. It features a tropane bicyclic core. At the 3-position of the tropane ring, there is a methyl group (-CH₃). At the 10-position, there is a propyl group (-CH₂CH₂CH₃). At the 11-position, there is a 2-methoxyphenyl group (-CH₂-C₆H₄-OCH₃). Additionally, there is a methylamino group (-NHCH₃) attached to the tropane ring.</p>	Exelon	Acetyl cholinesterase inhibitor
Memantine ^{46, 47}	 <p>The structure of Memantine is a tropane alkaloid. It features a tropane bicyclic core. At the 3-position of the tropane ring, there is a methyl group (-CH₃). At the 10-position, there is a propyl group (-CH₂CH₂CH₃). At the 11-position, there is a methylamino group (-NHCH₃).</p>	Namenda	NMDA receptor antagonist
Memantine + Donepezil ⁴⁸	-	Namzaric	Combined effect

Donepezil, galantamine and rivastigmine inhibit the acetylcholinesterase-mediated hydrolysis of acetylcholine therefore increasing the concentration of the neurotransmitter in the brain.

Glutamate-induced overstimulation is a cause of excitotoxicity for neuronal cells, which leads to disruptions of calcium homeostasis⁴⁷, which has been implicated in AD. Memantine functions as a N-methyl-D-aspartate (NMDA) receptor antagonist that prevents glutamate overstimulation of the receptor⁴⁷. However, pharmacological treatments of AD are thus far unable to prevent the progression of AD.

Non-pharmacological treatments rely mostly on patient care, preservation of the patient's quality of life and their ability to perform social and daily tasks, as well as maintenance of cognitive function. Life expectancy varies (3-10 years), depending on the age of diagnosis, manifestation of the disease and various other parameters⁴⁹. AD affects both the patient and the care-giver who are family members in their majority. From a financial perspective, AD burdens the health systems as well as the patients' families. Considering the wide-spread nature of AD and its predicted rise in following years, disease-modifying treatments are in tremendous demand⁵⁰.

1.5. Association of Alzheimer's disease with the amyloid- β peptide

Three primary findings distinguish the pathology of AD: (i) amyloid plaques surrounded by neurons that contain neurofibrillary tangles (NFTs), (ii) neuronal cell death and (iii) vascular damage caused by the amyloid plaques. In 1984, Glenner and Wong reported that a protein found in the β -sheet-rich amyloids of AD, might be central to the pathology of the disease⁵¹. This consisted the first mention of the amyloid A4 peptide, later referred to as the amyloid- β (A β) peptide, which would soon come to be considered the mediator of Alzheimer's pathology. Indeed, A β became the subject of a significant number of scientific studies in the years following its discovery. This focus on how A β linked to AD pathogenesis resulted to important discoveries concerning the A β role in AD pathogenesis.

One of these discoveries rose from the suggestion that the A β peptide was the product of post-translational processing of a larger protein⁵². Kang *et al.*, and Robakis *et al.*, managed to identify the A β peptide as the processed product of the Amyloid Precursor Protein (APP)^{53, 54}, which will be discussed in more detail in the following paragraphs.

Additionally, the fact that the APP-encoding gene locus lies in the 21st chromosome and that Down syndrome (DS) patients carry three replicates of the same gene (chromosome 21 trisomy)⁵², hinted at a possible role of A β as a core component of the amyloid formations in the brains of persons with DS. This discovery was instrumental in ascertaining the role of A β deposition in the initiation of AD pathogenesis^{51, 55}, as it led to an important hypothesis: cerebral samples from deceased DS patients would aid in defining a sequence of events in AD pathogenesis. Indeed, brains belonging to young DS patients showed only diffuse amyloid formations, lacking the distinctive senile plaques that characterize AD⁵⁶. On the contrary, DS patients that died at a more advanced age, exhibited the expected senile plaques, surrounded by neurons⁵⁶. The combination of these findings suggested a progressive appearance of the hallmarks of AD. Discovering the precursor protein for A β and observing the progressive deposition of the same peptide in DS brains provided the first clues towards the mechanism that underlies AD pathogenesis. However, the importance of these findings emphasized the need to identify the exact mechanism of A β production. Therefore, studies on APP production, post-translational modifications and function became central to understanding A β .

1.6. The Amyloid Precursor Protein (APP)

APP is a transmembrane glycoprotein with a relatively elusive function. It is expressed throughout the body in various isoforms, the main ones of which are APP₆₉₅, APP₇₅₁ and APP₇₇₀^{57, 58}. Out of these isoforms, APP₆₉₅ is the most abundantly produced in the cortex⁵⁹ while APP₇₅₁ and APP₇₇₀ are upregulated in AD⁶⁰, therefore there is no definite relation between AD pathogenesis and APP₆₉₅

prevalence in the brain. Post-translationally, APP localizes in various membranes such as those of Golgi compartments and the plasma membrane⁶¹⁻⁶³. The N terminus of APP is located in the outer part of the membrane and composes the greater part of the protein while its C-terminus lies internally to the membrane. The sequence of A β consists part of the ecto- and trans-membrane domains of APP.

Initial studies regarding the production of A β , showed two distinct proteolytic events that initiated different pathways. The secondary structure of APP proximal to the exterior side of the membrane can be recognized by a set of enzymes named APP- or α -secretases⁶⁴⁻⁶⁶. α -Secretases cleave APP within the A β sequence⁶⁴⁻⁶⁶, resulting to what was labelled the non-amyloidogenic pathway, since it does not lead to amyloid deposition.

In the early 1990s the second distinct proteolytic event that yielded full-length A β was credited to the internalization of the APP molecule and processing, through the endosomal-lysosomal pathway⁶⁷⁻⁷¹. The enzymes that were actually responsible for the amyloidogenic cleavage of APP, β - and γ -secretases had not yet been identified. In brief, β -secretase is an aspartyl protease that cleaves the APP sequence before Asp1 and γ -secretase is an enzymic complex with aspartyl protease activity that yields A β sequences ranging from 37 to 43 amino acids. These enzymes will be discussed more extensively in the following paragraphs.

1.7. The amyloid cascade hypothesis

The findings on the progression of AD pathology, on the A β peptide and on APP, provided a footing for the development of a hypothesis about the pathogenesis of AD. The years 1991 and 1992 saw the publication of several scientific articles, describing the amyloid cascade hypothesis⁷²⁻⁷⁵. The hypothesis posited that the deposition of A β is at the root of AD pathogenesis⁷³. It suggested that A β is neurotoxic although there were multiple suspected causes for its neurotoxicity. These involved A β exerting

neurotoxicity on its own or as an A β -containing cleavage product of APP^{76, 77}, as well as A β indirectly causing neurotoxicity by disrupting calcium homeostasis and rendering neurons more susceptible to excitotoxic damage from glutamate neurotoxicity^{78, 79}. The link between amyloid deposition and neurofibrillary tangle (NFT) formation was not determined at the time and it remains unclear today, with contradicting results as to which histological finding initiates the pathogenesis sequence. However, back in 1991 Hardy and Higgins proposed a link between amyloid deposition and NFTs, based on the knowledge that A β disrupts calcium homeostasis⁷⁹. NFTs are composed of paired helical filaments (PHFs) which are formed by hyper-phosphorylated forms of the protein tau^{80, 81}. Tau is a microtubule component, whose phosphorylation is controlled by intracellular calcium⁸². Therefore, increased calcium concentrations could promote phosphorylation which could lead to NFT formation⁷⁵.

The amyloid cascade hypothesis has been the prevalent model of AD pathogenesis since its publication, albeit with a number of necessary revisions owed to new discoveries on the components of the hypothesis.

1.8. Re-evaluation of the amyloid cascade hypothesis

Revisiting the amyloid cascade hypothesis, Hardy and Selkoe deliberated the discoveries since its first publication^{83, 84}. In the years following the original hypothesis, the A β peptide was shown to be a product of physiological cleavage of APP⁸⁵⁻⁸⁷. Additionally, APP mutations responsible for inconsistencies on physiological APP cleavage were discovered⁸⁸⁻⁹³. Such mutations are usually located in the parts of the APP sequence cleaved by α -, β and γ -secretases. Understanding the mechanisms that are disturbed by those APP mutations was helped by the characterization of the three secretases.

1.8.1. Findings concerning APP processing

As mentioned previously, α -secretase recognizes the sequence of the secondary structure of APP close to the membrane and leads to the non-amyloidogenic pathway of APP cleavage. Its existence was first suggested by Esch *et al.* in 1990, who showed that APP was cleaved between the 15th and 16th amino acid of the A β sequence by a membrane-bound protease that was uncharacterized at the time. The identity of this protease was later discovered in the ADAM (A Disintegrin And Metalloproteinase) family of proteases. Three members of the ADAM family exhibited the ability to cleave APP in the α -site (α -secretase cleavage site): ADAM9^{94, 95}, ADAM10^{96, 97} and ADAM17^{98, 99}. Out of the three metalloproteinases, ADAM10 has been recognized as the constitutive α -secretase of neurons¹⁰⁰⁻¹⁰³.

β -Secretase was the name given to the protease hypothesized to cleave the APP sequence at Asp1 of the A β peptide. This aspartyl protease was eventually identified as the Beta-site Amyloid precursor protein Cleaving Enzyme 1 (BACE1) or ASP2¹⁰⁴. BACE1 levels in the cortex are elevated in AD although the reasons as to why this elevation occurs are unclear¹⁰⁵. BACE1 inhibition presents an interesting therapeutic strategy, as it appears to play a primary role in the sequence of events leading up to AD pathology. However, BACE1 has a more complicated function than just APP cleavage, as substrates are still being discovered and its physiological function is not fully determined. Nevertheless, BACE poses a very attractive target for therapeutics development, with various BACE inhibitors undergoing clinical trials^{106, 107}.

γ -Secretase is the enzyme complex responsible for the sequence length of the produced A β peptide. It is compiled of Presenilin 1 (PS1) or Presenilin 2 (PS2), Nicastrin (NCT), the Anterior Pharynx defective domain APH-1a or APH-1b and a Presenilin enhancer domain, PEN-2¹⁰⁸. The cleavage of APP by γ -secretase is suggested to take place in a stepwise manner with the proposed cleavage sites ϵ -, ζ -, and γ -, being approximately three amino acids apart¹⁰⁹⁻¹¹². The cleavage sequence moves from ϵ to γ , shortening the C-terminus of the peptide. Of the three sites, γ is the least specific, yielding A β lengths

of 37 to 43 amino acids. The significance of these findings is that they place γ -secretase at a very central position in AD pathogenesis, since it determines what isoform of $A\beta$ will be produced. Depending on the position of the ϵ -cleavage site, the product of γ -secretase proteolysis can lead to $A\beta$ isoforms with different amyloidogenic properties. For instance, ϵ -site cleavage at the 49th amino acid of the $A\beta$ sequence produces mostly $A\beta_{40}$, an $A\beta$ isoform with much milder amyloidogenic properties. However, cleavage at the 48th amino acid of the $A\beta$ sequence, leads primarily to $A\beta_{42}$ ¹¹³. Modulation of γ -secretase cleavage poses another attractive pharmacological goal. In a review by Schenk *et al.*, the authors extensively discuss the challenges and possibilities of targeting either β - or γ -secretase for the inhibition of $A\beta_{42}$ production¹¹⁴.

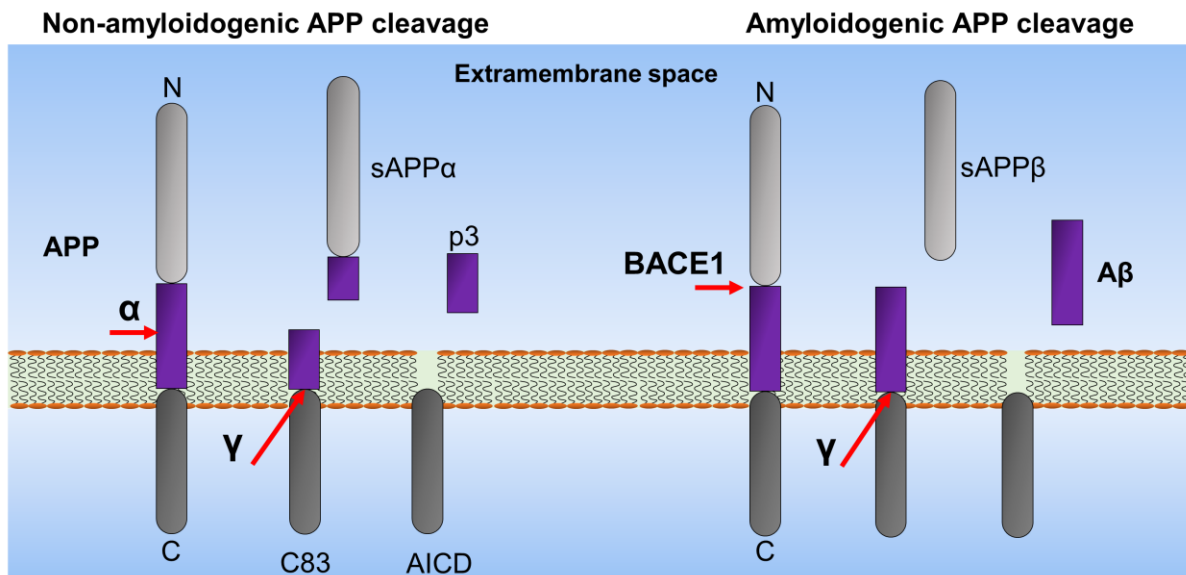


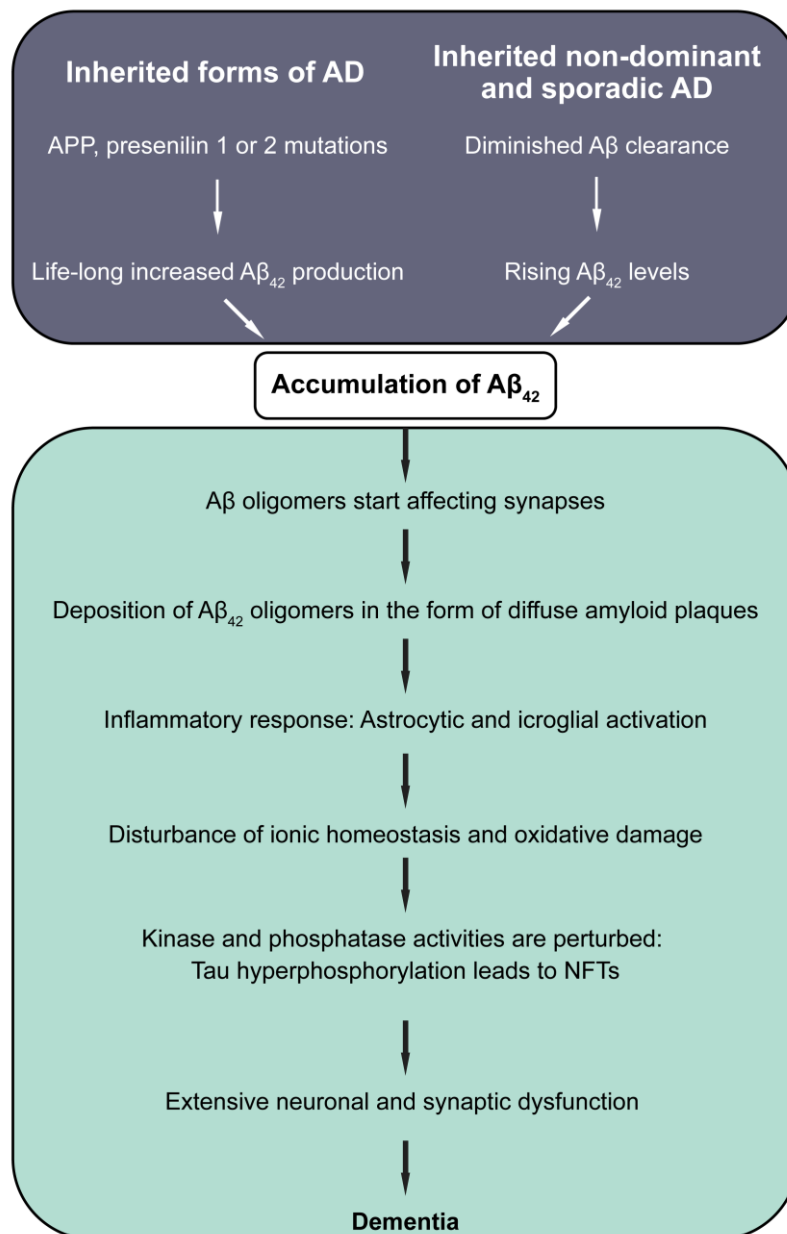
Figure 1.4. APP processing. Cleavage by α -secretase leads to proteolytic products that do not exhibit amyloidogenic properties. BACE1 cleavage followed by sequential cleavage from γ -secretase yields amyloidogenic products with varying aggregation propensities.

1.8.2. Findings in support of the amyloid cascade hypothesis

The elucidation of the mechanism of $A\beta$ generation through APP cleavage consisted a fraction of the discoveries on the complicated pathology of AD. An interesting finding was the identification of

varying segmental microduplications of chromosome 21. In rare occasions of translocation DS involving microduplication of a part of the 21st chromosome that did not contain the APP locus, the individuals showed DS features but did not develop AD¹¹⁵. On the contrary, humans with microduplications of only the APP locus did not have DS. However, these people did develop AD during their sixth decade¹¹⁶. These findings provided a clear indication that life-long production of APP at increased levels causes AD. More evidence supporting the amyloid cascade hypothesis was provided by the identification of the APP A673T mutation (or A β A2T). This mutation resulted in decreased β -secretase cleavage¹¹⁷ and therefore decreased A β production. Furthermore, β -secretase cleavage of APP A673T resulted in an A β peptide, which lacked the aggregation propensity of the wild-type peptide thus lowering the risk for the development of AD¹¹⁸⁻¹²⁰.

The amyloid cascade hypothesis remains until today a valid attempt to characterize the sequence of events that take place in AD. Increasing efforts to understand and treat the disease lead to findings that add to the hypothesis or even suggest different courses of disease development. Importantly, certain issues regarding this hypothesis remain unsolved: whether AD pathology spreads in a prion-disease manner, the determination of APP and A β functions, the characterization of the toxic A β species and many more. Despite the many unanswered questions surrounding AD pathology, the importance of the amyloid cascade hypothesis lies in the fact that it has aided the characterization and categorization of some of the components of a very complicated pathology, as will be shown in the following paragraphs that discuss A β -induced toxicity.



Adapted from: Selkoe, D.J. & Hardy, J. The amyloid hypothesis of Alzheimer's disease at 25 years. *EMBO Mol. Med.* **8**, 595-608 (2016).

Figure 1.5. The updated amyloid cascade hypothesis. More than 25 years since the original hypothesis, the amyloid cascade has been enriched with new data on the toxic Aβ species and on the interactions between the components of the disease.

1.9. AD pathogenesis

1.9.1. Toxicity-inducing A β variants

The physiological function of the A β peptide is not yet clarified; however, A β_{40} is the physiologically abundant isoform while the A β_{42} isoform appears in increased concentrations in AD pathogenesis. A β_{42} displays different aggregation kinetics compared to A β_{40} , which translate to an increased tendency to aggregate¹²¹⁻¹²³. Likewise, many familial AD-related A β mutants also show an increased tendency for aggregation¹²⁴. For instance, the aggregation kinetics of the Arctic (E22G), Dutch (E22Q), and Italian (E22K) mutants point to faster aggregation compared to the wild-type peptide. On the contrary, the Flemish (A21G) mutant exhibits slower aggregation kinetics^{125, 126}.

The complexity of A β -induced toxicity is highlighted by the discoveries of additional cytotoxic A β peptides, such as the highly amyloidogenic pyroglutamylated forms of A β ^{127, 128}. Nussbaum *et al.* suggested that the pyroglutamylated isoform A $\beta_{3(pE)-42}$, co-aggregates with A β_{42} and forms distinct low-n oligomers that are significantly more toxic than those formed by A β_{42} alone. Interestingly, the authors also showed that the observed neurotoxicity was Tau-dependent¹²⁹. Another amyloidogenic isoform, A β_{43} was shown to be even more aggregation prone and neurotoxic than A β_{42} ¹³⁰.

1.9.2. Toxicity related to oligomeric and aggregated A β species

1.9.2.1. Fibril toxicity

Historically, large insoluble fibrils were the first candidates for the neurotoxic effect of A β since they constitute an important component of amyloid plaques and studies of *in vitro* assemblies allowed easier and more complete characterization of amyloid fibrils, relative to other species. Indeed, Puzzo and Arancio showed that fibrils of synthetic A β impaired the late phase of long term potentiation

(LTP)¹³¹. In addition, fibril conformations differ both in structure and in the induced toxicity. Yoshiike *et al.* showed that apart from the β -sheet structure, the physicochemical properties of the fibrillary surface also contribute to fibril toxicity¹³². A problem with studies involving fibrils was the difficulty to acquire pure samples that do not contain prefibrillar species. Furthermore, *in vitro* toxicity studies failed to explain the weak correlation of the severity of AD with the density of fibrillary deposits¹³³⁻¹³⁵.

1.9.2.2. Oligomer toxicity

Contrary to fibrils, prefibrillar or soluble A β species showed a stronger correlation with cognitive impairment¹³⁶⁻¹³⁸. Walsh and Selkoe describe soluble A β species as those that remain in aqueous solutions even after ultracentrifugation. Using non-chaotropic solutions and molecular weight cut-off filters Kuo *et al.* were able to assess the low-n content A β species in the AD brain. The authors showed the presence of four classes of A β species judging on their ability to cross the molecular weight cut-offs: >100 kDa, <100 kDa and >30kDa, <30 kDa and >10 kDa, <10kDa. These results implied the involvement of oligomeric species by declaring them present during the course of the disease. Shankar *et al.* managed to provide a clearer link between oligomers and toxicity. The authors found that AD cortex-derived soluble oligomers of A β could impair synaptic plasticity and memory in healthy adult rats¹³⁹. The authors also showed that long-term potentiation was not affected by amyloid plaque cores, isolated from the same cortices and washed to remove any soluble A β oligomers. When such plaque cores were denatured however, the negative effect on long term potentiation was again noticeable¹³⁹. Additional evidence of the toxicity of soluble A β oligomers is provided by a study on post-mortem AD brains that looked into the oligomeric A β /amyloid plaque ratio, in non-demented and in mildly demented patients with rich plaque formations. The study showed that low oligomer/plaque ratios appeared in the cases of patients with plaque formations but with no signs of dementia¹⁴⁰.

The specific nature of oligomeric A β species responsible for toxicity remains elusive. However, certain oligomers have been recognized as more toxic than others. For instance, Lesne *et al.*, based on an APP transgenic mouse model, showed that the A β nonamer and dodecamer (A β *56) levels correlated with the impairment of spatial memory in the Tg2576 mouse model¹⁴¹. Additionally, similar impairment was noticed when isolated A β *56 was administered to young rats¹⁴¹. This finding was complemented by a work that compared various oligomeric species shown to exert toxic effects, in which A β *56 was again shown to induce cognitive impairment in rats¹⁴².

A very prominent synaptotoxic oligomeric A β species is the SDS-stable dimer. Klyubin *et al.* showed that untreated human cerebrospinal fluid containing dimeric A β , inhibited long term potentiation in rats¹⁴³. The same dimeric species was shown to induce cellular death in rat-derived hippocampal cultures, only in the presence of microglia¹⁴⁴. However, the attribution of synaptotoxicity to one species has been questioned through the use of a mouse model of AD, which showed that other aggregate formations appeared prior to the detection of the dimeric species, implying a more complicated A β assembly during the course of the disease¹⁴⁵.

A β -derived diffusible ligands (ADDLs) form a class of toxicity-inducing small A β oligomers that exhibit resistance to low SDS concentrations. Formation of ADDLs through the interaction of A β ₄₂ with the protein clusterin led to species that caused oxidative stress in PC12 cells¹⁴⁶. ADDLs are also known to quickly impair hippocampal long term potentiation in young adult rats¹⁴⁷, thin the membrane bilayer and increase its conductance¹⁴⁸, while they are also shown to reduce cell viability for various cell lines¹⁴⁹.

Another soluble species of A β oligomers named amylospheroids were reported by Hoshi *et al.* The authors showed that amylospheroids (ASPDs) had a perfectly spherical shape and were highly toxic to cultures containing septal and basal cholinergic neurons, derived from rats¹⁵⁰.

Protofibrils are an intermediate category of A β species that fall between fibrils and oligomers. Their involvement in AD pathology is limited possibly because of their quick transition to mature fibrils¹⁵¹.

The mechanism of oligomer-induced toxicity is unknown. However, it has been shown that A β oligomer formation can be seeded by A β bound on GM1 gangliosides^{152, 153}. The resulting oligomers were not identified by a monoclonal A β antibody specific to seeds, indicating a different formation pathway than that of fibrillary structures¹⁵². In fact, binding GM1 gangliosides leads to rapid sequestering of A β by cell membranes¹⁵⁴. The authors showed that by blocking the GM1 gangliosides, A β oligomer-mediated impairment of long-term potentiation in mice was prevented, suggesting that this might be part of a mechanism that leads to oligomer-induced toxicity¹⁵⁴. Moreover, the heat-shock protein α B-crystallin, found in increased concentrations in the AD brain, has been shown to prevent the formation of fibrillary structures of A β . However it was also shown that it increased *in vitro* toxicity for cultured neurons¹⁵⁵. This finding suggested that an alteration of the A β fibrillization pathway led to highly toxic non-fibrillary structures. Mechanisms that rely on neuronal receptor-mediated toxicity include the nerve growth factor (NGF) receptor¹⁵⁶, the N-methyl-D-aspartate glutamate (NMDA) receptor^{157, 158}, the insulin receptor¹⁵⁹⁻¹⁶¹ and the frizzled (Fz) receptor¹⁶². The NMDA receptor-mediated mechanism in particular, causes disruption of the calcium homeostasis discussed in the amyloid cascade hypothesis as one of the causes of toxicity^{157, 158}. The cellular prion protein (PrP^C) has also been shown to function as a receptor for A β oligomers, which could possibly disrupt PrP^C interactions with co-receptors, hindering neuronal signal transduction pathways.

Apart from binding A β species and leading to neurotoxicity, a number of receptors are responsible for internalizing A β oligomers that have been produced extracellularly. Internalization of A β as well as internally produced A β , lead to accumulation of the peptide within the cell. As expected, intracellular accumulation has been linked to various disruptions, such as inhibition of the proteasome¹⁶³ and destabilization of intracellular membranes¹⁶⁴.

Additional interactions of A β oligomers that lead to toxicity, include the oligomer-specific impairment of presynaptic P/Q calcium currents¹⁶⁵. Furthermore, A β oligomers have been suggested to destabilize the cell membrane, creating pores that could alter the ion concentrations and cause toxicity^{166, 167}.

1.9.3. Alternative causes of A β -induced toxicity

Mutations in APP and PS¹⁶⁸⁻¹⁷² that have been associated with familial AD can alter the physiological cleavage of APP and lead to increased A β levels or to production of highly toxic A β isoforms. Concerning sporadic AD, factors such as metal ion concentrations, oxidation of A β Met35, oxidative stress have been implicated in disease development.

An important protein related to AD, ApoE, is known to promote proteolytic degradation of A β ¹⁷³, however, the E4 allele of apolipoprotein E (ApoE4) represents a genetic risk factor that leads to early onset AD. The three alleles deposition (E2, E3 and E4) are known to exhibit differential modulation of A β deposition, through varying effects on the concentration of A β in the interstitial fluid (ISF). Out of these three *APOE* genotypes, the one corresponding to ApoE4 has been associated with increased A β deposition¹⁷⁴⁻¹⁷⁷.

1.10. Aggregation kinetics

The scientific consensus concerning the formation of mature fibrils, points to a nucleation-dependent process. As discussed previously, amyloid fibril formation requires the existence of an initial rate-limiting step. Kinetic studies of A β aggregation revealed the existence of such a rate-limiting step termed "lag phase". The lag phase represents the formation of the nuclei that seed A β aggregation. The accumulation of thermodynamically stable nuclei is then followed by a rapid elongation step,

which concludes in mature fibrils. Yet, the events that occur within the lag phase remain elusive. It has been suggested that the monomeric A β ₄₀ has to transition to an activated state in order to commence nuclei formation by associating with other monomers¹⁷⁸.

Furthermore, the A β monomer is known to exist in equilibrium between two conformations: an α -helical and a β -sheet conformation. It has been suggested that only the β -sheet can lead to β -sheet-rich oligomers^{179, 180}. Benseny-Cases *et al.* have showed for a specific sample of aggregating A β , that the oligomeric species that compose the lag phase are a mix of unordered, helical, and intermolecular non-fibrillar β -sheet structures¹⁷⁹. Additionally, since Walsh *et al.* showed that A β initially forms an α -helical prior to a β -sheet conformation¹⁸¹, more aggregation intermediates that contain α -helical components have been reported^{125, 182-184}. Moreover, stabilization of α -helical components of A β in a *Drosophila melanogaster* model, led to inhibition of aggregation as well as improved locomotor activity¹⁸⁴. Another debated aspect of the β -sheet involving pathway is the extent to which this β -sheet state has to be populated in order for oligomerization to commence. Two studies on proteins with the ability to form amyloids, HypF-N¹⁸⁵ and human lysozyme¹⁸⁶, have shown that only a small population in a partially folded state is necessary to cause the formation of amyloid fibrils.

An alternative pathway for A β aggregation was suggested by Necula *et al.* who showed that certain small-molecule inhibitors of A β aggregation inhibit oligomerization but not fibril formation¹⁸⁷. These findings suggested an aggregation pathway in which oligomerization was not required in order to seed mature amyloid fibril formation. However, the authors did not dismiss the possibility of oligomers ultimately forming such fibrils, rather they highlighted the possible existence of multiple aggregation pathways for A β ¹⁸⁷.

A β concentration is an important parameter of the peptide's aggregation. It directly affects the nucleation process, as it has to be in the low nM range in order to initiate nuclei formation and the consequent aggregation⁸⁶. On the contrary, nucleation dependent aggregation only occurs below a

maximum A β concentration of about $10^4 \times$ the physiological A β concentration, beyond which, aggregation does not occur in a nucleation dependent manner.

However, aggregation kinetics differ for each variant, a fact that translates to differences in amyloidogenic abilities as well as differences in the composition of the aggregation assembly between the variants of A β . Accordingly, variances in aggregation kinetics among variants suggest that A β aggregation proceeds through varying pathways¹⁸⁸⁻¹⁹¹.

1.11. Structural data for A β oligomers and aggregates

Despite several efforts to produce structural data for the A β monomer, it appears that the equilibrium between the α -helical and β -sheet conformations is its main structural characteristic. Prior to APP cleavage though, A β displays more defined characteristics. Located between the extramembrane and the transmembrane domain of APP, it predictably displays amphipathic properties. Its extramembrane section is hydrophilic, containing the first 28 amino acids of its sequence. The rest of the sequence forms part of the APP α -helical intramembrane domain^{192, 193}, without fully traversing the membrane.

Following APP cleavage and release of A β , the peptide tends to form β -sheet-rich aggregates, though as discussed previously, intermediates with α -helical components have also been shown to exist^{125, 182-}

¹⁸⁴.

Oligomeric species have been studied extensively with regard to their toxic effects. Existing structural data on oligomers, lack the definition required to distinguish them on a structural basis. However, structural information on such oligomers can be inferred through techniques such as electron microscopy, atomic force microscopy (AFM), circular dichroism (CD) and ion-mobility mass spectroscopy (IM-MS), which yield mostly morphological and secondary structure information, rather than detailed three-dimensional conformations. One interesting example of such work involved

IM-MS and provided qualitative structural data for A β ₄₀ and A β ₄₂ oligomers, by comparing the computational and experimental cross-sections of these aggregates¹⁵¹.

In an effort to offer a structural classification of toxic amyloid oligomers, Glabe categorized oligomer structures based on the ability of conformation-dependent antibodies to distinguish amyloid structures¹⁹⁴. Antibodies that bind fibrils also bind certain oligomers, therefore indicating structural similarities between them. These similarities form the basis for the classification of oligomeric structures into fibrillar-type oligomers and prefibrillar-type oligomers. According to the author, fibril structures present steric zipper motifs that consist of a single amino acid side chain, traversing the length of the fibril^{195, 196}. Steric zippers could provide the necessary epitopes shared between many amyloid fibrils. Therefore, recognition of both fibrils and oligomers from the same antibody would indicate that these epitopes exist in the oligomers in question. This hypothesis of a shared epitope between amyloid fibrils and certain oligomers, infers shared structural properties between the species. However, structural characterization of the oligomers would still be required to validate this assumption.

Unfortunately, the lack of high-resolution structural data impedes the development of rationally designed therapeutics. However, recently published atomic-resolution structures of A β fibrils gave valuable clues on the nature of these larger aggregates and could possibly aid the development of anti-aggregation approaches^{28, 29, 197}.

1.12. A β -related therapeutic strategies against AD

Efforts to treat AD by targeting the A β peptide include reduction of A β levels, inhibition of A β aggregation and disaggregation of A β species. Reducing A β levels can be addressed by preventing its production or by assisting A β clearance. Interestingly, increased β -secretase activity, and decreased

A β clearance have been observed in AD pathology^{105, 198-201}. Additionally, multiple aggregated A β species have been linked to AD pathology^{139, 202-204} and, therefore, preventing the formation of such species through early regulation or inhibition of their formation, presents a desirable course of action against AD. Furthermore, this strategy could assist in maintaining a form of A β homeostasis in cases of ApoE4-related problematic A β clearance^{114, 175, 201}.

1.12.1. BACE1 and BACE2 inhibitors

Regulation of A β production through the inhibition of β -secretases is an appealing prospect, since reduced cleavage of APP by these enzymes could lead to a reduced A β production. Indeed, until early 2017, five such drug candidates were undergoing clinical trials^{205, 206}. BACE1 and BACE2 are, structurally, two typical type I aspartic proteases with 59% identity and a bilobal structure. Docking on the membrane takes place through a type I transmembrane domain²⁰⁷. However, the physiological roles of BACE1 and BACE2 are still being discovered while the existing literature already provides a clear indication that the role of these enzymes is complicated and diverse²⁰⁸. In theory, targeting BACE1 and BACE2 could yield impressive results against AD pathogenesis since it could lead to treatments that regulate or block A β generation altogether. This approach however, has to take into consideration possible side effects from long-term inhibition of these enzymes. Verubecestat was a candidate drug that inhibited BACE1 in a specific manner, while administration in high dosages in rats and monkeys avoided problems associated with BACE1 inhibition, such as hepatotoxicity, neuromyelination, neurodegeneration and alterations in glucose homeostasis²⁰⁹. Despite promising results, it was halted in phase 2/3 of clinical trials. Failed attempts at drug development and continued efforts to fully characterize BACE1 and BACE2, will aid in developing inhibitors for β -secretases that minimize interference with the physiological functions of the enzymes.

1.12.2. γ -Secretase inhibitors

The compounds that have been developed as inhibitors of γ -secretase can be categorized into three groups: the non-selective inhibitors, the cleavage modulators and the APP selective/Notch sparing inhibitors¹¹⁴. The best-known example of non-selective inhibitors was semagacestat²¹⁰. This drug candidate reached phase 3 of clinical trials with mild to moderate AD patients, before being halted due to severe side effects²¹⁰. The main obstacle facing γ -secretase inhibitors is that the enzymic complex also cleaves the Notch protein of the homonymous signaling pathway. Inhibition of Notch cleavage has been shown to cause adverse effects in adult animals^{211, 212}, validating the concerns about inhibition of γ -secretase.

The second class consists of compounds that take a more subtle approach and aim to modulate γ -secretase cleavage rather than inhibit it, reducing the side effects. This class includes compounds such as ibuprofen and R-flurbiprofen (or tarenfluril). Both lowered brain $A\beta_{42}$ levels, with R-flurbiprofen reaching phase 3 clinical trial with mild to moderate AD patients, where it failed to produce positive clinical results^{213, 214}.

The third class of compounds inhibit or modulate γ -secretase in a substrate specific manner, i.e. they preferentially inhibit APP cleavage by γ -secretase, reducing side effects from the Notch pathway inhibition. Avagacestat and begacestat were two representative drug candidates reported to exhibit increased selectivity for APP against Notch that advanced to clinical trials. However, avagacestat was shown to not be as specific to APP²¹⁵ and was halted due to adverse effects in clinical trials²¹⁶. Begacestat reduced the levels of $A\beta$ in the CSF and went in phase 1 trials in combination with donepezil, however the results have not been published^{216, 217}. Another compound termed NIC5-15 is the natural compound pinitol²¹⁸ that went into phase 2 trials, with results indicating good tolerability and stabilization of cognition measured by the ADAS-Cog protocol. No published work exists on this

trial and the compound became available as a food supplement. Information on pinitol trials was sourced from alzforum.org/therapeutics/nic5-15.

1.12.3. Inhibition of A β aggregation

Monitoring the aggregation of A β *in vitro* is possible through a number of techniques, allowing the identification of compounds that interfere and modify this process. Thioflavin fluorescence, atomic force microscopy, surface plasmon resonance, circular dichroism and others have been used to identify compounds with the ability to inhibit the formation of aggregates or cause their disaggregation²¹⁹⁻²²⁵. Such molecules range from synthetic peptides, to natural products and small synthetic molecules²²⁶⁻²²⁹. A large number of compounds have been discovered in this manner; however, few have advanced to clinical trials. Tramiprosate is an example of such a compound that reached phase 3 trials. In mild to moderate cases of AD it exhibited insignificant results and was therefore halted²²⁹. Scyllo-inositol²²⁷ is a compound that advanced to phase 2 of clinical trials, while very recently a guanidine-appended derivative of the compound showed improved delivery to the brain of 5xFAD mice, as well as reduced gliosis, and improvement of behavioural memory²³⁰. Also recently, an approach that identified an anti-cancer drug in an A β -specific fragment-based library, showed that the drug in question inhibited A β ₄₂ nucleation, slowing aggregation both *in vitro* and *in vivo*²³¹.

1.12.4. Enhancing A β clearance

A β clearance can be mediated through the activation of enzymes that are known to degrade the peptide, such as neprilysin, endothelin-converting enzyme 1, insulin-degrading enzyme and plasmin²³². A second approach involves the relocation of A β to the periphery using the receptor for

advanced glycation end products (RAGE) and the low-density lipoprotein receptor-related protein 1 (LRP-1). RAGE is responsible for the influx of A β to the brain, from the central nervous system (CNS) while LRP-1 mediates the efflux of A β from the brain to the CNS²³³. A small molecule inhibitor of RAGE was tested in phase 2 trials where it did not produce enough evidence to continue its development²³⁴. Two more trials for TTP488 (azeliragon), an antagonist for RAGE, are currently in phase 1¹⁰⁷.

1.12.5. Immunotherapeutic strategies against A β aggregation

Immunotherapy is considered a very promising strategy for the development of AD therapeutics. Schenk *et al.* presented the concept of active immunotherapy in 1999. One year later, Bard *et al.* introduced passive immunotherapy^{235, 236}. Active immunotherapy involves the exposing of the subject to A β and generating a polyclonal antibody response²³⁵. In contrast, in passive immunotherapy the antibody is administered to the subject, bypassing the need of an immune response²³⁶.

So far, numerous antibodies have advanced into clinical trials. Solanezumab is an anti-A β monomer antibody that is believed to bind low-n toxic oligomers²³⁷. It went through three phase 3 trials, showing small benefits for mild AD cases. Eventually, its development was halted, provoking a discussion on the validity of the anti-amyloid strategies. However it is undergoing phase 3 clinical trials as a combination with a glycogen linked antigen binding antibody fragment (Fab) that targets soluble oligomers, named LY2599666²³⁸. Crenezumab is another antibody currently in phase 3 trials, which has shown positive results for early AD stage patients²³⁹. Finally, a drug candidate called Aducanumab is believed to bind A β fibrils and soluble oligomers. During clinical trials phase 1b, it caused a reduction in amyloid plaque levels of AD patients²⁴⁰⁻²⁴². It is currently undergoing phase 3 clinical trials.

Active immunotherapy has not yielded equally promising results. However, brains of patients that had taken part in a phase 1 trial for the AN-1792 vaccine, showed diminished neuronal dystrophy and

synaptic deficits²⁴³. Selkoe and Hardy posit that the polyclonal antibody response to a vaccine might be advantageous both for its biological effect and from a logistic aspect as AD affects a significant part of the human population⁸⁴.

1.12.6. Preclinical treatment

The aim of preclinical treatment of AD is to identify biomarkers that would help predict the onset of the disease and commence to address the pathology as early as possible. This strategy is based on the generally accepted estimation that AD begins more than a decade before cognitive symptoms appear. Existing methods for the identification of the preclinical stage include PET measurements of fibrillar A β , cerebrospinal fluid (CSF) measurement of A β , Tau and phosphorylated Tau, MRI to detect brain tissue shrinkage etc. Additionally, Escott-Price suggested the use of polygenic scores in order to identify patients whose disease might carry a larger genetic component²⁴⁴. This approach could identify at-risk individuals and aid the evaluation of their results in clinical trials. Langbaum *et al.* have reviewed established methods of MRI and fluid-based preclinical detection of AD, in an effort to present the current technologies that allow detection of the disease before cognitive damage²⁴⁵.

1.12.7. Drug discovery pipeline for AD

A review by Cummings *et al.* revealed that as of May 2017, there were 105 agents in clinical trials, out of which, 25 in phase 1, 52 in phase 2 and 28 in phase 3¹⁰⁷.

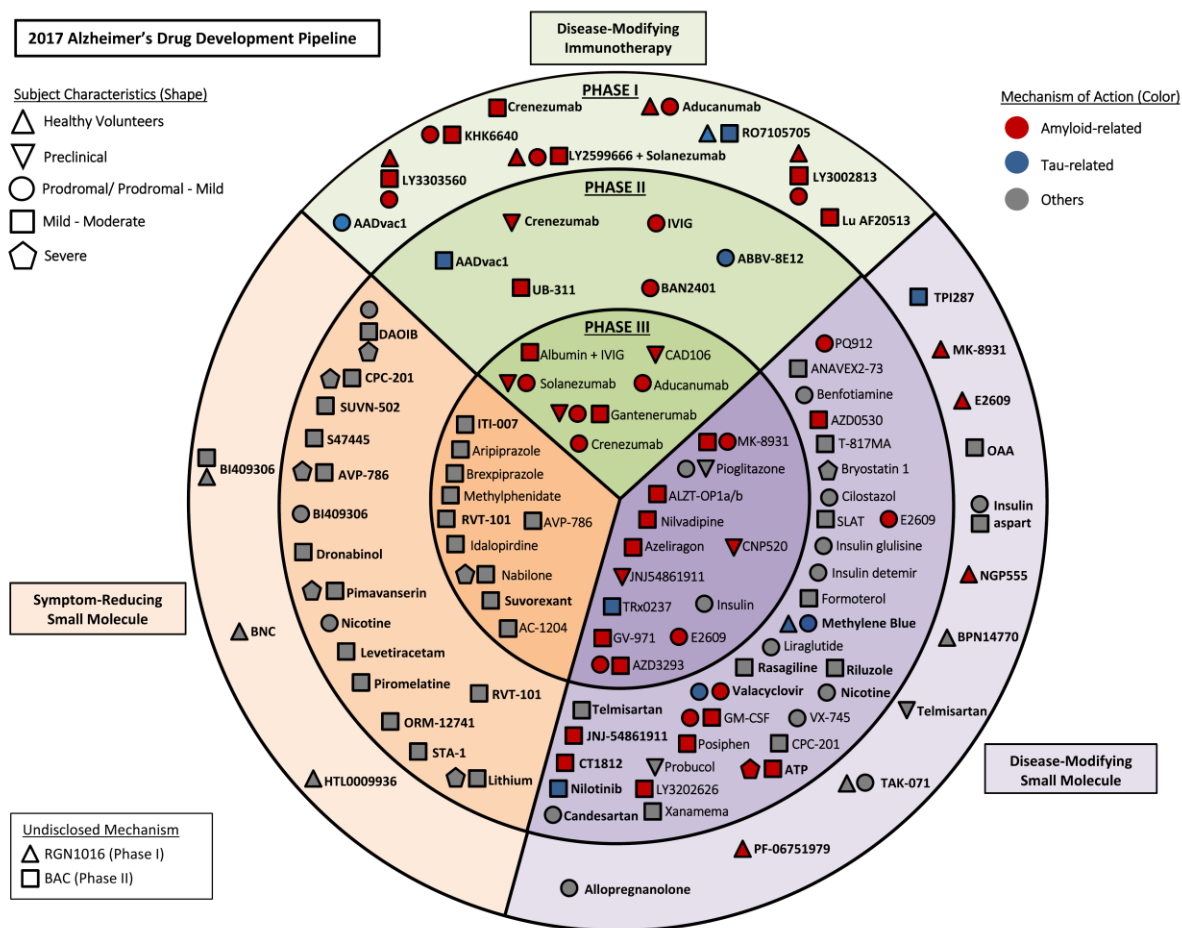


Figure 1.6. Agents in clinical trials for the treatment of Alzheimer's disease as of May 2017 (data acquired from clinicaltrials.gov on 1/5/2017). ATP: adenosinetriphosphate; BNC: bisnorcymserine; GM-CSF: granulocyte-macrophage colony-stimulating factor; OAA: oxaloacetate; IVIG: intravenous immunoglobulin; SLAT: simvastatin 1 L-arginine 1 tetrahydrobiopterin. Figure acquired without modifications from Cummings *et al*¹⁰⁷.

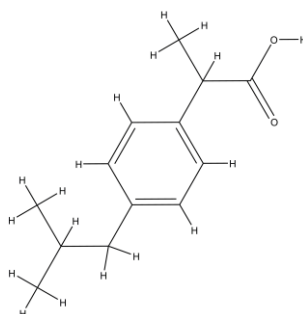
Among the possible agents against AD are many disease-modifying small molecules, which are of special interest to the pharmaceutical industry as they usually present desirable pharmacological properties. This category of potential anti-A β molecules have been extensively pursued by the pharmaceutical industry for the generation of drug candidates for AD, as well as other protein-misfolding diseases (PMDs). A successful example of a small molecule drug is Tafamidis, a stabilizer of the Transthyretin (TTR) tetramer, developed for the treatment of TTR amyloidosis^{246, 247}.

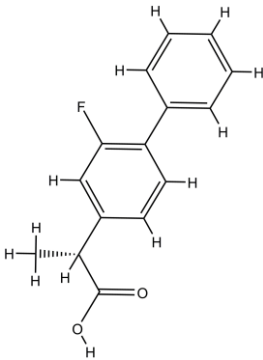
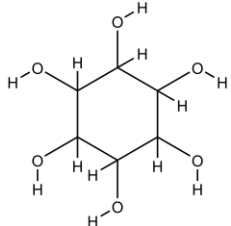
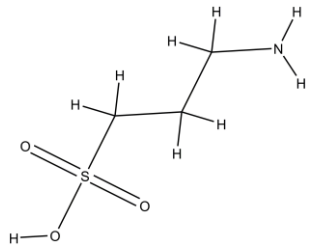
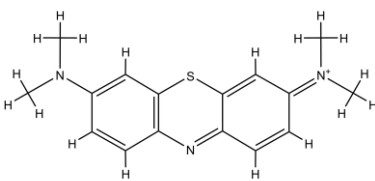
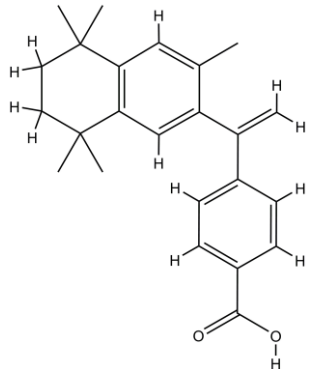
1.12.8. A β -targeting small molecules

The majority of available drugs is based on compact compounds with relatively small molecular weights, termed small molecules. Such compounds are more likely to display certain desirable pharmacological characteristics such as the ability to cross membranes and oral availability. Therefore, small molecules are of tremendous interest to the pharmaceutical industry and the major class of candidates in the search for AD treatments (**Figure 1.7**), along with antibodies. The challenge posed by the necessity to cross the blood-brain barrier (BBB) in order to co-localize with amyloid formations suggests that a small molecule lead would be more likely to develop into an AD drug, considering the large number of small molecule drug paradigms.

Many small-molecule candidates have been identified so far for AD. Out of these, molecules that exhibit anti-amyloid effects include linear peptides with homology to specific segments of the A β sequence²²⁶ as well as chemical compounds such as scyllo-inositol²²⁷, tramiprosate²²⁸, methylene blue²⁴⁸, ibuprofen^{249, 250}, R-flurbiprofen^{251, 252} and bexarotene²³¹, which have shown direct or indirect modulatory effects on A β aggregation and neurotoxicity. Notably, a number of amyloid-related small molecules against AD have advanced to clinical trials^{106, 107}. Small molecules with direct or indirect modulatory effects on the toxicity induced by the aggregation of A β , are presented in **Table 1.3**.

Table 1.3. Small molecules with anti-amyloid activity against A β .

Drug	Clinical Trials status	Mode of action	2D structure
Ibuprofen ^{249, 250}	N/A	γ -secretase modulation	

Drug	Clinical Trials status	Mode of action	2D structure
R-flurbiprofen ^{251, 252}	Phase 3, discontinued	γ -secretase modulation	
Scyllo-inositol ²²⁷	Phase 2	Aggregation prevention by stabilization of low-N oligomers	
Tramiprosate ²²⁸	N/A	Antagonizes the stabilization effect of proteoglycans ²⁵³	
Methylene blue ²⁴⁸	N/A	Promotes A β fibrillization	 Cl ⁻
Bexarotene ^{231, 239}	Phase 2	Increases ApoE production/ Suppresses primary aggregation interactions of A β	

Some of the molecules reported above have been shown to directly interact with A β and inhibit the toxicity associated with its aggregation, such as tramiprosate, bexarotene and scyllo-inositol. Typical protocols for the discovery of such molecules usually involve monitoring of A β aggregation. For instance, the ability of tramiprosate to stabilize an α -helical/random coil-rich A β conformation was studied with circular dichroism (CD)²²⁸. CD is a widely used method in the characterization of the secondary protein structure, and can report on β -sheet formation when aggregation occurs. Similarly, the initial identification of scyllo-inositol interactions with A β also relied on CD²⁵⁴. Biochemical assays such as fibril-staining with thioflavin, have been used extensively in the identification and study of molecules that interfere with the aggregation kinetics of A β .

Generally, microscopy methods like transmission electron microscopy (TEM), biochemical assays and spectroscopy are widely used in the discovery and study of potential aggregation modifiers of A β . Moreover, such biochemical and biophysical methods can be applied in the screening of chemical libraries, with the aim of identifying an A β aggregation inhibitor or modifier. However, these screens are usually laborious and slow, since they require separate sample preparation for each studied molecule. Accordingly, the screening of larger libraries with biochemical and biophysical assays raises the cost of the assay significantly. Naturally, the screening of libraries with these methods produces relatively low yields with regard to the number of hits, which could be traced to the small number of molecules that can be effectively screened.

A way to increase the throughput of screening protocols that yield anti-A β leads, involves screening DNA-encoded chemical libraries. Techniques such as mRNA-display²⁵⁵ and phage-display²⁵⁶ have been previously used in the discovery of peptides that interfere with the aggregation of A β ^{257, 258}. These technologies overcome the problem of small library sizes, since they allow the screening of extremely large populations. However, display assays do not allow selection of hits based on their ability to interfere with the process of aggregation; rather, they rely heavily on the hits' ability to bind A β . In

this way, display assays increase the chances of identifying hits with high affinity for the target that will require an evaluation of their ability to interfere with A β aggregation.

Unfortunately, none of the above methods for the discovery of anti-A β leads has been successfully tested in clinical trials. While this problem certainly does not invalidate the methods themselves, an alternative approach that involves more high-throughput functional assays might increase the chances of identifying a disease-modifying small molecule. Recently, various microbial systems have emerged as capable screening tools for aggregation inhibitors against a variety of protein targets, as well as A β ²⁵⁹⁻²⁶⁴. These systems usually rely on the heterologous production of the protein-target and the subsequent observation of a cellular phenotype that has been linked to the same protein. Moreover, the advancement of biosynthetic applications has enabled the easy adaptation of microbial screens for high-throughput screening, as will be discussed in the following paragraphs.

1.13. Microbial screens for the detection of A β aggregation inhibitors

The briefly presented assays below, serve to highlight the applicability of microbial platforms in the discovery of anti-aggregation molecules or agents, with the ability to target A β . Nevertheless, microbial assays are being used for the detection of possible therapeutics for many protein-misfolding diseases (PMDs)^{259-262, 264, 265}.

1.13.1. Dye-based bacterial screens for aggregation inhibitors

Thioflavin can penetrate biological membranes and bind amyloid formations of various misfolding or aggregating proteins, allowing the detection of those formations through a shift in fluorescence. Thioflavin-based assays have been successfully tested in multiple screens for anti-amyloid

inhibitors²⁶⁶⁻²⁶⁸. Such assays rely on the fact that intrinsically disordered or misfolding-prone proteins produced in the bacterium tend to aggregate and form inclusion bodies that may contain amyloid-like formations²⁶⁹. Thioflavin can bind such inclusion bodies, forming the basis of the screening assay (**Figure 1.7**). Subsequently, any agent that corrects the misfolding or prevents the aggregation of the target protein would inhibit inclusion body formation, and the absence of the characteristic shift in thioflavin fluorescence would indicate the presence of the agent.

A drawback of using thioflavin dyes in the cell is the potential fluorescence interference by cellular constituents. To counter this problem, ProteoStat was suggested as an alternative dye, proven in a bacterial screen of low activity inhibitors of A β ₄₂²⁷⁰.

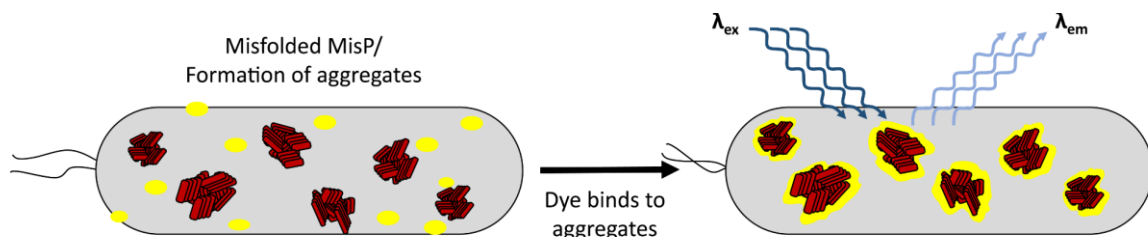


Figure 1.7. Dye-based assay for the detection of aggregates. The dye penetrates the microbial cell, binding protein aggregates. Bound dyes exhibit specific fluorescence profiles, allowing detection and quantification of aggregate formations within the cell.

1.13.2. Bacterial screens using target-protein fusions with a reporter protein

Fusing A β to a reporter protein can result in a chimera that tends to aggregate upon expression²⁷¹, since the aggregation tendency of A β affects the folding status of the whole fusion, forcing it to aggregate. Therefore, it is possible to use such chimeric fusions for the discovery of aggregation inhibitors for A β : any agents that block or slow A β aggregation would allow the reporter protein to fold properly and exhibit an observable phenotype, indicating the presence of the agents in question.

1.13.2.1. Screening assays based on end-to-end A β fusions with GFP

The green fluorescent protein (GFP) fusion of A β presents a very useful tool for the discovery of therapeutic candidates. GFP was first isolated from the jellyfish *Aequorea victoria* and was linked to the jellyfish's bioluminescence. Its green fluorescence is owed to a chromophore, whose structure was proposed by Shimomura in 1979²⁷². Following the work of Prasher *et al.* in the early 1990s, the scientific community began to realize GFP's potential as a reporter protein²⁷³.

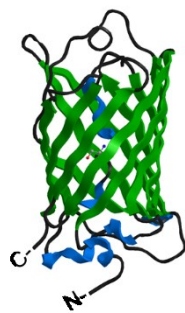


Figure 1.8. Crystal structure of wild-type GFP as solved by Yang *et al.* The structure consists of a cylinder of β -sheets engulfing an α -helix, and short α -helices at the bottom of the cylinder. This structure was first observed for GFP and was aptly named “ β -can” by the authors²⁷⁴.

In 1999, Waldo *et al.* suggested the use of GFP as a folding reporter in an *E. coli* expression system²⁷⁵. Specifically, it was shown that fusing a target-protein to the N-terminus of enhanced GFP (EGFP) through a GSAGSAAGSGEF linker, could report on proper target-protein folding²⁷⁵. The authors anticipated that any target-protein misfolding, would directly hinder EGFP's fluorescence by not allowing it to form a functional chromophore, and the fusion would accumulate in insoluble formations²⁷⁵. In contrast, a well-folded protein partner should allow EGFP to fluoresce.

A β_{42} inhibits EGFP fluorescence when the two proteins are fused²⁷⁶. Overexpression of A β_{42} -EGFP in *Escherichia coli* led to insoluble fusion accumulations in which EGFP fluorescence was severely inhibited, leading to a non-fluorescent cell phenotype²⁷⁶. This assay was used to study the effect of A β_{42} sequence mutations on its aggregation of A β_{42} ²⁷⁶. Briefly, the assay identified less amyloidogenic

mutants of A β ₄₂ in fusion with EGFP, because of their enhanced bacterial fluorescence and fusion solubility. The same bacterial assay was used to screen a triazine analogue library of ~1000 compounds²⁶⁰ and a more extended library of 65,000 compounds, identifying a compound that inhibited the formation of A β ₄₂ aggregates and improved the lifespan and locomotive ability of a *Drosophila* model of AD²⁷⁷. In another study, the bacterial A β ₄₂-EGFP fluorescence assay was used to screen two semi-rationally designed libraries of linear peptides²⁷⁸. Co-expression of the generated peptide library and A β ₄₂-EGFP was followed by selection of the bacteria exhibiting increased levels of fluorescence. The authors identified three anti-aggregation oligopeptides, one of which was also shown to disaggregate preformed A β aggregates²⁷⁸.

Overexpression of the A β -GFP fusion in the yeast *Saccharomyces cerevisiae* displayed a cellular phenotype that resembled that of the equivalent bacterial assay: production of insoluble aggregates of fusion protein with decreased fluorescence²⁷⁹. Screening of a yeast gene deletion library revealed several genetic factors that affected the A β -GFP fluorescence and localization²⁷⁹. One such modulator was the MAP kinase PBS2, a homologue of human MAP2K4. This enzyme is known to activate in the presence of oligomeric A β species in cortical neurons²⁸⁰. The significance of this finding was highlighted upon overexpression of MAP2K4 in a *Caenorhabditis elegans* model of AD, which resulted in intensified neuronal loss²⁷⁹.

1.13.2.2. Screening assays based on end-to-end A β fusions with Tat-pathway substrates

The twin-arginine (Tat) translocation pathway is one of three pathways responsible for the translocation of proteins into the periplasmic space of *E. coli* cells. This pathway only recognizes well-folded proteins as substrates for translocation, through the quality-control function of a translocase^{281, 282}. This translocase is composed of three membrane proteins (TatA, TatB, and TatC), and is versatile with regard to the size and the type of its substrates²⁸³.

The fusion of a misfolding or aggregating protein with a signal sequence (N-terminal) that allows identification by the Tat machinery, and with β -lactamase (Bla) (C-terminal), could be used as a screening assay that correlates A β aggregation with resistance to β -lactam antibiotics. Fusions that fold properly will be transported to the bacterial periplasm more efficiently, granting resistance to β -lactam antibiotics²⁸⁴. Likewise, fusion of A β with the signal sequence from the substrate protein TorA (ssTorA) and Bla, allowed the selection of A β variants with decreased tendencies to aggregate²⁸⁴. A similar fusion was successful in screening a library of \sim 1,000 derivatives of the triazine scaffold, for inhibitors of A β aggregation²⁸⁵. This system utilized the hydrolysis-sensitive fluorescent Bla substrate CCF2/AM, which exhibits a fluorescence shift upon hydrolysis by Bla²⁸⁶. As a result, four compounds that enabled the translocation of the A β -Bla fusion into the periplasm were identified and their anti-aggregation activity was verified *in vitro*, in a thioflavin T (ThT) fibril-staining assay.

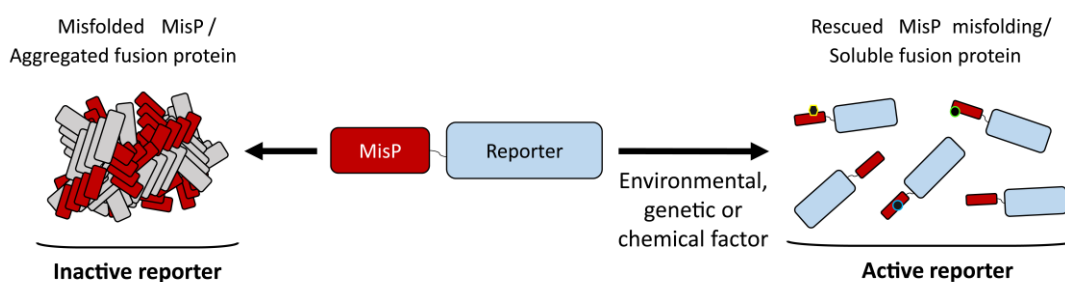


Figure 1.9. Representations of types of misfolding/agggregating protein fusions. Agents that prevent misfolding/agggregation of the misfolded protein (MisP) lead to an active reporter protein, which then leads to the display of an observable microbial phenotype.

Overall, microbial screens present the advantage of correlating a microbial phenotype to the folding status of the target-protein. Moreover, microbial screens are easy to use and provide a straightforward means of screening various types of chemical libraries, in a very high-throughput manner. The microbial screens described above are only representative examples of such screening systems. A more detailed presentation of those systems can be found in in a review by Kostelidou *et al.* 2018²⁸⁷. The most significant advantage of these microbial tools however, is that they can be

genetically modified to incorporate various biosynthetic protocols that allow the generation of highly diverse libraries, which significantly improve the hit identification capabilities of the microbial system, as discussed below.

1.14. Increasing the diversity of chemical libraries in microbial screens

As discussed, small molecules carry a special appeal for the pharmaceutical industry because of the pharmacological properties usually associated with this class: small molecular weights, membrane permeability etc. However, the tremendous diversity of this class of molecules (10^{60} molecules, with less than 30 heavy atoms^{288, 289}) is very difficult to access with typical synthetic libraries, since these usually reach sizes of up to $\sim 10^5$ - 10^7 members or they are based on the derivatization of a common scaffold, offering relatively small diversities.

A very convenient alternative to synthetic libraries is the *in vivo* generation of libraries, which has become a very effective approach with the advancement of synthetic biology technologies. Currently available technologies allow the generation of various types of *in vivo* generated libraries, which exhibit a number of advantages. For instance, from a financial perspective, these libraries do not involve the costs of library creation and usage that are usually associated with synthetic libraries. On the contrary, biosynthetic libraries are extremely easy to use, since library generation usually requires a simple cell culture. However, the most significant advantage provided by these technologies relates to the size and diversity of the created libraries. These libraries can contain a tremendous number of members, offering a diversity that is impossible to reach with synthetic or natural product libraries. As expected, this diversity is particularly useful in the search for novel lead compounds and significantly increases the chances of their discovery. Another most critical advantage of *in vivo* generated libraries is that they can be seamlessly paired to microbial screening assays. This property is especially valuable as it allows the combination of the high-throughput efficiency of microbial

screens with the ability of biosynthetic protocols to produce large and extremely diverse compound libraries. A subsequent advantage of combining a microbial screen to a biosynthetic library protocol is that hit deconvolution is largely facilitated, since the genetic information responsible for any discovered hits is carried within the cells²⁹⁰. Finally, an additional benefit of this combination is that selection takes place in a complex biological environment, requiring high selectivity from hits, in order to avoid random binding and specifically identify their protein target²⁹⁰. Altogether, the incorporation of biosynthetic libraries in microbial screening assays presents a tremendous opportunity for the discovery of small-molecular hits that target the aggregation of A β and subsequently inhibit its neurotoxic effect. Such highly diverse libraries permit the investigation of large portions of the chemical space occupied by small molecules, significantly increasing the chances of hit discovery.

Notably, one of the most pharmacologically interesting classes of molecules that can be easily produced via biosynthetic protocols in microbial cells involves peptides. A variety of peptide molecules can be produced through biosynthetic protocols in various biological systems. Furthermore, the advantages associated with this class of molecules make peptides a very convincing option for the development of drug candidates.

1.15. Synthetic biology in the discovery of cyclic peptide pharmaceuticals

1.15.1. General introduction to peptide pharmaceuticals

Peptides have been in the focus of the pharmaceutical industry since the 1980s, with more than 60 peptide-based therapeutics already in the market and more than 150 in development²⁹¹. Peptides are considered attractive therapeutic agents because of their potency and their high selectivity for a biological target *in vivo*. Nevertheless, peptides usually exhibit severely decreased oral bioavailability, poor membrane permeability and short half-life in plasma, compared to other small-molecule

therapeutics²⁹². Such weaknesses are however, addressable through modification strategies that modulate a peptide's pharmacokinetic properties, as well as with recombinant production of peptides with non-natural amino acids^{293, 294}.

A very appealing option for the amelioration of peptide pharmacological properties has been the development of cyclic peptide therapeutics. Cyclic peptides are naturally occurring compounds that have attracted significant interest from the pharmaceutical industry, with a total of nine cyclic peptide drugs having entered the market from 2006 to 2015 and numerous undergoing clinical trials²⁹⁵. This class of biological molecules assume less flexible conformations compared to their linear analogues, leading to energetically favourable target binding. Moreover, cyclic peptides are less likely to induce toxicity because of their benign amino acid composition and they are biochemically more stable than linear peptides.

1.15.2. Biosynthesis of cyclic peptides

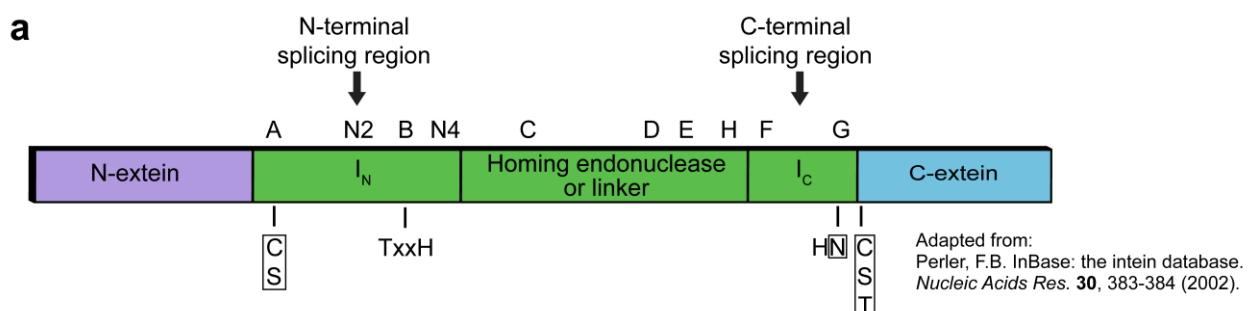
Adding to the advantages of cyclic peptides is the ease by which libraries can be produced biosynthetically. Biosynthesis of cyclic peptides is unmatched in its ability to generate extremely high-diversity libraries that are easily produced and screened²⁹⁶. Cyclic peptides can be produced via DNA-encoded technologies that involve cyclization through disulphide bond formation^{297, 298}, genetic reprogramming for the introduction of non-natural amino acids that assist cyclization²⁹⁸ and chemically-induced side-chain reactions, often used in libraries screened by mRNA display²⁵⁵. The latter, allows the generation of libraries of extreme diversity. However, their screening process occurs *in vitro*, impeding hit deconvolution. Other biosynthetic methods for cyclic peptide library construction also include non-ribosomal peptide synthetase-based protocols²⁹⁹, which, as suggested by their name, do not rely on ribosomal synthesis of the peptide, rather, they consist of a series of

enzyme-mediated amino acid additions and modifications that result in a cyclization reaction which is catalyzed by a thioesterase domain³⁰⁰.

One of the most convenient approaches for the simple biosynthesis of cyclic peptides entails the use of an intein-based technology, termed split intein-mediated circular ligation of peptides and proteins, or SICLOPPS^{296, 301, 302}. This technology allows the *in vivo* generation of exceptionally diverse libraries of cyclic peptides, exceeding library sizes of 10^9 members in *E. coli*. Importantly, this approach is very easily combined with a microbial screening assay, allowing the simultaneous generation and screening of the library. A brief introduction to inteins and a description of the SICLOPPS technology follows below.

1.15.2.1 Introduction to intein-mediated protein splicing

Inteins are a class of polypeptides found in all three taxonomic domains: archaea, bacteria and eukarya. They belong to the superfamily of proteins termed Hedgehog/Intein domains (Hint) which consists of inteins, bacterial intein-like (BIL) domains, and Hedgehog auto-processing (Hog) domains³⁰³. Inteins are functional components for protein precursors that usually consist of an intervening sequence (intein) and two flanking sequences (exteins) (**Figure 1.10a**). In these precursors, the intein domain is responsible for protein splicing: the creation of a peptide bond between the flanking exteins that yields the mature protein (**Figure 1.10b**)³⁰⁴.



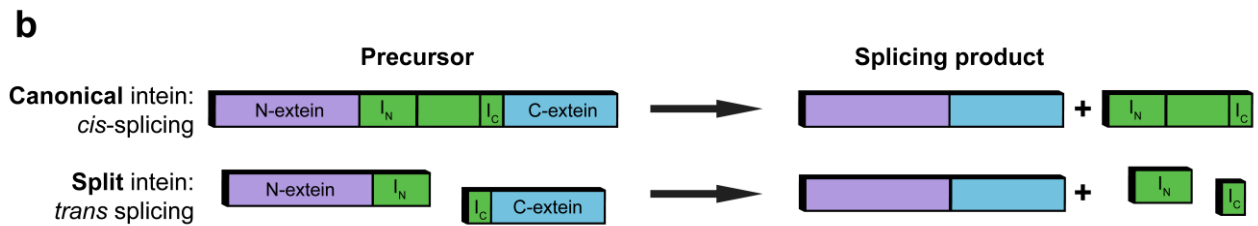


Figure 1.10. Naturally occurring inteins. a) Amino acid motifs that compose an intein-containing precursor. I_N and I_C signify the N- and C-terminal domains of the intein, respectively. Motifs responsible for protein splicing consist of blocks A, N2, B, N4, F and G³⁰⁵. Boxed amino acids are conserved and involved in the four nucleophilic substitutions that take place during splicing. Vertically grouped amino acids signify conserved nucleophilic residue alternatives. Threonine and Histidine in block B facilitate splicing. Blocks C, D, E, and H do not contribute to protein splicing and form the four conserved helices of the homing endonuclease domain. The majority of reported inteins contain a homing endonuclease however some inteins contain linkers of unknown function³⁰⁵. **b)** Two types of naturally occurring intein precursors have been reported. In canonical inteins, splicing involves a precursor that contains the complete intein sequence flanked by the extein domains. These precursors are *cis*-splicing. On the contrary, split inteins involve two separate precursors and splicing is achieved in a *trans* manner.

The commonly accepted and most widely encountered (canonical) mechanism of intein-mediated protein splicing takes place in four steps as described in **Figure 1.11**:

1. N-O or N-S acyl rearrangement: The -SH group of Cys or the -OH group of Thr or Ser in block A (**Figure 1.10a**) activate the N-terminal domain of the intein (I_N) by forming a thioester or an ester respectively, with the preceding carbonyl group.
2. A second nucleophilic substitution takes place through the conserved Cys, Ser or Thr residues of the C-extein. This residue attacks the above thioester/ester resulting in a transesterification reaction that transfers the N-extein to the side chain of the Cys, Ser or Thr first residue of the C-extein.
3. The next step involves the cyclization of the Asn preceding the initial Cys, Ser or Thr residues of the C-extein. This cyclization results in the release of the N-extein/C-extein complex.
4. The final step consists of an O-N shift that leads to a native peptide bond between the N-extein and the C-extein. A detailed mechanism can be found in the intein database: InBase³⁰⁵.

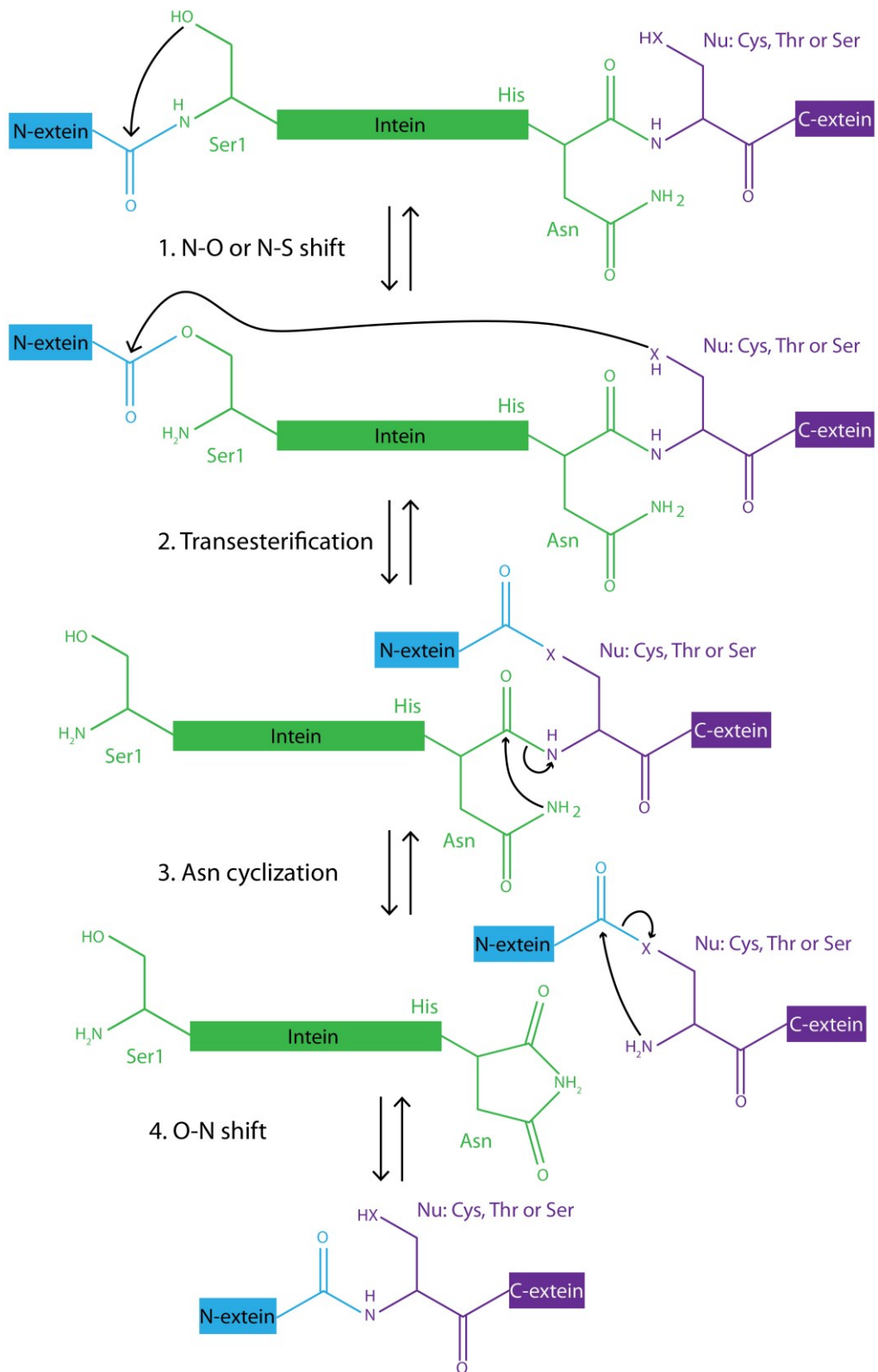
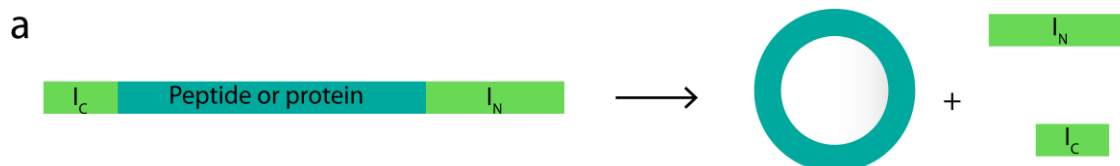


Figure 1.11. The canonical splicing mechanism of inteins. Nu represents the conserved nucleophilic residues at the N-terminus of the C-extein domain. X in the sidechain of the nucleophilic residue is either an –OH or an –SH.

1.15.2.2. SICLOPPS

The usefulness of inteins in peptide or protein cyclization first became evident after the observations of Wu *et al.*, who identified a *trans*-splicing intein (**Figure 1.10b**)³⁰⁶. The authors showed that two genes separately encoded the N- and C-termini of DnaE (the catalytic subunit α , of DNA polymerase III in *Synechocystis* sp. PCC6803)³⁰⁶. The N-terminus of one part of the DnaE extein was followed by the N-terminus of the intein (I_N) and the C-terminus of DnaE was preceded by a C-terminus of the intein (I_C)³⁰⁶. These findings led Benkovic and co-workers to suggest that inteins could mediate the cyclization of peptides or proteins, if the amino acid sequence in question was placed between the I_C and I_N domains of an intein: I_C -(protein or peptide)- I_N (**Figure 1.12a**)^{301, 302}.

The proposed mechanism of splicing for SICLOPPS involves a nucleophilic substitution performed by the first residue of the intervening sequence. The mechanism of SICLOPPS is very similar to the mechanism of *cis*-splicing inteins³⁰⁵, although it appears to be a less complicated version. An essential element of SICLOPPS is the asparagine as the last residue of the I_C splicing domain, as its absence does not allow the necessary residue cyclization to occur, prohibiting the release of the cyclic peptide (**Figure 1.12b**)^{301, 302}.



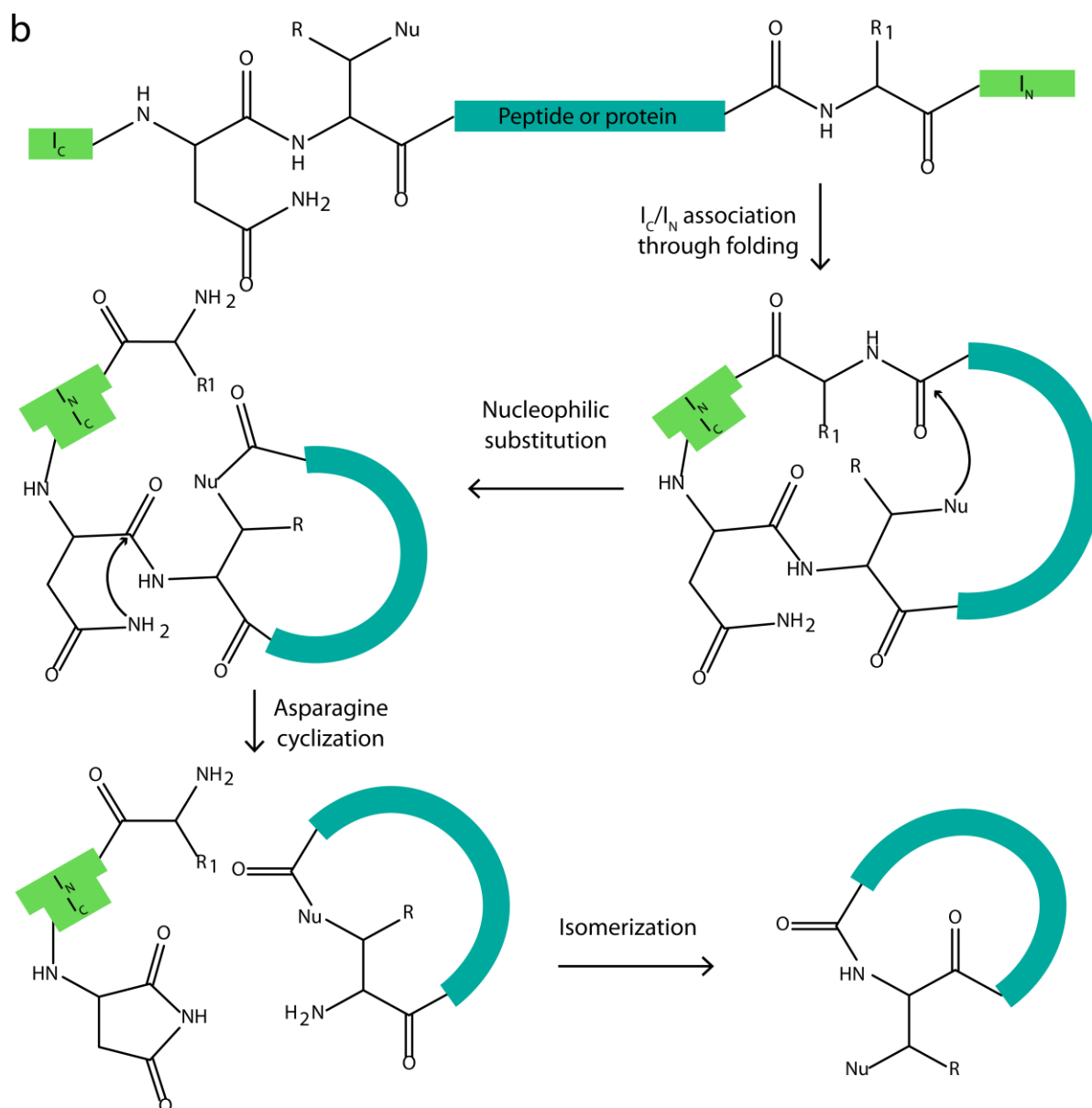


Figure 1.12. Production of cyclic peptides or proteins with SICLOPPS. a) The protein fusion consists of the splicing domains of an intein, flanking the protein or peptide that will undergo cyclization. In order for splicing to occur, the I_C splicing domain must be upstream of the intervening sequence and the I_N domain. **b)** Suggested mechanism for the cyclization of the intervening sequence. Nu: the nucleophilic side group necessary to perform the nucleophilic substitution. This is the first residue of the peptide or protein sequence. The product of the nucleophilic attack, a lariat, undergoes asparagine cyclization, followed by an isomerization that yields the final cyclic peptide.

Thus far, SICLOPPS libraries have facilitated the identification of several drug candidates, such as cyclic peptides that inhibit Dam methyltransferase or peptides that reduce the toxicity of α -synuclein^{262, 307-}

³⁰⁹. Moreover, its ability for straightforward combination with microbial screening assays^{262, 307-309} and its capacity to generate highly diverse libraries establish SICLOPPS as a valuable element for microbial systems that aim to facilitate the discovery of inhibitors of A β aggregation and toxicity. Additionally, SICLOPPS libraries are used mostly in functional screening assays rather than extracellular affinity assays, such as phage or mRNA display²⁹⁶. Therefore, produced hits do not only display affinity for the target-protein, but also affect the target-protein's folding and aggregation. This combination allows easy hit deconvolution, since the information regarding the hit is DNA-encoded within the cell. Finally, SICLOPPS can easily generate libraries of peptides with short backbones that display very high diversities, as shown in the following chapter. The relatively small molecular weights of those peptides as well as their compact structure due to their cyclic nature, provide them with small molecule-like characteristics. As such, SICLOPPS presents a great tool for the investigation of the chemical space occupied by small molecules

1.16. Thesis outline

The work presented in this thesis describes the development of a bacterial platform for the discovery of inhibitors of A β aggregation and neurotoxicity, with potentially therapeutic properties against AD. This platform enables the biosynthesis of combinatorial libraries of cyclic peptides with extended chemical diversities in *E. coli* cells, and their simultaneous screening for A β aggregation inhibitors through an ultrahigh-throughput microbial genetic assay. Furthermore, this assay has the ability to detect bioactive cyclic peptides by simply monitoring the bacterial fluorescence of engineered cells that co-express a chimeric A β fusion with the green fluorescent protein (GFP), along with the cyclic peptide libraries under investigation. The selection process is based on the detection and isolation of the individual bacterial clones that exhibit enhanced fluorescence, via fluorescence-activated cell sorting (FACS). FACS complements the described bacterial system in a very efficient manner as it

allows ultrahigh-throughput sorting capabilities that can easily accommodate the screening of extremely diverse microbially produced libraries.

Through use of this engineered *E.coli* system, it was possible to biosynthesize and rapidly screen a cyclic peptide library of more than 10 million different short cyclic peptides. This process allowed the identification of hundreds of bioactive peptides with putative modulatory effects on A β aggregation. Among the multitude of isolated hits, two cyclic pentapeptides displaying the sequences cyclo-SASPT and cyclo-TAFDR, termed A β C5-34 and A β C5-116 respectively, were chemically synthesized and purified. Subsequently, these cyclic peptides were used in a variety of biochemical, biophysical and biological assays *in vitro*, in order to investigate their effect on the aggregation and neurotoxicity of A β . The outcome of those analyses revealed that the selected cyclic peptides interfered with the normal course of A β aggregation, leading to the formation of atypical A β aggregates with reduced neurotoxicity, compared to that of typical A β fibrillar structures. The protective effects of A β C5-34 and A β C5-116 against the cytotoxic aggregation of A β were also evaluated *in vivo*, in established *Caenorhabditis elegans* models of AD. Indeed, these assays substantiated the protective effect of A β C5-34 and A β C5-116 against the cytotoxic aggregation of A β . Combined, the *in vitro* and *in vivo* assays established A β C5-34 and A β C5-116 as attractive leads, which upon further evaluation and development, could lead to anti-amyloid therapies against AD. Finally, a combination of site-directed mutagenesis data and next-generation sequencing analysis of hits derived from the screening process, led to the discovery of structure-activity relationships within the selected cyclic peptide population. In doing so, these analyses were very helpful in the characterization of distinct families of cyclic peptides with bioactivity against the aggregation of A β .

Chapter 2. Screening of a biosynthetic cyclic peptide library and identification of bioactive hits with inhibitory effects against the aggregation of A β

2.1. *In vivo* generation of a diverse library of cyclic peptides

In order to discover potential inhibitors of the aggregation of A β , it was decided to investigate recombinant libraries of backbone-cyclized, small cyclic peptides. As discussed previously, cyclic peptides acquire more stable conformations compared to their linear analogues, they are more resistant to proteolysis and generally display higher affinities for the target protein²⁹⁵. Moreover, the small size of peptides included in these libraries provides them with small molecule-like characteristics, as their molecular weights resemble those of usual small molecule pharmaceuticals.

An important factor in the decision to investigate cyclic peptide libraries was the ability to produce them *in vivo*, which was facilitated by the SICLOPPS technology. SICLOPPS libraries are easily produced and simultaneously screened in microbial systems, and depending on the system, they can contain significant chemical and structural diversity. Indeed, the limit to the size of the encoded library is usually defined by the host's transformation efficiency. SICLOPPS library sizes can reach billions of members when produced in bacteria, since the bacterial host allows for such transformation efficiencies.

2.1.1. A SICLOPPS-generated cyclic peptide library

For this work, the constructed SICLOPPS libraries were designed to produce tetra-, penta- and hexapeptides, with the aim of combining small-molecule-like behaviour to expanded library diversity. In order to achieve the maximum library diversity, cyclic peptides were designed according to the formula cyclo-NuX₁X₂...X_n, where Nu stands for any of the three amino acids with nucleophilic side-

chains (Cys, Thr, Ser), X is any of the 20 natural amino acids and n=3-5. Theoretical library sizes based on all possible amino acid combinations, are shown in **Table 2.1**.

Table 2.1. Theoretical library sizes. Nu is any of the three amino acids with nucleophilic side-chains (Cys, Thr, Ser), X is any of the 20 natural amino acids. n=3-5.

Formula	Peptide sequence length	Fully randomized positions (n)	Library size
cyclo-NuX ₁ X ₂ X ₃	4	3	24,000
cyclo-NuX ₁ X ₂ X ₄	5	4	480,000
cyclo-NuX ₁ X ₂ X ₅	6	5	9,600,000
	combined		10,104,000

Library size = 3 × 20ⁿ

The combination of all three libraries was expected to contain 10,104,000 cyclic peptides (**Table 2.1**). Construction of this highly diverse peptide library first required the construction of an equally diverse plasmid vector library. For that, the plasmid vector library pSICLOPPS-NuX₁X₂X₃₋₅ was constructed using degenerate codons (see **Methods and Materials**). Briefly, the nucleophilic residues (Nu: Cys, Ser, Thr) were encoded by the codons TGC, AGC, and ACC respectively, and the X residues (X: any of the 20 naturally occurring amino acids) were encoded by NNS codons (N: A, T, G, C and S: G, C)³¹⁰. The NNS sequence allowed the random generation of all 20 amino acids through 32 codons excluding the UAA (ochre) and UGA (opal) stop codons³¹⁰. Thus, the possibility of introducing stop-codons in the peptide sequence was minimized. In the constructed plasmid vector, the peptide-encoding DNA sequence was placed between the C-terminal (I_C) and N-terminal (I_N) splicing domains of the *Synechocystis* sp PCC6803 DnaE intein (**Figure 2.4a**). In addition, a chitin binding domain (CBD) was located downstream of I_N. This CBD domain was originally inserted in the fusion to facilitate purification, as described in Tavassoli and Benkovic, 2007³¹⁰. In this occasion however, its purpose was to aid in the detection of fusion production and intein splicing via western blots (**Figure 2.5**).

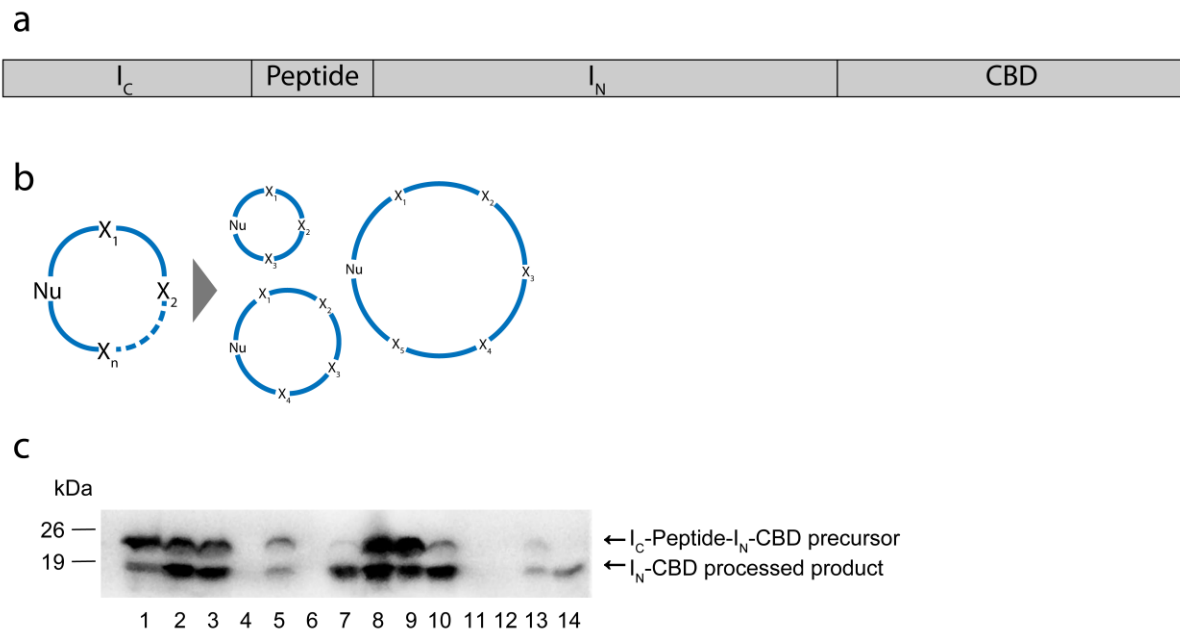


Figure 2.4. Split-intein fusion of the *Ssp DnaE* intein and encoded peptide size classes. **a)** The split-intein tetra-partite fusion as encoded by the combined pSICLOPPS-NuX₁X₂X₃-X₅ plasmid vector library. **b)** Three peptide size classes are produced by this combined plasmid vector library: tetrapeptides, pentapeptides and hexapeptides. Nu is any of the three amino acids with nucleophilic side-chains (Cys, Thr, Ser), X is any of the 20 natural amino acids. n=3-5. **c)** Anti-CBD western blot for library clones. Representative clones were used to indicate the appearance of the precursor band as well as the processed band that signifies peptide cyclization. Nu is any of the three amino acids with nucleophilic side-chains (Cys, Thr, Ser), X is any of the 20 natural amino acids. n=3-5. Library construction and characterization were performed by D. C. Delivoria, at NHRF.

Following transformation of the combined pSICLOPPS-NuX₁X₂X₃-X₅ library in *E. coli* cells (see **Methods and Materials**), colonies were selected for PCR with intein-specific primers in order to identify the percentage of cells containing correct inserts. Out of the 124 colonies, 88 (~71%) showed the expected insert. Moreover, when the intein fusion was overexpressed in the presence of 0.002% arabinose, 99 out of 150 clones (66.00%) exhibited production of the non-spliced tetra-partite fusion (splicing precursor). Expression of the precursor was observed by western blotting with a mouse anti-CBD primary antibody in a 1:100,000 dilution, followed by a goat anti-mouse HRP-conjugated secondary antibody in a 1:4,000 dilution (**Figure 2.4c**). The observed bands corresponding to the precursor appeared at ~25 kDa. Additionally, 82 out of the 99 precursor-producing clones (82.83%) also showed a lower molecular-weight band of ~20 kDa that corresponded to the slicing product I_N-CBD, indicating

successful intein processing and possible peptide cyclization. This band is indicative of cyclic peptide release following asparagine cyclization (**Figure 1.13b**) and the dissociation of the I_N and I_C domains in the suggested SICLOPPS mechanism. Estimation of viable colonies showed that approximately 20,760,000 clones (66.00%) were capable of producing the tetra-partite splicing precursor, out of which, 82.83% also produced cyclic peptide sequences. Considering the library was designed to contain 10,104,000 unique peptide sequences, the production of the tetra-partite precursor by 20,760,000 clones ensured an approximate two-fold coverage of the theoretical library size.

Finally, sequencing of 23 randomly selected clones confirmed the existence of all necessary nucleophilic amino acids in the first position (Cys, Ser, Thr) of the peptide-encoding sequence as well as an early estimation of the distribution of the twenty amino acids in the remaining positions of the sequence (**Figure 2.5**).

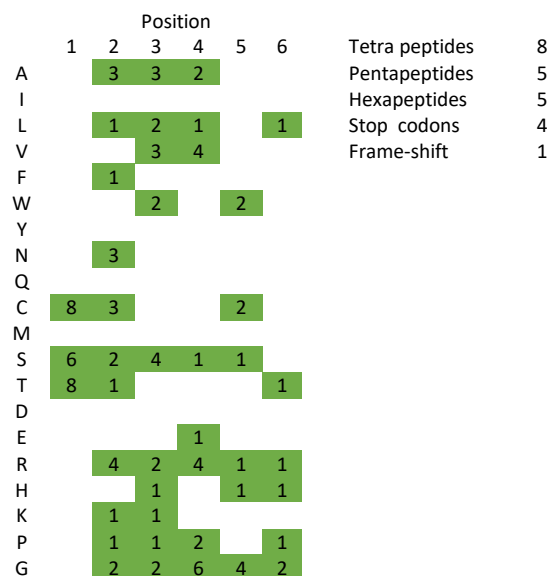


Figure 2.5. Amino acid distribution following DNA sequencing of 23 randomly selected clones. The majority of the 20 natural amino acids are represented in this initial qualitative characterization of the combined library.

2.2. Identification of cyclic peptide inhibitors of A β aggregation

In order to identify molecular inhibitors of A β aggregation, the combined cyclic peptide library described previously, was screened using a microbial genetic assay that monitors A β_{42} aggregation through a fluorescence phenotype. This bacterial assay expresses an end-to-end fusion of A β_{42} with EGFP. The increased aggregation propensity of A β_{42} forces the A β_{42} -EGFP fusion to accumulate in inclusion bodies, while EGFP chromophore formation is disturbed, preventing the emission of green fluorescence (**Figure 2.6**)^{123, 260, 276}. Accordingly, any condition or molecule that inhibits the aggregation propensity of A β_{42} results in the formation of soluble and fluorescent A β_{42} -EGFP, also causing the emission of bacterial fluorescence (**Figure 2.6**).

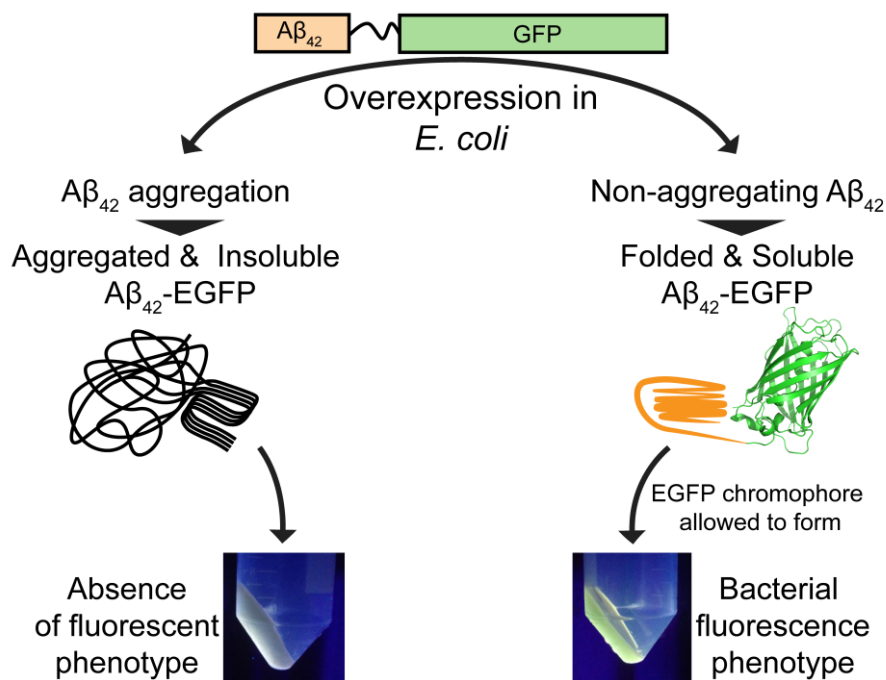


Figure 2.6. The bacterial A β_{42} -EGFP assay. A β_{42} transfers its aggregation propensity on its fusion with EGFP, leading to fusion aggregation and loss of solubility. In this scenario, bacteria produce only background levels of fluorescence (left). When the A β_{42} -EGFP fusion is overexpressed in the presence of an inhibitor for the aggregation of A β_{42} , the fusion is found in a folded and soluble state, which allows the EGFP chromophore to form. These bacterial clones are able to produce fluorescence and are therefore detectable.

In the bacterial A β ₄₂-EGFP assay, the production of the fusion is controlled by the T7 promoter and induced with isopropyl β -D-1-thiogalactopyranoside (IPTG). In order to establish the optimal concentration of IPTG for recombinant A β ₄₂-EGFP production, *E. coli* BL21(DE3) cells were transformed with pETA β ₄₂-EGFP or pETA β ₄₂(F19S;L34P)-EGFP (see **Methods and Materials**). The F19S;L34P “green” mutant of A β ₄₂ in fusion with EGFP (A β ₄₂(F19S;L34P)-EGFP or GM6)²⁷⁶ was used as a positive fluorescence control. These mutations in the A β ₄₂ sequence lead to enhanced bacterial and fusion solubility. Protein production was induced in liquid cultures, with IPTG concentrations of 0 μ M to 1 μ M. Even though A β ₄₂-EGFP forms mainly insoluble aggregates in *E. coli*, a low level of background fluorescence was expected and was indeed observed. Fluorescence measurements of equal-cell-number culture samples were obtained as described in **Methods and Materials**. Both A β ₄₂-EGFP and A β ₄₂(F19S;L34P)-EGFP showed a cell fluorescence maximum at 0.1 μ M IPTG (**Figure 2.7**). In fact, cellular fluorescence started to drop slightly at 1 mM, possibly due to excessive bacterial stress caused by the amyloidogenic properties of the produced fusions or by excessive protein overproduction³¹¹. Eventually, the 0.1 mM concentration was selected for fluorescence experiments that included cyclic peptide production, in order to ensure that high levels of A β ₄₂-EGFP were produced without significant cost to the cell.

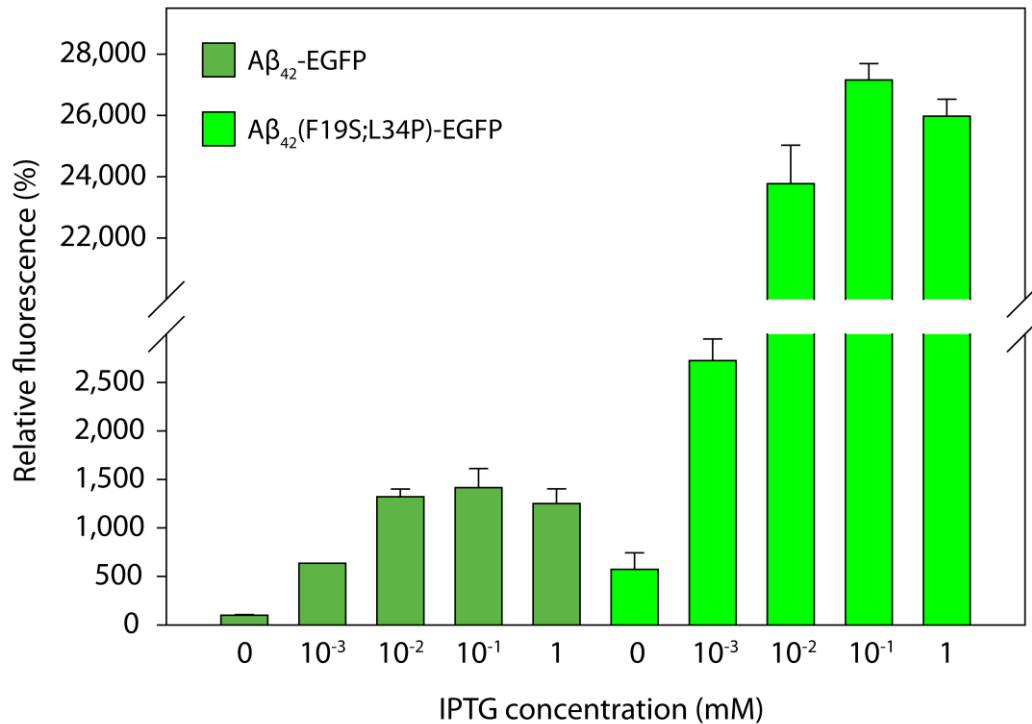


Figure 2.7. Effect of IPTG concentration on bacterial Aβ₄₂-EGFP fluorescence. Cells overexpress Aβ₄₂-EGFP or Aβ₄₂(F19S;L34P)-EGFP. Cell fluorescence was presented in relation to uninduced (0 mM IPTG) bacteria that carry the pETAβ₄₂-EGFP plasmid, which is arbitrarily set at 100%. Mean fluorescence ± sd shown, *n*=1 (triplicate cultures) for all samples.

To determine the optimal temperature for production of Aβ₄₂-EGFP, *E. coli* BL21(DE3) cells were transformed with either pETAβ₄₂-EGFP or pETAβ₄₂(F19S;L34P)-EGFP. Protein overexpression took place in liquid cultures, following induction with 0.1 mM IPTG, at an OD₆₀₀=0.4-0.5. Two temperatures were used to evaluate the effect on cell fluorescence: 20 °C and 37 °C. Noticeably, protein production at 20 °C led to enhanced cell fluorescence. This result lies in agreement with the findings of de Groot *et al.*, which showed that the fluorescence of *E. coli* expressing Aβ₄₂-EGFP is increased at lower temperatures, suggesting a competition between the aggregation and folding of Aβ₄₂-EGFP³¹². The authors advocated that lower temperatures affected the molecular interactions that led to the formation of stable inclusion bodies (IBs). Therefore, such IBs would contain non-denatured EGFP moieties with the ability to exhibit fluorescence³¹². However, the screening assay had to produce stable, non-fluorescent and amyloid-like IBs³¹³⁻³¹⁵. Furthermore, screening at higher temperatures

should lead to the selection of cyclic peptides with a more robust anti-aggregation effect, since aggregation is accelerated under such conditions³¹⁶, thus, protein production at 37 °C was selected.

Interestingly, A β ₄₂(F19S;L34P)-EGFP-producing cells showed enhanced fluorescence at 37 °C. An explanation for this phenomenon could be found in the reduced aggregation potential of A β ₄₂(F19S;L34P). Since, A β (F19S;L34P) is less prone to aggregate, intermolecular interactions are reduced in this fusion allowing the EGFP moiety to transition to a native and fluorescent state. The expected increase in protein production at higher temperatures could then explain the enhanced bacterial fluorescence at 37 °C.

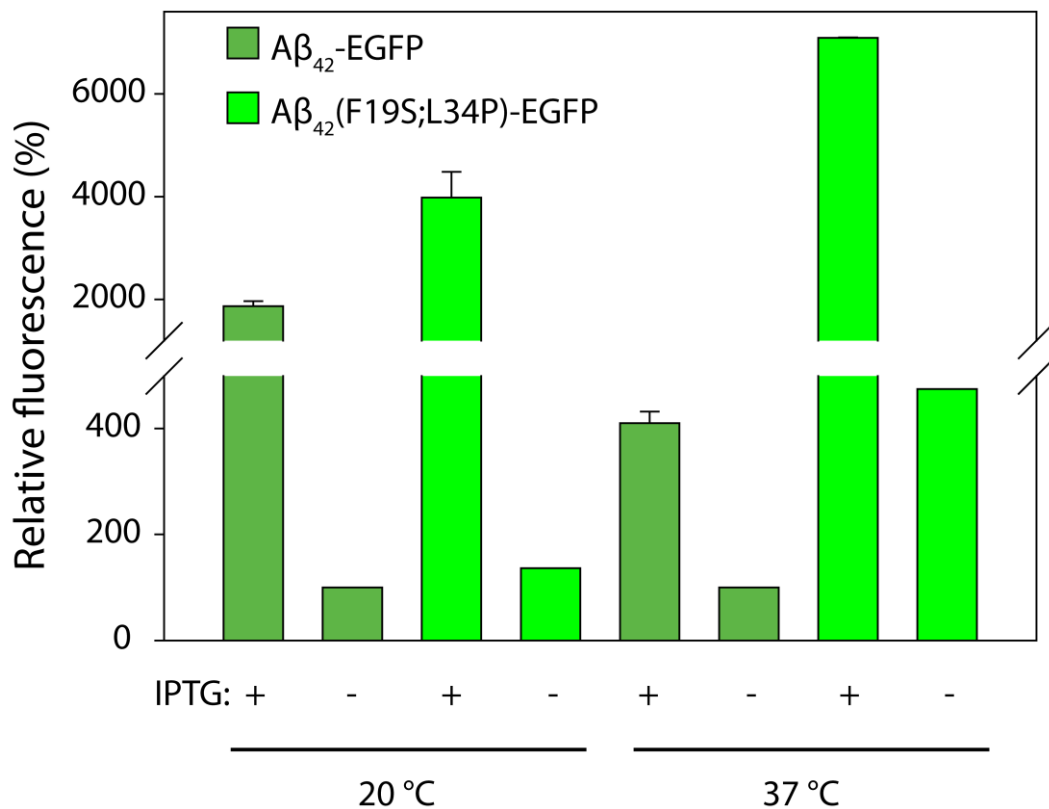


Figure 2.8. Effect of temperature on bacterial fluorescence. Fluorescence of cells producing either A β ₄₂-EGFP or A β ₄₂(F19S;L34P)-EGFP. Cellular fluorescence was compared to the uninduced (-IPTG) A β ₄₂-EGFP culture sample, for each temperature, which was set arbitrarily at 100%. Mean fluorescence \pm sd shown, $n=1$ (triplicate cultures) for all samples.

2.2.1. FACS-enabled screening of the combined library of cyclic peptides

Having established these initial parameters for the production of the A β ₄₂-EGFP it was possible to adapt the bacterial A β ₄₂-EGFP fluorescence assay to perform ultrahigh-throughput screening for cyclic peptide inhibitors of A β ₄₂ aggregation. This was done by modifying the bacteria to produce both the A β ₄₂-EGFP fusion as well as the combined SICLOPPS library. In this way, any produced peptides that inhibited the aggregation of A β would be easily identified by the emission of cellular A β ₄₂-EGFP fluorescence. Fittingly, the method that was chosen for the detection and selection of the fluorescence-emitting bacterial clones was fluorescence-activate cell sorting (FACS) (**Figure 2.7**). FACS presented the advantage of being able to perform very efficient sorting of large populations of cells in small time-frames.

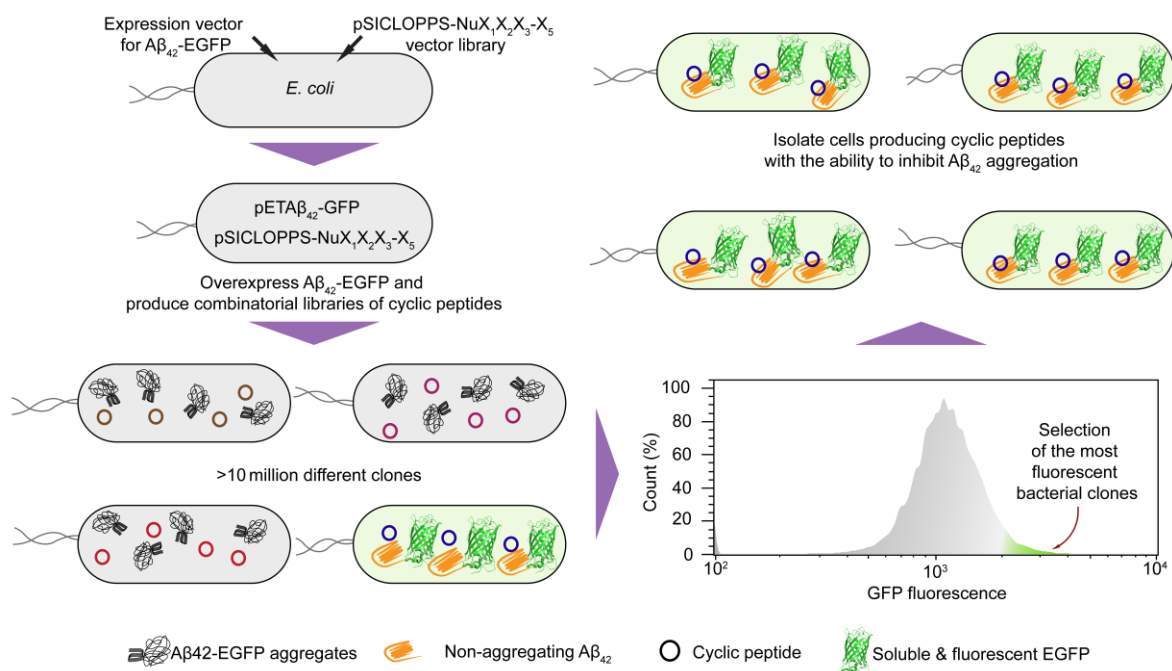


Figure 2.9. Representation of the integrated bacterial system utilized in the identification of aggregation inhibitors for A β ₄₂. Bacterial cells were transformed with plasmids encoding A β ₄₂-EGFP (pA β ₄₂-EGFP) and the combined cyclic peptide library (pSICLOPPS-NuX₁X₂X₃-X₅, Nu: Nucleophilic residue, X is any of the 20 natural amino acids). Following protein overexpression, the population of bacteria underwent fluorescence-activated cell sorting (FACS), where the most fluorescent clones were selected and used for additional sorting rounds. The final/sorted bacterial population was enriched in fluorescent bacterial clones containing the cyclic peptides responsible for fluorescence enhancement.

In order to screen the combined SICLOPPS library, *E. coli* BL21(DE3) cells that expressed the combined cyclic peptide library and A β_{42} -EGFP cells were screened with FACS. The initial cell population (initial library) exhibited a mean fluorescence of ~1200 a.u. (FITC-H: fluorescein isothiocyanate-height). Selection of the top 1-3% of the most fluorescent bacterial clones was performed for two consecutive sorting rounds, after which the mean bacterial fluorescence had been increased by ~2.5-fold (**Figure 2.10**). The resulting sorted population pool was subsequently used for individual clone isolation and cyclic peptide identification and evaluation.

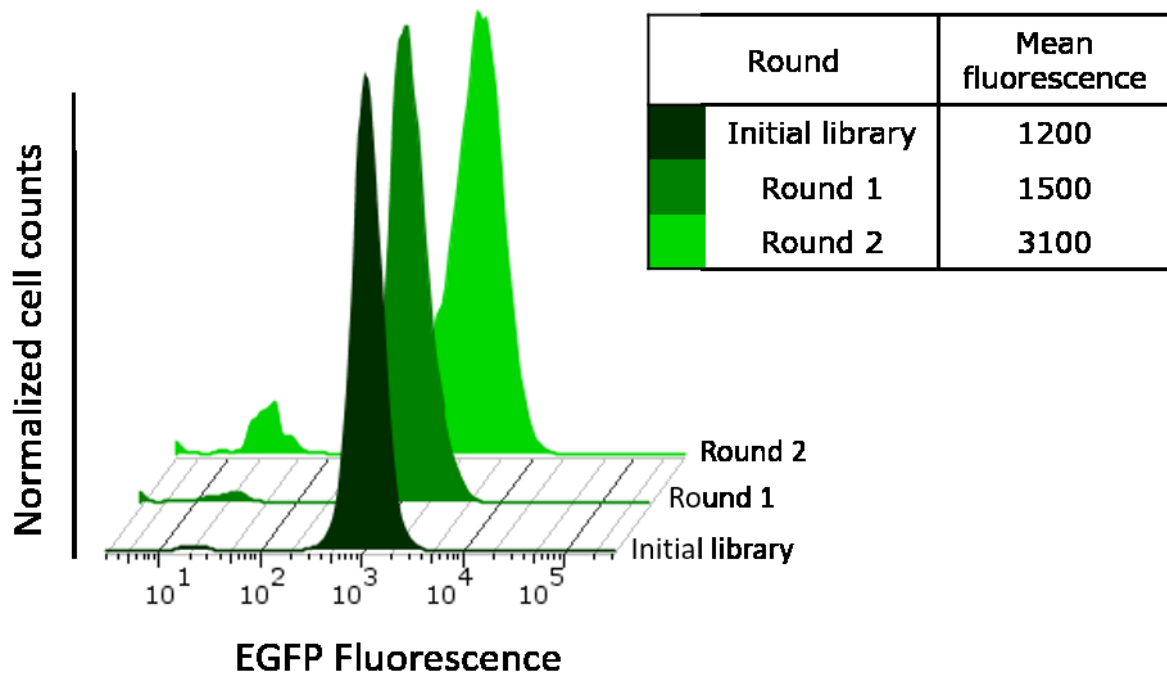


Figure 2.10. Fluorescence-activated cell sorting (FACS). Sorting of the initial library (grey) resulted in a population of cells with marginally increased fluorescence (dark green, Round 1). This population was subjected to a second round of sorting which resulted in a cell population with a $\sim 2.5 \times$ Mean fluorescence increase, compared to the initial library (green, Round 2). In both rounds, the instrument was set to sort 10,000 cells among the top 1-3% of the fluorescent population.

2.2.2. Isolation and evaluation of individual cyclic peptide-producing clones from the sorted pool

Ten bacterial clones were randomly selected from the sorted bacterial pool, as described in **Methods and Materials**, and upon plasmid vector isolation, they were Sanger-sequenced with intein-specific primers in order to identify the peptide-encoding DNA sequence (**Table 2.2**).

Table 2.2. DNA and amino acid sequences of peptides isolated from FACS Round 2.

Clone	Peptide-encoding region	Backbone Length	Peptide Sequence
1	ACC ACC GTG GAC CGG	5	TTVDR
2	ACC ACG TAC GCC AGG	5	TTYAR
3	ACC ACC ACG GCC CGG	5	TTTAR
4	ACC CCG GTC TGG TTC GAC	6	TPVWFD
5	ACC CCG GTC TGG TTC GAC	6	TPVWFD
6	ACC ACG TAC GCC AGG	5	TTYAR
7	AGC GCC TCG CCG ACG	5	SASPT
8	ACC GCG TGG TGC CGC	5	TAWCR
9	ACC ACC TGG TGC CGG	5	TTWCR
10	ACC GCG TTC GAC CGG	5	TAFDR

Subsequently, *E. coli* BL21(DE3) cells were transformed with each isolated cyclic peptide-encoding plasmid along with pETA β_{42} -EGFP, in order to examine the observed fluorescence phenotype in fresh bacteria. Following protein overexpression, fluorescence was observed in equal-cell-number samples (**Figure 2.11**). Cell fluorescence was enhanced in the presence of all ten isolated peptides, verifying the individual cyclic peptide effect on bacterial A β_{42} -EGFP fluorescence. Clones termed Random 1 & 2 were isolated from the “unsorted”, initial cyclic peptide library, and had no effect on the bacterial A β_{42} -EGFP fluorescence. These were used as negative controls in the following assays.

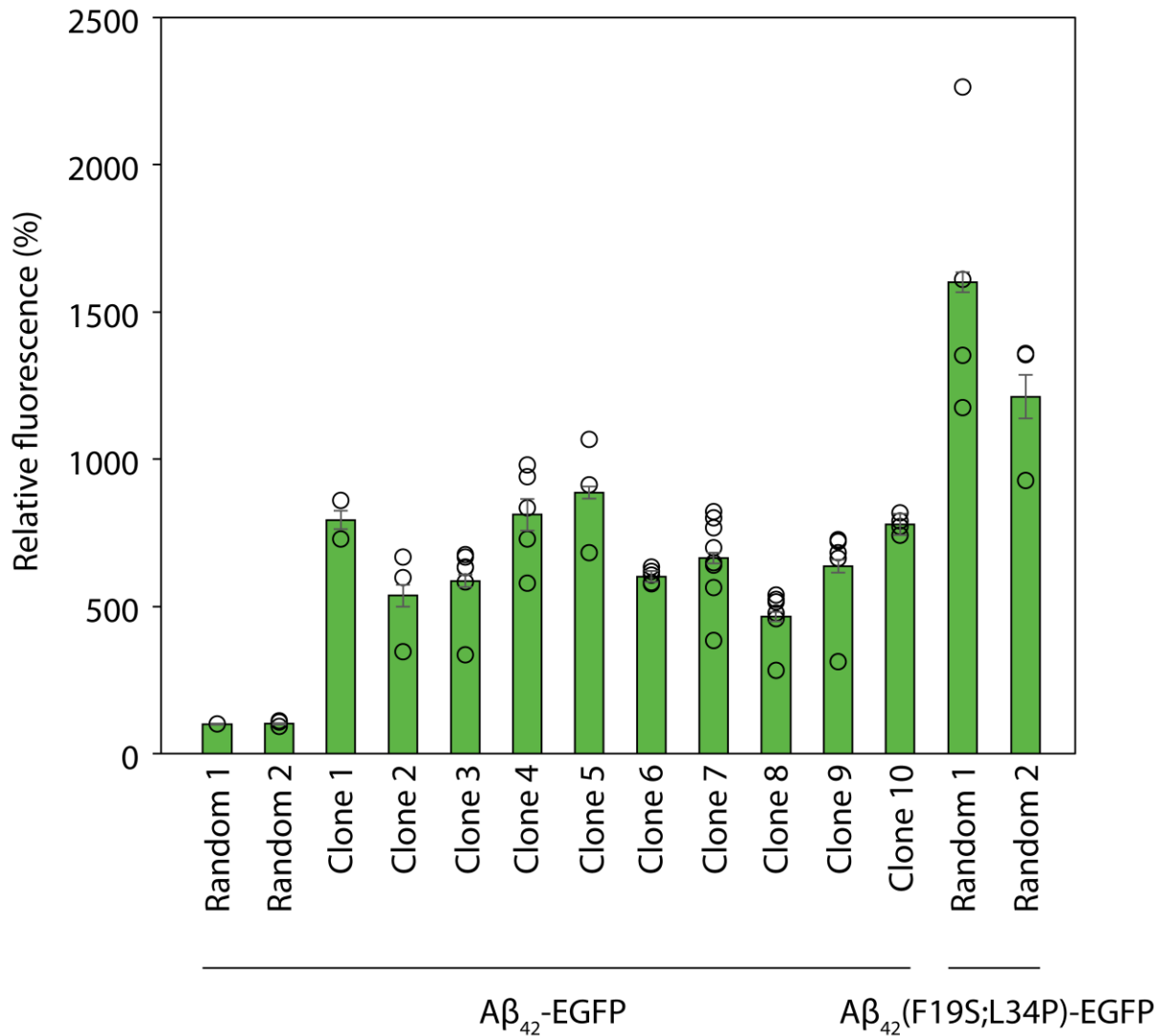


Figure 2.11. Selected peptides individually enhance bacterial fluorescence. Fluorescence of *E. coli* BL21(DE3) cells co-expressing $A\beta_{42}$ -EGFP along with the peptides Clones 1-10. Cellular fluorescence was compared to the $A\beta_{42}$ -EGFP/Random 1-producing sample, which was arbitrarily set at 100%. Mean fluorescence \pm sem shown, $n \geq 3$ for all samples except Clone 1 where $n=2$

In addition, anti-CBD western blots on lysates of cells producing $A\beta_{42}$ -EGFP along with any of the ten isolated peptide clones (Clones 1-10) revealed the presence of a lower band at ~ 19 kDa. All the clones showed the ability to produce the tetra-partite splicing precursor as well as the ability to undergo intein splicing, as evidenced by the presence of band corresponding to the I_N -CBD product of splicing. (Figure 2.12).

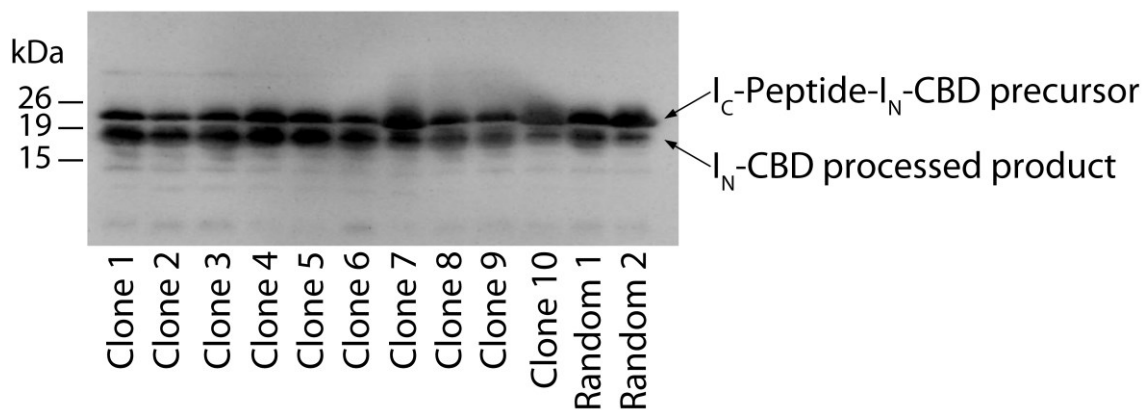


Figure 2.12. The intein tetra-partite precursor undergoes splicing. Western blot of individual isolated Clones 1-10 and two random peptide clones. Two main bands were visible: The precursor intein fusion (I_C -Peptide- I_N -CBD), and a ~19-kDa band that corresponded to I_N -CBD.

The significance of cyclization on peptide bioactivity was investigated for each of the ten cyclic peptide, using non-splicing variants of the isolated peptide-encoding plasmid vectors. The His24 and Phe26 residues of the C-intein domain (I_C) of the Ssp-DnaE intein, have been shown to promote asparagine cyclization^{317, 318}, which is the final step before the release of a cyclic peptide (**Figure 1.13b**)²⁶². Based on this, an H24L;F26A mutant of the I_C domain eliminates intein splicing^{319, 320}. These plasmid constructs were then used in a bacterial $A\beta_{42}$ -EGFP fluorescence assay, leading to baseline levels of fluorescence for all the tested clones, showing that intein processing and peptide cyclization was essential to the observed bacterial $A\beta_{42}$ -EGFP fluorescence.

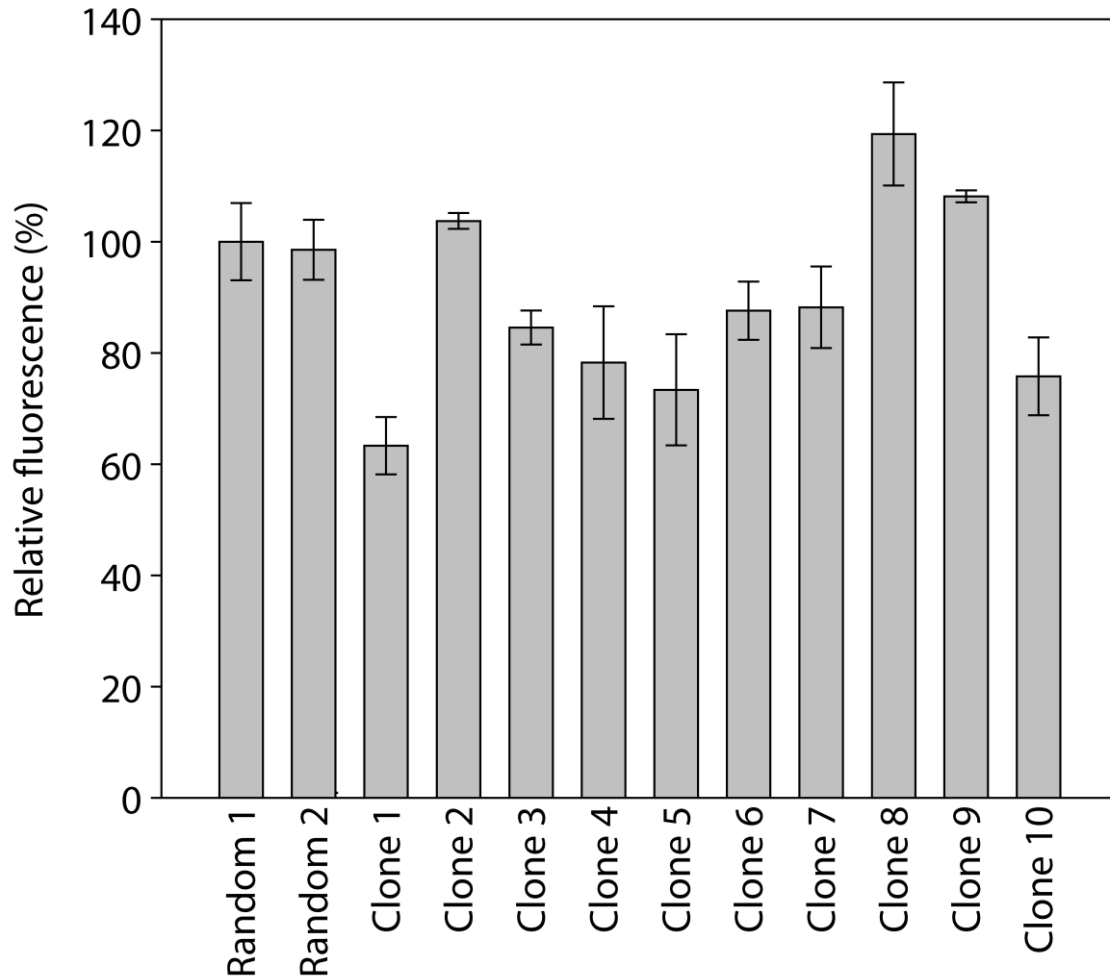


Figure 2.13. Peptide cyclization is necessary for the production of bacterial fluorescence. Fluorescence of *E. coli* BL21(DE3) cells co-expressing A β ₄₂-EGFP along with the non-splicing H24L;F26A peptide Clones 1-10. The two random peptide-producing clones were also carrying the H24L;F26A mutations. Cellular fluorescence was compared to the A β ₄₂-EGFP/Random 1 sample, which was arbitrarily set at 100%. Mean fluorescence \pm sd shown, $n=1$ (triplicate cultures) for all samples.

2.2.3. Individual peptide clones increase the soluble fraction of A β ₄₂-EGFP

The ten isolated peptide clones were also evaluated for their ability to increase the solubility of the A β ₄₂-EGFP fusion. Co-expression of the isolated cyclic peptides (clones 1-10) with the A β ₄₂-EGFP fusion showed that a relatively increased portion of A β ₄₂-EGFP was found in a soluble state (**Figure 2.14**). Total protein levels were similar for samples producing individual selected peptides (Clones 1-10) or

random-control peptides, with the exception of samples that produced A β_{42} (F19S;L34P)-EGFP. These differences in total protein production could be attributed to the increased fitness-cost associated with the more amyloidogenic A β_{42} isoform³¹¹.

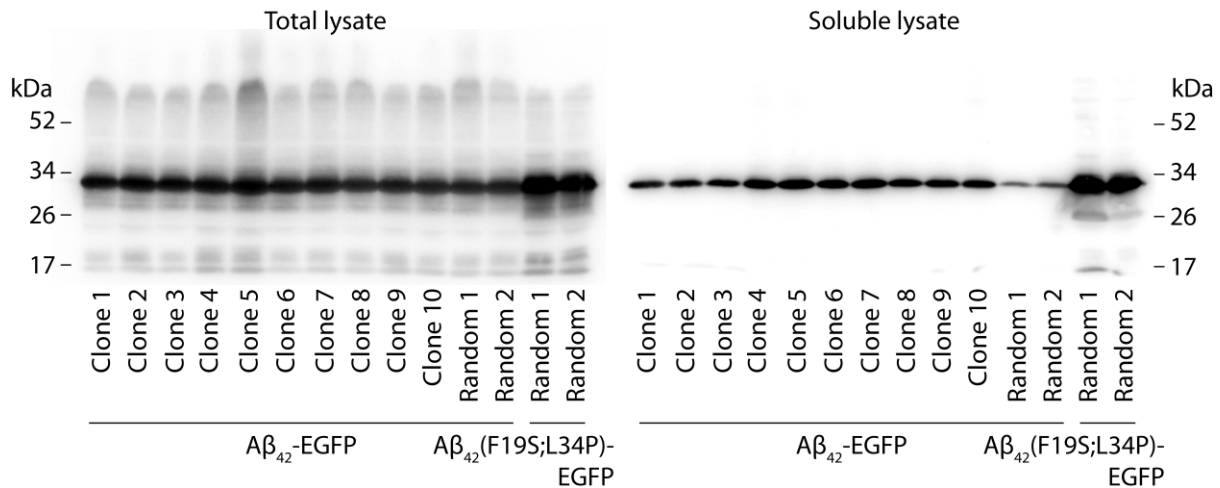


Figure 2.14. Isolated cyclic peptides enhance A β_{42} -EGFP solubility. Western blots of total and soluble cell lysates of *E. coli* BL21(DE3) cells co-expressing A β_{42} -EGFP along with the peptide Clones 1-10. Co-expression of A β_{42} (F19S;L34P)-EGFP with two randomly selected peptides was used as positive control and A β_{42} -EGFP co-expression with the two random peptides served as negative control. Samples were stained with the anti-A β (6E10) antibody. Total cell lysates (left): The main band at ~31 kDa corresponds to A β_{42} -EGFP or A β_{42} (F19S;L34P)-EGFP. Soluble protein fraction (right): Peptide clones 1-10 show enhanced solubility of the A β_{42} -EGFP fusion, compared to the random peptide-encoding clones.

Native-PAGE followed by western blots led to similar findings regarding the increase of the A β_{42} -EGFP soluble fraction. A characteristic difference between native and denaturing methods was that A β_{42} -EGFP was unable to enter the gel during native-PAGE under normal conditions, as seen in the control samples Random 1 & 2 (**Figure 2.15**). In this case, production of the A β_{42} -EGFP was verified by denaturing western blotting, performed in parallel native western blotting (**Figure 2.14**). The native blots revealed that the isolated peptide clones 1-10 increased the levels of soluble A β_{42} -EGFP again, allowing it to enter the gel and subsequently become detected (**Figure 2.15**). While it is difficult to calculate the degree of oligomerization of A β_{42} -EGFP in native blots, comparison of the main (lowest) band migration for A β_{42} -EGFP –producing samples with A β_{42} (F19S;L34P)-EGFP –producing samples

suggested that the main band was possibly a monomer or a low molecular weight oligomer, considering the severely reduced aggregation propensity of $A\beta_{42}(F19S;L34P)$ -EGFP²⁷⁶.

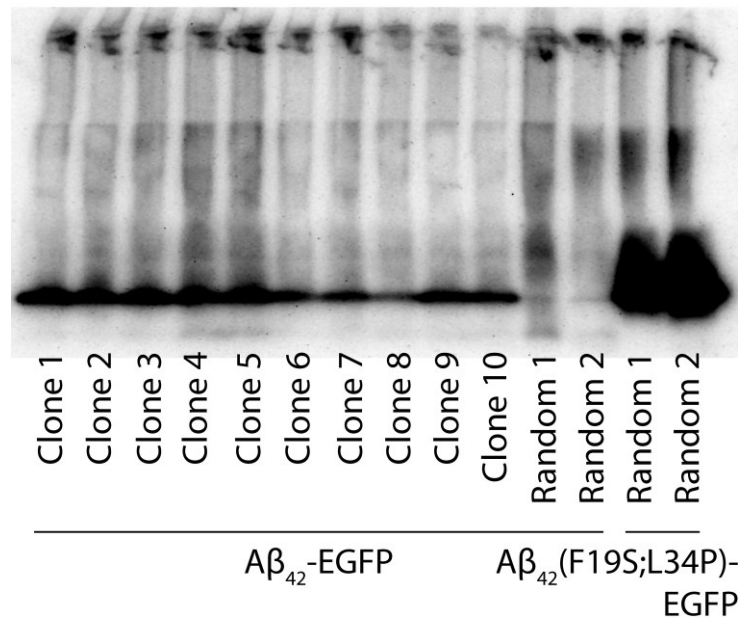


Figure 2.15. Isolated cyclic peptides enhance $A\beta_{42}$ -EGFP solubility. Native-PAGE followed by western blots of total cell lysates of *E. coli* BL21(DE3) cells co-expressing $A\beta_{42}$ -EGFP along with the peptide Clones 1 to 10. Co-expression of $A\beta_{42}(F19S;L34P)$ -EGFP with two randomly selected peptides was used as positive control and $A\beta_{42}$ -EGFP co-expression with the two random peptides served as negative control. Samples were stained with the anti- $A\beta$ (6E10) antibody.

Native in-gel fluorescence assays were also used, to provide a link between the phenotypes of protein solubility and cell fluorescence. For peptide clones 1-10, fluorescence could be attributed almost exclusively to the soluble (S) protein fraction (**Figure 2.16**). Moreover, in-gel fluorescence was almost absent for random control-peptide clones 1 & 2, in agreement with fluorescence measurements for liquid cultures producing $A\beta_{42}$ -EGFP/Random 1 & 2 (**Figure 2.11**). Likewise, fluorescence was mostly absent in samples containing the insoluble protein fraction.

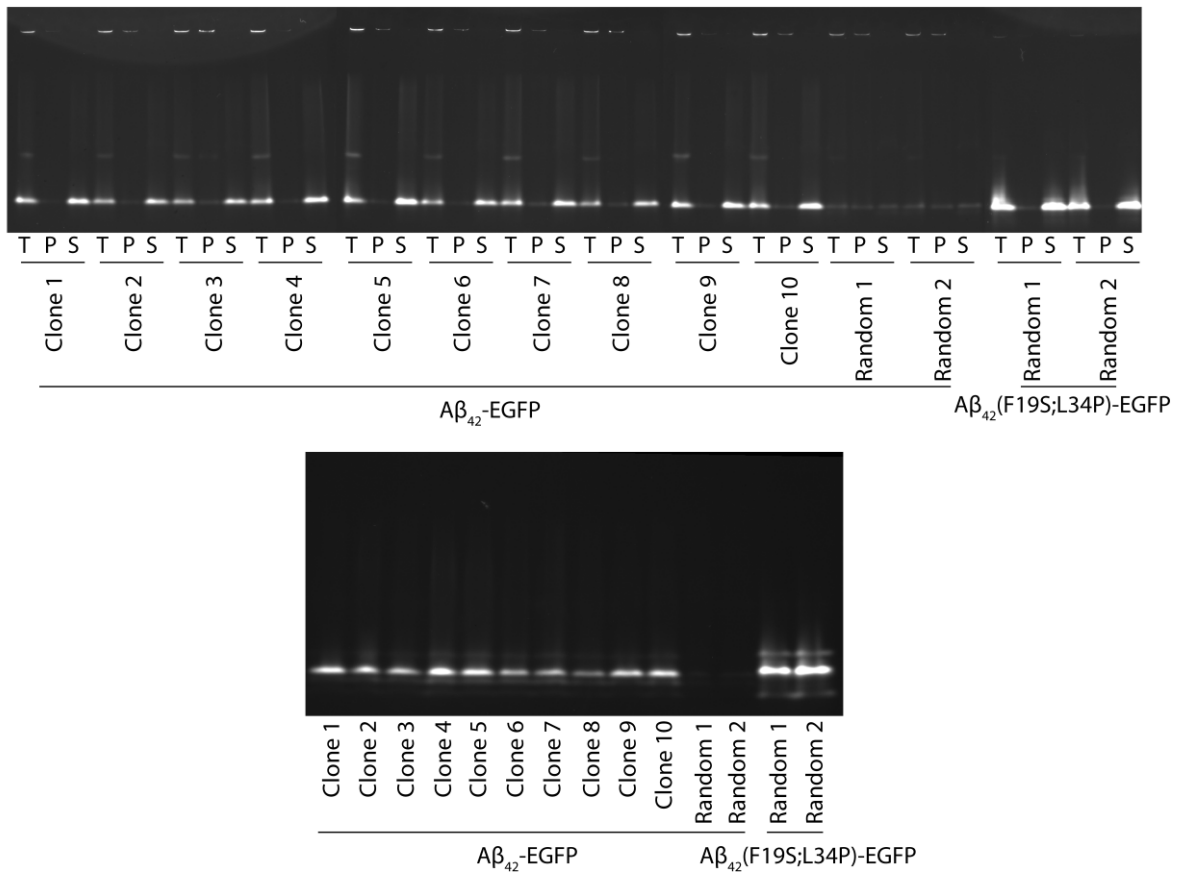


Figure 2.16. Isolated cyclic peptides enhance $A\beta_{42}$ -EGFP solubility. In-gel fluorescence of cell lysates of *E. coli* BL21(DE3) cells co-expressing $A\beta_{42}$ -EGFP along with the peptide Clones 1 to 10. Co-expression of $A\beta_{42}$ (F19S;L34P)-EGFP with two randomly selected peptides was used as positive control and $A\beta_{42}$ -EGFP co-expression with the two random peptides served as negative control. Samples were stained with the anti- $A\beta$ (6E10) antibody. Total cell lysates (T), insoluble protein fractions (pellets, P) and soluble protein fractions (S) are presented. Lower gel shows only total cell lysates for comparison of total fluorescence.

Overall, studies of the cyclic peptide effect on $A\beta_{42}$ -EGFP solubility enabled the observation of a second peptide-associated phenotype, fusion solubility, which developed in parallel with the bacterial fluorescence phenotype. Since accumulation into IBs does not always impose a loss of fluorescence³²¹, it was necessary to study both phenotypes. Moreover, it is possible that the cyclic peptides interact with $A\beta_{42}$ early in its oligomerization process, in order for $A\beta_{42}$ -EGFP to remain in the soluble protein fraction and thus produce fluorescence, which lies in agreement with the screening assay's premise for the production of fluorescence²⁶⁰.

2.3. Selection of two cyclic peptides for further evaluation

Two peptide sequences among the isolated cyclic peptide clones were chosen for evaluation in *in vitro* and *in vivo* studies, as well as structure activity analyses. Specifically, clones 7 & 10, corresponding to peptides cyclo-SASPT and cyclo-TAFDR respectively, were produced in mg quantities via solid-phase synthesis. Both chosen peptides were pentapeptides, representing the majority of sorted peptide clones, as evident by the individual peptide clone studies. Cyclo-TAFDR was chosen as representative of the dominant cyclo-TXXXXR pentapeptide motif, and cyclo-SASPT was chosen because of its unique sequence among individual selected peptides (**Figure 2.17**).

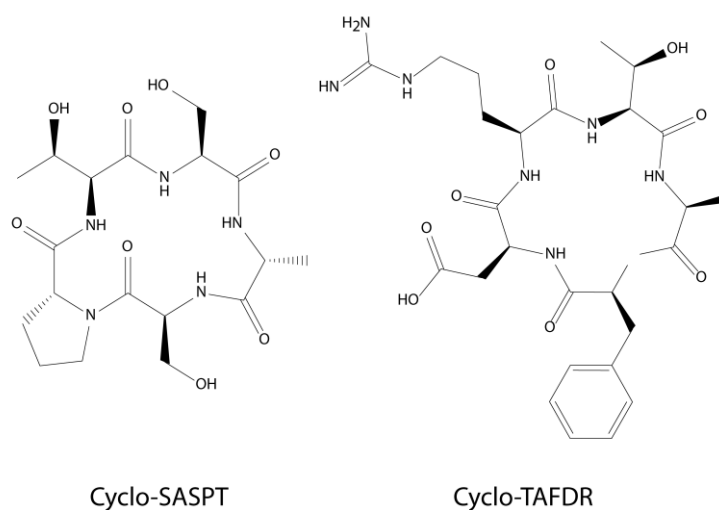


Figure 2.17. Cyclo-SASPT and cyclo-TAFDR. The two peptides were chosen for further evaluation

2.3.1. The cyclo-SASPT and cyclo-TAFDR effect is A β -specific

Enhancement of bacterial A β_{42} -EGFP fluorescence in the presence of the cyclo-SASPT (Clone 7) and cyclo-TAFDR (Clone10) peptides (**Table 2.2**) was shown to be A β -specific. Neither cyclo-TAFDR, nor the unique cyclo-SASPT peptide enhanced bacterial fluorescence when the A β_{42} moiety in the

A β_{42} -EGFP fusion was replaced with either the DNA-binding (core) domain of the human p53 containing a Tyr220Cys substitution (p53C(Y220C))¹⁴, or a mutant of human Cu/Zn superoxide dismutase 1 (SOD1(A4V))³²². Both alternative target-proteins are known to misfold, and have been linked to various forms of cancer and familial amyotrophic lateral sclerosis (ALS), respectively. Furthermore, expression of their fusions with EGFP in *E. coli* causes a similar bacterial phenotype as A β_{42} -EGFP expression. Overexpression of the isolated cyclic peptide clones along with either p53C(Y220C)-EGFP or SOD1(A4V)-EGFP revealed that the cyclic peptides inhibited A β_{42} aggregate formation specifically, and did not exhibit a universal inhibitory activity against the aggregation of misfolding proteins expressed in *E. coli*.

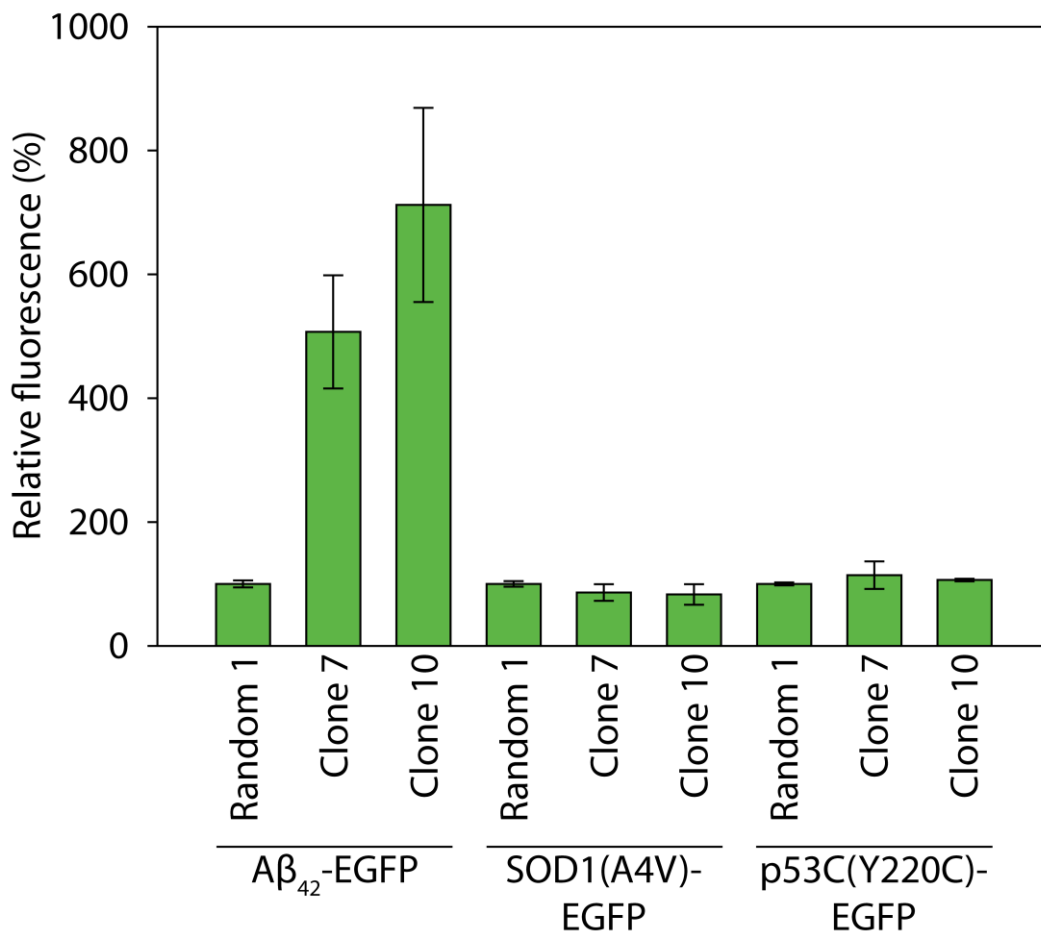


Figure 2.18. Cyclo-SASPT and cyclo-TAFDR are A β -specific. Fluorescence of *E. coli* BL21(DE3) cells co-expressing A β_{42} -EGFP, SOD1(A4V)-EGFP or p53C(Y220C)-EGFP along with peptide Clone 7 or 10. The Random 1-producing clone was arbitrarily set as the fluorescence baseline at 100% in all three occasions. Mean fluorescence \pm sd shown, $n=1$ (triplicate cultures) for all samples

2.3.2. The cyclo-SASPT and cyclo-TAFDR effect on A β isoforms/variants of A β

Peptide specificity towards A β was not limited to the A β_{42} isoform/variant. Cyclo-SASPT (Clone 7) and cyclo-TAFDR (Clone 10) also enhanced the fluorescence of A β_{40} -EGFP and A β_{42} (E22G)-EGFP (A β_{42} “arctic” mutant³²³). Expression of the three fusions in *E. coli* showed a small fluorescence increase for A β_{40} -EGFP, which was expected due to the smaller aggregation propensity of A β_{40} (Figure 2.19, left). Likewise, the differences in the fluorescence of A β -EGFP fusions in the presence of peptides could possibly be explained by varying aggregation kinetics between the isoforms/variants, however, the positive effects of the peptides were apparent in all three cases (Figure 2.19, right).

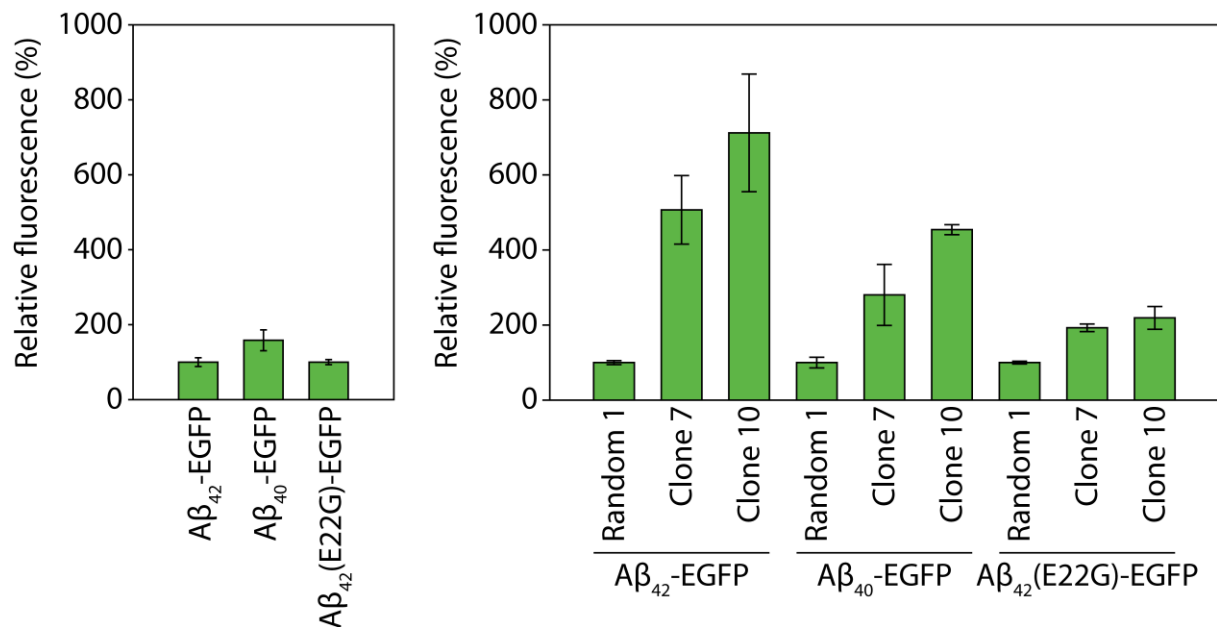


Figure 2.19. Cyclo-SASPT and cyclo-TAFDR effect on bacterial fluorescence of A β -EGFP variants. Left: Background bacterial fluorescence of cells producing A β_{42} -EGFP, A β_{40} -EGFP or A β_{42} (E22G)-EGFP fusions. Fluorescence was compared against that of the A β_{42} -EGFP sample, which was arbitrarily set at 100%. Right: Fluorescence of *E. coli* BL21(DE3) cells co-expressing A β_{42} -EGFP, A β_{42} -EGFP or A β_{42} (E22G)-EGFP along with peptide Clones 7 and 10. Fluorescence was compared against that of the A β_{42} -EGFP & Random 1-producing clone, which was arbitrarily set at 100%. Mean fluorescence \pm sd shown, $n=1$ (triplicate cultures) for all samples

2.4. Remarks

This presented bacterial system was effective in sorting a large biosynthetic library of cyclic peptides and yielded a sorted population of cyclic peptides, which upon initial examination showed preference for pentapeptides. The isolated cyclic peptides that were subsequently investigated were shown to be specific to A β , while their bioactivity was dependent on peptide cyclization. Notably, all investigated peptides exerted a dual effect on the bacterially produced A β ₄₂-EGFP fusion, by enhancing both its fluorescence and solubility. This dual effect could possibly be traced to an early interaction between A β ₄₂ and cyclic peptide, which stabilizes monomeric or low molecular weight oligomers of the fusion, allowing EGFP to fold and remain in a soluble state.

Following this investigation, it was necessary to evaluate whether the bioactivity of hits reached beyond the setting of the screening system. Thus, the following chapter provides a description of the biochemical, biophysical and biological assays in support of the bacterial system's ability to discover hits that not only interfere with the aggregation of A β , but also display the desirable effect of suppressing the cytotoxicity related to A β aggregation.

Chapter 3. Effect of the selected peptides on A β aggregation and neurotoxicity

Cyclo-SASPT and cyclo-TAFDR were chosen from the sorted library of cyclic peptides to be synthesized in mg quantities by solid-phase synthesis in order to study their effect on A β aggregation and toxicity, as mentioned previously. Cyclo-SASPT and cyclo-TAFDR will be henceforth referred to as A β C5-34 and A β C5-116, respectively (A β -targeting cyclic 5-peptide number 34 and 116). A more detailed explanation of the peptide nomenclature will be presented in the next chapter. Along with the two peptides selected against A β , a SOD1(A4V)-targeting synthetic cyclic pentapeptide was used as control. The sequence of this control peptide is cyclo-TWSVW and is referred to as SOD1C5-4³²⁴.

3.1. *In vitro* evaluation of the effect of the selected peptides on A β aggregation

3.1.1. Evaluation of the effect of the selected peptides on A β aggregation using circular dichroism spectroscopy and thioflavin T staining

Circular dichroism (CD) spectroscopy was used to assess the effect of the selected pentapeptides on the aggregation process of A β_{40} and A β_{42} , since it allowed monitoring of the formation of β -sheets in a protein sample aggregating over time. The hypothesis was that synthetic A β exposed to synthetic cyclic peptides would display a different progression of β -sheet formation, compared to peptide-free A β . Indeed, in the presence of equimolar concentrations of A β C5-116 (50 μ M), samples containing A β_{40} showed inhibition of the aggregation of the amyloid peptide forcing it to remain in a random coil conformation for the duration of the experiment. On the contrary, equimolar concentrations of A β C5-34 (50 μ M) resulted in accelerated β -sheet formation, relative to the “no peptide” A β_{40} sample. The same 30 d-matured samples were finally stained with thioflavin-T (ThT), in an assay that measured amyloid fibril formation³²⁵. In the case of A β C5-116, A β_{40} fibril formation was reduced. However, in

agreement with CD results, A β C5-34 failed to reduce the formation of amyloid-like fibrils (**Figure 3.1a, b**).

For samples containing synthetic A β_{42} , both A β C5-34 and A β C5-116 were shown to alter the normal aggregation pathway of the amyloid peptide. In the presence of equimolar concentrations of either cyclic peptide, unchanging negative peaks appeared at \sim 218 nm, suggesting stabilization of a β -sheet structure. However, ThT staining of the 30 d-matured samples revealed that typical amyloid fibril formation was reduced in the presence of either cyclic peptide. Interestingly, this finding suggested that the β -sheet formations of A β_{42} that were created in the presence of either cyclic peptides were less amyloid-like (**Figure 3.1a, b**).

Similar results were observed when A β C5-34 and A β C5-116 were added at higher concentrations in the samples (100 μ M). Negative peaks at \sim 218 nm were more pronounced in this case, thus suggesting a dose-dependent effect, while ThT staining showed that amyloid-like fibril formation was avoided altogether (**Figure 3.1c**). The addition of the SOD1-targeting cyclic pentapeptide SOD1C5-4 (**Figure 3.1a, b**) did not affect A β_{40} or A β_{42} aggregation in any way.

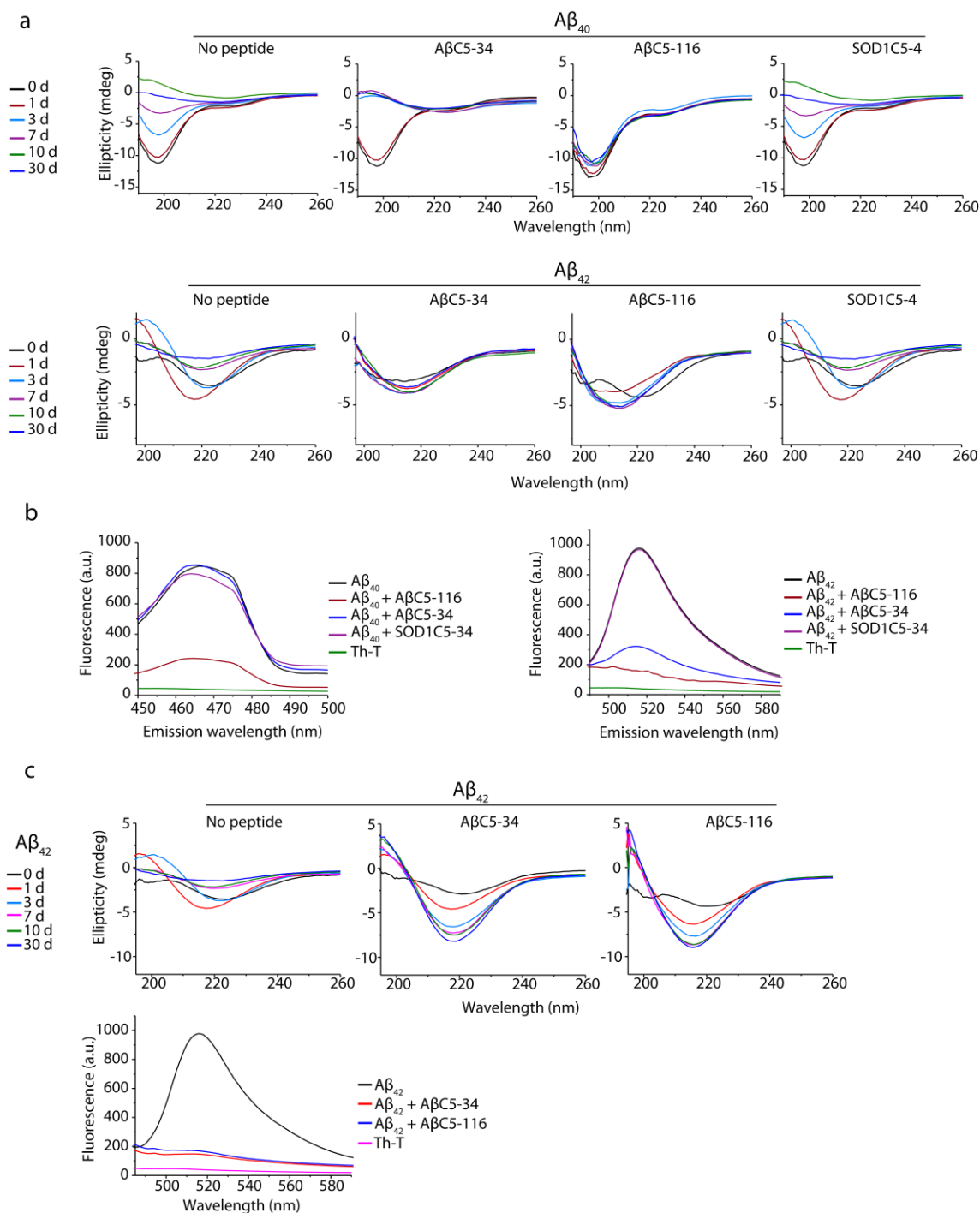


Figure 3.1. AβC5-34 and AβC5-116 interfere with the aggregation of Aβ₄₀ and Aβ₄₂. **a**) CD spectra of 50 μM Aβ₄₂ in phosphate buffer (10 mM, pH 7.33), in the presence of 50 μM of either AβC5-34 or AβC5-116. Spectra were collected for a period of 30 days, at 33 °C. **b**) Thioflavin-T (ThT) staining of the 30 d-matured samples used in (a). **c**) CD spectra of 50 μM Aβ₄₂ in phosphate buffer (10 mM, pH 7.33), in the presence of 100 μM of either AβC5-34 or AβC5-116. Spectra were collected for a period of 30 days at 33 °C. CD experiments and ThT staining were conducted in the laboratory of Dr. Maria Pelecanou at NCSR Demokritos.

3.1.2. Evaluation of the effect of the selected peptides on A β aggregation using transmission electron microscopy

Samples containing A β_{42} (50 μ M) were matured as in the CD experiments in the presence or absence of A β C5-34 or A β C5-116 (100 μ M), and were used to obtain the transmission electron microscopy (TEM) images shown in **Figure 3.2**. A β_{42} incubated without peptides exhibited the typical dense network of intertwined fibrils^{326, 327}. On the contrary, presence of either A β C5-34 or A β C5-116 caused the formation of considerably fewer, shorter and ill-developed fibrils.

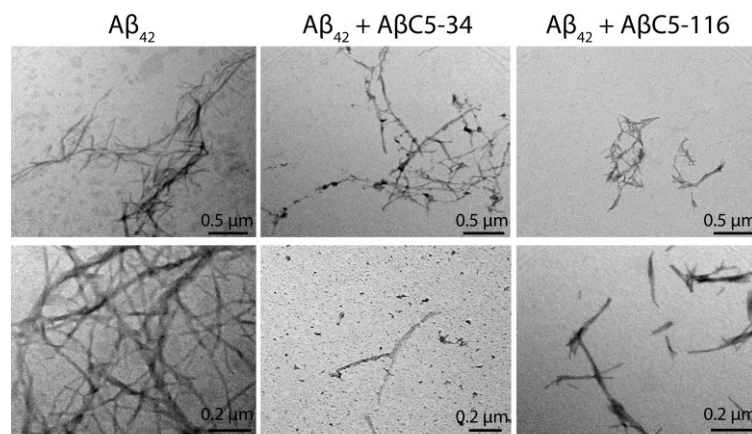
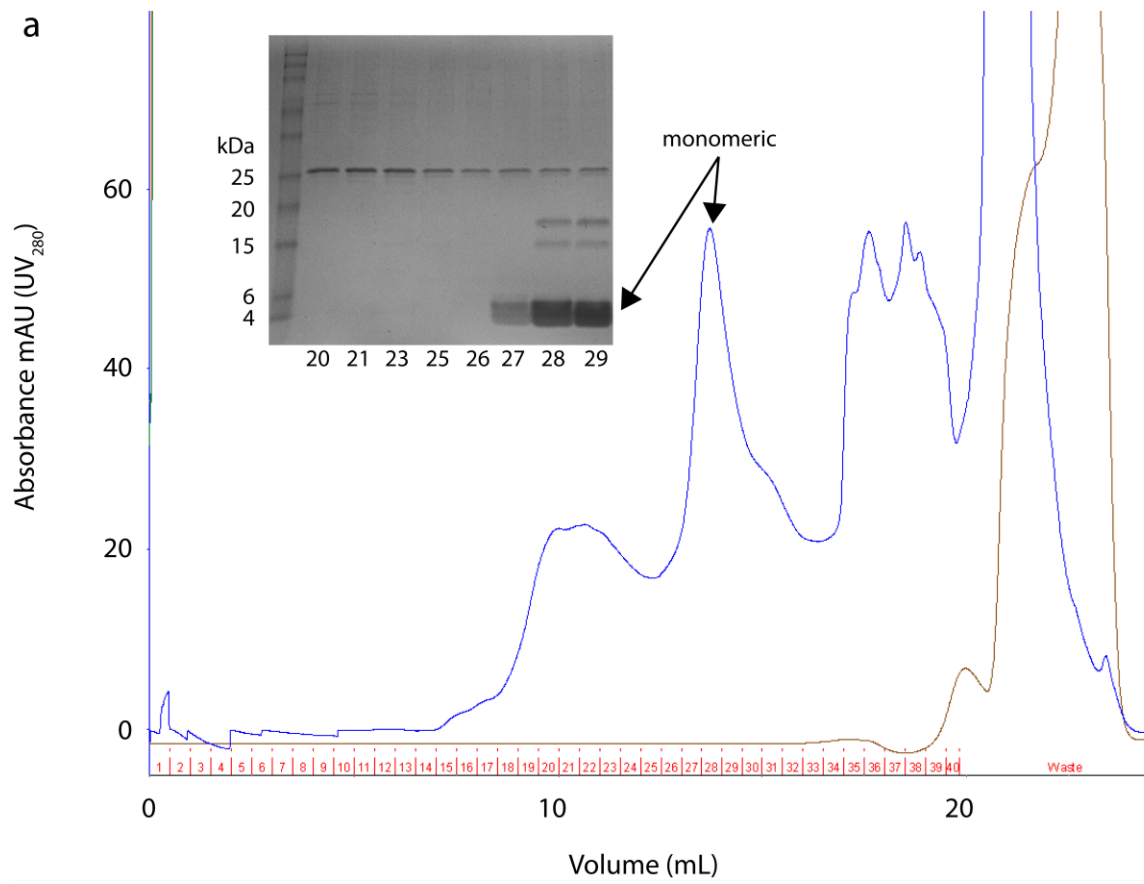


Figure 3.2. Evaluation of the effect of the selected peptides on A β_{42} aggregation by transmission electron microscopy. TEM images of the matured 50 μ M A β_{42} samples described in Figure 3.1, both in the presence and absence of A β C5-34 and A β C5-116 (100 μ M). Images are representative from the two independent experiments that were conducted ($n=2$). TEM analyses were carried out in the laboratory of Dr. Nikos Boukos at NCSR Demokritos.

3.1.3. Evaluation of the effect of the selected peptides on A β aggregation using dynamic light scattering.

For the evaluation of the effect of the selected peptides on A β aggregation using dynamic light scattering, purified A β_{42} was used. A β_{42} was produced recombinantly in *E. coli* cells by adapting a previously described protocol³²⁸. Recombinant Met1-A β_{42} was purified by size-exclusion

chromatography (SEC) (**Figure 3.3a**) and monomeric/oligomeric species were immediately used to record dynamic light scattering spectra (DLS) in the presence or absence of synthetic A β C5-35 or A β C5-116 (**Materials and Methods**). All samples showed high polydispersity, with distributions comprised of numerous peaks. This method was used for a strictly qualitative assessment of the peptide effect on A β ₄₂ aggregation, since even a monodisperse population of non-globular formations such as fibrils would yield a complicated distribution of peaks in DLS³²⁹. As a result, the precise nature and size of the species that constituted the observed distributions could not be determined. However, the initial spectra (0 d) indicated the presence of particles with a hydrodynamic radius of 2-5 nm, which migrated towards 10 nm as the sample maturation progressed (**Figure 3.3b**). Interestingly, the peptide-containing samples showed slower peak-migration towards the 10 nm hydrodynamic radius, compared to peptide-free Met1-A β ₄₂. This slower peak-migration became more evident at longer maturation times (30 days) and was more pronounced for the A β C5-34 (1:1) and A β C5-116 samples. Naturally, the progress of aggregation was evident in all samples, with the appearance of a complicated ensemble of particles with hydrodynamic radii larger than 10 nm. A peak at 1 nm for the A β C5-116 1:10 sample was shown to correspond to the cyclic peptide since this also appeared in A β -free, peptide-only samples (not shown). Overall, this assay suggested the formation of altered oligomers/aggregates in the presence of either one of the selected cyclic peptides, though additional assays would be necessary for the characterization of A β formations within the samples.



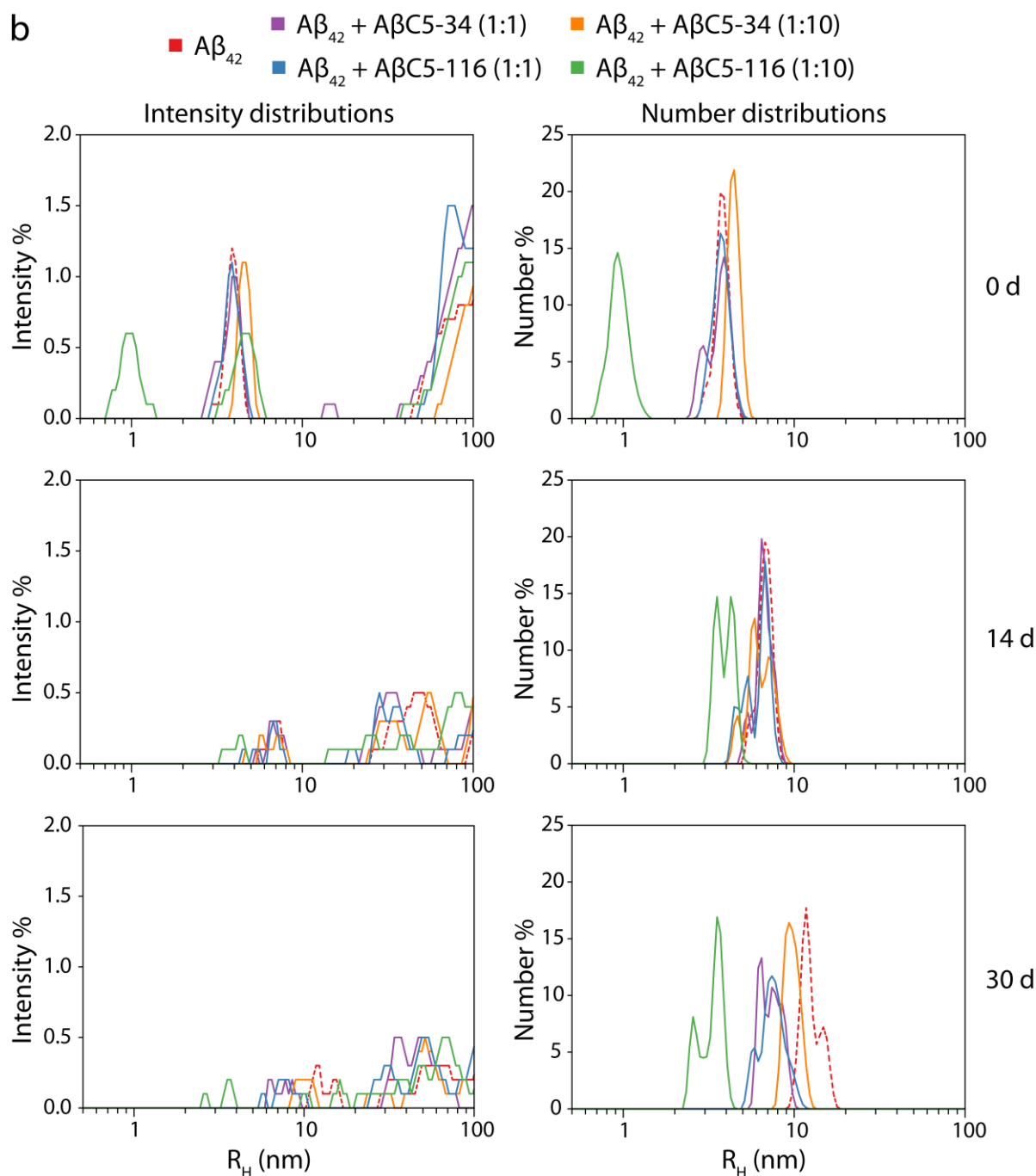


Figure 3.3. Purification and size distributions of $A\beta_{42}$ samples. a) Representative chromatogram for size exclusion chromatography of Met- $A\beta_{42}$. SDS-PAGE of elution fractions, in 15% polyacrylamide gel. Fractions 28 and 29 were used in all DLS assays. b) DLS spectra taken at the indicated time points. Distributions are presented by intensity (left) and by number of particles (right). $MA\beta_{42}$ concentration was measured by UV_{280} absorption, at 42.9 μ M. Samples were matured at 37 $^{\circ}$ C. DLS spectra were recorded as the average of six distinct runs per sample, each consisting of 10 sequential measurements of 30 sec.

Taken together, the presented biochemical/biophysical analyses on the effect of the A β C5-34 and A β C5-116 selected cyclic pentapeptides demonstrate that these interfere with the aggregation pathway of A β , causing the formation of atypical aggregates.

3.2. Evaluation of the effect of the selected peptides on A β -induced neurotoxicity *in vitro*

3.2.1. Evaluation of the effect of the selected peptides on A β -induced neurotoxicity in primary mouse hippocampal neurons

In order to study the effect of A β C5-34 and A β C5-116 on A β -induced toxicity, MTT assay conditions were first optimized by exposing primary mouse hippocampal neurons to A β ₄₀ and A β ₄₂. As a result, it was decided that neuron cultures would be exposed to three-day-matured synthetic A β ₄₀ and one-day-matured A β ₄₂ (**Supplementary Figure 2**). Consequently, an MTT assay on neuronal cultures that were exposed to A β ₄₀ or A β ₄₂ containing either A β C5-34 or A β C5-116, showed that the cyclic peptides inhibited the neurotoxicity of the amyloid peptide in a dose-responsive manner significantly (**Figure 3.4**).

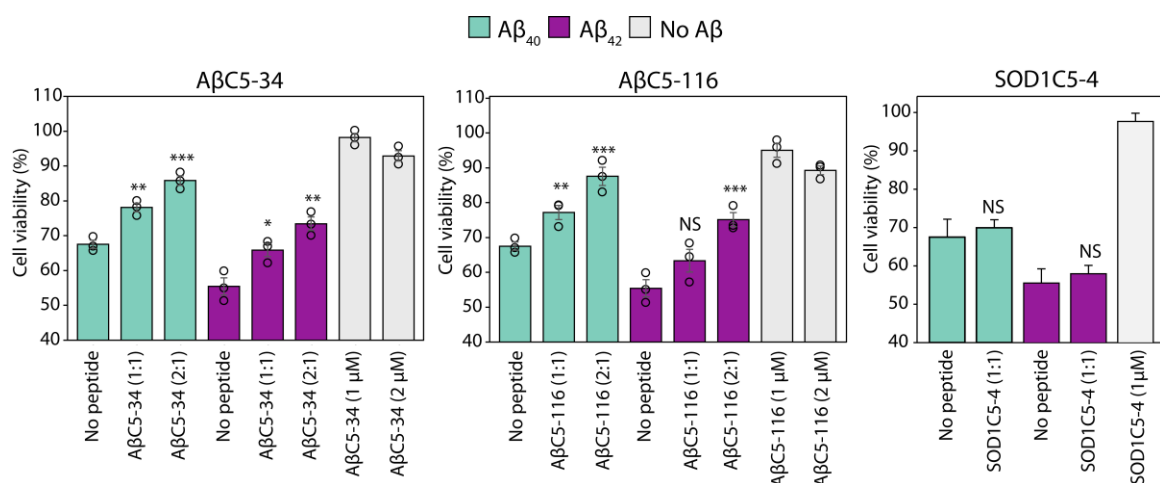


Figure 3.4. Cyclic peptides A β C5-34 and A β C5-116 inhibit A β -induced neurotoxicity *in vitro*. Effect of the selected cyclic pentapeptides A β C5-34 (left) and A β C5-116 (middle) on the cytotoxicity of A β ₄₀ or A β ₄₂ (1 μ M) in primary mouse hippocampal neurons as determined using the MTT assay. Cell

viability was also tested for A β samples pre-aggregated in the presence or absence of 1 μ M SOD1C5-4 for 3d (right). Results are expressed as the percentage of MTT reduction, assuming that the absorbance of control (untreated) cells was 100%. Mean values \pm s.e.m. of three independent experiments ($n=3$) with six replicate wells for each condition are reported. Statistical significances of the differences in the levels of viability between cells untreated and treated with A β or between cells treated with A β in the presence and absence of the selected cyclic peptides are presented. * $P \leq 0.05$, ** $P \leq 0.01$, *** $P \leq 0.001$; NS, not significant ($P > 0.05$). MTT assays were performed in the lab of Dr. Maria Pelecanou at the National Center for Scientific Research “Demokritos”.

Furthermore, phase contrast microscopy was used to assess the effect of A β C5-34 and A β C5-116 on the physiology of the primary neurons exposed to A β . Pre-aggregated A β caused reductions in the population of attached neurons while detached and rounded cells were observed floating in the supernatant. Indicators of neuron degeneration such as shrinkage, membrane blebbings, fragmented neurites and ill-developed axons were also observable in the preparations. On the contrary, presence of the A β C5-34 and A β C5-116 cyclic peptides alleviated A β -induced toxicity, as evident in the images of **Figure 3.5**.

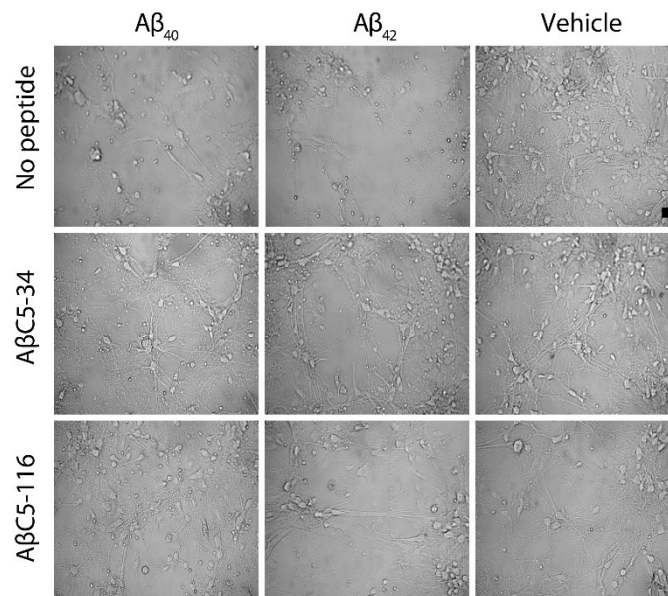


Figure 3.5. Phase-contrast microscopy images of primary mouse hippocampal neurons. Neuronal cultures were exposed to A β_{40} or A β_{42} (1 μ M), which had been pre-aggregated in the presence or absence of either A β C5-34 or A β C5-116 (2 μ M), at 37 $^{\circ}$ C for 24 h. Scale bar, 50 μ m. Representative images from $n=2$ independent experiments are presented. Experiments were performed in the lab of Dr. Maria Pelecanou at the National Center for Scientific Research “Demokritos”.

3.2.2. Evaluation of the effect of the selected peptides on A β -induced neurotoxicity in neuronal cell lines

In a similar manner, A β -induced toxicity on the glioblastoma cell line U87MG was diminished in the presence of either A β C5-34 or A β C5-116. The toxicity suppressing effects were noticeable for both A β ₄₀- and A β ₄₂-induced toxicity. The presence of the SOD1C5-4 cyclic peptide during pre-aggregation of the amyloid peptide did not affect cell viability, as was the case for the selected peptides in the absence of an A β isoform (**Figure 3.6**).

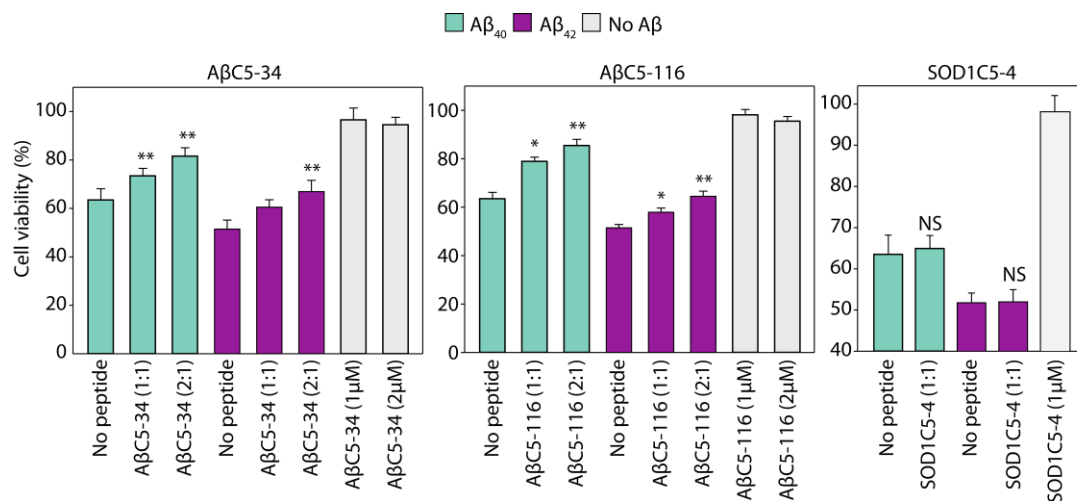


Figure 3.6. Cell viability as determined by the MTT assay of serum-starved U87MG cells. Cells were incubated without A β or with 1 μ M preparations of A β ₄₀ or A β ₄₂ for 24 h at 37 °C. Both A β preparations had been previously aggregated in the presence or absence of 1 and 2 μ M of either A β C5-34 or A β C5-116 respectively. All A β ₄₀ solutions were pre-aggregated for 3 d while A β ₄₂ solutions were pre-aggregated for 1 d. MTT stock solution in DMEM complete medium was added in each well to a final concentration of 1 mg/mL, prior to a 4 h incubation at 37 °C. *P \leq 0.05, **P \leq 0.01, NS, not significant (P > 0.05). MTT assays were performed in the lab of Dr. Maria Pelecanou at the National Center for Scientific Research “Demokritos”.

3.2.3. Evaluation of the effect of the selected peptides on the neurotoxicity caused by naturally secreted A β oligomers

So far, the cytotoxicity suppressive effects of A β C5-34 and A β C5-116 have been observed against the toxicity of synthetic A β aggregation. 7PA2 Chinese hamster ovary cells are known to secrete A β oligomers that inhibit hippocampal long-term potentiation³³⁰. Thus, primary mouse cortical neurons were exposed to A β_{42} oligomers secreted from 7PA2 cells. These are Chinese hamster ovary (CHO) cells, that have been transfected with an APP751 cDNA, which carries the V717F mutation. The medium of those cells contains monomeric, dimeric and trimeric A β species³³¹. Neuronal growth in the presence of A β C5-116 revealed a reduced oligomer neurotoxic effect while A β C5-34 and the control peptide SOD1C5-4 did not affect oligomer neurotoxicity (**Figure 3.7**).

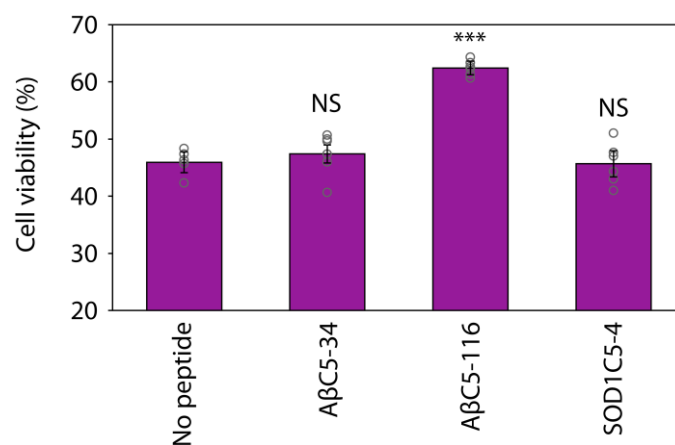


Figure 3.7. Cell viability of primary mouse cortical neurons exposed to A β oligomers derived from 7PA2 cells. Cultures were grown in the presence or absence of either A β C5-34, A β C5-116 or SOD1C5-4 (10 μ M). Mean values \pm s.e.m. are presented ($n=6$ independent experiments, each one performed in three replicates). *** $P \leq 0.001$; NS, not significant ($P > 0.05$). The viability assay was performed in the lab of Dr. Kostas Vekrellis at Biomedical Research Foundation of the Academy of Athens.

3.2.4. Evaluation of the effect of the selected peptides on the binding of A β ₄₂ aggregates to the neuronal surface

Extracellular A β ₄₂ cytotoxicity is thought to originate from the binding of aggregated A β ₄₂ to the neuronal membrane surface and the subsequent uptake of these aggregated A β structures³³². In order to evaluate the effects of A β C5-34 and A β C5-116 on the membrane-binding ability of A β species, cortical neurons were exposed briefly to pre-aggregated A β preparations and binding to the neuronal surface was assessed by immunocytochemistry. The A β ₄₀ isoform was chosen for this assay as evidence supports its ability for physicochemical interactions with synaptic plasma membranes in both soluble and aggregated forms³³³. Moreover, A β ₄₀ is less toxic to neurons, which would aid in maintaining an intact neuronal network, essential to this assay. Indeed, following a brief neuronal exposure to A β ₄₀ and staining with the anti-amyloid precursor protein antibody R1(57), which does not interact with A β , it was evident that neurite formation was not significantly affected and the neuronal network remained intact. Staining with the anti-A β 6E10 antibody and comparison of the images to those for R1(57) staining, revealed the presence of large, 6E10-reactive, patch-like assemblies of aggregated A β , which were attached to the neuron surface, in accordance with previous findings^{157, 161}. Presence of either A β C5-34 or A β C5-116 to the A β ₄₀ aggregation reaction, however, resulted in decreased 6E10-reactive, patch-like staining. Therefore, results suggested that the selected peptides inhibit the formation of aggregated A β species capable of attaching to the neuronal surface.

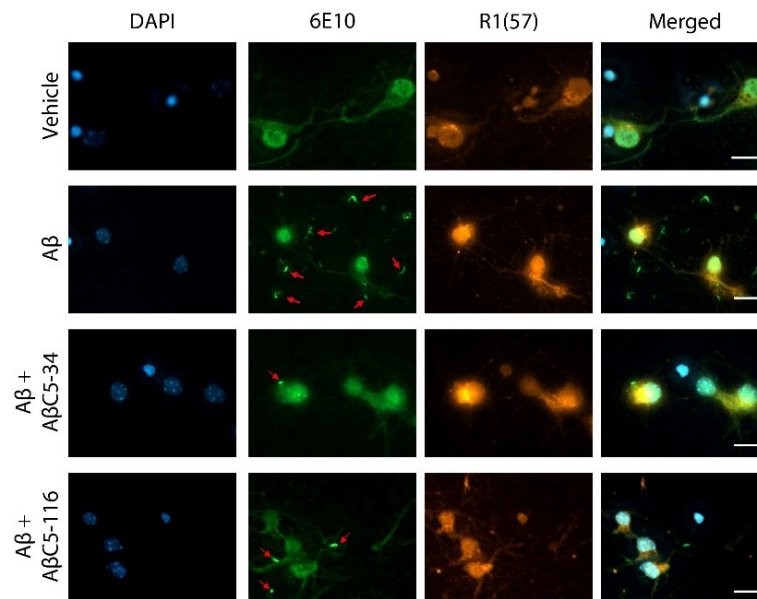


Figure 3.8. Immunofluorescence labelling of mouse cortical neurons. Cells were treated with either the anti-A β 6E10 antibody or the anti-APP R1(57) antibody, for 1 h at 37 °C with pre-aggregated A β_{40} (1 μ M). A β_{40} was previously aggregated for 3 d in the presence or absence of A β C5-34 and A β C5-116 (1 μ M). Red arrows show the patch-like assemblies of aggregated A β_{40} , which are specifically recognized by the anti-A β antibody 6E10 but not by the anti-APP antibody R1(57). Scale bar, 30 μ m. Representative images from $n=2$ independent experiments are presented. The surface binding assay was performed in the lab of Dr. Spiros Efthimiopoulos at the University of Athens.

Taken together, the results presented in this section indicated that the examined cyclic peptides decrease the toxic effect of A β on neuronal cultures. Additionally, A β C5-116 is capable of reducing the toxic effect of A β_{42} oligomers, secreted by 7PA2 cells, while both cyclic peptides led to the formation of fewer A β_{40} species capable of binding the neuronal surface.

3.3. Evaluation of the effect of the selected peptides on A β aggregation and neurotoxicity *in vivo*

The nematode *C. elegans* was employed in order to assess the peptides' protective effect against A β aggregation and toxicity *in vivo*, as various strains of *C. elegans* are established models of AD³³⁴. A paralysis assay was initially conducted in the CL2006 strain, in which human A β_{42} is constitutively

expressed in the body-wall muscle cells of the worm. In this model, A β aggregation is followed by adult-onset paralysis. *E. coli* OP50 cells were transformed with the pSICLOPPS plasmids responsible for encoding A β C5-34 or A β C5-116 and protein production was induced. Following growth to an OD₆₀₀≈0.7, the bacteria were used to feed the CL2006 worms. These worms exhibited significant delays in the appearance of the paralysis phenotype, compared to animals that were fed OP50 bacteria that produced a random peptide (**Figure 3.9a**, left). A parallel paralysis assay used CL4176 worms, which express human A β ₄₂ under the control of a heat-inducible promoter. This assay allowed the investigation of cyclic peptide concentration effect on paralysis. Indeed, paralysis-delaying effects occurred in a dose-responsive manner for both A β C5-34 and A β C5-116 (**Figure 3.9a, b, c**).

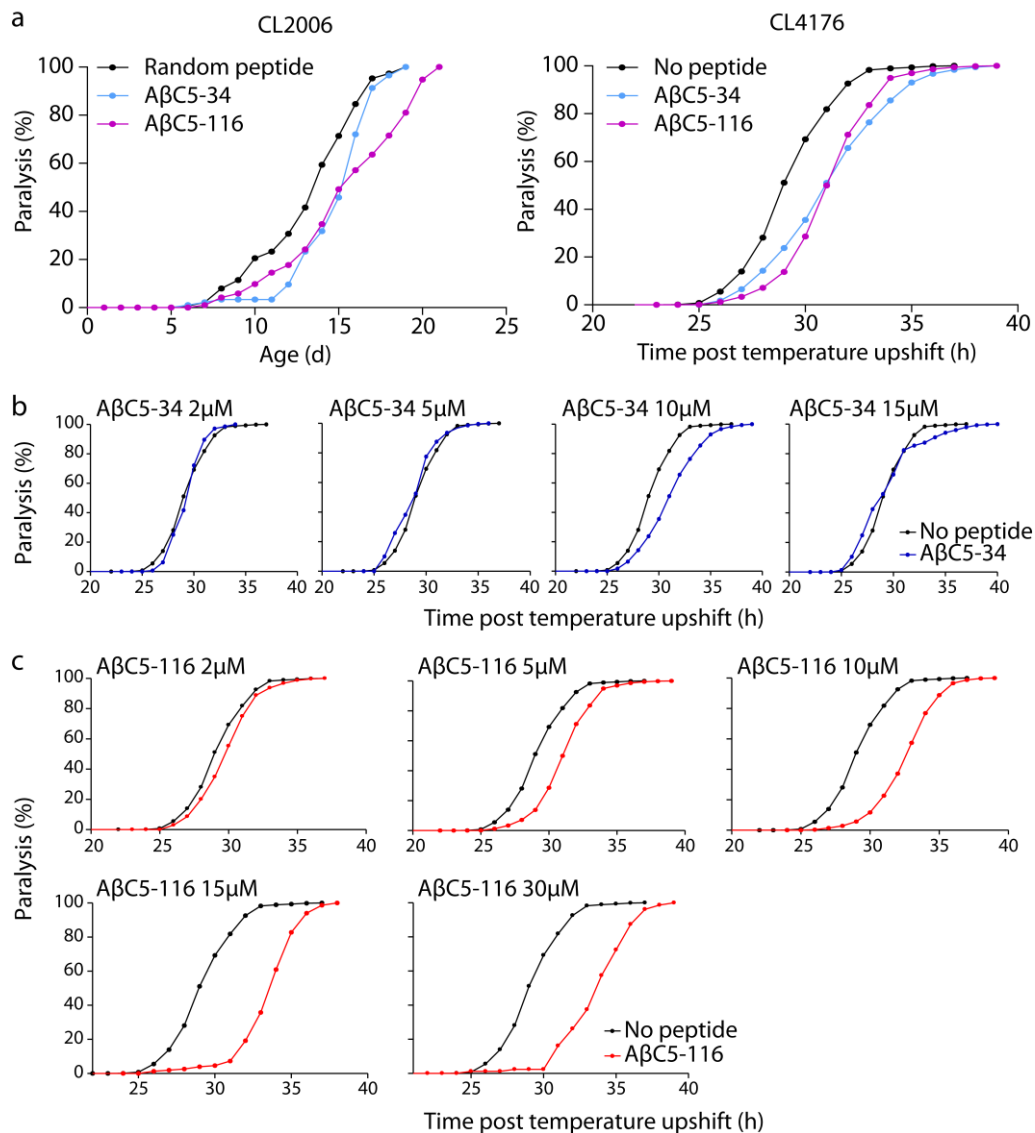


Figure 3.9. Paralysis curves of *C. elegans* strains CL2006 and CL4176. a) (Left) CL2006 worms were fed with *E. coli* OP50 cells biosynthetically producing AβC5-34, AβC5-116, or a randomly selected cyclic peptide from the vectors pSICLOPPS-AβC5-34, pSICLOPPS-AβC5-116, and pSICLOPPS-Random1, respectively. Statistical significances for differences compared to the “Random peptide” samples are as follows: Random peptide: mean=13.44±0.2, $n=213/305$; AβC5-34: mean=15.06±0.1, $n=46/105$, $P<0.0001$; AβC5-116: mean=15.1±0.2, $n=119/218$, $P<0.0001$. (Right) Paralysis curves of *C. elegans* CL4176 treated with synthetic AβC5-34 (10 μM), AβC5-116 (5 μM), or without peptide. Statistical significances for differences compared to the “No peptide” samples are as follows: No peptide: mean=29.2±0.2, $n=1309/1325$. AβC5-34: mean=31.0±0.2, $n=789/806$, $P<0.0001$; AβC5-116: mean=31.0±0.1, $n=733/743$, $P<0.0001$. b, c) Selected cyclic pentapeptides AβC5-34 and AβC5-116 inhibit Aβ aggregation and Aβ-induced toxicity *in vivo*, dose-dependently. Paralysis curves of the CL4176 strain producing human Aβ₄₂, which was treated with synthetic AβC5-34 (b) or AβC5-116 (c), at the concentration indicated in each graph. The “No peptide” sample was used as control for all experiments while it contains a volume of DMSO that is equivalent to that of the corresponding sample containing synthetic cyclic peptide (0.26% final plate concentration). No peptide: mean=29.00±0.1, $n=651/659$. AβC5-34 (2 μM): mean=29.20±0.1, $n=144/147$, NS; AβC5-34 (5 μM): mean= 28.78±0.2,

$n=600/606$, $P<0.01$; A β C5-34 (10 μ M): mean= 31.0 ± 0.2 , $n=789/806$, $P<0.0001$; A β C5-34 (15 μ M): mean= 28.79 ± 0.2 , $n=588/599$, NS. A β C5-116 (2 μ M): mean= 29.77 ± 0.2 , $n=726/736$, $P<0.001$; A β C5-116 (5 μ M): mean= 31.0 ± 0.1 , $n=733/743$, $P<0.0001$; A β C5-116 (10 μ M): mean= 32.65 ± 0.2 , $n=564/571$, $P<0.001$; A β C5-116 (15 μ M): mean= 33.57 ± 0.1 , $n=151/153$, $P<0.001$; A β C5-116 (30 μ M): mean= 33.66 ± 0.2 , $n=80/87$, $P<0.001$. *C. elegans* assays were performed in the lab of Dr. Niki Chondrogianni at the National Hellenic Research Foundation.

The state of A β aggregation in the worms was visualized using the CL2331 strain, which commences expression of the A β_{3-42} -GFP fusion in its body-wall muscle cells following a temperature up-shift. Worms that were treated with either A β C5-34 or A β C5-116 showed a significant reduction of A β deposits (Figure 3.10a). Additionally, the biochemical analysis of A β levels in CL4176 worms, revealed a significant reduction of both total and oligomeric species upon treatment with A β C5-34 and A β C5-116 (Figure 3.10b).

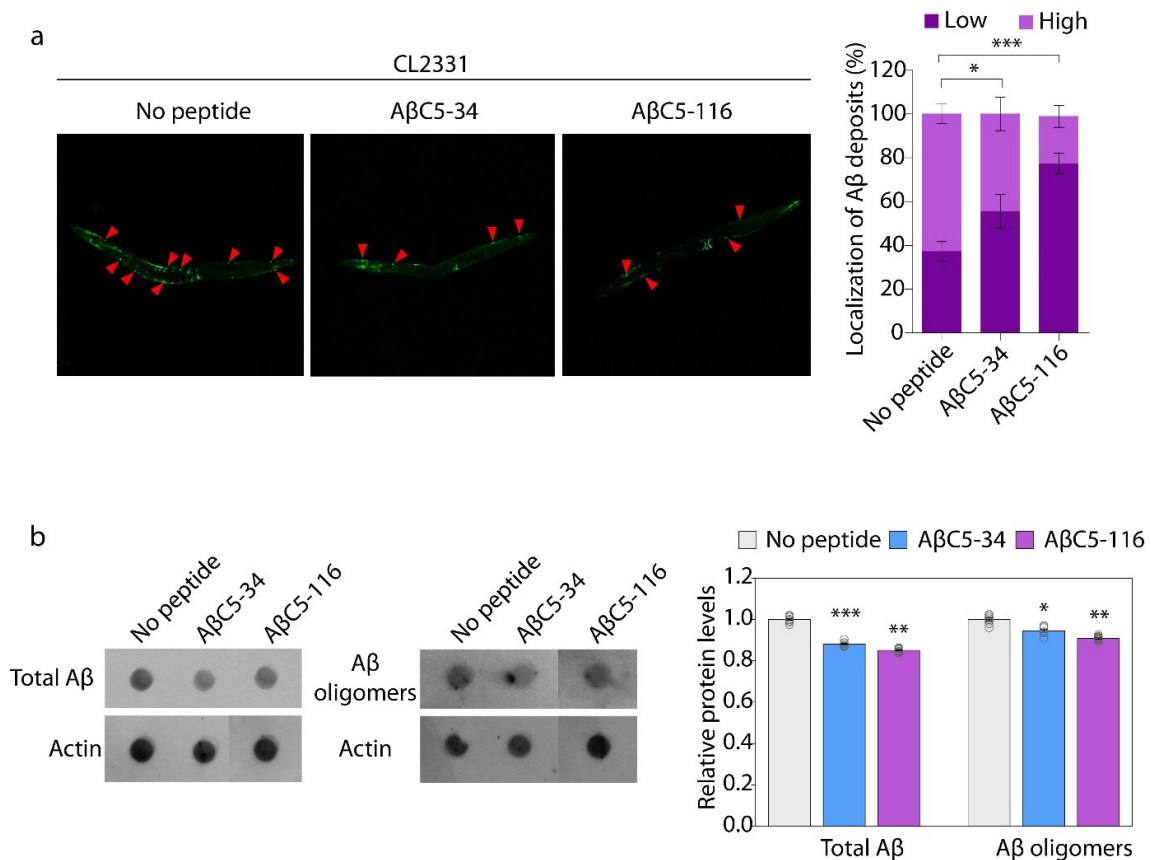


Figure 3.10. Selected cyclic pentapeptides A β C5-34 and A β C5-116 inhibit A β -induced aggregation *in vivo*. a) Representative fluorescence micrographs of CL2331 worms that were treated with synthetic

A β C5-34 (10 μ M), A β C5-116 (5 μ M), or without peptide for 48 h. Red arrows show deposits of aggregated A β_{3-42} -GFP (left). Count of A β_{3-42} -GFP deposits in the body of CL2331 worms treated with synthetic A β C5-34 (10 μ M, $n=79$), A β C5-116 (5 μ M, $n=95$), or without peptide ($n=190$). **b**) Dot blot (left) and quantification (right) of total and oligomeric A β_{42} . The antibodies 6E10 and AB9234 were used for total and oligomeric A β_{42} respectively, in CL4176 worms that were treated with synthetic A β C5-34 (10 μ M), A β C5-116 (5 μ M), or without peptide. The worms were collected when 50% of the control population was paralyzed (Figure 4.24a). Actin served as loading control. Statistical significance is indicated for differences compared to the “No peptide” samples. Mean values \pm s.e.m. are reported in all cases. * $P \leq 0.05$, ** $P \leq 0.01$, *** $P \leq 0.001$. *C. elegans* assays were performed in the lab of Dr. Niki Chondrogianni at the National Hellenic Research Foundation.

3.4. Remarks

Other than proving the bioactivity of two therapeutic candidates, the assays presented in this chapter aimed at validating the combined bacterial system’s ability to discover such candidates. As discussed in the previous chapters, the bacterial system displayed a set of properties that should increase the chances of identifying biologically relevant hits. Indeed, the *in vitro* biochemical assays indicated that the synthetic A β C5-34 and A β C5-116 cyclic peptides interfered with the aggregation pathways of A β and in the process led to the generation of fewer and atypical A β aggregates. Interestingly, CD experiments revealed that A β C5-34 and A β C5-116 displayed different effects on A β_{40} , as only A β C5-116 appeared to delay the formation of β -sheets. This observation was not surprising, as the two A β isoforms have been known to follow different aggregation pathways^{188,190}. Nonetheless, both selected cyclic peptides were capable of preventing pre-aggregated A β_{40} from binding the neuronal membrane. Furthermore, the *in vitro* neuronal viability assays with pre-aggregated A β showed that the aggregates formed in the presence of either A β C5-34 or A β C5-116 were less neurotoxic than the aggregates of cyclic peptide-free A β . However, exposure of mouse primary cortical neurons to 7PA2-produced A β_{42} oligomers³³⁰ showed that only A β C5-116 inhibited cytotoxicity. The inability of A β C5-34 to inhibit the cytotoxic effect of preformed oligomers is an indication of a different mode of action than A β C5-116, since these oligomeric species might evade the neuroprotective effect of A β C5-34. Finally, *in vivo*

experiments in CL4176 worms showed that the decelerated paralysis upon administration of the selected cyclic peptides was correlated to reductions in total and oligomeric A β levels. In total, the utilized *C. elegans* models helped place the toxicity- and aggregation-inhibiting effect of these peptides in a more complex biological setting. Combined with the biochemical findings, these results demonstrated the bacterial system's capability of identifying hits with biological activity against the aggregation and toxicity of A β .

Chapter 4. Structure-activity analysis of the selected A β -targeting cyclic peptides

4.1. Identification of residues critical for bioactivity in A β C5-34 and A β C5-116 by site-directed mutagenesis

In order to identify the residues that contributed to the bioactivity of the A β C5-34 and A β C5-116 peptides, their sequences were subjected to targeted mutagenesis. Specifically, the nucleophilic amino acid at position 1 was substituted with either one of the two alternative nucleophilic amino acids, while positions 3-5 underwent alanine mutagenesis. Substitution of Ser1 in A β C5-34 with either Cys or Thr led to background levels of fluorescence in the bacterial A β ₄₂-EGFP assay. Similarly, alanine-scanning mutagenesis of positions 3 to 5 also led to background fluorescence levels for all tested variants. In total, the cyclo-SASPT sequence of A β C5-34 did not tolerate any of the attempted alterations (**Figure 4.1**). This suggests that all amino acids in the cyclo-SASPT sequence are critical for its A β -targeting activity.

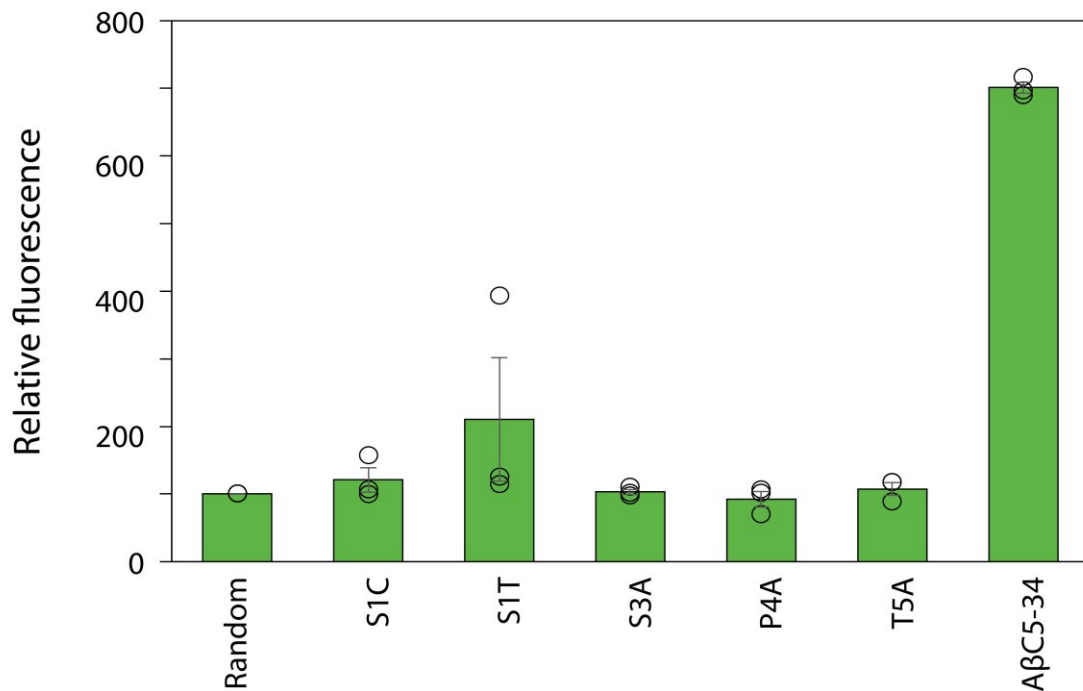


Figure 4.1. Bacterial Aβ₄₂-EGFP fluorescence for AβC5-34 sequence variants. Fluorescence of *E. coli* BL21(DE3) cells co-expressing Aβ₄₂-EGFP along with the selected cyclic pentapeptide AβC5-34, the indicated nucleophile variants in position 1, or the position 3-5 Ala variants. The Aβ₄₂-EGFP fluorescence of the cell population producing a random cyclic peptide was arbitrarily set at 100%. Mean values ± s.e.m. are reported ($n=3$ independent experiments, each one performed in triplicate).

On the contrary, AβC5-116 proved to be much more flexible with regard to the composition of its sequence, as indicated by the ability of the generated variants to enhance bacterial Aβ₄₂-EGFP fluorescence (**Figure 4.2**). Nucleophilic residue substitutions for AβC5-116 revealed a characteristic preference for the original Thr1, though Cys1 and Ser1 variants did lead to marginally enhanced bacterial Aβ₄₂-EGFP fluorescence. Alanine substitutions at positions 3 and 4 for AβC5-116 were tolerated, as indicated by the significant enhancement in bacterial fluorescence resulting from the production of the AβC5-116(F3A) and AβC5-116(D4A) variants. On the other hand, substitution of Arg5 by Ala caused bacterial Aβ₄₂-EGFP fluorescence to drop to background levels, confirming the residue's contribution to peptide functionality (**Figure 4.2**). The data derived from the above substitutions were in agreement with the prevalence of Thr1 and Arg5 among the isolated pentapeptides.

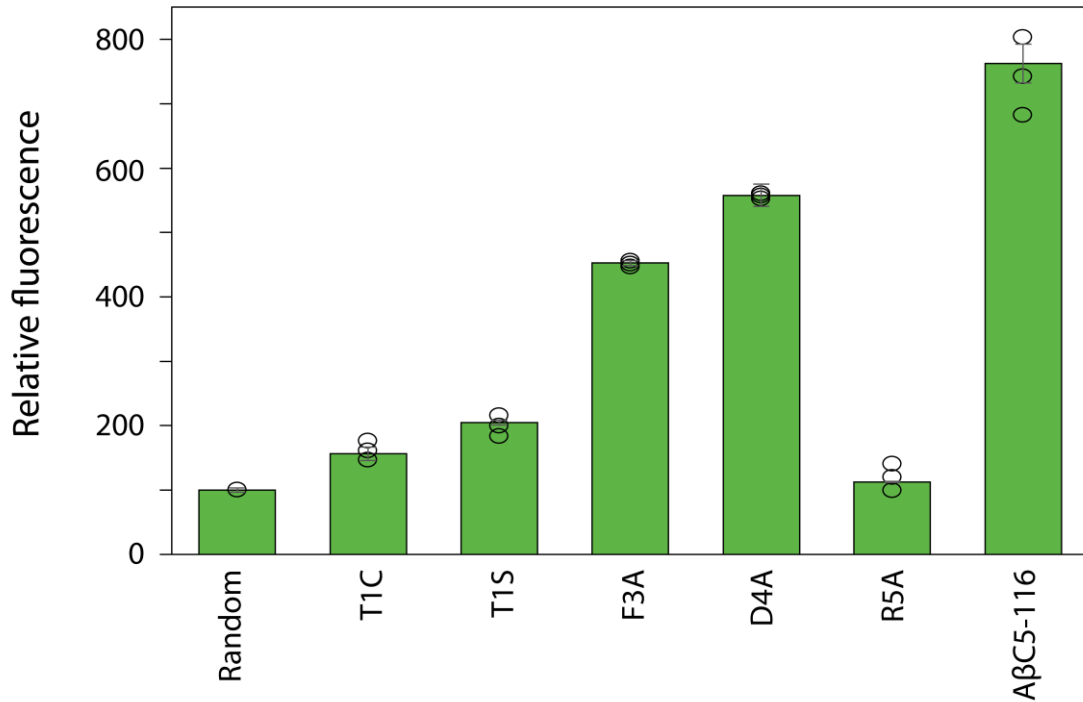


Figure 4.2. Bacterial A β ₄₂-EGFP fluorescence for A β C5-116 sequence variants. Fluorescence of *E. coli* BL21(DE3) cells co-expressing A β ₄₂-EGFP along with the selected cyclic pentapeptide A β C5-116, the indicated position 1 nucleophile variants, or the position 3-5 Ala variants. The A β ₄₂-EGFP fluorescence of the cell population producing a random cyclic peptide was arbitrarily set at 100%. Mean values \pm s.e.m. are reported ($n=3$ independent experiments, each one performed in triplicate).

Finally, semi-saturation mutagenesis of position 2 for A β C5-116 using representative amino acids of various chemical classifications revealed that only Thr and Ser could occupy position 2 and maintain an enhanced bacterial A β ₄₂-EGFP fluorescence (**Figure 4.3**). Consistent with this, Thr2 was observed in five of the ten individually isolated peptide clones (**Table 2.3**), which also contained Thr1 and Arg5.

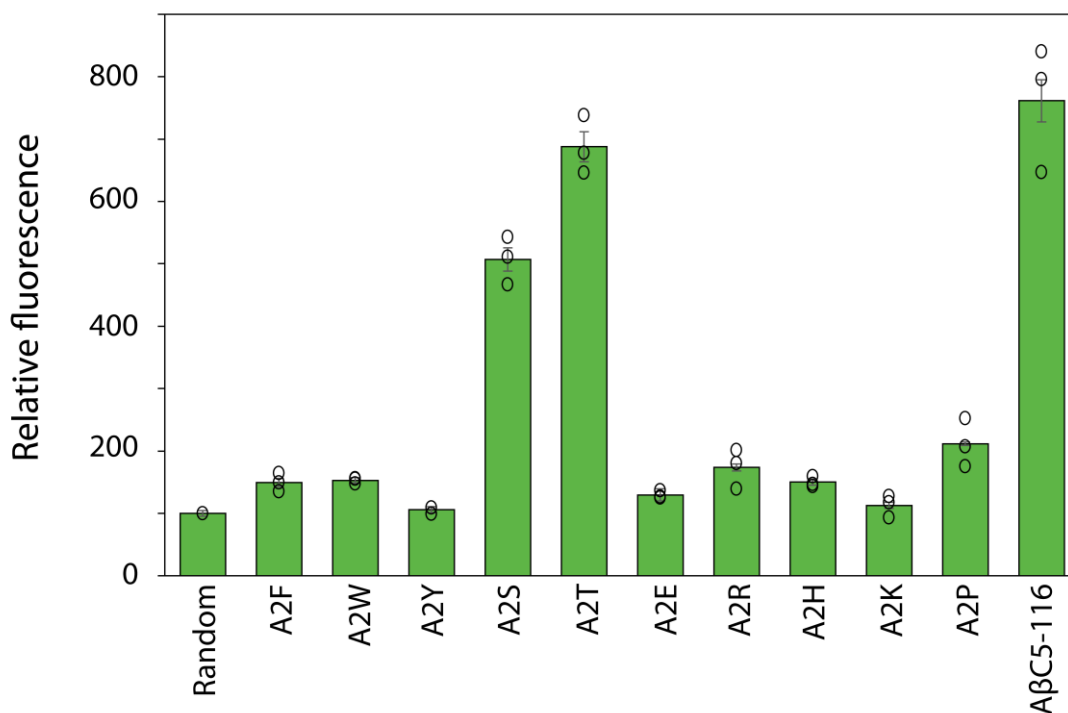


Figure 4.3. Bacterial A β ₄₂-EGFP fluorescence for A2 variants of A β C5-116. Fluorescence of *E. coli* BL21(DE3) cells co-expressing A β ₄₂-EGFP along with the selected cyclic pentapeptide A β C5-116 and semi-saturation mutagenesis variants of A β C5-116. The A β ₄₂-EGFP fluorescence of the cell population producing a random cyclic peptide was arbitrarily set at 100%. Mean values \pm s.e.m. are reported ($n=3$ independent experiments, each one performed in triplicate).

4.2. High-throughput sequencing of the sorted peptide library

The results presented in the previous section suggested that a significant number of A β -targeting pentapeptides that resemble A β C5-116 and very few similar to A β C5-34, should appear in the sorted peptide library. In order to test this hypothesis, the sorted bacterial population underwent deep sequencing analysis

An Ion Torrent™ PROTON (Thermo Fisher Scientific, USA) Next Generation Sequencing (NGS) system was used to determine the DNA sequences of the peptide-encoding region of the pSICLOPPS vectors contained in the selected bacterial clones after the 2nd round of FACS sorting (**Figure 2.10**), as described in **Methods and Materials**. Sequencing and raw data analysis were performed at the

Genomics core facility of the Biomedical Sciences Research Center “Alexander Fleming” (Athens, Greece), yielding a list of DNA sequences. Sequencing coverage allowed most peptide-encoding DNA sequences to be identified multiple times, since NGS results showed $\sim 5.6 \times 10^6$ sequence reads. This number of sequence reads was enough to cover the maximum theoretical diversity of the sorted library (10^3), derived from $10^7 \times 0.01 \times 0.01$; 10^7 refers to the predicted actual diversity of the unsorted combined library and 0.01 refers to the selection of approximately 1% of the bacterial population that exhibited the highest fluorescence in each of the two sorting rounds. Therefore, it was assumed that practically all cyclic peptide-encoding sequences within the sorted library were represented in the produced library.

4.2.1. Identification of valid cyclic peptide-encoding DNA sequences

Initially, all the DNA sequences derived from NGS were translated to their respective linear peptide sequences using Bioedit software³³⁵. Since the initial combined library was designed to contain tetra- penta- and hexapeptides, only DNA sequences of 12, 15 or 18 base pairs (4, 5 or 6 codons) were considered. As a result, approximately 10^6 sequence reads were rejected and the total number of analyzed sequences was reduced to 4.55×10^6 . The existence of sequence lengths other than the prescribed can be attributed to various causes like misreads during sequencing, PCR errors, mutations etc.

Next, the ability of the sequence to be translated to a peptide capable of undergoing cyclization was examined. As discussed previously, cyclization requires a nucleophilic amino acid (Cys, Ser or Thr) in position 1 of the peptide sequence. While the library was designed to carry only nucleophilic amino acids in position 1, a small number of reads (16,358) corresponded to DNA sequences that did not encode the required nucleophilic amino acid. These sequences were discarded, leaving the total

number of reads essentially unchanged. Finally, peptide-encoding sequences that contained stop codons were also discarded, reducing the total number of reads to 4,550,719 (Table 4.1).

Table 4.1. Unique cyclic peptide-encoding DNA sequences. Sequences that correspond to tetra-, penta- and hexapeptides are displayed. Only DNA sequences that encode peptides with a nucleophilic amino acid in position 1 (Cys, Ser or Thr) were considered.

bp	Peptide	With STOP-codons		<i>Without</i> STOP-codons	
		Unique DNA sequences	Total reads	Unique DNA sequences	Total reads
12	Tetra	276	47,013	254	20,675
15	Penta	1,832	4,075,655	1,766	4,073,710
18	Hexa	316	458,369	298	456,334
Combined		2,424	4,581,037	2,318	4,550,719

The restrictions described above only reduced the total DNA reads by a relatively small number. However, they were necessary in order to obtain only the unique DNA sequences with the ability to encode a cyclic peptide. The remaining 4.55×10^6 DNA sequence reads corresponded to 2,318 unique DNA sequences.

Nevertheless, the cyclic peptide library plasmids were constructed using codon degeneracy. As a result, many of these 2,318 unique DNA sequences were expected to encode identical peptide sequences. Since the analysis of the sorted peptide library would focus on amino acid composition, it became necessary to build a list of unique peptide sequences. Hence, duplicate peptide sequences were consolidated using Microsoft EXCEL: identical peptide sequences were combined and their corresponding DNA sequence reads were summed. As a result, the sorted library was found to contain 1,663 unique peptide sequences.

Table 4.2. Unique peptide sequences encoded by the sorted library.

	Unique sequences	Number of reads
Tetrapeptides	180	20,675
Pentapeptides	1,208	4,073,710
Hexapeptides	275	456,334
Combined	1,663	4,550,719

4.2.2. Cyclic-peptide nomenclature

In order to distinguish between unique peptides, a nomenclature system was established. Specifically, all cyclic peptides were given the acronym A β referencing the target protein, followed by the letter “C”, referring to the cyclic nature of the peptide. A β C is then followed by the numbers 4, 5 or 6 that correspond to the three sequence lengths (tetra-, penta- or hexapeptide). An additional number describes the order in which the peptide appears, in a decreasing read-number order. For instance, A β C5-34 refers to the A β -targeting cyclic pentapeptide with the 34th highest read number among selected pentapeptides.

4.2.3. Initial observations on NGS results

The abundance of pentapeptides among the selected peptide pool was immediately apparent (**Figure 4.4**), consistent with the initial observations on the individually isolated clones from the sorted library (**Table 2.2**). Pentapeptides comprised 72.64% of the unique peptide sequences and 89.52% of the total sequence reads while their prevalence among NGS-identified sequences was also reflected in the individually isolated peptide clones (80% of all isolated peptides). Considering that pentapeptides theoretically consisted only 4.75% of the theoretical number of peptides in the initial combined library, this selection outcome was quite unexpected.

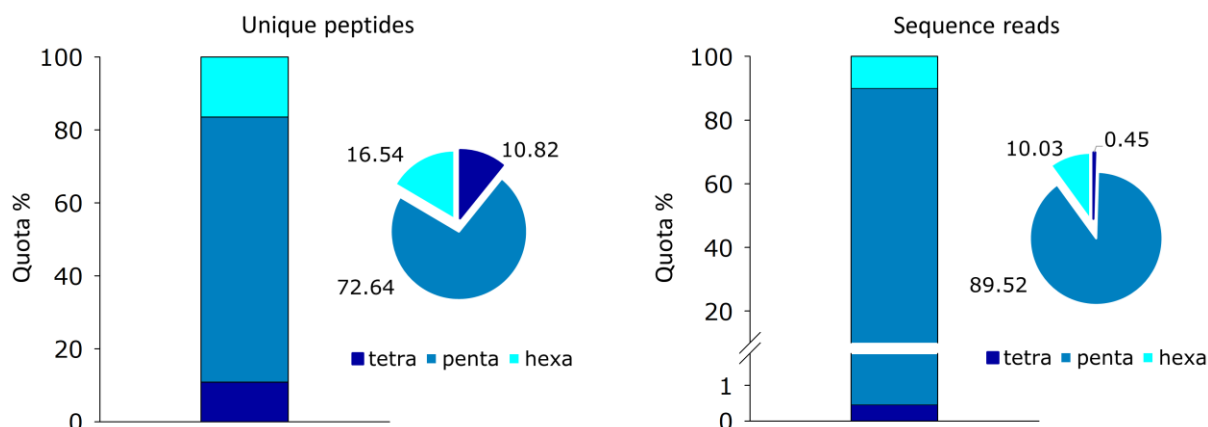


Figure 4.4. Distribution of the different sizes of cyclic peptides among the sorted library. Only unique cyclic peptides were considered. Left: Distribution of unique peptide sequences regardless of read number. Right: Distribution tetra-, penta and hexapeptide reads.

After an initial examination of the sorted peptide pool, the following were observed: Thr1 was emphatically the most favoured of the nucleophilic residues while Arg5 was the preferred amino acid among the selected pentapeptides (**Figure 4.5**). Similarly, Arg4 or Arg6 were among the most dominant amino acids for tetrapeptides and hexapeptides, respectively. In addition, Thr2 was a very frequently occurring amino acid in pentapeptides, as well as tetrapeptides (**Figure 4.5**). A more exhaustive analysis of prevalent amino acids is presented in the following paragraphs.

Amino acid composition as determined by sequences

Tetrapeptides					Pentapeptides						
	1	2	3	4		1	2	3	4	5	
A	0.00	16.67	5.56	10.56	A	0.00	10.76	4.30	12.33	1.24	A
I	0.00	3.33	2.78	0.00	I	0.00	8.77	4.47	0.91	5.55	I
L	0.00	5.00	4.44	10.00	L	0.00	5.30	11.26	3.06	15.31	L
V	0.00	15.56	15.00	16.67	V	0.00	11.84	11.26	11.75	6.95	V
F	0.00	0.00	1.67	1.11	F	0.00	0.41	2.73	3.56	1.08	F
W	0.00	1.67	8.89	6.67	W	0.00	1.32	6.37	14.90	4.30	W
Y	0.00	1.11	1.67	0.00	Y	0.00	0.17	2.57	1.57	0.00	Y
N	0.00	1.11	0.00	1.11	N	0.00	1.66	1.41	1.66	0.25	N
Q	0.00	0.00	1.67	3.33	Q	0.00	0.75	1.74	0.58	5.55	Q
C	5.00	2.22	7.78	0.56	C	19.54	1.32	2.65	4.88	0.75	C
M	0.00	1.11	0.00	1.11	M	0.00	1.90	2.57	1.74	0.17	M
S	2.22	7.22	15.00	4.44	S	8.61	7.78	9.69	6.04	2.32	S
T	92.78	30.00	4.44	2.22	T	71.85	31.46	8.36	5.88	1.57	T
D	0.00	0.00	3.89	5.00	D	0.00	0.33	5.30	5.88	0.58	D
E	0.00	0.00	4.44	3.89	E	0.00	0.00	8.94	2.73	1.90	E
R	0.00	8.33	11.67	22.78	R	0.00	6.21	7.95	7.95	44.87	R
H	0.00	1.67	1.11	0.56	H	0.00	5.71	1.66	4.88	0.00	H
K	0.00	0.00	0.00	0.00	K	0.00	0.00	0.41	0.00	0.25	K
P	0.00	5.00	1.67	8.33	P	0.00	2.73	3.06	3.23	3.64	P
G	0.00	0.00	8.33	1.67	G	0.00	1.57	3.31	6.46	3.73	G
Sum	100.00	100.00	100.00	100.00	Sum	100.00	100.00	100.00	100.00	100.00	Sum

Hexapeptides							
	1	2	3	4	5	6	
A	0.00	2.18	11.27	2.91	6.55	1.09	A
I	0.00	5.45	1.45	1.82	1.45	6.55	I
L	0.00	14.18	7.27	5.82	6.55	16.73	L
V	0.00	9.82	14.18	9.09	15.27	6.55	V
F	0.00	0.36	1.82	2.91	16.36	0.73	F
W	0.00	1.45	1.45	16.36	2.18	2.55	W
Y	0.00	2.18	0.00	1.45	0.73	0.36	Y
N	0.00	2.91	0.73	0.73	0.00	4.73	N
Q	0.00	5.45	1.45	2.55	0.36	3.64	Q
C	3.27	1.09	0.73	6.55	4.36	0.73	C
M	0.00	1.45	0.73	0.73	1.45	0.36	M
S	8.36	1.82	10.18	14.18	11.64	1.09	S
T	88.36	8.36	11.27	6.91	6.55	5.82	T
D	0.00	0.36	2.91	0.00	1.82	17.82	D
E	0.00	0.36	3.64	1.09	1.82	2.55	E
R	0.00	9.45	11.64	10.18	10.91	20.00	R
H	0.00	13.09	3.64	0.73	1.82	0.36	H
K	0.00	0.36	2.55	0.00	0.00	1.09	K
P	0.00	18.18	6.91	7.64	7.64	1.45	P
G	0.00	1.45	6.18	8.36	2.55	5.82	G
Sum	100.00	100.00	100.00	100.00	100.00	100.00	Sum

Figure 4.5. Amino acid composition analyses for tetra-, penta- and hexapeptides. The amino acid compositions of tetrapeptides, pentapeptides and hexapeptides have been analyzed using only unique sequences: each amino acid is counted once per peptide sequence in which it appears, and the sum is divided by the total number of unique peptide sequences.

4.2.4. Sequence read numbers and amino acid composition analyses

The number of times each sequence within the sorted pool was identified varied significantly. This was not believed to be a random phenomenon as the more frequent appearance of certain peptide sequences could possibly signify increased peptide activities and therefore higher levels of A β ₄₂-EGFP fluorescence. Villar-Pique *et al.* suggested that A β isoforms show variable aggregation propensities even within the A β -EGFP fusion, leading to varying impairment effects on bacterial aging and division: A β isoforms with a higher aggregation propensity exerted a more toxic effect³¹¹. Likewise, when one of the selected peptides interferes with A β ₄₂-EGFP aggregation, it could also exert a positive cellular fitness effect, providing the bacterium with an evolutionary advantage. As such, peptides that were identified more frequently were suspected to display relatively increased activities against the aggregation of A β .

Naturally, the variations in the frequency of occurrence for each peptide had to be considered in order to identify sequence motifs that would not be biased towards the more frequently occurring sequences. Equally, it was necessary to avoid extended randomization of the amino acid composition, as this would “blur” the analyses, obstructing the characterization of motifs. As a result, it was decided to exclude sequences with less than 50 reads. This cut-off did not significantly alter the composition analysis as shown in **Figure 4.6**; rather it highlighted the already prevalent amino acids simplifying motif characterization. Most importantly, the majority of rare residues were represented, ensuring that these would be considered in the discovery of sequence motifs (**Figure 4.6**). One notable exception was Arg6 for hexapeptides, which was prevalent for the amino acid composition of sequences with ≥ 1 reads, though practically non-existent in sequences with ≥ 50 reads. This unique discrepancy will be discussed in a following paragraph.

Amino acid composition determined by sequences

Tetrapeptides										Pentapeptides									
≥1 read					≥50 reads					≥1 read					≥50 reads				
	1	2	3	4	1	2	3	4	1	2	3	4	5	1	2	3	4	5	
A	0.00	16.67	5.56	10.56	0.00	7.32	0.00	12.20	A	0.00	10.76	4.30	12.33	1.24	0.00	10.56	3.11	13.25	0.62
I	0.00	3.33	2.78	0.00	0.00	0.00	4.88	0.00	I	0.00	8.77	4.47	0.91	5.55	0.00	8.07	4.35	0.83	8.90
L	0.00	5.00	4.44	10.00	0.00	7.32	7.32	7.32	L	0.00	5.30	11.26	3.06	15.31	0.00	6.83	12.22	4.35	16.77
V	0.00	15.56	15.00	16.67	0.00	19.51	14.63	7.32	V	0.00	11.84	11.26	11.75	6.95	0.00	14.70	12.42	11.39	6.63
F	0.00	0.00	1.67	1.11	0.00	0.00	0.00	2.44	F	0.00	0.41	2.73	3.56	1.08	0.00	0.41	3.31	4.97	1.24
W	0.00	1.67	8.89	6.67	0.00	4.88	7.32	7.32	W	0.00	1.32	6.37	14.90	4.30	0.00	1.24	8.28	18.43	0.83
Y	0.00	1.11	1.67	0.00	0.00	4.88	4.88	0.00	Y	0.00	0.17	2.57	1.57	0.00	0.00	0.21	2.69	1.86	0.00
N	0.00	1.11	0.00	1.11	0.00	2.44	0.00	0.00	N	0.00	1.66	1.41	1.66	0.25	0.00	0.83	1.45	1.24	0.62
Q	0.00	0.00	1.67	3.33	0.00	0.00	0.00	2.44	Q	0.00	0.75	1.74	0.58	5.55	0.00	0.83	1.45	0.21	3.52
C	5.00	2.22	7.78	0.56	14.63	2.44	4.88	2.44	C	19.54	1.32	2.65	4.88	0.75	28.36	2.48	2.90	2.90	0.21
M	0.00	1.11	0.00	1.11	0.00	0.00	0.00	0.00	M	0.00	1.90	2.57	1.74	0.17	0.00	1.04	4.76	1.45	0.41
S	2.22	7.22	15.00	4.44	9.76	2.44	14.63	2.44	S	8.61	7.78	9.69	6.04	2.32	10.77	5.18	5.80	4.55	1.04
T	92.78	30.00	4.44	2.22	75.61	41.46	9.76	4.88	T	71.85	31.46	8.36	5.88	1.57	60.87	36.85	10.35	8.90	1.24
D	0.00	0.00	3.89	5.00	0.00	0.00	2.44	7.32	D	0.00	0.33	5.30	5.88	0.58	0.00	0.00	4.55	6.83	0.21
E	0.00	0.00	4.44	3.89	0.00	0.00	0.00	4.88	E	0.00	0.00	8.94	2.73	1.90	0.00	0.00	8.70	1.45	0.62
R	0.00	8.33	11.67	22.78	0.00	7.32	14.63	17.07	R	0.00	6.21	7.95	7.95	44.87	0.00	2.07	5.80	3.73	53.42
H	0.00	1.67	1.11	0.56	0.00	0.00	2.44	0.00	H	0.00	5.71	1.66	4.88	0.00	0.00	3.52	1.24	7.25	0.00
K	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	K	0.00	0.00	0.41	0.00	0.25	0.00	0.00	0.62	0.00	0.21
P	0.00	5.00	1.67	8.33	0.00	0.00	0.00	14.63	P	0.00	2.73	3.06	3.23	3.64	0.00	3.52	2.90	2.07	1.45
G	0.00	0.00	8.33	1.67	0.00	0.00	12.20	7.32	G	0.00	1.57	3.31	6.46	3.73	0.00	1.66	3.11	4.35	2.07
Sum	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	Sum	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00

Hexapeptides												
≥1 read						≥50 reads						
	1	2	3	4	5	6	1	2	3	4	5	6
A	0.00	2.18	11.27	2.91	6.55	1.09	0.00	1.23	12.35	1.23	1.23	1.23
I	0.00	5.45	1.45	1.82	1.45	6.55	0.00	8.64	2.47	1.23	2.47	7.41
L	0.00	14.18	7.27	5.82	6.55	16.73	0.00	22.22	7.41	6.17	14.81	14.81
V	0.00	9.82	14.18	9.09	15.27	6.55	0.00	13.58	13.58	12.35	20.99	14.81
F	0.00	0.36	1.82	2.91	16.36	0.73	0.00	1.23	2.47	4.94	27.16	0.00
W	0.00	1.45	1.45	16.36	2.18	2.55	0.00	2.47	2.47	25.93	0.00	2.47
Y	0.00	2.18	0.00	1.45	0.73	0.36	0.00	1.23	0.00	3.70	2.47	0.00
N	0.00	2.91	0.73	0.73	0.00	4.73	0.00	0.00	0.00	1.23	0.00	4.94
Q	0.00	5.45	1.45	2.55	0.36	3.64	0.00	0.00	2.47	4.94	1.23	3.70
C	3.27	1.09	0.73	6.55	4.36	0.73	7.41	1.23	2.47	1.23	2.47	1.23
M	0.00	1.45	0.73	0.73	1.45	0.36	0.00	4.94	1.23	0.00	0.00	1.23
S	8.36	1.82	10.18	14.18	11.64	1.09	11.11	2.47	9.88	17.28	8.64	1.23
T	88.36	8.36	11.27	6.91	6.55	5.82	81.48	2.47	23.46	9.88	1.23	3.70
D	0.00	0.36	2.91	0.00	1.82	17.82	0.00	0.00	1.23	0.00	1.23	30.86
E	0.00	0.36	3.64	1.09	1.82	2.55	0.00	0.00	6.17	2.47	6.17	3.70
R	0.00	9.45	11.64	10.18	10.91	20.00	0.00	4.94	1.23	3.70	8.64	2.47
H	0.00	13.09	3.64	0.73	1.82	0.36	0.00	1.23	1.23	0.00	0.00	0.00
K	0.00	0.36	2.55	0.00	0.00	1.09	0.00	1.23	1.23	0.00	0.00	3.70
P	0.00	18.18	6.91	7.64	7.64	1.45	0.00	25.93	1.23	1.23	0.00	1.23
G	0.00	1.45	6.18	8.36	2.55	5.82	0.00	4.94	7.41	2.47	1.23	1.23
Sum	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00

Figure 4.6. Comparison between amino acid compositions. Sequence-based amino acid composition of tetra-, penta- and hexapeptides. Compositions are presented for peptide sequences with ≥1 read (left) and for peptides with ≥50 reads (right).

Following the application of the 50-read cut-off, 605 distinct sequences composed the selected peptide pool, which were used in the identification of sequence motifs. This population consisted of 41 tetrapeptides, 483 pentapeptides and 81 hexapeptides (Figure 4.7).

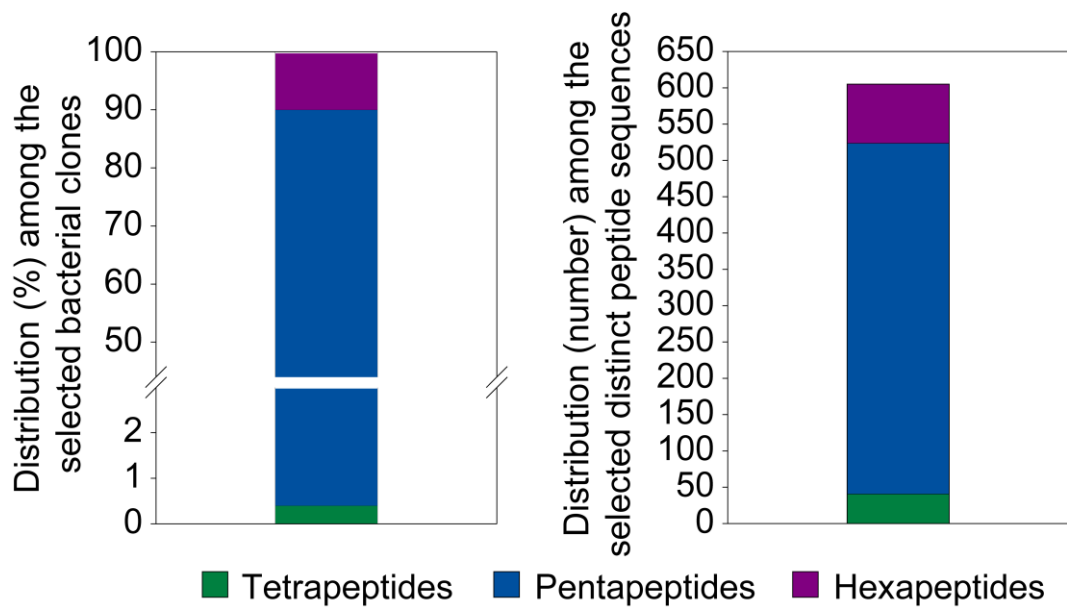


Figure 4.7. Peptide distribution within the sorted population. (Left). Peptide distribution given as a percentage of the total sorted population. (Right). Distribution of unique peptide sequences for the 605 members of the sorted peptide library.

Furthermore, it was investigated whether infrequently occurring peptide sequences enhanced the bacterial A β ₄₂-EGFP fluorescence. Four pentapeptide sequences with a very low sequence reads among the selected peptide pool were tested to that effect and all four were shown to enhance bacterial fluorescence significantly, suggesting that peptides with lower read numbers are also active modulators of A β aggregation (**Figure 4.8**).

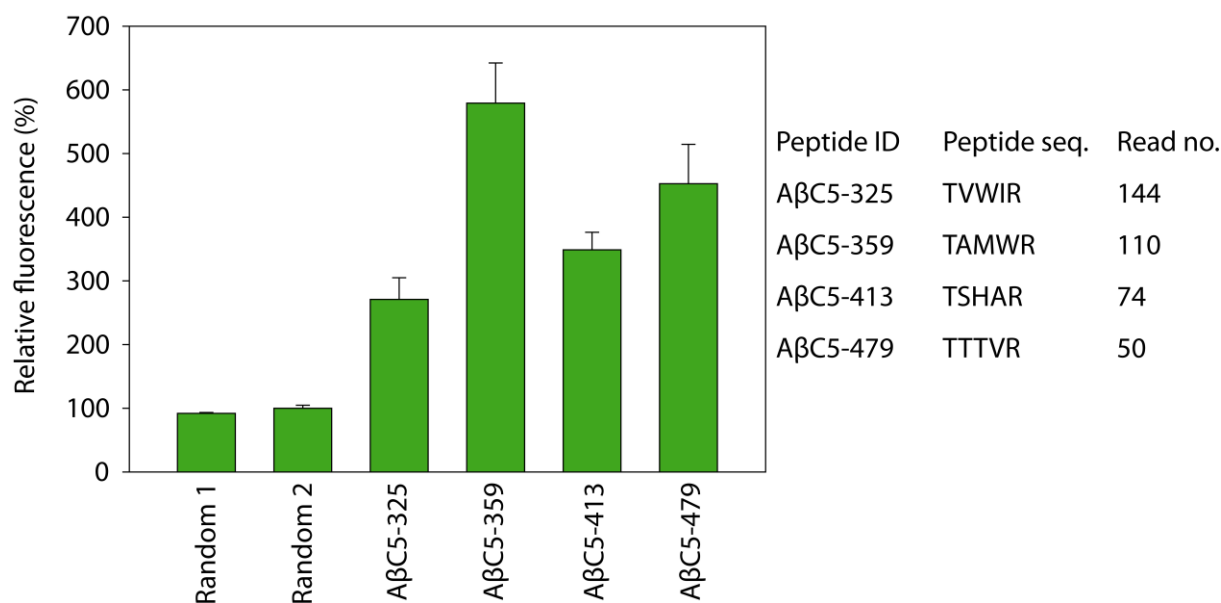


Figure 4.8. Low-read-number peptide effect on bacterial Aβ₄₂-EGFP fluorescence. Fluorescence of *E. coli* BL21(DE3) cells co-expressing Aβ₄₂-EGFP along with the selected cyclic pentapeptides AβC5-325, AβC5-359, AβC5-413 and AβC5-479 compared against the Aβ₄₂-EGFP fluorescence of the cell population producing a random cyclic peptide was arbitrarily set at 100%. Mean values ± s.d. are reported. (*n*=1 independent experiments, performed in triplicate).

4.2.5. Identification of distinct pentapeptide motifs within the sorted peptide library

Pentapeptides comprised 73% of the unique peptide sequences identified by NGS and the list of 483 pentapeptides with ≥50 sequence reads was used to create a sequence alignment in MEGA7 (<http://www.megasoftware.net/>)³³⁶. Subsequently, the alignment was used to produce a phylogenetic tree whose purpose was to simplify the identification of possible peptide motifs. Since sequence alignments require a starting and an ending point in the involved sequences, the alignment only included the linear interpretations (as translated) of the selected peptides. Naturally, this excluded the various linear interpretations that occurred upon rotation of each cyclic peptide (**Figure 4.8**). These were generated by sequentially rotating each sequence in an anticlockwise manner, a process termed circular permutation, (EXCEL, Microsoft, USA) yielding 2415 linear sequences for pentapeptides alone. However, small sequences are already difficult to align as they do not usually contain extended

segments of conserved sequence, thus, a complete alignment of both “as translated” and rotated sequences (**Figure 4.8**) would lead to a very complicated and difficult to interpret phylogenetic tree. On the contrary, phylogenetic analysis of the “as translated” linear sequences allowed easier identification of possible sequence motifs. Naturally, upon initial identification of a sequence motif, both as translated and rotated sequences would be considered in its analysis.

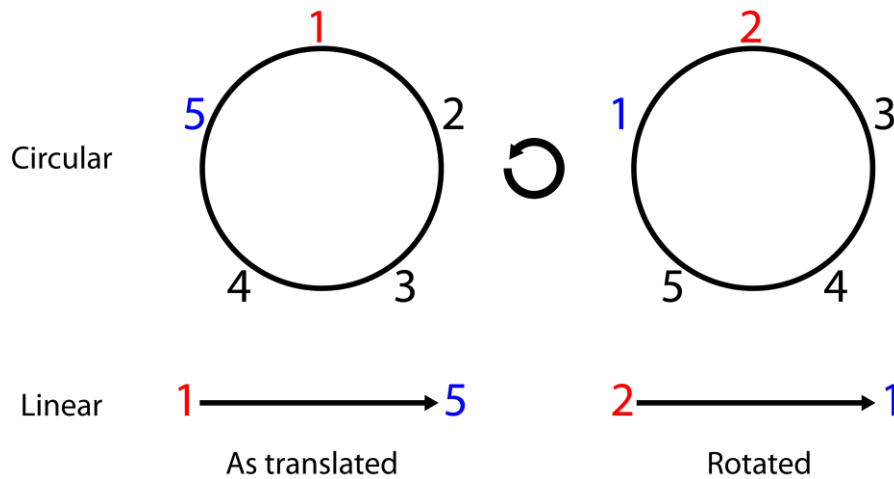


Figure 4.8. Anticlockwise rotation of a cyclic pentapeptide. Sequential rotations afforded five linear sequence interpretations of the same peptide, to be used for sequence alignment purposes. The “as translated” sequence refers to the peptide sequence that was derived from the translation of the NGS results.

A representative phylogenetic tree that was produced with MEGA7 for the top 120 pentapeptides among the selected pool (according to sequence read numbers) is presented in **Figure 4.9**. Less populated trees were generated for pentapeptide groups (families) with conserved residues and used as described in the following paragraphs. The complete pentapeptide list and the complete phylogenetic tree for pentapeptides in **Supplementary Table 3** and **Supplementary Figure 1**.

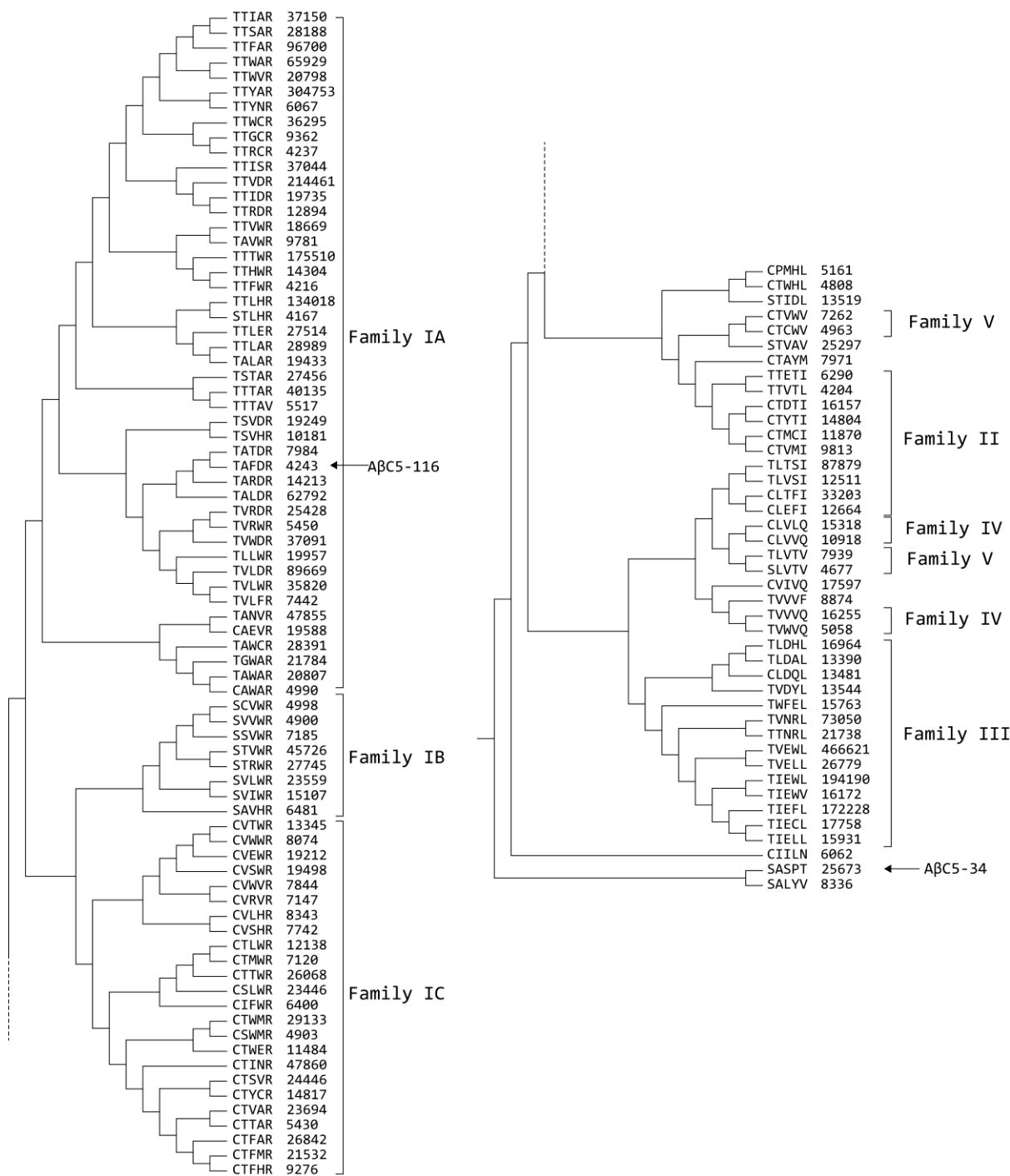


Figure 4.9. Representative phylogenetic tree for pentapeptide sequences. The phylogenetic tree contains the top 120 pentapeptide sequences, with regard to sequence reads. Tree was generated using the UPGMA statistical. Read numbers are attached next to the peptide sequences. Brackets designate possible pentapeptide families.

Initial observations derived from the phylogenetic tree presented in **Figure 4.9** suggested the existence of certain clusters of pentapeptides sharing sequence similarities. However, the inherent

difficulty of aligning a large population of short peptide sequences, the complicated nature of the respective phylogenetic trees and finally, the circular backbone of the investigated peptides, meant that any resulting observations on the phylogenetic tree of **Figure 4.9** would be used as initial guidelines for the more exact definition of sequence similarities throughout the pentapeptide population. Indeed, the combination of observations resulting from the phylogenetic analysis in **Figure 4.9** and the complete phylogenetic analysis for pentapeptides (**Supplementary Figure 1**) revealed the existence of apparently distinct pentapeptide families. This definition of certain common characteristics among groups of pentapeptides aided in the generation of separate and more targeted analyses for the hypothesized pentapeptide families, which revealed additional family-specific characteristics.

From this analysis, five distinctive clusters of selected pentapeptides were observed:

- Thr1/Arg5-containing sequences, termed Family IA
- Ser1/Arg5-containing sequences, termed Family IB
- Cys1/Arg5-containing sequences, termed Family IC
- Nu1/Ile5-containing sequences, termed Family II
- Thr1/Leu5-containing sequences, termed Family III
- Nu1/Gln5-containing sequences, termed Family IV
- Nu1/Val5-containing sequences, termed Family V
- A β C5-34 –like pentapeptides

Nu signifies the three possible nucleophilic residues for position 1 (Cys, Ser, Thr). The defining sequence characteristics of these clusters are investigated individually and in detail below. Interestingly, substitution of Arg5 in the A β C5-116 (cyclo-TAFDR) peptide by aliphatic amino acids led to bacterial A β ₄₂-EGFP fluorescence phenotypes that were practically diminished (**Figure 4.2** and **Figure 4.10**), suggesting a differentiation between the modes of bioactivity inherent to each one of

the Family I-IV. Importantly, none of the constructed peptide sequences was identified during NGS suggesting that these four families do not exhibit overlapping bioactivities.

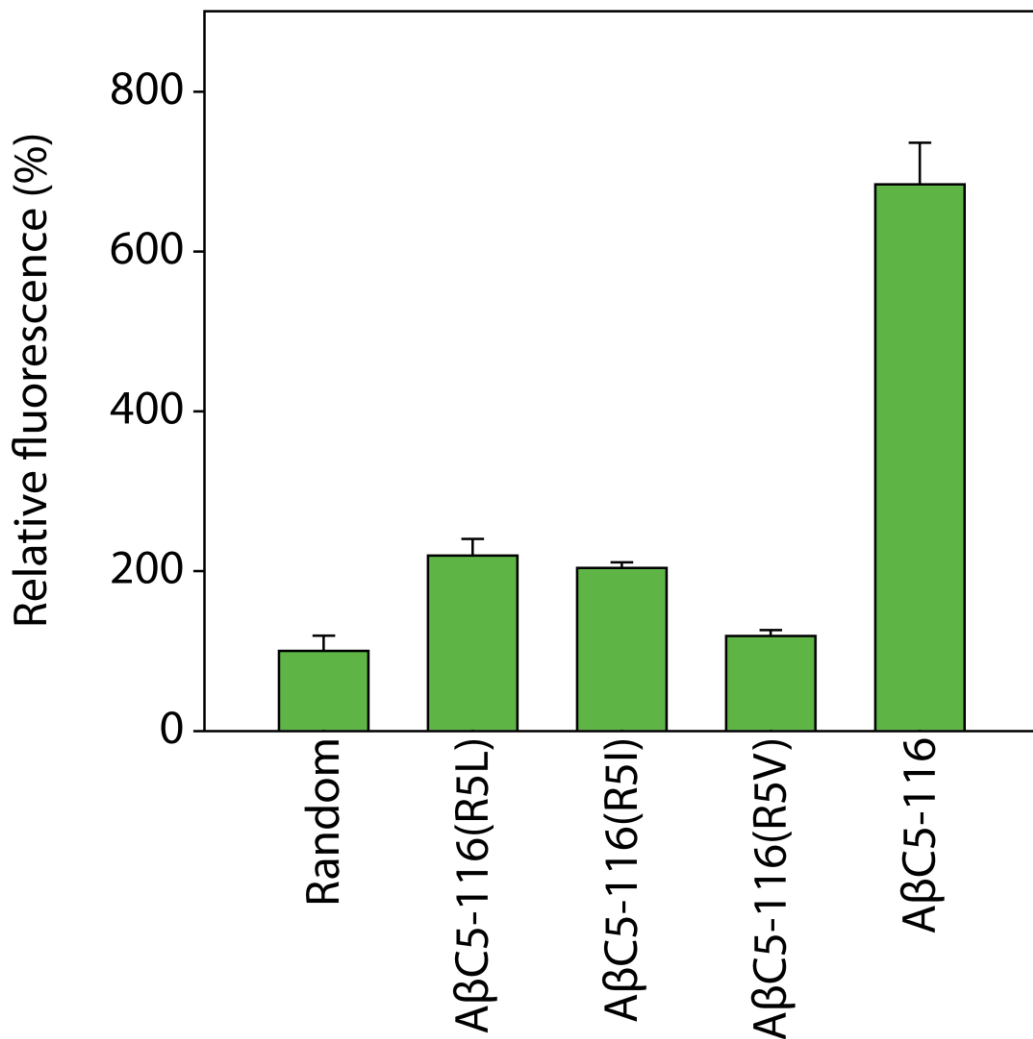


Figure 4.10. Substitutions of Arg5 in AβC5-116. Fluorescence of *E. coli* BL21(DE3) cells co-expressing Aβ₄₂-EGFP along with the cyclic pentapeptide AβC5-116 or the Arg5 variants. Measurements of fluorescence were compared to the “Random” peptide sample, which was arbitrarily set at 100%. Mean values ± sd are reported (*n*=1 independent experiments, performed in triplicate).

4.2.5.1. Family I pentapeptides: Nu1/Arg5

Upon examination of the representative phylogenetic tree (**Figure 4.9**), the complete pentapeptide phylogenetic tree (**Supplementary Figure 1**) and the pentapeptide amino acid composition analysis

(Figure 4.5), Arg5 was shown to be the most common residue after Thr1. Further examination of the complete phylogenetic tree for pentapeptides revealed that Arg5 was combined with all three possible nucleophilic amino acids at position 1, however, with a characteristic preference for Thr1. The characteristics of each combination of Arg5 as well as their similarities are investigated below.

4.2.5.1.1 Family IA pentapeptides: Thr1/Arg5

Thr1 appeared much more frequently than Cys1 or Ser1: 60.87% of pentapeptides contained Thr1, versus 28.36 for Cys1 and 10.77% for Ser1 (Figure 4.6). Moreover, the “as translated” population of Thr1/Arg5-containing pentapeptides consisted of 158 sequences, or 32.7% of all selected pentapeptides. This group of sequences, hereafter termed Family IA, was also the most abundant within the complete pool of sorted peptides (26.12% of all unique peptides).

Having identified a possible sequence motif, in which Thr1 and Arg5 were the most conserved residues it was possible to examine the list of rotated pentapeptides in order to identify any similar sequences. Only one such sequence was identified (AβC5-190: CLRTL, derived from the “as translated” sequence TLCLR), with more than 50 reads. The complete list of cyclo-TXXXR sequences (Family IA) sequences is presented in Table 4.3.

Table 4.3. Peptide sequences that match the Family IA pentapeptide motif. Sequences and read numbers of the selected Family IA pentapeptides are presented, as determined by NGS of the selected pSICLOPPS-NuX₁X₂X₃-X₅ vectors after the second round of bacterial sorting for enhanced Aβ₄₂-EGFP fluorescence.

No.	Peptide name	Amino acid sequence	Number of reads	Reads/Total TXXXR reads (%)	Reads/Total pentapeptide reads (%)	Reads/Total peptide reads (%)
1	AβC5-2	T T Y A R	304,753	16.023	7.506	6.727
2	AβC5-3	T T V D R	214,461	11.276	5.282	4.734
3	AβC5-5	T T T W R	175,510	9.228	4.323	3.874
4	AβC5-7	T T L H R	134,018	7.046	3.301	2.958
5	AβC5-8	T T F A R	96,700	5.084	2.382	2.134
6	AβC5-9	T V L D R	89,669	4.715	2.209	1.979

No.	Peptide name	Amino acid sequence	Number of reads	Reads/Total TXXR reads (%)	Reads/Total pentapeptide reads (%)	Reads/Total peptide reads (%)
7	AβC5-12	T T W A R	65,929	3.466	1.624	1.455
8	AβC5-13	T A L D R	62,792	3.301	1.547	1.386
9	AβC5-15	T A N V R	47,855	2.516	1.179	1.056
10	AβC5-17	T T T A R	40,135	2.110	0.989	0.886
11	AβC5-18	T T I A R	37,150	1.953	0.915	0.820
12	AβC5-19	T V W D R	37,091	1.950	0.914	0.819
13	AβC5-20	T T I S R	37,044	1.948	0.912	0.818
14	AβC5-21	T T W C R	36,295	1.908	0.894	0.801
15	AβC5-22	T V L W R	35,820	1.883	0.882	0.791
16	AβC5-25	T T L A R	28,989	1.524	0.714	0.640
17	AβC5-26	T A W C R	28,391	1.493	0.699	0.627
18	AβC5-27	T T S A R	28,188	1.482	0.694	0.622
19	AβC5-29	T T L E R	27,514	1.447	0.678	0.607
20	AβC5-30	T S T A R	27,456	1.444	0.676	0.606
21	AβC5-35	T V R D R	25,428	1.337	0.626	0.561
22	AβC5-41	T G W A R	21,784	1.145	0.537	0.481
23	AβC5-44	T A W A R	20,807	1.094	0.512	0.459
24	AβC5-45	T T W V R	20,798	1.094	0.512	0.459
25	AβC5-46	T L L W R	19,957	1.049	0.492	0.440
26	AβC5-47	T T I D R	19,735	1.038	0.486	0.436
27	AβC5-50	T A L A R	19,433	1.022	0.479	0.429
28	AβC5-51	T S V D R	19,249	1.012	0.474	0.425
29	AβC5-53	T T V W R	18,669	0.982	0.460	0.412
30	AβC5-66	T T H W R	14,304	0.752	0.352	0.316
31	AβC5-67	T A R D R	14,213	0.747	0.350	0.314
32	AβC5-73	T T R D R	12,894	0.678	0.318	0.285
33	AβC5-80	T S V H R	10,181	0.535	0.251	0.225
34	AβC5-82	T A V W R	9,781	0.514	0.241	0.216
35	AβC5-83	T T G C R	9,362	0.492	0.231	0.207
36	AβC5-89	T A T D R	7,984	0.420	0.197	0.176
37	AβC5-94	T V L F R	7,442	0.391	0.183	0.164
38	AβC5-102	T T Y N R	6,067	0.319	0.149	0.134
39	AβC5-105	T V R W R	5,450	0.287	0.134	0.120
40	AβC5-116	T A F D R	4,243	0.223	0.105	0.094
41	AβC5-117	T T R C R	4,237	0.223	0.104	0.094
42	AβC5-118	T T F W R	4,216	0.222	0.104	0.093
43	AβC5-121	T I K D R	3,970	0.209	0.098	0.088
44	AβC5-123	T T V H R	3,371	0.177	0.083	0.074
45	AβC5-126	T T L L R	3,016	0.159	0.074	0.067
46	AβC5-129	T T L F R	2,630	0.138	0.065	0.058
47	AβC5-130	T A Y H R	2,594	0.136	0.064	0.057
48	AβC5-136	T A L H R	2,026	0.107	0.050	0.045
49	AβC5-139	T T S P R	1,904	0.100	0.047	0.042
50	AβC5-146	T T W S R	1,612	0.085	0.040	0.036
51	AβC5-147	T A M H R	1,611	0.085	0.040	0.036
52	AβC5-155	T S L D R	1,251	0.066	0.031	0.028
53	AβC5-158	T T G A R	1,172	0.062	0.029	0.026
54	AβC5-162	T S V W R	1,094	0.058	0.027	0.024
55	AβC5-173	T T H A R	953	0.050	0.023	0.021
56	AβC5-176	T A G W R	945	0.050	0.023	0.021
57	AβC5-177	T A T A R	925	0.049	0.023	0.020
58	AβC5-184	T V L A R	818	0.043	0.020	0.018
59	AβC5-185	T T F N R	800	0.042	0.020	0.018
60	AβC5-188	T G M R R	768	0.040	0.019	0.017
61	AβC5-189	T T V A R	757	0.040	0.019	0.017
62	AβC5-190	T L C L R	739	0.039	0.018	0.016
63	AβC5-192	T G L A R	720	0.038	0.018	0.016
64	AβC5-198	T S W C R	679	0.036	0.017	0.015
65	AβC5-209	T T R A R	580	0.030	0.014	0.013
66	AβC5-215	T T P W R	524	0.028	0.013	0.012
67	AβC5-218	T V L H R	497	0.026	0.012	0.011
68	AβC5-223	T G L D R	464	0.024	0.011	0.010
69	AβC5-230	T T S D R	442	0.023	0.011	0.010
70	AβC5-239	T T M H R	384	0.020	0.009	0.008
71	AβC5-242	T T S T R	376	0.020	0.009	0.008
72	AβC5-244	T T R V R	366	0.019	0.009	0.008
73	AβC5-245	T T R F R	364	0.019	0.009	0.008
74	AβC5-248	T T T H R	339	0.018	0.008	0.007
75	AβC5-250	T H A W R	334	0.018	0.008	0.007

No.	Peptide name	Amino acid sequence	Number of reads	Reads/Total TXXXR reads (%)	Reads/Total pentapeptide reads (%)	Reads/Total peptide reads (%)
76	AβC5-252	T V I W R	331	0.017	0.008	0.007
77	AβC5-253	T T W F R	327	0.017	0.008	0.007
78	AβC5-255	T T S R R	325	0.017	0.008	0.007
79	AβC5-258	T T S C R	301	0.016	0.007	0.007
80	AβC5-260	T T W T R	295	0.016	0.007	0.007
81	AβC5-262	T T S S R	286	0.015	0.007	0.006
82	AβC5-263	T H L A R	284	0.015	0.007	0.006
83	AβC5-264	T S G A R	282	0.015	0.007	0.006
84	AβC5-266	T T L R R	274	0.014	0.007	0.006
85	AβC5-270	T A T W R	266	0.014	0.007	0.006
86	AβC5-272	T C M W R	254	0.013	0.006	0.006
87	AβC5-275	T A H V R	249	0.013	0.006	0.005
88	AβC5-276	T S W A R	249	0.013	0.006	0.005
89	AβC5-278	T T W L R	241	0.013	0.006	0.005
90	AβC5-291	T T L D R	213	0.011	0.005	0.005
91	AβC5-294	T T P H R	207	0.011	0.005	0.005
92	AβC5-298	T T R G R	201	0.011	0.005	0.004
93	AβC5-299	T T V G R	200	0.011	0.005	0.004
94	AβC5-301	T T T R R	191	0.010	0.005	0.004
95	AβC5-304	T S I N R	182	0.010	0.004	0.004
96	AβC5-305	T T A D R	181	0.010	0.004	0.004
97	AβC5-315	T T S E R	158	0.008	0.004	0.003
98	AβC5-316	T T C A R	157	0.008	0.004	0.003
99	AβC5-317	T T A W R	156	0.008	0.004	0.003
100	AβC5-320	T T V E R	150	0.008	0.004	0.003
101	AβC5-321	T T T F R	148	0.008	0.004	0.003
102	AβC5-323	T A V D R	147	0.008	0.004	0.003
103	AβC5-325	T V W I R	144	0.008	0.004	0.003
104	AβC5-329	T T V R R	141	0.007	0.003	0.003
105	AβC5-333	T H V R R	137	0.007	0.003	0.003
106	AβC5-343	T N L D R	125	0.007	0.003	0.003
107	AβC5-344	T T P G R	125	0.007	0.003	0.003
108	AβC5-348	T T L T R	119	0.006	0.003	0.003
109	AβC5-355	T A T V R	115	0.006	0.003	0.003
110	AβC5-359	T A M W R	110	0.006	0.003	0.002
111	AβC5-361	T T K W R	108	0.006	0.003	0.002
112	AβC5-362	T T W D R	107	0.006	0.003	0.002
113	AβC5-364	T T M A R	106	0.006	0.003	0.002
114	AβC5-365	T T G G R	106	0.006	0.003	0.002
115	AβC5-366	T T M V R	105	0.006	0.003	0.002
116	AβC5-375	T N L A R	97	0.005	0.002	0.002
117	AβC5-376	T I R D R	96	0.005	0.002	0.002
118	AβC5-378	T T T G R	96	0.005	0.002	0.002
119	AβC5-379	T R L G R	95	0.005	0.002	0.002
120	AβC5-381	T T H T R	93	0.005	0.002	0.002
121	AβC5-382	T T I T R	92	0.005	0.002	0.002
122	AβC5-384	T T Y T R	90	0.005	0.002	0.002
123	AβC5-385	T T L Y R	90	0.005	0.002	0.002
124	AβC5-389	T H L D R	89	0.005	0.002	0.002
125	AβC5-391	T L L I R	88	0.005	0.002	0.002
126	AβC5-392	T T C D R	87	0.005	0.002	0.002
127	AβC5-393	T T G R R	87	0.005	0.002	0.002
128	AβC5-394	T T V S R	86	0.005	0.002	0.002
129	AβC5-395	T T Q H R	85	0.004	0.002	0.002
130	AβC5-396	T T T P R	84	0.004	0.002	0.002
131	AβC5-399	T A F A R	82	0.004	0.002	0.002
132	AβC5-405	T T S H R	78	0.004	0.002	0.002
133	AβC5-410	T V L G R	76	0.004	0.002	0.002
134	AβC5-411	T T Q R R	75	0.004	0.002	0.002
135	AβC5-413	T S H A R	74	0.004	0.002	0.002
136	AβC5-415	T T T C R	74	0.004	0.002	0.002
137	AβC5-422	T A W R R	72	0.004	0.002	0.002
138	AβC5-428	T T C G R	69	0.004	0.002	0.002
139	AβC5-434	T T S G R	65	0.003	0.002	0.001
140	AβC5-438	T T T S R	62	0.003	0.002	0.001
141	AβC5-440	T A T G R	61	0.003	0.002	0.001
142	AβC5-441	T A W D R	61	0.003	0.002	0.001
143	AβC5-443	T T H H R	60	0.003	0.001	0.001
144	AβC5-448	T A Y A R	58	0.003	0.001	0.001

No.	Peptide name	Amino acid sequence					Number of reads	Reads/Total TXXXR reads (%)	Reads/Total pentapeptide reads (%)	Reads/Total peptide reads (%)
145	AβC5-449	T	A	N	A	R	58	0.003	0.001	0.001
146	AβC5-450	T	R	D	V	R	58	0.003	0.001	0.001
147	AβC5-452	T	H	V	D	R	58	0.003	0.001	0.001
148	AβC5-453	T	L	F	W	R	57	0.003	0.001	0.001
149	AβC5-459	T	T	A	A	R	55	0.003	0.001	0.001
150	AβC5-463	T	V	V	D	R	54	0.003	0.001	0.001
151	AβC5-464	T	T	P	A	R	54	0.003	0.001	0.001
152	AβC5-469	T	T	I	G	R	53	0.003	0.001	0.001
153	AβC5-472	T	M	Y	A	R	51	0.003	0.001	0.001
154	AβC5-473	T	H	V	A	R	51	0.003	0.001	0.001
155	AβC5-474	T	T	W	P	R	51	0.003	0.001	0.001
156	AβC5-475	T	T	G	D	R	51	0.003	0.001	0.001
157	AβC5-479	T	T	T	V	R	50	0.003	0.001	0.001
158	AβC5-481	T	V	F	G	R	50	0.003	0.001	0.001
159	AβC5-483	T	R	V	G	R	50	0.003	0.001	0.001
		Sum					1,901,945	100	46.847	41.980

In order to reach a more precise characterization of the Family IA motif, an amino acid composition analysis and a new Family IA phylogenetic tree were generated, both of which were based on pentapeptides that combined Thr1 with Arg5. Their examination revealed the following: Thr2 was the most commonly encountered amino acid among Family IA sequences (88 sequences, 55.35%) (**Figure 4.11**). Ala2 was also highly represented with 26 unique sequences (15.72%), making it the second most prevalent amino acid at position 2. Val2 was the final prevalent residue at this position, with 13 sequences (8.18%) (**Figure 4.11**). Position 3 exhibited higher amino acid variability with most amino acids being represented. Leu3 was slightly more prevalent, with 28 sequences (17.61%). Val3 and Trp3 with 18 sequences (11.32%) each were also relatively frequent in position 3 (**Figure 4.11**). However, the increased number of residues that occurred in more than 10 sequences suggested that only Leu3 was indeed prevalent. Moreover, position 3 appears randomized in most subtrees in the phylogenetic analysis suggesting that it should be considered more flexible, with regard to its amino acid composition (**Figure 4.11**). Interestingly, β -branched amino acids (Leu, Val, and Thr) appeared frequently at positions 2 and 3. Such amino acids are known to appear preferentially in β -sheets³³⁷ and their appearance could imply the existence of β -sheet-like interactions between the selected cyclic peptides of this nature and A β . Nevertheless, the participation of β -branched amino acids in β -sheets is contextual. Therefore this hypothesis would require confirmation from biophysical assays for

the examination of β -sheet content. Finally, position 4 was predominantly occupied by Ala (32 sequences, 20.13%), Asp (27 sequences, 16.98%) and Trp (19 sequences, 11.95%) (**Figure 4.11**). Investigation of subtrees in the phylogenetic tree (**Figure 4.11**) revealed that most Family IA peptides contained two or more prevalent amino acids with most combinations appearing at random, in relatively small subtrees. In contrast, Thr2 was the third most conserved residue of Family IA and was very frequently partnered with the most prevalent amino acids for positions 3 and 4. In fact, 20 Family IA sequences combined Thr2 with prevalent amino acids of position 4, suggesting the existence of a possible secondary motif within this family of sequences.

Amino acid composition as determined by:
Sequences

	1	2	3	4	5		1	2	3	4	5	
A		25	4	32		Counts		15.72	2.52	20.13		%
I		2	7	2				1.26	4.40	1.26		
L		4	28	3				2.52	17.61	1.89		
V		13	18	8				8.18	11.32	5.03		
F		0	7	5				0.00	4.40	3.14		
W		0	18	19				0.00	11.32	11.95		
Y		0	6	1				0.00	3.77	0.63		
N		2	2	3				1.26	1.26	1.89		
Q		0	2	0				0.00	1.26	0.00		
C		1	4	7				0.63	2.52	4.40		
M		1	7	0				0.63	4.40	0.00		
S		10	10	5				6.29	6.29	3.14		
T	159	88	16	6			100.00	55.35	10.06	3.77		
D		0	1	27				0.00	0.63	16.98		
E		0	0	3				0.00	0.00	1.89		
R		3	10	9	159			1.89	6.29	5.66	100.00	
H		6	6	13				3.77	3.77	8.18		
K		0	2	0				0.00	1.26	0.00		
P		0	4	3				0.00	2.52	1.89		
G		4	7	13				2.52	4.40	8.18		
Sum	159	159	159	159	159		100.00	100.00	100.00	100.00	100.00	Sum
							1	2	3	4	5	
Positive charge (R, H, K)							0.00	5.66	11.32	13.84	100.00	
Negative charge (D, E)							0.00	0.00	0.63	18.87	0.00	
Aliphatic (A, I, L, V)							0.00	27.67	35.85	28.30	0.00	
Aromatic (F, W, Y)							0.00	0.00	19.50	15.72	0.00	
β -branched (L, V, T)							100.00	66.04	38.99	10.69	0.00	
H-bond donor (Y, N, Q, S, T, R, K)							100.00	66.67	27.04	36.48	0.00	
H-bond acceptor (Y, N, Q, S, T, D, E, H)							100.00	64.78	30.19	15.09	100.00	
Pro content							0.00	0.00	2.52	1.89	0.00	
Gly content							0.00	2.52	4.40	8.18	0.00	
Cys content							0.00	0.63	2.52	4.40	0.00	

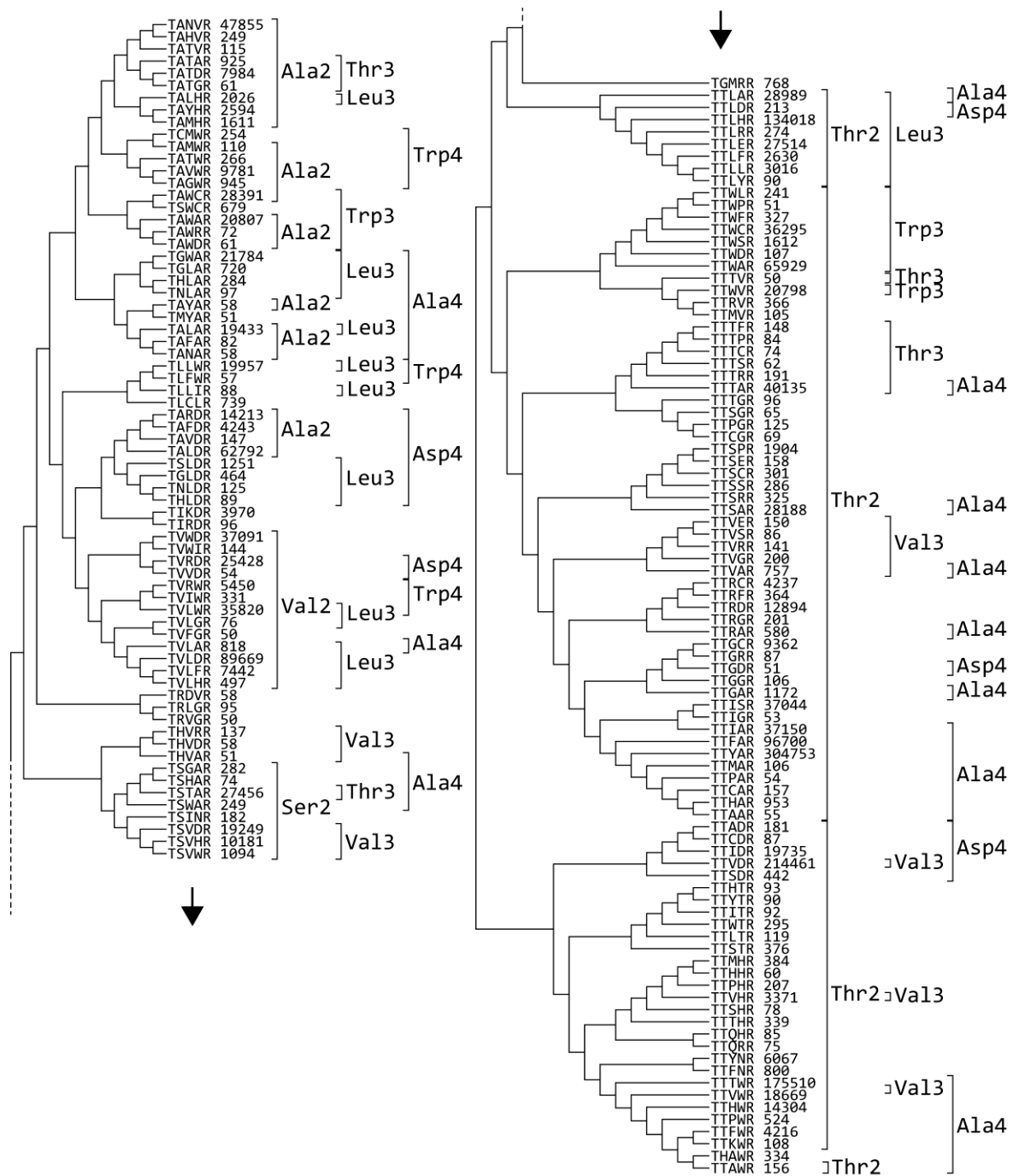


Figure 4.11. Amino acid composition and phylogenetic tree for Family IA. Sequence-based amino acid compositions are presented for the 159 sequences that compose this motif. All rotated and non-rotated sequences with more than 50 reads that contain Thr1 and Arg5 were used in the analysis. The phylogenetic tree was generated using the Maximum Likelihood statistical method. Read numbers are attached next to the peptide sequences. Brackets designate groups of sequences with conserved amino acids in the indicated position.

Additional observations on the amino acid composition in Family IA pentapeptides included the lack of negatively charged residues in positions 2 and 3, as well as the absence of aromatic residues at position 2 (**Figure 4.11**). Consistent with these observations, previously presented data from the semi-saturation mutagenesis analysis of position 2 of AβC5-116 (**Figure 4.3**) have excluded aromatic amino acids from this position. In addition, charged residues were strongly disfavored among Family IA peptides, with the notable exception of Asp4. Similarly, the observation concerning charged amino acids was also consistent with the semi-saturation studies at A2 of AβC5-116 (**Figure 4.3**). Finally, Lys, Gln and the β-sheet-breaking Pro²²⁶, were either absent or insignificantly represented.

In the search for number-of-reads-related biases in the amino acid composition, an analysis based on read numbers for Family IA sequences was performed (**Figure 4.12**).

Amino acid composition as determined by:

Read numbers

	1	2	3	4	5		1	2	3	4	5	
A		224,879	726	698,907		DNA sequence reads		11.82	0.04	36.75		A
I		4,066	94,587	232				0.21	4.97	0.01		I
L		20,841	438,606	3,996				1.10	23.06	0.21		L
V		202,870	278,637	69,596				10.67	14.65	3.66		V
F		0	106,148	10,911				0.00	5.58	0.57		F
W		0	234,933	287,886				0.00	12.35	15.14		W
Y		0	313,613	90				0.00	16.49	0.00		Y
N		222	47,913	7,049				0.01	2.52	0.37		N
Q		0	160	0				0.00	0.01	0.00		Q
C		254	1,052	79,339				0.01	0.06	4.17		C
M		51	3,338	0				0.00	0.18	0.00		M
S		60,697	32,123	39,090				3.19	1.69	2.06		S
T	1,901,945	1,363,173	253,496	1,065			100.00	71.67	13.33	0.06		T
D		0	58	515,155				0.00	0.00	27.09		D
E		0	0	27,822				0.00	0.00	1.46		E
R		203	63,829	2,070	1,901,945			0.01	3.36	0.11	100.00	R
H		953	15,733	155,451				0.05	0.83	8.17		H
K		0	4,078	0				0.00	0.21	0.00		K
P		0	910	2,039				0.00	0.05	0.11		P
G		23,736	12,005	1,247				1.25	0.63	0.07		G
Sum	1,901,945	1,901,945	1,901,945	1,901,945	1,901,945		100.00	100.00	100.00	100.00	100.00	Sum

Figure 4.12. Amino acid composition for Family IA, as determined by read numbers. Read-based amino acid compositions are presented for the 159 sequences that compose this motif. All rotated and non-rotated sequences with more than 50 reads were used.

In this analysis, the number of times each amino acid in a specific position was read was divided by the total number of reads for the respective peptide class, which in this case was pentapeptides. This analysis revealed no alternative amino acid preferences, with Tyr3 and Thr3 being the only exceptions. These exceptions were expected, as these residues were contained in a small number of sequences with very high read numbers.

Overall, the analysis of Family IA revealed the bioactive motif T(T,V,A) Ψ (A,D,W)R where Ψ stands for all non-negatively charged amino acids.

A tetrapeptide variation of Family IA.

The flexibility of positions 3 and 4 hinted that the bioactive sequence length for Family IA could possibly be reduced. Therefore, truncated variants of the Family IA representative A β C5-116 (cyclo-TAFDR) were constructed and tested for their effect on bacterial A β ₄₂-EGFP fluorescence (**Figure 4.13**).

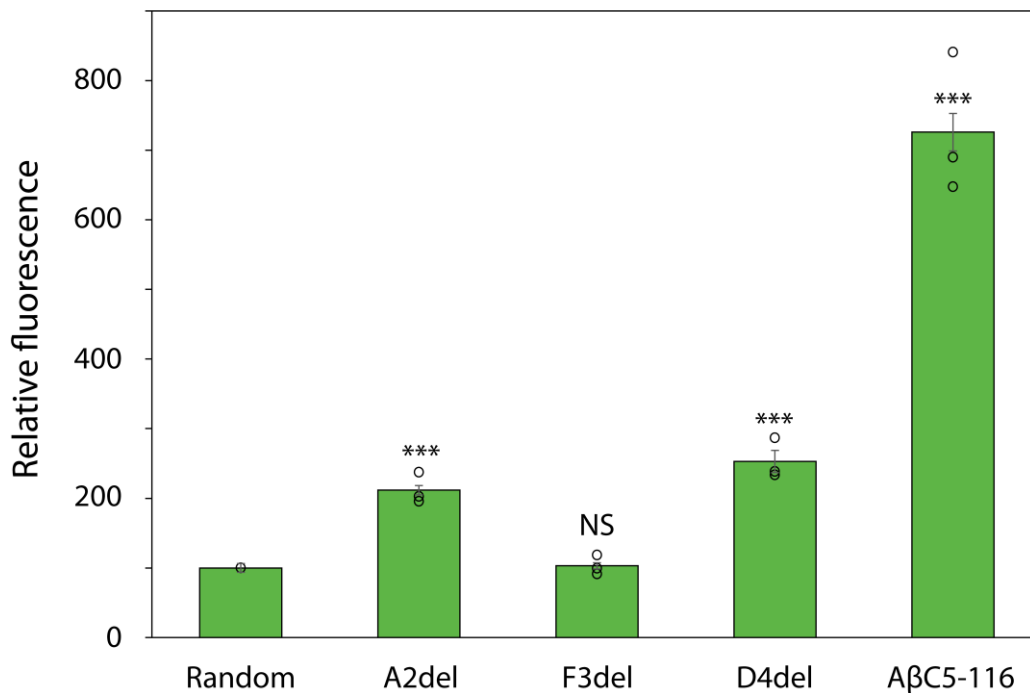


Figure 4.13. Bacterial A β ₄₂-EGFP fluorescence for truncated A β C5-116 variants. Fluorescence of *E. coli* BL21(DE3) cells co-expressing A β ₄₂-EGFP along with the truncated variants of the selected cyclic pentapeptide A β C5-116. The A β ₄₂-EGFP fluorescence of the cell population producing a random cyclic peptide was arbitrarily set at 100%. Mean values \pm s.e.m. are reported ($n=3$ independent experiments, each one performed in triplicate). Statistical significance is indicated for differences compared to the “random peptide” sample. * $P \leq 0.05$, ** $P \leq 0.01$, *** $P \leq 0.001$, NS: not significant.

Tetrapeptide variants lacking Ala2 and Asp4 showed nearly two-fold increases in bacterial A β ₄₂-EGFP fluorescence, suggesting that a tetrapeptide variant of A β C5-116 could be a functional alternative to the original peptide (**Figure 4.13**). Consequently, it was expected that cyclo-TXXR-like sequences would be identified within the tetrapeptide pool. Indeed, six sequences were discovered with characteristic similarity to a truncated form of Family IA pentapeptides. Interestingly, bypassing the cut-off rule of ≥ 50 reads allowed 35 additional similar tetrapeptides to be discovered, shown in

Table 4.4.

Table 4.4. Peptide sequences that match the cyclo-TXXR sequence. Sequences and read numbers of the selected tetrapeptides as determined by NGS of the isolated pSICLOPPS-NuX₁X₂X₃-X₅ vectors after the second round of bacterial sorting for enhanced A β ₄₂-EGFP fluorescence. Double horizontal bars separate sequences with more than 50 reads from those with more than 1 read.

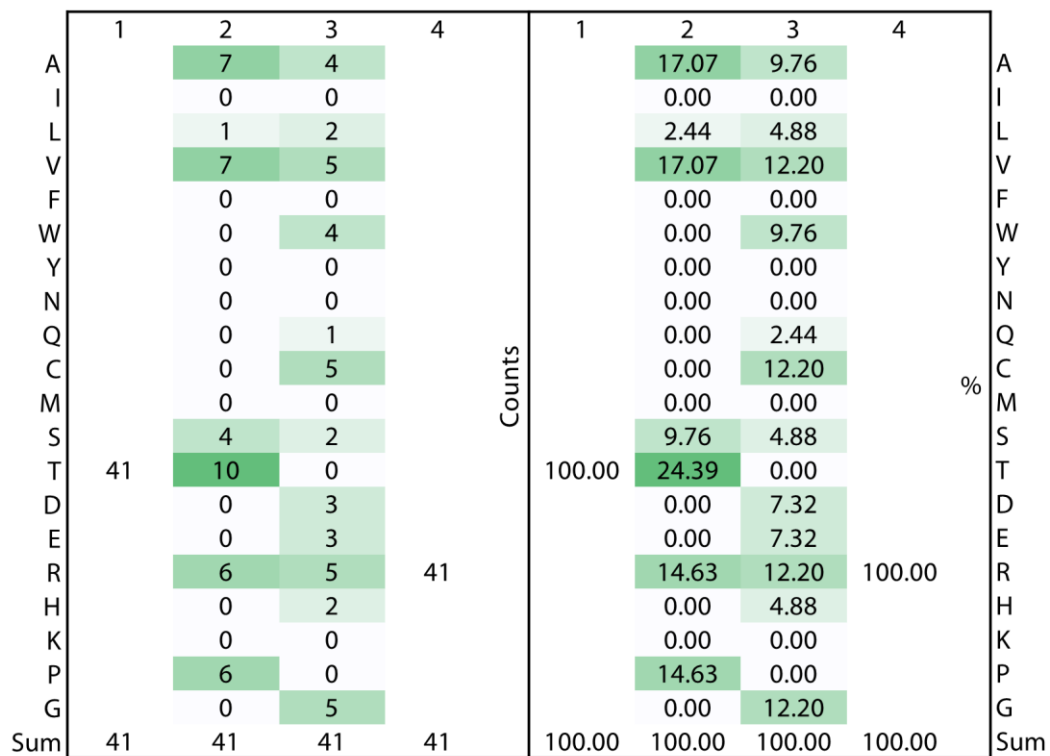
Number	Peptide name	Aminoacid sequence				Number of reads	Reads/Total TXXR reads (%)	Reads/Total tetrapeptide reads (%)
1	A β C4-9	T	T	C	R	258	17.492	1.248
2	A β C4-11	T	T	R	R	248	16.814	1.200
3	A β C4-31	T	T	S	R	67	4.542	0.324
4	A β C4-34	T	R	G	R	63	4.271	0.305
5	A β C4-35	T	T	G	R	61	4.136	0.295
6	A β C4-41	T	R	R	R	51	3.458	0.247
7	A β C4-44	T	T	V	R	45	3.051	0.218
8	A β C4-48	T	V	C	R	41	2.780	0.198
9	A β C4-49	T	V	W	R	40	2.712	0.193
10	A β C4-50	T	A	G	R	39	2.644	0.189
11	A β C4-55	T	P	R	R	36	2.441	0.174
12	A β C4-57	T	T	W	R	34	2.305	0.164
13	A β C4-58	T	R	V	R	34	2.305	0.164
14	A β C4-69	T	T	Q	R	31	2.102	0.150
15	A β C4-68	T	T	H	R	31	2.102	0.150
16	A β C4-71	T	T	A	R	30	2.034	0.145
17	A β C4-76	T	R	W	R	26	1.763	0.126
18	A β C4-77	T	A	R	R	26	1.763	0.126
19	A β C4-78	T	A	C	R	25	1.695	0.121
20	A β C4-83	T	P	C	R	23	1.559	0.111
21	A β C4-94	T	P	D	R	19	1.288	0.092
22	A β C4-95	T	R	D	R	18	1.220	0.087
23	A β C4-106	T	V	E	R	16	1.085	0.077
24	A β C4-110	T	P	V	R	15	1.017	0.073
25	A β C4-111	T	R	E	R	15	1.017	0.073
26	A β C4-116	T	S	G	R	14	0.949	0.068
27	A β C4-128	T	S	V	R	13	0.881	0.063
28	A β C4-129	T	P	H	R	13	0.881	0.063
29	A β C4-131	T	V	D	R	13	0.881	0.063
30	A β C4-132	T	A	A	R	13	0.881	0.063
31	A β C4-134	T	A	W	R	12	0.814	0.058
32	A β C4-136	T	A	E	R	12	0.814	0.058
33	A β C4-143	T	V	V	R	11	0.746	0.053
34	A β C4-144	T	V	R	R	11	0.746	0.053
35	A β C4-145	T	S	C	R	11	0.746	0.053

Number	Peptide name	Aminoacid sequence				Number of reads	Reads/Total TXXR reads (%)	Reads/Total tetrapeptide reads (%)
36	AβC4-148	T	P	A	R	11	0.746	0.053
37	AβC4-151	T	L	A	R	11	0.746	0.053
38	AβC4-152	T	S	L	R	11	0.746	0.053
39	AβC4-153	T	T	L	R	11	0.746	0.053
40	AβC4-154	T	V	G	R	11	0.746	0.053
41	AβC4-168	T	A	S	R	5	0.339	0.024
Sum						1,475	100.000	7.134

Amino acid composition analysis for the sequences presented in **Table 4.4**, as well as generation of a phylogenetic tree showed that Thr2 was the most prevalent intermediate residue, while position 3 exhibited a higher degree of diversity (**Figure 4.14**).

Amino acid composition as determined by:

Sequences



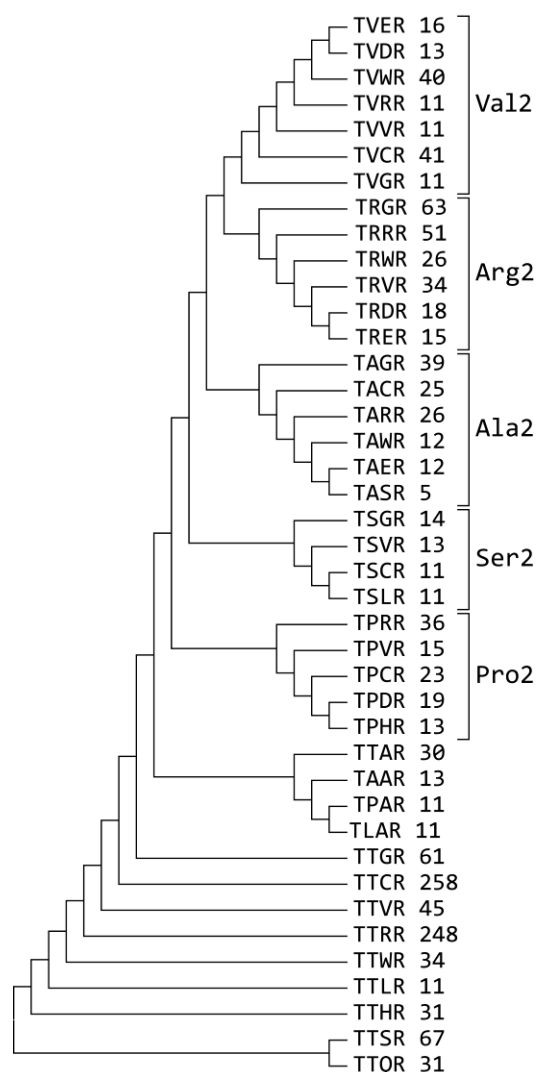


Figure 4.14. Amino acid composition and phylogenetic tree for cyclo-TXXR tetrapeptides. Sequence-based amino acid compositions are presented for 41 tetrapeptides. The phylogenetic tree was generated using the Maximum Likelihood statistical method. Read numbers are attached to the peptide sequences. Brackets designate groups of sequences with conserved amino acids in the indicated position.

Combined, the above results from the truncated variants of A β C5-116 and from the identified sequences within the tetrapeptide pool, indicated that cyclo-TXXR tetrapeptides resembling Family IA peptides are capable of exhibiting bioactivity. As such, these findings suggest that Family IA peptides can tolerate truncated variations of their original sequences.

4.2.5.1.2. Family IB: Ser1/Arg5 pentapeptides

As discussed previously, Arg5 was frequently combined with Cys1 and Ser1 as well. Sequences that contained Ser1 and Arg5 (Family IB) are presented in **Table 4.5**. No matching rotated sequences were identified for this family.

Table 4.5. Pentapeptide sequences of Family IB. Sequences and read numbers of the selected pentapeptides are presented, as determined by NGS of the selected pSICLOPPS-NuX₁X₂X₃-X₅ vectors after the second round of bacterial sorting for enhanced A β ₄₂-EGFP fluorescence. All rotated and non-rotated sequences were considered for the list, although no rotated sequences above 50 reads matched Family IB.

No.	Peptide name	Amino acid sequence					Number of reads	Reads/Total SXXXR reads (%)	Reads/Total pentapeptide reads (%)	Reads/Total peptide reads (%)
1	A β C5-16	S	T	V	W	R	45726	30.433	1.126	1.009
2	A β C5-28	S	T	R	W	R	27745	18.466	0.683	0.612
3	A β C5-39	S	V	L	W	R	23559	15.680	0.580	0.520
4	A β C5-63	S	V	I	W	R	15107	10.054	0.372	0.333
5	A β C5-96	S	S	V	W	R	7185	4.782	0.177	0.159
6	A β C5-99	S	A	V	H	R	6481	4.313	0.160	0.143
7	A β C5-109	S	C	V	W	R	4998	3.326	0.123	0.110
8	A β C5-113	S	V	V	W	R	4900	3.261	0.121	0.108
9	A β C5-120	S	T	L	H	R	4167	2.773	0.103	0.092
10	A β C5-140	S	S	W	A	R	1852	1.233	0.046	0.041
11	A β C5-141	S	C	W	C	R	1754	1.167	0.043	0.039
12	A β C5-145	S	A	V	W	R	1637	1.089	0.040	0.036
13	A β C5-168	S	T	I	N	R	1000	0.666	0.025	0.022
14	A β C5-194	S	A	F	F	R	703	0.468	0.017	0.016
15	A β C5-199	S	A	M	W	R	678	0.451	0.017	0.015
16	A β C5-204	S	V	W	C	R	640	0.426	0.016	0.014
17	A β C5-206	S	A	W	W	R	609	0.405	0.015	0.013
18	A β C5-246	S	A	L	W	R	357	0.238	0.009	0.008
19	A β C5-254	S	T	L	V	R	326	0.217	0.008	0.007
20	A β C5-336	S	H	L	A	R	133	0.089	0.003	0.003
21	A β C5-346	S	H	L	D	R	123	0.082	0.003	0.003
22	A β C5-363	S	T	I	V	R	106	0.071	0.003	0.002
23	A β C5-369	S	N	L	W	R	104	0.069	0.003	0.002
24	A β C5-388	S	H	L	H	R	89	0.059	0.002	0.002
25	A β C5-403	S	T	L	W	R	81	0.054	0.002	0.002
26	A β C5-408	S	T	W	V	R	76	0.051	0.002	0.002
27	A β C5-435	S	R	R	V	R	64	0.043	0.002	0.001
28	A β C5-466	S	S	L	W	R	53	0.035	0.001	0.001
						Sum	150253	100	3.701	3.316

The amino acid composition analysis and the phylogenetic tree for Family IB, revealed preference for similar residues as in Family IA. Specifically, Thr (8 sequences, 28.57%), Ala (6 sequences, 21.43%) were the prevalent amino acids for position2 while Val (4 sequences, 14.29%) was also slightly prevalent. Position 3 was mostly occupied by Leu (10 sequences, 35.71%), followed by Val (6 sequences, 21.43%) and Trp (5 sequences, 17.86%). The most frequently occurring amino acid for this population however,

was Trp in position 4 with 14 sequences (50.00%). Position 4 also showed a very slight preference for Val (4 sequences, 14.29%) (**Figure 4.15**). Despite the fact that the sequences combining Ser1 and Arg5 were few (28), the similarities with Family IA were striking.

Amino acid composition as determined by:

		Sequences											
		1	2	3	4	5	1	2	3	4	5		
A			6	0	2		21.43	0.00	7.14			A	
I			0	3	0		0.00	10.71	0.00			I	
L			0	10	0		0.00	35.71	0.00			L	
V			4	6	4		14.29	21.43	14.29			V	
F			0	1	1		0.00	3.57	3.57			F	
W			0	5	14		0.00	17.86	50.00			W	
Y			0	0	0		0.00	0.00	0.00			Y	
N			1	0	1		3.57	0.00	3.57			N	
Q			0	0	0		0.00	0.00	0.00			Q	
C			2	0	2		7.14	0.00	7.14			C	
M			0	1	0		0.00	3.57	0.00			M	
S	28		3	0	0		10.71	0.00	0.00			S	
T			8	0	0		28.57	0.00	0.00			T	
D			0	0	1		0.00	0.00	3.57			D	
E			0	0	0		0.00	0.00	0.00			E	
R			1	2	0	28	3.57	7.14	0.00	100.00		R	
H			3	0	3		10.71	0.00	10.71			H	
K			0	0	0		0.00	0.00	0.00			K	
P			0	0	0		0.00	0.00	0.00			P	
G			0	0	0		0.00	0.00	0.00			G	
Sum	28	28	28	28	28		100.00	100.00	100.00	100.00	100.00	Sum	
		1	2	3	4	5							
Positive charge (R, H, K)		0.00	14.29	7.14	10.71	100.00							
Negative charge (D, E)		0.00	0.00	0.00	3.57	0.00							
Aliphatic (A, I, L, V)		0.00	35.71	67.86	21.43	0.00							
Aromatic (F, W, Y)		0.00	0.00	21.43	53.57	0.00							
β -branched (L, V, T)		0.00	42.86	57.14	14.29	0.00							
H-bond donor (Y, N, Q, S, T, R, K)		100.00	53.57	0.00	17.86	0.00							
H-bond acceptor (Y, N, Q, S, T, D, E, H)		100.00	46.43	7.14	3.57	100.00							
Pro content		0.00	0.00	0.00	0.00	0.00							
Gly content		0.00	0.00	0.00	0.00	0.00							
Cys content		0.00	7.14	0.00	7.14	0.00							

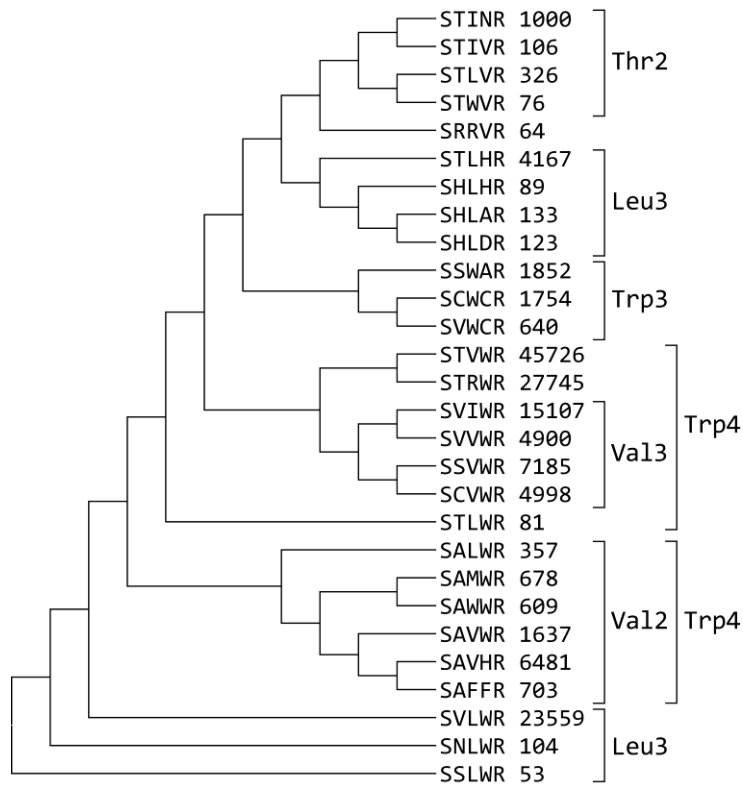


Figure 4.15. Amino acid composition and phylogenetic tree for Family IB as determined by unique sequences. Sequence-based amino acid composition is presented for 28 sequences. All rotated and non-rotated sequences with more than 50 reads were considered in the analysis. The phylogenetic tree was generated using the Maximum Likelihood statistical method. Read numbers are attached to peptide sequences.

Contrary to Family IA, the read-based amino acid composition analysis of Family IB was very slightly different from the sequence-based analysis. Specifically, position 3 showed an increased preference for Arg, which thus far occurs randomly in the middle positions (2-4) (**Figure 4.16**). This Arg occurrence was heavily influenced by cyclo-TSRWR, the second most prevalent Family IB pentapeptide.

Amino acid composition as determined by:

		Read numbers											
		1	2	3	4	5	1	2	3	4	5		
A			10,465	0	1,985		6.96	0.00	1.32			A	
I			0	16,213	0		0.00	10.79	0.00			I	
L			0	28,992	0		0.00	19.30	0.00			L	
V			44,206	70,927	572		29.42	47.21	0.38			V	
F			0	703	703		0.00	0.47	0.47			F	
W			0	4,931	132,739		0.00	3.28	88.34			W	
Y			0	0	0		0.00	0.00	0.00			Y	
N			104	0	1,000		0.07	0.00	0.67			N	
Q			0	0	0		0.00	0.00	0.00			Q	
C			6,752	0	2,394		4.49	0.00	1.59			C	
M			0	678	0		0.00	0.45	0.00			M	
S	150,253		9,090	0	0		100.00	6.05	0.00	0.00		S	
T			79,227	0	0		52.73	0.00	0.00			T	
D			0	0	123		0.00	0.00	0.08			D	
E			0	0	0		0.00	0.00	0.00			E	
R			64	27,809	0	150,253	0.04	18.51	0.00	100.00		R	
H			345	0	10,737		0.23	0.00	7.15			H	
K			0	0	0		0.00	0.00	0.00			K	
P			0	0	0		0.00	0.00	0.00			P	
G			0	0	0		0.00	0.00	0.00			G	
Sum	150,253	150,253	150,253	150,253	150,253		100.00	100.00	100.00	100.00	100.00	Sum	

Figure 4.16. Amino acid composition for Family IB as determined by unique sequence read numbers. The amino acid composition is presented for 28 sequences. All rotated and non-rotated sequences with more than 50 reads were considered in the analysis.

4.2.5.1.3. Family IC: Cys1/Arg5 pentapeptides

Compared to Family IB, pentapeptides combining Cys1 with Arg5 (Family IC) occurred more frequently within the pentapeptide population. Only one rotated peptide matching Family IC was identified and used in the following analysis. All Family IC sequences are presented in **Table 4.6**.

Table 4.6. Pentapeptide sequences of Family IC. Sequences and read numbers of the selected pentapeptides are presented, as determined by NGS of the selected pSICLOPPS-NuX₁X₂X₃-X₅ vectors after the second round of bacterial sorting for enhanced Aβ₄₂-EGFP fluorescence. All rotated and non-rotated sequences have been included in the list.

No.	Peptide name	Amino acid sequence	Number of reads	Reads/Total CXXXR reads (%)	Reads/Total pentapeptide reads (%)	Reads/Total peptide reads (%)
1	AβC5-14	C T I N R	47860	10.628	1.179	1.056
2	AβC5-24	C T W M R	29133	6.469	0.718	0.643
3	AβC5-31	C T F A R	26842	5.961	0.661	0.592
4	AβC5-33	C T T W R	26068	5.789	0.642	0.575
5	AβC5-37	C T S V R	24446	5.429	0.602	0.540

No.	Peptide name	Amino acid sequence	Number of reads	Reads/Total CXXR reads (%)	Reads/Total pentapeptide reads (%)	Reads/Total peptide reads (%)
6	AβC5-38	C T V A R	23694	5.262	0.584	0.523
7	AβC5-40	C S L W R	23446	5.206	0.578	0.518
8	AβC5-43	C T F M R	21532	4.781	0.530	0.475
9	AβC5-48	C A E V R	19588	4.350	0.482	0.432
10	AβC5-49	C V S W R	19498	4.330	0.480	0.430
11	AβC5-52	C V E W R	19212	4.266	0.473	0.424
12	AβC5-64	C T Y C R	14817	3.290	0.365	0.327
13	AβC5-72	C V T W R	13345	2.963	0.329	0.295
14	AβC5-76	C T L W R	12138	2.695	0.299	0.268
15	AβC5-78	C T W E R	11484	2.550	0.283	0.253
16	AβC5-84	C T F H R	9276	2.060	0.228	0.205
17	AβC5-86	C V L H R	8343	1.853	0.205	0.184
18	AβC5-88	C V W W R	8074	1.793	0.199	0.178
19	AβC5-92	C V W V R	7844	1.742	0.193	0.173
20	AβC5-93	C V S H R	7742	1.719	0.191	0.171
21	AβC5-97	C V R V R	7147	1.587	0.176	0.158
22	AβC5-98	C T M W R	7120	1.581	0.175	0.157
23	AβC5-100	C I F W R	6400	1.421	0.158	0.141
24	AβC5-106	C T T A R	5430	1.206	0.134	0.120
25	AβC5-110	C A W A R	4990	1.108	0.123	0.110
26	AβC5-112	C S W M R	4903	1.089	0.121	0.108
27	AβC5-117	C R T T R	4237	0.941	0.104	0.094
28	AβC5-122	C A T A R	3420	0.759	0.084	0.075
29	AβC5-124	C T T M R	3129	0.695	0.077	0.069
30	AβC5-125	C T W V R	3050	0.677	0.075	0.067
31	AβC5-128	C T S A R	2885	0.641	0.071	0.064
32	AβC5-133	C T R M R	2170	0.482	0.053	0.048
33	AβC5-138	C T W L R	1951	0.433	0.048	0.043
34	AβC5-143	C S T W R	1683	0.374	0.041	0.037
35	AβC5-149	C A V H R	1576	0.350	0.039	0.035
36	AβC5-159	C T C H R	1143	0.254	0.028	0.025
37	AβC5-160	C V V W R	1133	0.252	0.028	0.025
38	AβC5-166	C T M A R	1022	0.227	0.025	0.023
39	AβC5-167	C T I H R	1012	0.225	0.025	0.022
40	AβC5-171	C A Q W R	960	0.213	0.024	0.021
41	AβC5-181	C C M W R	846	0.188	0.021	0.019
42	AβC5-183	C T I R R	830	0.184	0.020	0.018
43	AβC5-186	C C A W R	786	0.175	0.019	0.017
44	AβC5-187	C A R A R	773	0.172	0.019	0.017
45	AβC5-202	C T M M R	648	0.144	0.016	0.014
46	AβC5-207	C I G W R	605	0.134	0.015	0.013
47	AβC5-210	C V L L R	576	0.128	0.014	0.013
48	AβC5-213	C V K F R	532	0.118	0.013	0.012
49	AβC5-227	C A A V R	451	0.100	0.011	0.010
50	AβC5-229	C A L V R	444	0.099	0.011	0.010
51	AβC5-234	C C R V R	425	0.094	0.010	0.009
52	AβC5-240	C S W I R	382	0.085	0.009	0.008
53	AβC5-241	C T W T R	376	0.083	0.009	0.008
54	AβC5-259	C V L V R	297	0.066	0.007	0.007
55	AβC5-268	C V W A R	272	0.060	0.007	0.006
56	AβC5-269	C T T C R	270	0.060	0.007	0.006
57	AβC5-277	C S T V R	247	0.055	0.006	0.005
58	AβC5-280	C T A A R	240	0.053	0.006	0.005
59	AβC5-288	C T T V R	220	0.049	0.005	0.005
60	AβC5-307	C S W A R	175	0.039	0.004	0.004
61	AβC5-326	C A A W R	143	0.032	0.004	0.003
62	AβC5-334	C T P Y R	135	0.030	0.003	0.003
63	AβC5-337	C T V V R	132	0.029	0.003	0.003
64	AβC5-340	C V I V R	126	0.028	0.003	0.003
65	AβC5-356	C V R I R	114	0.025	0.003	0.003
66	AβC5-357	C C T W R	113	0.025	0.003	0.002
67	AβC5-374	C A L W R	97	0.022	0.002	0.002
68	AβC5-429	C T T Y R	68	0.015	0.002	0.002
69	AβC5-461	C T F T R	54	0.012	0.001	0.001
70	AβC5-468	C V M V R	53	0.012	0.001	0.001
71	AβC5-471	C T P W R	51	0.011	0.001	0.001
72	AβC5-477	C G A W R	50	0.011	0.001	0.001
73	AβC5-478	C V T F R	50	0.011	0.001	0.001
		Sum	450324	100	11.092	9.940

Phylogenetic tree analysis and amino acid composition analysis revealed that Thr2 was the most prevalent amino acid within Family IC (32 sequences, 43.84%) (Figure 4.17).

Amino acid composition as determined by:
Sequences

	1	2	3	4	5		1	2	3	4	5	
A		10	5	11		Counts	100.00	13.70	6.85	15.07		%
I		2	4	2			2.74	5.48	2.74			
L		0	7	2			0.00	9.59	2.74			
V		17	4	14			23.29	5.48	19.18			
F		0	5	2			0.00	6.85	2.74			
W		0	12	20			0.00	16.44	27.40			
Y		0	1	2			0.00	1.37	2.74			
N		0	0	1			0.00	0.00	1.37			
Q		0	1	0			0.00	1.37	0.00			
C	73	4	1	2			5.48	1.37	2.74			
M		0	5	6			0.00	6.85	8.22			
S		6	4	0			8.22	5.48	0.00			
T		32	13	3			43.84	17.81	4.11			
D		0	0	0			0.00	0.00	0.00			
E		0	2	1			0.00	2.74	1.37			
R		1	5	1	73		1.37	6.85	1.37	100.00		
H		0	0	6			0.00	0.00	8.22			
K		0	1	0			0.00	1.37	0.00			
P		0	2	0			0.00	2.74	0.00			
G		1	1	0			1.37	1.37	0.00			
Sum	73	73	73	73	73	100.00	100.00	100.00	100.00	100.00	100.00	
							1	2	3	4	5	
							0.00	1.37	8.22	9.59	100.00	
							0.00	0.00	2.74	1.37	0.00	
							0.00	39.73	27.40	39.73	0.00	
							0.00	0.00	24.66	32.88	0.00	
							0.00	67.12	32.88	26.03	0.00	%
							0.00	52.05	28.77	17.81	0.00	
							0.00	53.42	34.25	9.59	100.00	
							0.00	0.00	2.74	0.00	0.00	
							0.00	1.37	1.37	0.00	0.00	
							100.00	5.48	1.37	2.74	0.00	

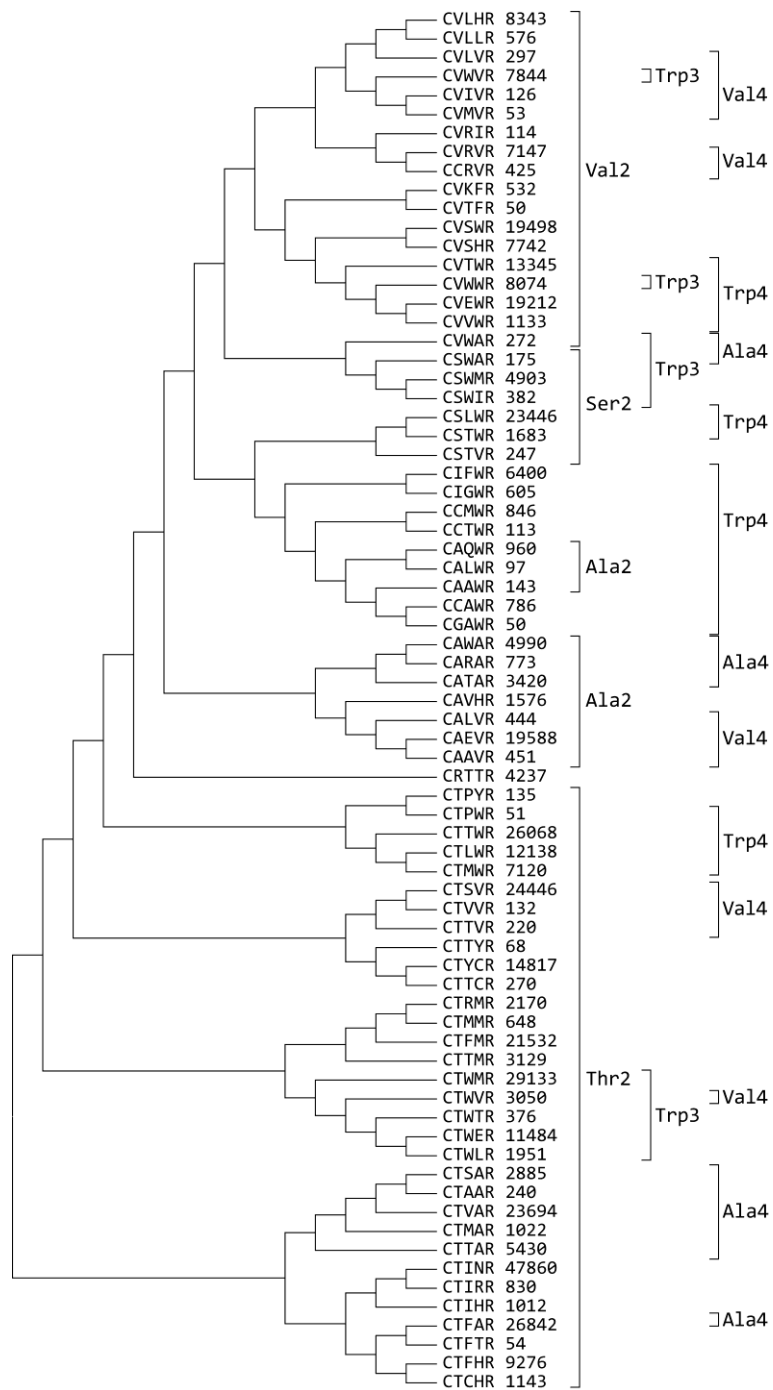


Figure 4.17. Amino acid composition and phylogenetic tree for Family IC. Sequence-based amino acid compositions are presented for 73 sequences. All rotated and non-rotated sequences with more than 50 reads were considered in the analysis. The phylogenetic tree was generated using the Maximum Likelihood statistical method. Read numbers are attached next to the peptide sequences. Brackets designate groups of sequences with conserved amino acids in the indicated position.

Similarly to Family IA, position 2 was also frequently occupied by Val (17 sequences, 23.29%) and Ala (10 sequences, 13.70%). Position 3 was again the one with the highest levels of amino acid diversity, showing a small preference for Thr (13 sequences, 17.81%) and Trp (12 sequences, 16.44%). (Figure 4.17). Position 4 however, presented the most notable composition. Similarly to Family IA, Trp (20 sequences, 27.40%) and Ala (11 sequences, 15.07%) were prevalent. Nevertheless, Asp4, which was very dominant in Family IA, was completely missing and Val4 had taken its place (14 sequences, 19.18%). This divergence was further highlighted by the insignificant presence of Val4 in Family IA. Detailed examination of the phylogenetic tree in Figure 4.17 revealed that the prevalent residues of Family IC were rather randomly combined. Overall, Family IC showed great resemblance to Family IA, with the obvious exception of Cys1, and Asp4.

Amino acid composition as determined by:

Read numbers

	1	2	3	4	5		1	2	3	4	5		
A		32,442	1,670	69,743		DNA sequence reads		7.20	0.37	15.49		A	
I		7,005	49,828	496				1.56	11.06	0.11		I	
L		0	45,341	2,527				0.00	10.07	0.56		L	
V		94,358	26,535	64,470				20.95	5.89	14.32		V	
F		0	64,104	582				0.00	14.24	0.13		F	
W		0	72,634	141,768				0.00	16.13	31.48		W	
Y		0	14,817	203				0.00	3.29	0.05		Y	
N		0	0	47,860				0.00	0.00	10.63		N	
Q		0	960	0				0.00	0.21	0.00		Q	
C	450,324	2,170	1,143	15,087				100.00	0.48	0.25	3.35		C
M		0	9,689	61,515					0.00	2.15	13.66		M
S		30,836	54,571	0					6.85	12.12	0.00		S
T		279,226	58,280	4,667					62.01	12.94	1.04		T
D		0	0	0					0.00	0.00	0.00		D
E		0	38,800	11,484					0.00	8.62	2.55		E
R		4,237	10,629	830	450,324				0.94	2.36	0.18	100.00	R
H		0	0	29,092					0.00	0.00	6.46		H
K		0	532	0					0.00	0.12	0.00		K
P		0	186	0					0.00	0.04	0.00		P
G		50	605	0					0.01	0.13	0.00		G
Sum	450,324	450,324	450,324	450,324	450,324		100.00	100.00	100.00	100.00	100.00	Sum	

Figure 4.18. Amino acid composition for Family IC as determined by read numbers. Amino acid composition is presented for the 73 sequences that compose Family IC. All rotated and non-rotated sequences with more than 50 reads were considered in the analysis.

Finally, investigation of the read-based amino acid compositions revealed a similar amino acid content, although certain residues (Ser3, Met4) became more prevalent because of the heavy influence of the read number in this type of analysis (**Figure 4.18**).

In total, the three subfamilies that contained Arg5 showed similar preferences in their amino acid compositions, albeit with a divergence in position 4. This observation was consistent with the fact that T1S and T1C substitutions in A β C5-116 (cyclo-TAFDR) was significantly less tolerated compared to the original Thr (**Figure 4.2**), suggesting that the Family I subfamilies IA, IB and IC are indeed closely related but also distinct.

4.2.5.2. Family II: Ile5 pentapeptides

Family I was immediately apparent upon initial investigation of the pentapeptide phylogenetic tree. Other possible families however, required a deeper examination of the pentapeptide population. Observation of the compact phylogenetic tree in **Figure 4.9** allowed the observation of a smaller cluster of peptides containing Ile5. Further examination of the complete phylogenetic tree for the pentapeptide population revealed that more Ile5-containing pentapeptides appeared in small clusters spread throughout the tree (**Supplementary Figure 1**).

In order to gain a more detailed perspective on the pentapeptide population that contained Ile5, all such peptides were identified and an amino acid composition analysis and a phylogenetic tree were generated. Despite the small number of sequences composing this group, a degree of preference was shown for specific residues. Threonine was especially prevalent as it occurred in positions 1 to 4 at high frequencies (32.56% for position 1 and 23.26% for positions 2-4). However, the most prevalent residue for position 1 was Cys (23 sequences, 53.49%), which presented a clear deviation from Family I which favoured Thr1. Position 2 was also occupied by the aliphatic residues Leu (11 sequences,

25.58%) and Val to a lesser extent (7 sequences, 16.28%). Interestingly, Pro2 also appeared in 6 sequences (13.95%). Position 3 also showed preference for Asp, with 7 sequences or 16.28%. Finally, position 4 was also frequently occupied by Ser (9 sequences, 20.93%) (Figure 4.19).

Amino acid composition as determined by:

		Sequences											
		1	2	3	4	5	1	2	3	4	5		
A		0	2	2	2		0.00	4.65	4.65	4.65		A	
I		0	3	0	0	43	0.00	6.98	0.00	0.00	100.00	I	
L		0	11	3	5		0.00	25.58	6.98	11.63		L	
V		0	7	5	0		0.00	16.28	11.63	0.00		V	
F		0	1	0	5		0.00	2.33	0.00	11.63		F	
W		0	0	0	4		0.00	0.00	0.00	9.30		W	
Y		0	0	2	0		0.00	0.00	4.65	0.00		Y	
N		0	0	0	0		0.00	0.00	0.00	0.00		N	
Q		0	0	1	0		0.00	0.00	2.33	0.00		Q	
C		23	1	3	1		53.49	2.33	6.98	2.33		C	
M		0	0	4	1		0.00	0.00	9.30	2.33		M	
S		6	2	0	9		13.95	4.65	0.00	20.93		S	
T		14	10	10	10		32.56	23.26	23.26	23.26		T	
D		0	0	7	1		0.00	0.00	16.28	2.33		D	
E		0	0	3	0		0.00	0.00	6.98	0.00		E	
R		0	0	0	0		0.00	0.00	0.00	0.00		R	
H		0	0	0	5		0.00	0.00	0.00	11.63		H	
K		0	0	0	0		0.00	0.00	0.00	0.00		K	
P		0	6	2	0		0.00	13.95	4.65	0.00		P	
G		0	0	1	0		0.00	0.00	2.33	0.00		G	
Sum		43	43	43	43	43	100.00	100.00	100.00	100.00	100.00	Sum	
							1	2	3	4	5		
	Positive charge (R, H, K)						0.00	0.00	0.00	11.63	0.00		
	Negative charge (D, E)						0.00	0.00	23.26	2.33	0.00		
	Aliphatic (A, I, L, V)						0.00	53.49	23.26	16.28	100.00		
	Aromatic (F, W, Y)						0.00	2.33	4.65	20.93	0.00		
	β -branched (L, V, T)						32.56	65.12	41.86	34.88	0.00		%
	H-bond donor (Y, N, Q, S, T, R, K)						46.51	27.91	53.49	58.14	0.00		
	H-bond acceptor (Y, N, Q, S, T, D, E, H)						46.51	27.91	30.23	44.19	0.00		
	Pro content						0.00	13.95	4.65	0.00	0.00		
	Gly content						0.00	0.00	2.33	0.00	0.00		
	Cys content						53.49	2.33	6.98	2.33	0.00		

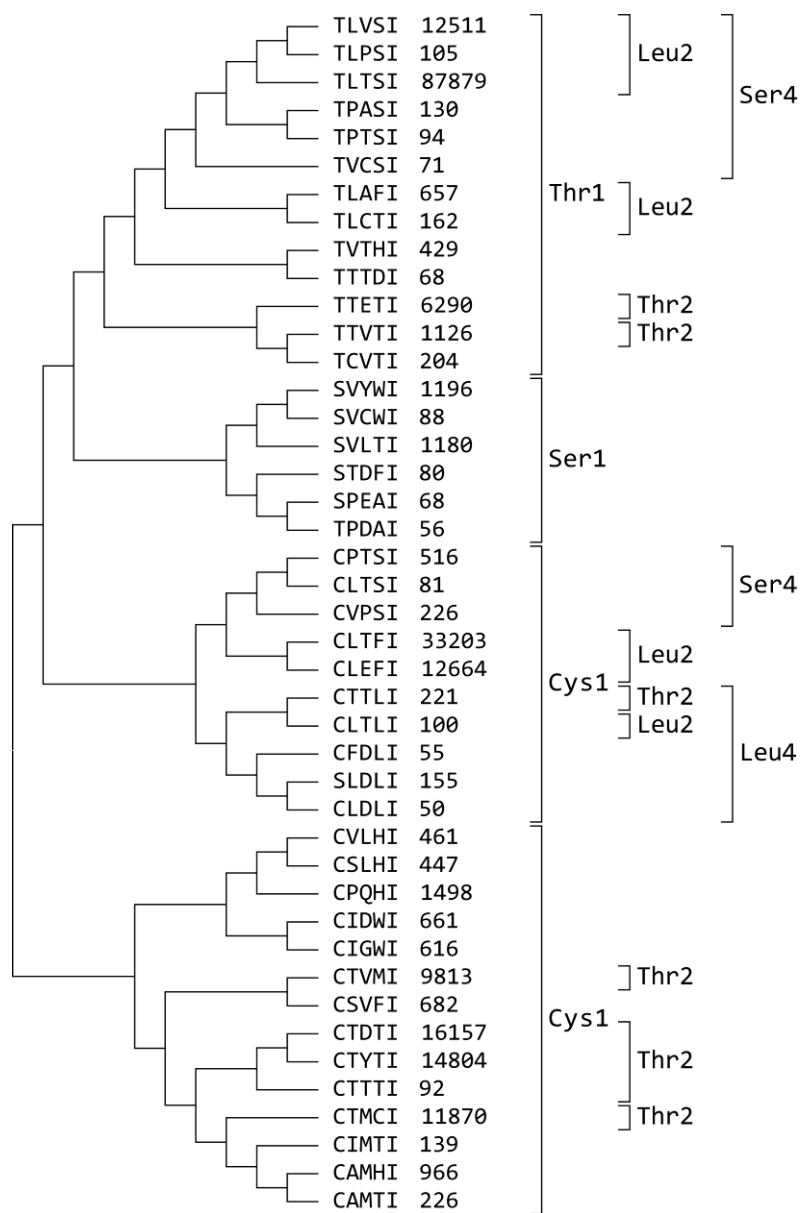


Figure 4.19. Amino acid composition and phylogenetic tree for Family II. Sequence-based amino acid compositions are presented for 43 sequences. Only linear, “as translated” sequences were used in the analysis. The phylogenetic tree was generated using the Maximum Likelihood statistical method. Read numbers are attached to the peptide sequences. Brackets designate groups of sequences with conserved amino acids in the indicated position.

Examination of the phylogenetic tree showed various small subtrees, with the greatest conservation being observed in sequences that contained Cys1. In total, the amino acid composition of Ile5-containing pentapeptides (Family II) was characterized primarily by the prevalence of Cys at position 1, a preference for Thr in positions 2-4 and for Leu in position 2. A Family II bioactive motif can be

represented as (C,T)(L,T,V)(T,D)(T,S)I. The relatively increased diversity of the amino acid composition for Family II suggests that possibly diverging subfamilies could exist within the original family, though these would require a larger Ile5-containing pentapeptide population to characterize.

Finally, the read number-based amino acid composition of Family II pentapeptides revealed similar amino acid prevalence, with the characteristic exception of Phe4 (**Figure 4.20**), owed to the high read numbers of cyclo-CLTFI and cyclo-CLEFI (**Figure 4.19**).

Amino acid composition as determined by:

		Reads					%				
		1	2	3	4	5	1	2	3	4	5
A	0	1,192	787	124		0.00	0.55	0.36	0.06		
I	0	1,416	0	0	218,097	0.00	0.65	0.00	0.00	100.00	
L	0	147,567	2,088	581		0.00	67.66	0.96	0.27		
V	0	3,651	24,336	0		0.00	1.67	11.16	0.00		
F	0	55	0	47,286		0.00	0.03	0.00	21.68		
W	0	0	0	2,561		0.00	0.00	0.00	1.17		
Y	0	0	16,000	0		0.00	0.00	7.34	0.00		
N	0	0	0	0		0.00	0.00	0.00	0.00		
Q	0	0	1,498	0		0.00	0.00	0.69	0.00		
C	105,548	204	321	11,870		48.39	0.09	0.15	5.44		
M	0	0	13,201	9,813		0.00	0.00	6.05	4.50		
S	2,767	1,129	0	101,613		1.27	0.52	0.00	46.59		
T	109,782	60,521	122,683	40,380		50.34	27.75	56.25	18.51		
D	0	0	17,214	68		0.00	0.00	7.89	0.03		
E	0	0	19,022	0		0.00	0.00	8.72	0.00		
R	0	0	0	0		0.00	0.00	0.00	0.00		
H	0	0	0	3,801		0.00	0.00	0.00	1.74		
K	0	0	0	0		0.00	0.00	0.00	0.00		
P	0	2,362	331	0		0.00	1.08	0.15	0.00		
G	0	0	616	0		0.00	0.00	0.28	0.00		
Sum	218,097	218,097	218,097	218,097	218,097	100.00	100.00	100.00	100.00	100.00	

Figure 4.20. Amino acid composition for Family II. Read-based amino acid compositions are presented for the 43 sequences that compose this motif. All rotated and non-rotated sequences with more than 50 reads that contain Ile5 were used in the analysis.

4.2.5.3. Family III: Thr1/Leu5 pentapeptides

One of the more extended subtrees observed in **Figure 4.9** displayed a conservation of Thr1 and Leu5. Indeed, examination of the complete pentapeptide phylogenetic tree revealed a number of subtrees with the same characteristics. Out of the 483 pentapeptides with more than 50 reads, 81 contained Leu5 (16.77%). The amino acid composition analysis and the phylogenetic tree for these peptides revealed that Thr was the major nucleophilic amino acid at position 1 with 58 sequences (71.60%), followed by Cys with 17 sequences (20.99%). Ser1 was insignificantly represented (6 sequences). Position 2 was predominantly occupied by aliphatic amino acids: Val (15 sequences, 18.52%), Ile (15 sequences, 18.52%) and Leu (13 sequences, 16.05%). Thr2 also appeared in 11 sequences (13.58%) with a less prevalent role compared to Family I motifs. Position 3 however revealed the most dramatic change from previous pentapeptide motifs, as Glu was the predominant amino acid with 23 sequences (28.40%), while it was excluded from Family I motifs. Position 4 was occupied mostly by Trp (16 sequences, 19.75%) or Thr (13 sequences, 16.05%) (**Figure 4.21**). Previously, Thr4 was insignificantly represented in Family I motifs.

Amino acid composition as determined by:

		Sequences											
		1	2	3	4	5	1	2	3	4	5		
A		0	3	1	4		0.00	3.70	1.23	4.94		A	
I		0	15	1	0		0.00	18.52	1.23	0.00		I	
L		0	13	4	4	81	0.00	16.05	4.94	4.94	100.00	L	
V		0	15	7	6		0.00	18.52	8.64	7.41		V	
F		0	0	1	9		0.00	0.00	1.23	11.11		F	
W		0	5	1	16		0.00	6.17	1.23	19.75		W	
Y		0	1	1	2		0.00	1.23	1.23	2.47		Y	
N		0	1	5	1		0.00	1.23	6.17	1.23		N	
Q		0	0	1	1		0.00	0.00	1.23	1.23		Q	
C		17	1	3	2		20.99	1.23	3.70	2.47		C	
M		0	2	3	0		0.00	2.47	3.70	0.00		M	
S		6	2	5	3		7.41	2.47	6.17	3.70		S	
T		58	11	5	13		71.60	13.58	6.17	16.05		T	
D		0	0	8	3		0.00	0.00	9.88	3.70		D	
E		0	0	23	2		0.00	0.00	28.40	2.47		E	
R		0	2	4	6		0.00	2.47	4.94	7.41		R	
H		0	3	0	7		0.00	3.70	0.00	8.64		H	
K		0	0	0	0		0.00	0.00	0.00	0.00		K	
P		0	7	4	0		0.00	8.64	4.94	0.00		P	
G		0	0	4	2		0.00	0.00	4.94	2.47		G	
Sum		81	81	81	81	81	100.00	100.00	100.00	100.00	100.00	Sum	

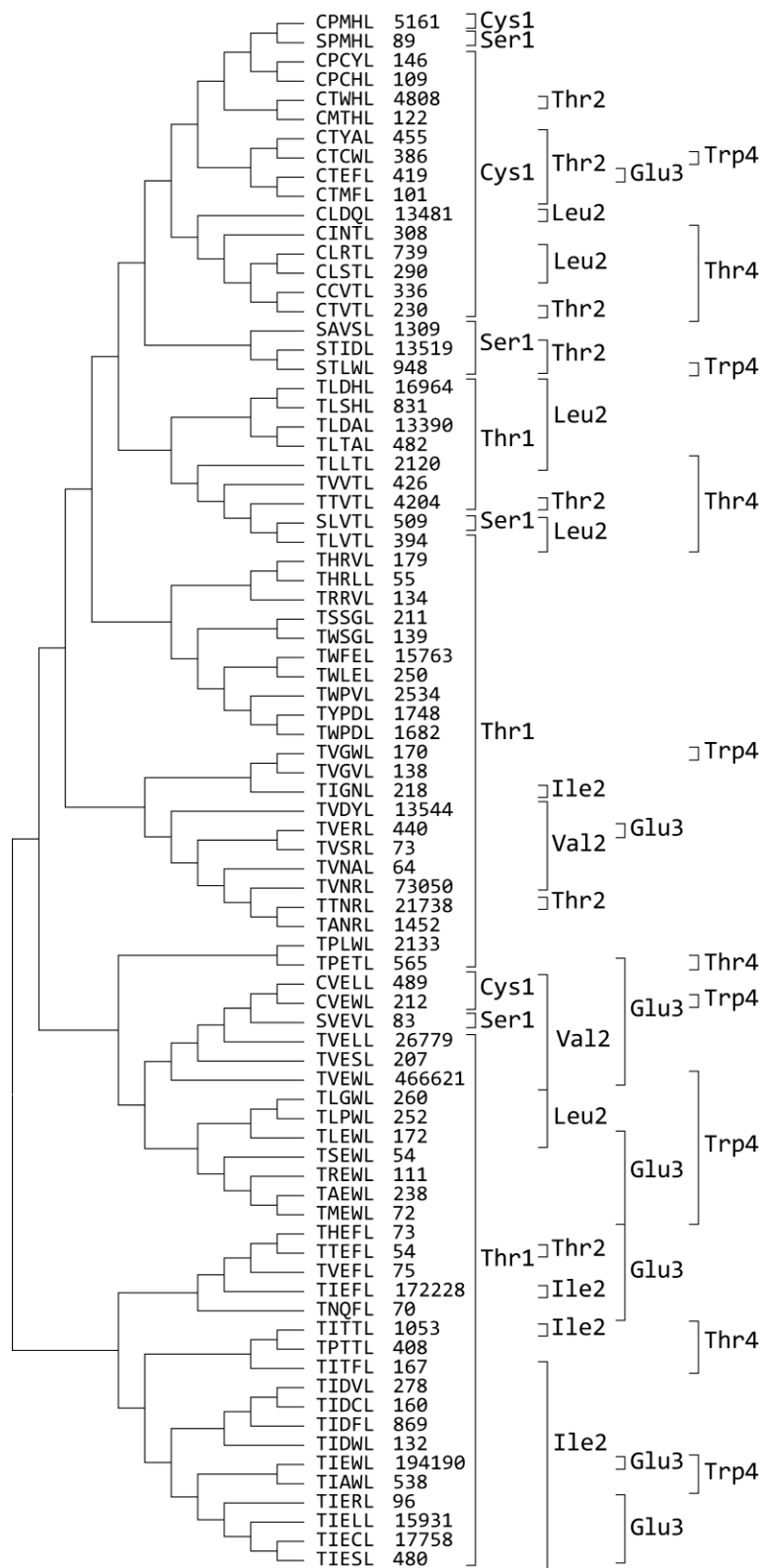


Figure 4.21. Amino acid composition and phylogenetic tree for Thr1/Leu5 pentapeptides. Sequence-based amino acid compositions are presented for 81 sequences. Only linear, “as translated” sequences were used in the analysis. The phylogenetic tree was generated using the Maximum Likelihood statistical method. Read numbers are attached to the peptide sequences. Brackets designate groups of sequences with conserved amino acids in the indicated position.

Ser1 was practically absent from Leu5-containing pentapeptides. Similarly, Cys1 was found only in 18 Leu5-containing sequences (including one rotated peptide sequence). The amino acid composition for such a small number of peptides managed only to highlight Thr2 (7 sequences, 38.89%) and Thr4 (6 sequences, 33.33%) (**Figure 4.22**), while examination of the phylogenetic tree in **Figure 4.21** revealed that Cys1 and Leu5 pentapeptides only formed small subtrees with other prevalent residues, suggesting minimal conservation.

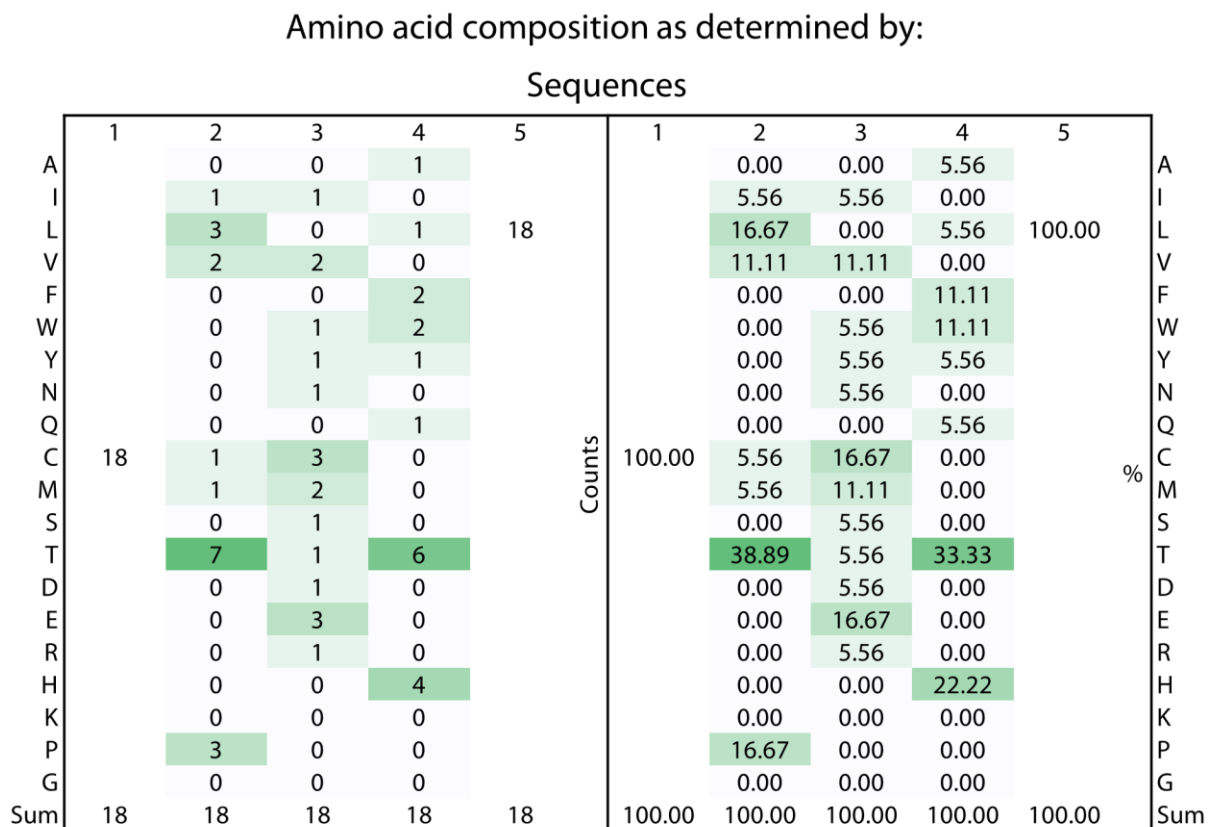


Figure 4.22. Amino acid composition for Cys1/Leu5 pentapeptides as determined by unique sequences. Amino acid compositions are presented for 18 pentapeptides. All rotated and non-rotated sequences were used in the analysis.

In total, it became apparent that Ser1 and Cys1 in combination with Leu5 could not be used to reach a specific Leu5-containing pentapeptide motif, because of limited representation and sequence randomization. Therefore, the effort to characterize Leu5-containing pentapeptides was based on

sequences that combined Thr1 and Leu5 (Family III). The peptides that matched the Family III motif are reported in **Table 4.7**, including 8 rotated sequences.

Table 4.7. Peptide sequences of Family III cyclic pentapeptides. Sequences and read numbers of the selected Family II pentapeptides are presented, as determined by NGS of the selected pSICLOPPS-NuX₁X₂X₃-X₅ vectors after the second round of bacterial sorting for enhanced A β ₄₂-EGFP fluorescence. All rotated and non-rotated sequences have been included in the list.

Number	Peptide name	Amino acid sequence	Number of reads	Reads/Total TXXXL reads (%)	Reads/Total penta-peptide reads (%)	Reads/Total peptide reads (%)
1	A β C5-1	T V E W L	466,621	38.899	11.493	10.299
2	A β C5-4	T I E W L	194,190	16.188	4.783	4.286
3	A β C5-6	T I E F L	172,228	14.357	4.242	3.801
4	A β C5-10	T S I T L	87,879	7.326	2.165	1.940
5	A β C5-11	T V N R L	73,050	6.090	1.799	1.612
6	A β C5-23	T F I C L	33,203	2.768	0.818	0.733
7	A β C5-32	T V E L L	26,779	2.232	0.660	0.591
8	A β C5-42	T T N R L	21,738	1.812	0.535	0.480
9	A β C5-54	T I E C L	17,758	1.480	0.437	0.392
10	A β C5-56	T L D H L	16,964	1.414	0.418	0.374
11	A β C5-60	T I E L L	15,931	1.328	0.392	0.352
12	A β C5-61	T W F E L	15,763	1.314	0.388	0.348
13	A β C5-68	T V D Y L	13,544	1.129	0.334	0.299
14	A β C5-71	T L D A L	13,390	1.116	0.330	0.296
15	A β C5-119	T T V T L	4,204	0.350	0.104	0.093
16	A β C5-131	T W P V L	2,534	0.211	0.062	0.056
17	A β C5-134	T P L W L	2,133	0.178	0.053	0.047
18	A β C5-135	T L L T L	2,120	0.177	0.052	0.047
19	A β C5-135	T L T L L	2,120	0.177	0.052	0.047
20	A β C5-142	T Y P D L	1,748	0.146	0.043	0.039
21	A β C5-144	T W P D L	1,682	0.140	0.041	0.037
22	A β C5-152	T A N R L	1,452	0.121	0.036	0.032
23	A β C5-157	T I S V L	1,180	0.098	0.029	0.026
24	A β C5-165	T I T T L	1,053	0.088	0.026	0.023
25	A β C5-180	T I D F L	869	0.072	0.021	0.019
26	A β C5-182	T L S H L	831	0.069	0.020	0.018
27	A β C5-211	T P E T L	565	0.047	0.014	0.012
28	A β C5-212	T I A W L	538	0.045	0.013	0.012
29	A β C5-220	T L T A L	482	0.040	0.012	0.011
30	A β C5-220	T A L T L	482	0.040	0.012	0.011
31	A β C5-221	T I E S L	480	0.040	0.012	0.011
32	A β C5-231	T V E R L	440	0.037	0.011	0.010
33	A β C5-233	T V V T L	426	0.036	0.010	0.009
34	A β C5-236	T P T T L	408	0.034	0.010	0.009
35	A β C5-237	T L V T L	394	0.033	0.010	0.009
36	A β C5-265	T I D V L	278	0.023	0.007	0.006
37	A β C5-271	T L G W L	260	0.022	0.006	0.006
38	A β C5-273	T L P W L	252	0.021	0.006	0.006
39	A β C5-274	T W L E L	250	0.021	0.006	0.006
40	A β C5-281	T A E W L	238	0.020	0.006	0.005
41	A β C5-289	T I G N L	218	0.018	0.005	0.005
42	A β C5-293	T S S G L	211	0.018	0.005	0.005
43	A β C5-295	T V E S L	207	0.017	0.005	0.005
44	A β C5-306	T H R V L	179	0.015	0.004	0.004
45	A β C5-309	T L E W L	172	0.014	0.004	0.004
46	A β C5-310	T V G W L	170	0.014	0.004	0.004
47	A β C5-311	T I T F L	167	0.014	0.004	0.004
48	A β C5-314	T I D C L	160	0.013	0.004	0.004
49	A β C5-331	T W S G L	139	0.012	0.003	0.003
50	A β C5-332	T V G V L	138	0.012	0.003	0.003
51	A β C5-335	T R R V L	134	0.011	0.003	0.003
52	A β C5-338	T I D W L	132	0.011	0.003	0.003
53	A β C5-348	T R T T L	119	0.010	0.003	0.003

Number	Peptide name	Amino acid sequence	Number of reads	Reads/Total TXXXL reads (%)	Reads/Total penta-peptide reads (%)	Reads/Total peptide reads (%)
54	AβC5-358	T R E W L	111	0.009	0.003	0.002
55	AβC5-372	T L I C L	100	0.008	0.002	0.002
56	AβC5-376	T I E R L	96	0.008	0.002	0.002
57	AβC5-402	T S I C L	81	0.007	0.002	0.002
58	AβC5-411	T V E F L	75	0.006	0.002	0.002
59	AβC5-419	T H E F L	73	0.006	0.002	0.002
60	AβC5-420	T V S R L	73	0.006	0.002	0.002
61	AβC5-422	T M E W L	72	0.006	0.002	0.002
62	AβC5-425	T N Q F L	70	0.006	0.002	0.002
63	AβC5-436	T V N A L	64	0.005	0.002	0.001
64	AβC5-458	T H R L L	55	0.005	0.001	0.001
65	AβC5-460	T T E F L	54	0.005	0.001	0.001
66	AβC5-462	T S E W L	54	0.005	0.001	0.001
Sum			1,199,581	100.000	29.547	26.478

Family III sequences showed preference for Ile (15 sequences, 22.73%), Val (12 sequences, 18.18%) and Leu (11 sequences, 16.67%) at position 2. The third most prevalent residue of Family I motifs, Thr2, was characteristically absent. Position 3 was mostly occupied by Glu (19 sequences, 28.79%), while position 4 was more diverse, though still showing a preference for Trp (13 sequences, 19.70%) and Thr (10 sequences, 15.15%) (Figure 4.23).

Amino acid composition as determined by:

Sequences

	1	2	3	4	5	
A		3	1	3		A
I		15	4	0		I
L		11	4	4	66	L
V		12	3	6		V
F		1	1	7		F
W		5	0	13		W
Y		1	0	1		Y
N		1	4	1		N
Q		0	1	0		Q
C		0	0	5		C
M		1	0	0		M
S		4	5	2		S
T	66	3	6	10		T
D		0	7	2		D
E		0	19	2		E
R		3	3	6		R
H		3	0	2		H
K		0	0	0		K
P		3	4	0		P
G		0	4	2		G
Sum	66	66	66	66	66	Sum

	1	2	3	4	5	
A		4.55	1.52	4.55		A
I		22.73	6.06	0.00		I
L		16.67	6.06	6.06	100.00	L
V		18.18	4.55	9.09		V
F		1.52	1.52	10.61		F
W		7.58	0.00	19.70		W
Y		1.52	0.00	1.52		Y
N		1.52	6.06	1.52		N
Q		0.00	1.52	0.00		Q
C		0.00	0.00	7.58		C
M		1.52	0.00	0.00		M
S		6.06	7.58	3.03		S
T	100.00	4.55	9.09	15.15		T
D		0.00	10.61	3.03		D
E		0.00	28.79	3.03		E
R		4.55	4.55	9.09		R
H		4.55	0.00	3.03		H
K		0.00	0.00	0.00		K
P		4.55	6.06	0.00		P
G		0.00	6.06	3.03		G
Sum	100.00	100.00	100.00	100.00	100.00	Sum

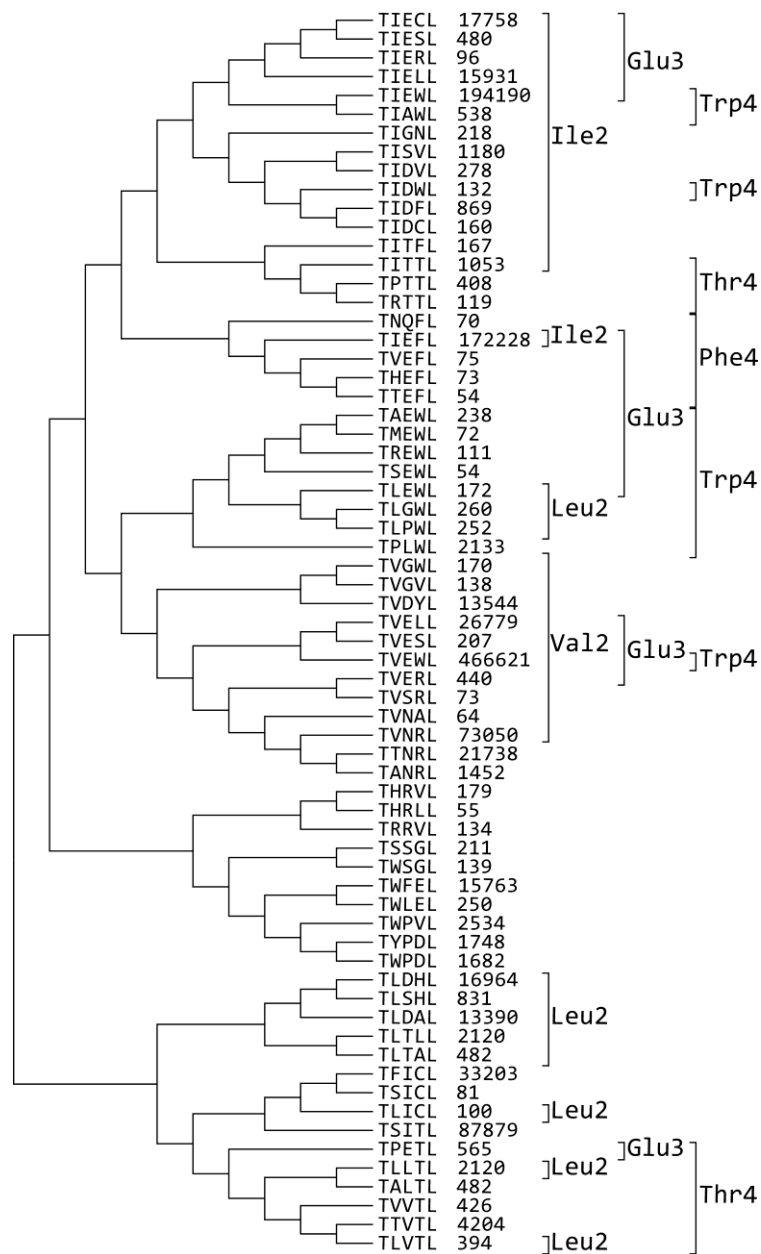


Figure 4.23. Amino acid composition and phylogenetic tree for Family III. Sequence-based amino acid compositions are presented for 66 pentapeptides. Both “as translated” and rotated sequences were used in the analysis. The phylogenetic tree was generated using the Maximum Likelihood statistical method with a Jones-Taylor-Thornton substitution model. Read numbers are attached to the peptide sequences. Brackets designate groups of sequences with conserved amino acids in the indicated position.

In an effort to further investigate the combinations of prevalent residues for the Family III motif, an amino acid composition analysis for Glu3 and Leu5-containing pentapeptides was generated, since these were the two most prevalent residues, excluding the nucleophilic Thr1. As a result, the analysis

revealed an obvious preference for Thr1. Position 2 was occupied most frequently by Val (8 sequences, 34.78%) and Ile (6 sequences, 26.09%). In addition, position 4 showed an enhanced preference for aromatic residues: Trp (8 sequences, 34.78%) and Phe (5 sequences, 21.74%) (**Figure 4.24**).

Amino acid composition as determined by:
Sequences

	1	2	3	4	5		1	2	3	4	5	
A	0	1		0		Counts	0.00	4.35		0.00		%
I	0	6		0			0.00	26.09		0.00		
L	0	1		3	23		0.00	4.35		13.04	100.00	
V	0	8		1			0.00	34.78		4.35		
F	0	0		5			0.00	0.00		21.74		
W	0	0		8			0.00	0.00		34.78		
Y	0	0		0			0.00	0.00		0.00		
N	0	0		0			0.00	0.00		0.00		
Q	0	0		0			0.00	0.00		0.00		
C	3	0		1			13.04	0.00		4.35		
M	0	1		0			0.00	4.35		0.00		
S	1	1		2			4.35	4.35		8.70		
T	19	2		1			82.61	8.70		4.35		
D	0	0		0			0.00	0.00		0.00		
E	0	0	23	0			0.00	0.00	100.00	0.00		
R	0	1		2			0.00	4.35		8.70		
H	0	1		0			0.00	4.35		0.00		
K	0	0		0			0.00	0.00		0.00		
P	0	1		0			0.00	4.35		0.00		
G	0	0		0			0.00	0.00		0.00		
Sum	23	23	23	23	23		100.00	100.00	100.00	100.00	100.00	

Figure 4.24. Amino acid composition for pentapeptide sequences that contain Leu5 and Glu3. Sequence-based amino acid compositions are presented for 23 Glu3/Leu5 pentapeptides. Both “as translated” and rotated sequences were used in the analysis.

The read number-based amino acid composition analysis revealed preference for the same amino acids that were highlighted by the analysis for Glu3/Leu5 pentapeptides (**Figure 4.25**). This result was expected, considering the very high read numbers of Thr1/Glu3/Leu5 pentapeptides (**Table 4.7**). Likewise, the most frequently occurring peptide in the sorted pool was cyclo-TVEWL (AβC5-1).

Amino acid composition as determined by:

		Reads					DNA sequence reads						
		1	2	3	4	5	1	2	3	4	5		
A			2,172	538	13,936			0.18	0.04	1.16		A	
I			405,278	121,263	0			33.78	10.11	0.00		I	
L			37,085	4,985	44,885	1,199,581		3.09	0.42	3.74	100.00	L	
V			581,587	5,024	4,443			48.48	0.42	0.37		V	
F			33,203	15,763	173,536			2.77	1.31	14.47		F	
W			20,368	0	664,943			1.70	0.00	55.43		W	
Y			1,748	0	13,544			0.15	0.00	1.13		Y	
N			70	96,304	218			0.01	8.03	0.02		N	
Q			0	70	0			0.00	0.01	0.00		Q	
C			0	0	51,302			0.00	0.00	4.28		C	
M			72	0	0			0.01	0.00	0.00		M	%
S			88,225	2,434	687			7.35	0.20	0.06		S	
T	1,199,581	25,996	4,349	97,650			100.00	2.17	0.36	8.14		T	
D			0	45,337	3,430			0.00	3.78	0.29		D	
E			0	896,144	16,013			0.00	74.70	1.33		E	
R			364	368	96,849			0.03	0.03	8.07		R	
H			307	0	17,795			0.03	0.00	1.48		H	
K			0	0	0			0.00	0.00	0.00		K	
P			3,106	6,216	0			0.26	0.52	0.00		P	
G			0	786	350			0.00	0.07	0.03		G	
Sum	1,199,581	1,199,581	1,199,581	1,199,581	1,199,581		100.00	100.00	100.00	100.00	100.00	Sum	

Figure 4.25. Amino acid composition for Family III, as determined by sequence reads. Read-based amino acid compositions are presented for all Thr1/Leu5 pentapeptides. Both “as translated” and rotated sequences were used in the analysis.

Examination of NGS results appeared to further support the argument that Arg5 and Leu5 are components of distinct motifs. The amino acid composition shown in **Figure 4.26** indicated that Glu3-containing pentapeptides showed a clear preference for Leu5 rather than Arg5. In addition, this analysis highlighted the prevalence of Ile2 and Val2 as well as Trp4 and Phe4, reinforcing the assumption that the very specific motif cyclo-T(V,I)E(W,F)L was the major component of Family III.

Amino acid composition as determined by:

Sequences

	1	2	3	4	5		1	2	3	4	5	
A	0	2		3	0	Counts	0.00	4.76		7.14	0.00	%
I	0	14		0	3		0.00	33.33		0.00	7.14	
L	0	2		4	23		0.00	4.76		9.52	54.76	
V	0	13		3	7		0.00	30.95		7.14	16.67	
F	0	0		8	1		0.00	0.00		19.05	2.38	
W	0	0		15	0		0.00	0.00		35.71	0.00	
Y	0	0		1	0		0.00	0.00		2.38	0.00	
N	0	0		0	0		0.00	0.00		0.00	0.00	
Q	0	0		0	0		0.00	0.00		0.00	0.00	
C	9	0		1	0		21.43	0.00		2.38	0.00	
M	0	1		0	0		0.00	2.38		0.00	0.00	
S	2	2		3	4		4.76	4.76		7.14	9.52	
T	31	3		2	0		73.81	7.14		4.76	0.00	
D	0	0		0	0		0.00	0.00		0.00	0.00	
E	0	0	42	0	0		0.00	0.00	100.00	0.00	0.00	
R	0	1		2	2		0.00	2.38		4.76	4.76	
H	0	2		0	0		0.00	4.76		0.00	0.00	
K	0	0		0	0		0.00	0.00		0.00	0.00	
P	0	2		0	2		0.00	4.76		0.00	4.76	
G	0	0		0	0		0.00	0.00		0.00	0.00	
Sum	42	42	42	42	42		100.00	100.00	100.00	100.00	100.00	

Figure 4.26. Amino acid composition for pentapeptide sequences that contain Glu3. Sequence-based amino acid compositions are presented for 42 Glu3-containing pentapeptides.

Finally, the distinction between Leu5 and Arg5-containing motifs was further verified by an amino acid composition analysis for all 483 pentapeptides, in which both Thr2 and Arg5-containing peptides had been removed, essentially removing the most characteristic residues of Family I from the analysis (**Figure 4.27**). As a result, the remaining 175 sequences showed preference for residues with that were also prevalent in Family III.

Amino acid composition as determined by:

Sequences

	1	2	3	4	5		1	2	3	4	5	
A	0	10	4	12	3	Counts	0.00	5.71	2.29	6.86	1.71	%
I	0	35	6	0	33		0.00	20.00	3.43	0.00	18.86	
L	0	30	10	15	70		0.00	17.14	5.71	8.57	40.00	
V	0	37	22	29	23		0.00	21.14	12.57	16.57	13.14	
F	0	2	3	12	6		0.00	1.14	1.71	6.86	3.43	
W	0	6	1	27	2		0.00	3.43	0.57	15.43	1.14	
Y	0	1	2	4	0		0.00	0.57	1.14	2.29	0.00	
N	0	1	4	1	1		0.00	0.57	2.29	0.57	0.57	
Q	0	4	4	1	16		0.00	2.29	2.29	0.57	9.14	
C	46	5	6	2	1		26.29	2.86	3.43	1.14	0.57	
M	0	4	7	0	1		0.00	2.29	4.00	0.00	0.57	
S	19	6	12	17	5		10.86	3.43	6.86	9.71	2.86	
T	110		13	23	6		62.86		7.43	13.14	3.43	
D	0	0	19	2	1		0.00	0.00	10.86	1.14	0.57	
E	0	0	37	2	2		0.00	0.00	21.14	1.14	1.14	
R	0	6	10	6			0.00	3.43	5.71	3.43		
H	0	8	0	11	0		0.00	4.57	0.00	6.29	0.00	
K	0	0	0	0	0		0.00	0.00	0.00	0.00	0.00	
P	0	17	8	6	3		0.00	9.71	4.57	3.43	1.71	
G	0	3	7	5	2		0.00	1.71	4.00	2.86	1.14	
Sum	175	175	175	175	175	100.00	100.00	100.00	100.00	100.00		

Figure 4.27. Amino acid composition for all pentapeptide sequences that do not contain either Thr2 or Arg5. Sequence-based amino acid compositions are presented for 175 pentapeptides.

In total, pentapeptides that contained Leu5 showed greater sequence conservation when Thr1 was present, while Glu3 appeared to be central to this family. The more frequent association of Glu3 with Val3, Ile3, Trp4 and Phe4 suggested a more specific representation for Family III, with preference for aliphatic amino acids in position 2 and aromatic amino acids for position 4: T(V,I)E(W,F)L.

4.2.5.4. Family IV: Nu1/Gln5 pentapeptides

A small cluster in the compact pentapeptide phylogenetic tree of **Figure 4.9** contained sequences that ended in Gln5. Gln5 pentapeptides were rather rare in the sorted pool, with only 17 “as translated” sequences (**Figure 4.28**)

Amino acid composition as determined by:

		Sequences											
		1	2	3	4	5	1	2	3	4	5		
A	0	1	0	3		0.00	5.88	0.00	17.65		A		
I	0	1	3	0		0.00	5.88	17.65	0.00		I		
L	0	2	1	4		0.00	11.76	5.88	23.53		L		
V	0	4	5	6		0.00	23.53	29.41	35.29		V		
F	0	1	0	0		0.00	5.88	0.00	0.00		F		
W	0	1	1	0		0.00	5.88	5.88	0.00		W		
Y	0	0	0	0		0.00	0.00	0.00	0.00		Y		
N	0	0	0	0		0.00	0.00	0.00	0.00		N		
Q	0	3	2	0	17	0.00	17.65	11.76	0.00	100.00	Q		
C	8	0	0	0		47.06	0.00	0.00	0.00		C		
M	0	0	0	0		0.00	0.00	0.00	0.00		M		
S	2	0	2	3		11.76	0.00	11.76	17.65		S		
T	7	1	1	0		41.18	5.88	5.88	0.00		T		
D	0	0	0	0		0.00	0.00	0.00	0.00		D		
E	0	0	0	0		0.00	0.00	0.00	0.00		E		
R	0	1	2	1		0.00	5.88	11.76	5.88		R		
H	0	0	0	0		0.00	0.00	0.00	0.00		H		
K	0	0	0	0		0.00	0.00	0.00	0.00		K		
P	0	2	0	0		0.00	11.76	0.00	0.00		P		
G	0	0	0	0		0.00	0.00	0.00	0.00		G		
Sum	17	17	17	17	17	100.00	100.00	100.00	100.00	100.00	Sum		

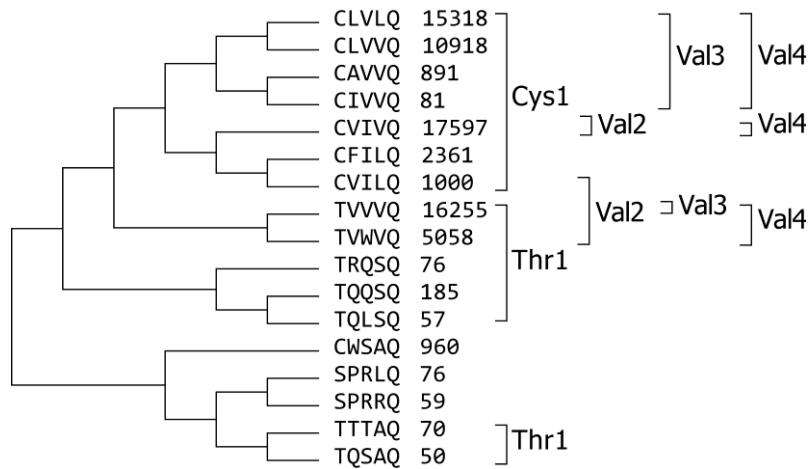


Figure 4.28. Amino acid composition and phylogenetic tree for Family IV. Sequence-based amino acid compositions are presented for 17 sequences. Only linear, “as translated” sequences were used in the analysis. The phylogenetic tree was generated using the Maximum Likelihood statistical method. Read numbers are attached to the peptide sequences. Brackets designate groups of sequences with conserved amino acids in the indicated position.

Examination of the amino acid composition and the phylogenetic tree in **Figure 4.28** showed that position 1 was occupied almost equally by Cys and Thr (8 and 7 sequences respectively). Despite the

limited number of sequences, Family IV exhibits a characteristic preference for Val in the middle positions 2-4 (4, 5 and 6 sequences respectively). An interesting observation derived from the phylogenetic tree referred to the fact that Val in positions 3 and 4 combined more frequently with Cys1 rather than Thr1. Finally, other than Val, residues Gln2, Ile3, and Leu4 also show an above average prevalence for their respective positions, suggesting a more descriptive presentation for Family IV: (C,T)(V,Q)(V,I)(V,L)Q.

4.2.5.5. Family V: Nu1/Val5 pentapeptides

Pentapeptide sequences ending in Val were also rare, though evident in the sorted peptide population. Their occurrence as incorporated sequences in clusters dominated by other families initially suggested that they might present variations to previously described motifs. Nevertheless, upon investigation of all 32 Val5-containing pentapeptide sequences, these exhibited some characteristics unique to this family. Specifically, position 1 of the sequence was predominantly occupied by Thr (20 sequences, 62.50%). Position 2 was occupied by Thr in 9 sequences (28.13%) and Val or Ile (6 sequences each, 18.75%). Position 3 showed preference for Glu (7 sequences, 21.88%) and Val 6 sequences, 18.75%). Finally, position 4 was occupied by Thr (8 sequences, 25.00%), and Trp or Val (6 sequences each, 18.75%) (**Figure 4.29**).

Amino acid composition as determined by:

Sequences

	1	2	3	4	5		1	2	3	4	5	
A	0	2	0	5			0.00	6.25	0.00	15.63		A
I	0	6	0	0			0.00	18.75	0.00	0.00		I
L	0	3	2	0			0.00	9.38	6.25	0.00		L
V	0	6	6	6	32		0.00	18.75	18.75	18.75	100.00	V
F	0	0	1	1			0.00	0.00	3.13	3.13		F
W	0	0	1	6			0.00	0.00	3.13	18.75		W
Y	0	0	1	3			0.00	0.00	3.13	9.38		Y
N	0	0	0	0			0.00	0.00	0.00	0.00		N
Q	0	0	0	0			0.00	0.00	0.00	0.00		Q
C	8	1	1	0		Counts	25.00	3.13	3.13	0.00		C
M	0	0	3	0			0.00	0.00	9.38	0.00		M
S	4	1	0	1			12.50	3.13	0.00	3.13		S
T	20	9	3	8			62.50	28.13	9.38	25.00		T
D	0	0	3	0			0.00	0.00	9.38	0.00		D
E	0	0	7	1			0.00	0.00	21.88	3.13		E
R	0	1	2	0			0.00	3.13	6.25	0.00		R
H	0	2	0	0			0.00	6.25	0.00	0.00		H
K	0	0	0	0			0.00	0.00	0.00	0.00		K
P	0	0	0	1			0.00	0.00	0.00	3.13		P
G	0	1	2	0			0.00	3.13	6.25	0.00		G
Sum	32	32	32	32	32		100.00	100.00	100.00	100.00	100.00	Sum

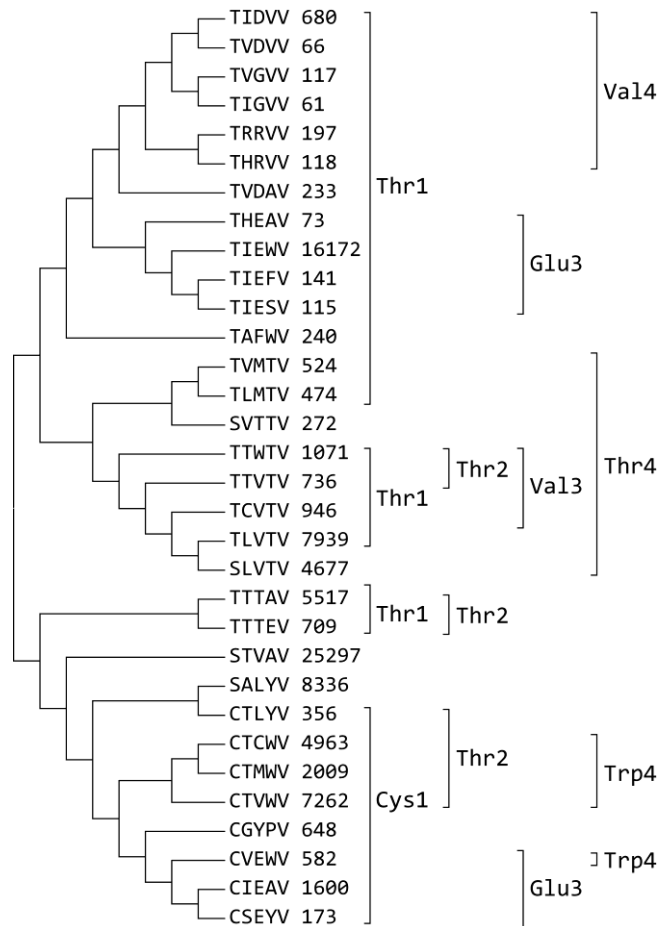


Figure 4.29. Amino acid composition and phylogenetic tree for Family V. Sequence-based amino acid compositions are presented for 32 sequences. Only linear, “as translated” sequences were used in the

analysis. The phylogenetic tree was generated using the Maximum Likelihood statistical method. Read numbers are attached to the peptide sequences. Brackets designate groups of sequences with conserved amino acids in the indicated position.

In total, Family V showed a combination of prevalent residues that differentiated it from the rest of the pentapeptide families. However, despite showing a more defined motif than Family IV, a larger sequence population would allow a more specific characterization of this family. Interestingly, this family shows preference for both Thr2 and Glu3, which are characteristic residues of Family I and Family III peptides. Furthermore, the phylogenetic tree shows that these residues are not combined in Family V, which allows for a degree of overlap between the Family V with Family I or Family III. Despite this, the combination of either Thr2 or Glu3 with Val5 is unique to this group of peptides, which are more specifically described as: T(T,V,I)(E,V)(T,W,V)V.

4.2.5.6. A β C5 34 (cyclo-SASPT) is a mostly unique A β -targeting cyclic pentapeptide

Unlike A β C5-116 (cyclo-TAFDR), which was part of a peptide family of 159 members (Family IA), A β C5-34 (cyclo-SASPT) was a more exclusive sequence. In order to identify all possible peptides with similar sequences to A β C5-34, all pentapeptide sequences were rotated and the complete list of 2,416 pentapeptides with ≥ 50 reads was examined. Previous results from alanine-scanning of the SASPT sequence indicated a limited tolerance for sequence alterations (**Figure 4.1**). Therefore, sequences were searched for amino acid content of similar chemistry to that of A β C5-34: i) Ser1, as its substitution resulted in inactive peptides (**Figure 4.1**) ii) Ser3, Thr3, Cys3, Pro3, Gln3, Asn3 (polar, uncharged) or Thr3, Val3, Leu3 (β -branched) for position 3. ii) Ser4, Thr4, Cys4, Pro4, Gln4, Asn4 (polar, uncharged) for position 4. iii) Ser5, Thr5, Cys5, Pro5, Gln5, Asn5 (polar, uncharged) and Thr5, Val5, Leu5 (β -branched). The sequences presented in **Table 4.8** were the ones most similar to cyclo-SASPT

and appeared more than 50 times during NGS of the sorted library. The significance of Pro4 was taken into special consideration because of its role as a β -sheet breaker²²⁶.

Table 4.8. Peptide sequences that resemble ABC5-34 (cyclo-SASPT). Sequences and read numbers of the selected cyclic pentapeptides resembling A β C5-34, as determined by NGS of the isolated pSICLOPPS-NuX₁X₂X₃-X₅ vectors after the second round of bacterial sorting for enhanced A β ₄₂-EGFP fluorescence.

Number	Peptide name	Amino acid sequence	Number of reads	Reads/Total SASPT-like reads (%)	Reads/Total pentapeptide reads (%)	Reads/Total peptide reads (%)
1	A β C5-34	S A S P T	25,673	97.349	0.632	0.567
2	A β C5-216	S I C P T	516	1.957	0.013	0.011
3	A β C5-380	S I T P T	94	0.356	0.002	0.002
4	A β C5-387	S H S P T	89	0.337	0.002	0.002
	Sum		26,372	100	0.645	0.578

4.2.6. Identification of distinct hexapeptide motifs within the sorted peptide library

The hexapeptide pool was composed of 81 occurring more than 50 times during NGS, presented in **Supplementary Table 4**. In order to investigate the hexapeptide population and identify underlying sequence motifs, all sequences with more than 50 reads were used to generate the phylogenetic tree presented in **Figure 4.30**. Subsequent analysis of the phylogenetic tree and of the amino acid composition of hexapeptides with ≥ 50 reads in **Figure 4.6**, allowed the following initial observations: Position 1 was mostly occupied by Thr (66 sequences, 81.48%). Position 2 was frequently occupied by Pro (21 sequences, 25.93%) and Leu (18 sequences, 22.22%) and position 3 showed prevalence of Thr (19 sequences, 23.46%). Residues Val3 and Ala3 occurred at lower percentages (11 sequences, 13.58% and 10 sequences, 12.35%, respectively). Trp (21 sequences, 25.93%) and Ser (14 sequences, 17.28%) were prevalent for position 4, while position 5 showed a preference for aromatic amino acids: Trp (22 sequences, 27.16%) and Phe (17 sequences, 20.99%). Finally, position 6 was frequently occupied by Asp (25 sequences, 30.86%) and aliphatic residues Leu (12 sequences, 14.81%) and Val (12 sequences,

14.81%) (Figure 4.30). Examination of the phylogenetic tree revealed that Thr1 and Asp6 were very frequently combined, supporting the existence of a dominant hexapeptide motif for Thr1/Asp6 sequences (Figure 4.30).

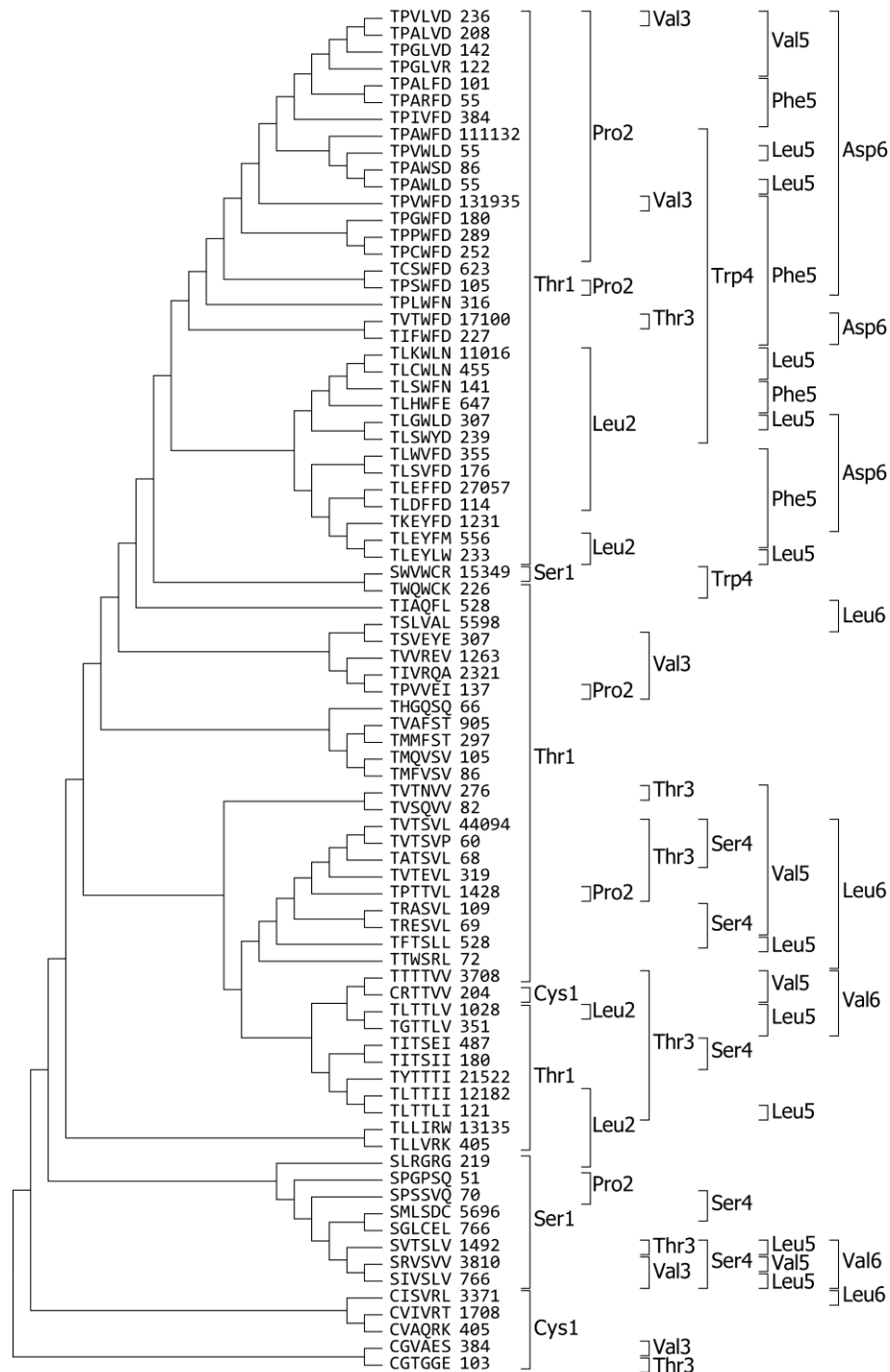


Figure 4.30. Phylogenetic tree for hexapeptide sequences. The phylogenetic tree was generated using the Maximum Likelihood statistical method. Read numbers are attached to the peptide sequences. Brackets designate groups of sequences with conserved amino acids in the indicated position.

4.2.6.1. Family I hexapeptides: Thr1/Asp6

In order to gain a more detailed view of positions 2-5 for Thr1/Asp6 containing hexapeptides (Family I-hexapeptides), an additional analysis was performed on the 25 hexapeptides with Thr1 and Asp6, presented in **Table 4.9**.

Table 4.9. Hexapeptide sequences of Family I-hexapeptides. Sequences and read numbers of the selected cyclo-TXXXD pentapeptides are presented, as determined by NGS of the selected pSICLOPPS-NuX₁X₂X₃-X₅ vectors after the second round of bacterial sorting for enhanced A β ₄₂-EGFP fluorescence. All rotated and non-rotated sequences have been included in the list.

Number	Peptide name	Amino acid sequence						Number of reads	Reads/Total TXXXD reads (%)	Reads/Total hexapeptide reads (%)	Reads/Total peptide reads (%)
1	A β C6-1	T	P	V	W	F	D	131,935	45.084	29.151	2.912
2	A β C6-2	T	P	A	W	F	D	111,132	37.975	24.555	2.453
3	A β C6-4	T	L	E	F	F	D	27,057	9.246	5.978	0.597
4	A β C6-6	T	V	T	W	F	D	17,100	5.843	3.778	0.377
5	A β C6-21	T	K	E	Y	F	D	1,231	0.421	0.272	0.027
6	A β C6-27	T	C	S	W	F	D	623	0.213	0.138	0.014
7	A β C6-36	T	P	I	V	F	D	384	0.131	0.085	0.008
8	A β C6-37	T	L	W	V	F	D	355	0.121	0.078	0.008
9	A β C6-42	T	L	G	W	L	D	307	0.105	0.068	0.007
10	A β C6-44	T	P	P	W	F	D	289	0.099	0.064	0.006
11	A β C6-46	T	P	C	W	F	D	252	0.086	0.056	0.006
12	A β C6-47	T	L	S	W	Y	D	239	0.082	0.053	0.005
13	A β C6-48	T	P	V	L	V	D	236	0.081	0.052	0.005
14	A β C6-50	T	I	F	W	F	D	227	0.078	0.050	0.005
15	A β C6-53	T	P	A	L	V	D	208	0.071	0.046	0.005
16	A β C6-55	T	P	G	W	F	D	180	0.062	0.040	0.004
17	A β C6-57	T	L	S	V	F	D	176	0.060	0.039	0.004
18	A β C6-58	T	P	G	L	V	D	142	0.049	0.031	0.003
19	A β C6-63	T	L	D	F	F	D	114	0.039	0.025	0.003
20	A β C6-65	T	P	S	W	F	D	105	0.036	0.023	0.002
21	A β C6-68	T	P	A	L	F	D	101	0.035	0.022	0.002
22	A β C6-69	T	P	A	W	S	D	86	0.029	0.019	0.002
23	A β C6-78	T	P	A	R	F	D	55	0.019	0.012	0.001
24	A β C6-79	T	P	A	W	L	D	55	0.019	0.012	0.001
25	A β C6-80	T	P	V	W	L	D	55	0.019	0.012	0.001
		Sum						292,644	100	64.660	6.459

As a result, it was shown that Pro was the most prevalent amino acid at position 2 (15 sequences, 60.00%) (Figure 4.31). Leu2, occurred in significantly fewer sequences (6 sequences, 24.00%). Position 3 was less well-defined, with only Ala being relatively noticeable (6 sequences, 24.00%), followed by Ser and Val with 4 and 3 sequences respectively. Position 4 was mostly occupied by Trp (14 sequences, 56.00%) while position 5 was occupied mainly by Phe (17 sequences, 68.00%). Examination of the phylogenetic tree showed that Pro2, Trp4 were the most frequently combined residues (9 sequences). The extended combination of Pro2, Trp4 and Phe5 was also moderately frequent, occurring in 6 sequences (Figure 4.31).

Amino acid composition as determined by:

		Sequences													
		1	2	3	4	5	6	1	2	3	4	5	6		
A			0	6	0	0			0.00	24.00	0.00	0.00		A	
I			1	1	0	0			4.00	4.00	0.00	0.00		I	
L			6	0	4	3			24.00	0.00	16.00	12.00		L	
V			1	3	3	3			4.00	12.00	12.00	12.00		V	
F			0	1	2	17			0.00	4.00	8.00	68.00		F	
W			0	1	14	0			0.00	4.00	56.00	0.00		W	
Y			0	0	1	1			0.00	0.00	4.00	4.00		Y	
N			0	0	0	0			0.00	0.00	0.00	0.00		N	
Q			0	0	0	0			0.00	0.00	0.00	0.00		Q	
C			1	1	0	0			4.00	4.00	0.00	0.00		C	
M			0	0	0	0			0.00	0.00	0.00	0.00		M	
S			0	4	0	1			0.00	16.00	0.00	4.00		S	
T	25		0	1	0	0		100.00	0.00	4.00	0.00	0.00		T	
D			0	1	0	0	25		0.00	4.00	0.00	0.00	100.00	D	
E			0	2	0	0			0.00	8.00	0.00	0.00		E	
R			0	0	1	0			0.00	0.00	4.00	0.00		R	
H			0	0	0	0			0.00	0.00	0.00	0.00		H	
K			1	0	0	0			4.00	0.00	0.00	0.00		K	
P			15	1	0	0			60.00	4.00	0.00	0.00		P	
G			0	3	0	0			0.00	12.00	0.00	0.00		G	
Sum	25	25	25	25	25	25		100.00	100.00	100.00	100.00	100.00	100.00	Sum	

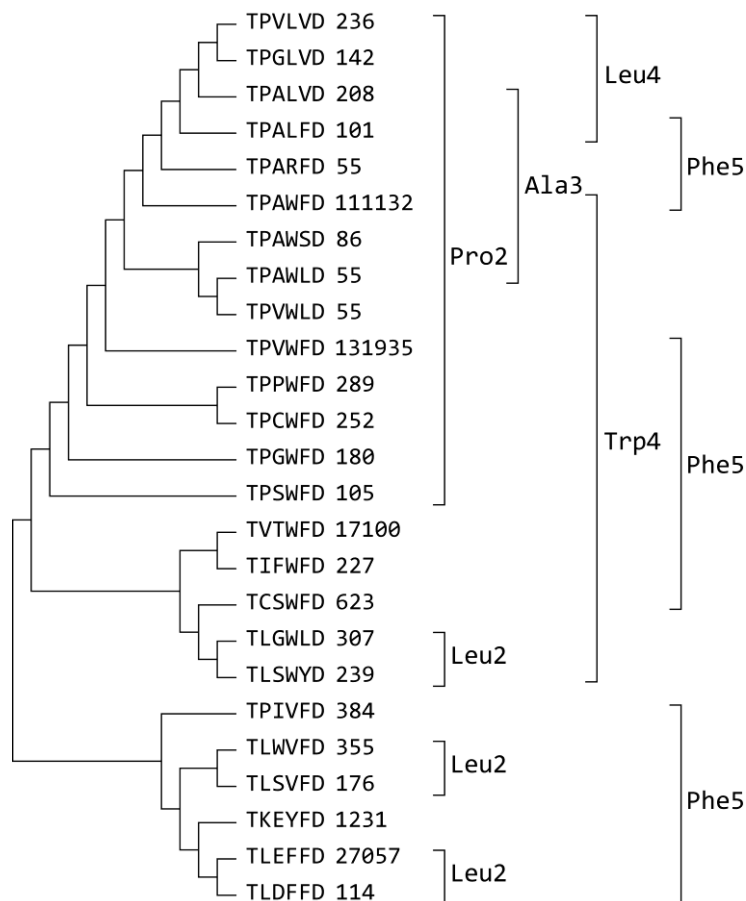


Figure 4.31. Amino acid composition and phylogenetic tree for Family I-hexapeptides. Sequence-based amino acid compositions are presented for 25 hexapeptides with read numbers ≥ 50 . The phylogenetic tree was generated using the Maximum Likelihood statistical method. Read numbers are attached to the peptide sequences. Brackets designate groups of sequences with conserved amino acids in the indicated position.

In total, the amino acid composition of Family I-hexapeptides showed remarkable resemblance to A β C6-1 and A β C6-2 (cyclo-TPVWFD and cyclo-TPAWFD), which exhibited the highest read numbers among hexapeptides. A more specific representation of this motif could be written: TP(A,S,V)WFD. The cyclo-TPXWFD motif (Family I-hexapeptides) was the most specific peptide motif identified within the sorted A β -targeting cyclic peptide pool.

The significance of Asp6 in the Family I-hexapeptides motif was verified by simply removing Asp6 from the amino acid composition analysis, as shown in **Figure 4.32**. As a result, positions 2, 5 and 6 showed

a slight preference for aliphatic amino acids while positions 3 and 4 preferred Thr and Ser (18 and 14 sequences respectively). More importantly, the occurrence of Pro2, Trp4 and Phe5 was notably decreased. Thus, the elimination of Asp6 from the analysis removed the Family I-hexapeptides motif, suggesting that the combination of Asp6 with the other prevalent amino acids of this motif is not coincidental.

Amino acid composition as determined by:

		Sequences						Counts						%						
		1	2	3	4	5	6	1	2	3	4	5	6	1	2	3	4	5	6	
A		0	1	4	1	1	1	0.00	1.79	7.14	1.79	1.79	1.79	0.00	1.79	7.14	1.79	1.79	1.79	
I		0	6	1	1	2	6	0.00	10.71	1.79	1.79	3.57	10.71	0.00	10.71	1.79	1.79	3.57	10.71	
L		0	12	6	1	9	12	0.00	21.43	10.71	1.79	16.07	21.43	0.00	21.43	10.71	1.79	16.07	21.43	
V		0	10	8	7	14	12	0.00	17.86	14.29	12.50	25.00	21.43	0.00	17.86	14.29	12.50	25.00	21.43	
F		0	1	1	2	5	0	0.00	1.79	1.79	3.57	8.93	0.00	0.00	1.79	1.79	3.57	8.93	0.00	
W		0	2	1	7	0	2	0.00	3.57	1.79	12.50	0.00	3.57	0.00	3.57	1.79	12.50	0.00	3.57	
Y		0	1	0	2	1	0	0.00	1.79	0.00	3.57	1.79	0.00	0.00	1.79	0.00	3.57	1.79	0.00	
N		0	0	0	1	0	4	0.00	0.00	0.00	1.79	0.00	7.14	0.00	0.00	1.79	0.00	7.14	0.00	
Q		0	0	2	4	1	3	0.00	0.00	3.57	7.14	1.79	5.36	0.00	0.00	3.57	7.14	1.79	5.36	
C		6	0	1	1	2	1	10.71	0.00	1.79	1.79	3.57	1.79	0.00	0.00	1.79	1.79	3.57	1.79	
M		0	4	1	0	0	1	0.00	7.14	1.79	0.00	0.00	1.79	0.00	0.00	7.14	1.79	0.00	1.79	
S		9	2	4	14	6	1	16.07	3.57	7.14	25.00	10.71	1.79	0.00	0.00	16.07	3.57	7.14	25.00	10.71
T		41	2	18	8	1	3	73.21	3.57	32.14	14.29	1.79	5.36	0.00	0.00	73.21	3.57	32.14	14.29	1.79
D		0	0	0	0	1	0	0.00	0.00	0.00	0.00	1.79	0.00	0.00	0.00	0.00	0.00	1.79	0.00	0.00
E		0	0	3	2	5	3	0.00	0.00	5.36	3.57	8.93	5.36	0.00	0.00	5.36	3.57	8.93	5.36	5.36
R		0	4	1	2	7	2	0.00	7.14	1.79	3.57	12.50	3.57	0.00	0.00	7.14	1.79	3.57	12.50	3.57
H		0	1	1	0	0	0	0.00	1.79	1.79	0.00	0.00	0.00	0.00	0.00	1.79	1.79	0.00	0.00	0.00
K		0	0	1	0	0	3	0.00	0.00	1.79	0.00	0.00	5.36	0.00	0.00	0.00	0.00	1.79	0.00	5.36
P		0	6	0	1	0	1	0.00	10.71	0.00	1.79	0.00	1.79	0.00	0.00	10.71	0.00	1.79	0.00	1.79
G		0	4	3	2	1	1	0.00	7.14	5.36	3.57	1.79	1.79	0.00	0.00	7.14	5.36	3.57	1.79	1.79
Sum		56	56	56	56	56	56	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00

Figure 4.32. Amino acid composition for hexapeptides without Asp6. All 56 non-rotated sequences with more than 50 reads were used in the analysis.

4.2.6.2. Investigation of Arg6-containing hexapeptides

Comparison of amino acid compositions for hexapeptides with ≥ 1 and ≥ 50 reads in **Figure 4.6** revealed that Arg6 occurred more frequently than Asp6 in the analysis for hexapeptides with ≥ 1 reads. As such, it was assumed that Arg6 was more prevalent in hexapeptides of smaller read numbers. Since this was the only notable occasion where the 50 reads cut-off did not include all dominant amino acids, the occurrence of Arg6 was investigated as it could lead to a possible peptide motif that applied to

sequences with low read numbers. Amino acid composition analysis for all hexapeptides that contained Arg6 revealed the following: Thr1 was the most prevalent amino acid (51 sequences, 93%). Thr2 and His2 were also prevalent with 15 sequences (27.27%) each. Position 3 was occupied more frequently by Arg and Leu (12 sequences, 21.82% and 9 sequences 16.36%, respectively). The most frequent amino acids in position 4 were Ser and Pro, both with 10 sequences (18.18%), while position 5 was occupied by Pro in 15 sequences (27.27%) and by Ala in 10 sequences (18.18%). As predicted, this motif corresponded mostly to hexapeptide sequences with low read numbers (≤ 50) (**Figure 4.33**).

Amino acid composition as determined by:

		Sequences						%							
		1	2	3	4	5	6	1	2	3	4	5	6		
A		0	1	6	2	10		0.00	1.82	10.91	3.64	18.18		A	
I		0	4	1	2	0		0.00	7.27	1.82	3.64	0.00		I	
L		0	0	9	5	0		0.00	0.00	16.36	9.09	0.00		L	
V		0	0	6	3	4		0.00	0.00	10.91	5.45	7.27		V	
F		0	0	1	0	0		0.00	0.00	1.82	0.00	0.00		F	
W		0	2	0	6	1		0.00	3.64	0.00	10.91	1.82		W	
Y		0	1	0	0	0		0.00	1.82	0.00	0.00	0.00		Y	
N		0	6	0	0	0		0.00	10.91	0.00	0.00	0.00		N	
Q		0	4	1	0	0		0.00	7.27	1.82	0.00	0.00		Q	
C		1	0	0	4	4		1.82	0.00	0.00	7.27	7.27		C	
M		0	0	0	0	0		0.00	0.00	0.00	0.00	0.00		M	
S		3	1	1	10	3		5.45	1.82	1.82	18.18	5.45		S	
T		51	15	0	3	5		92.73	27.27	0.00	5.45	9.09		T	
D		0	0	2	0	4		0.00	0.00	3.64	0.00	7.27		D	
E		0	0	2	0	0		0.00	0.00	3.64	0.00	0.00		E	
R		0	4	12	3	5	55	0.00	7.27	21.82	5.45	9.09	100.00	R	
H		0	15	5	2	1		0.00	27.27	9.09	3.64	1.82		H	
K		0	0	0	0	0		0.00	0.00	0.00	0.00	0.00		K	
P		0	2	6	10	15		0.00	3.64	10.91	18.18	27.27		P	
G		0	0	3	5	3		0.00	0.00	5.45	9.09	5.45		G	
Sum		55	55	55	55	55		100.00	100.00	100.00	100.00	100.00	100.00	Sum	

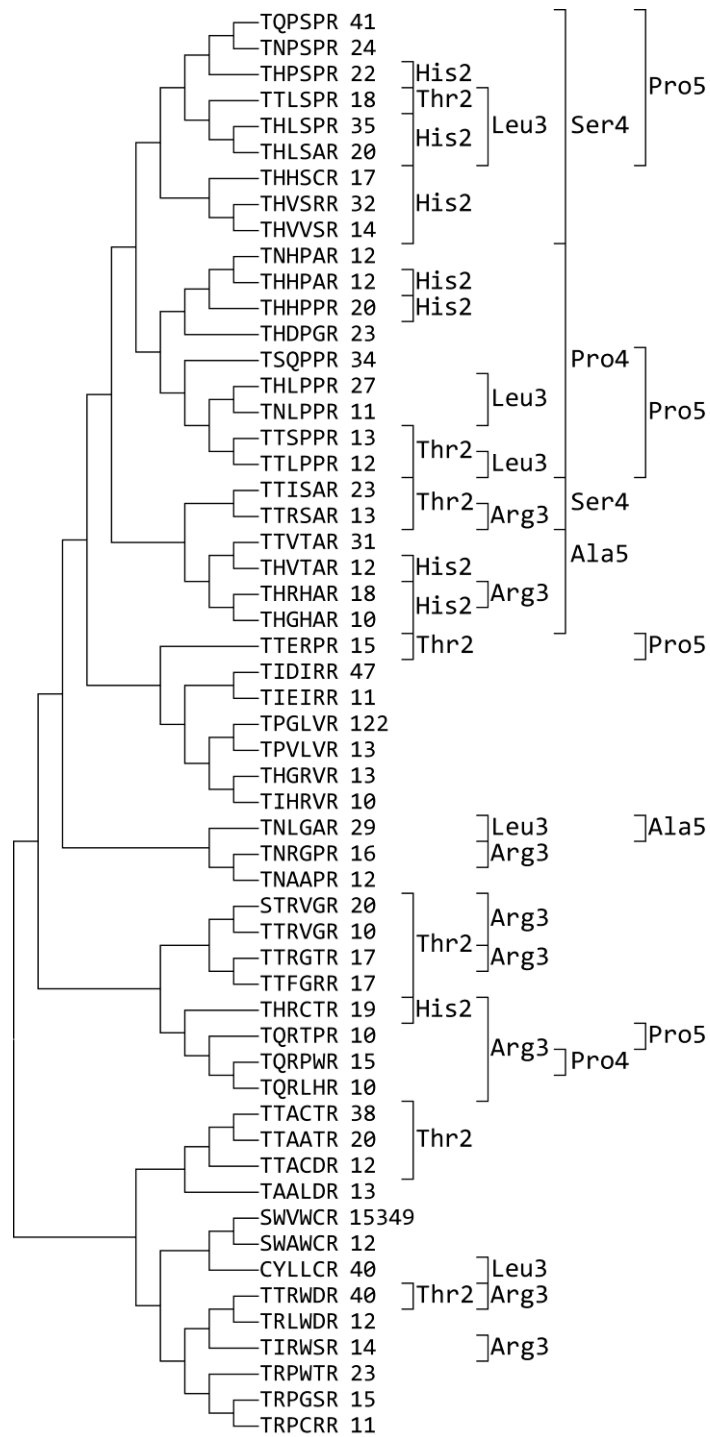


Figure 4.33. Amino acid composition and phylogenetic tree for Arg6-containing hexapeptide sequences. Sequence-based amino acid compositions are presented for 55 hexapeptides with read numbers ≥ 1 . The phylogenetic tree was generated using the Maximum Likelihood statistical method. Read numbers are attached to the peptide sequences. Brackets designate groups of sequences with conserved amino acids in the indicated position.

In addition, the small subtrees of the phylogenetic tree revealed limited conservation of the prevalent residues, which suggested that these were infrequently combined. Residues Ser4 and Pro4 showed the greatest conservation and appeared in various combinations with the other prevalent residues (**Figure 4.33**). A detailed representation of this motif is T(T,H)(R,L)(P,S)(P,A)R, however, the lack of experimental data on those peptides, other than their selection, did not permit further assumptions with regard to their bioactivity.

4.2.7. Investigation of the tetrapeptide pool of sequences

Tetrapeptides formed the smallest peptide group in the sorted library, with only 41 sequences having ≥ 50 reads and 181 sequences in total. The list of tetrapeptides with ≥ 50 reads is presented in **Supplementary Table 5**.

An initial examination of the amino acid composition (**Figure 4.6**) and phylogenetic tree of tetrapeptides (**Figure 4.34**) revealed that the most abundant residue among 41 sequences was Thr1, followed by Thr2 (17 sequences, 41.46%) and Val2 (8 sequences, 19.51%). Position 3 contained Ser, Val and Arg equally (6 sequences, 16.43%). The final position was composed primarily of Arg4 (7 sequences, 17.07%) and Pro4 (6 sequences, 14.63%).

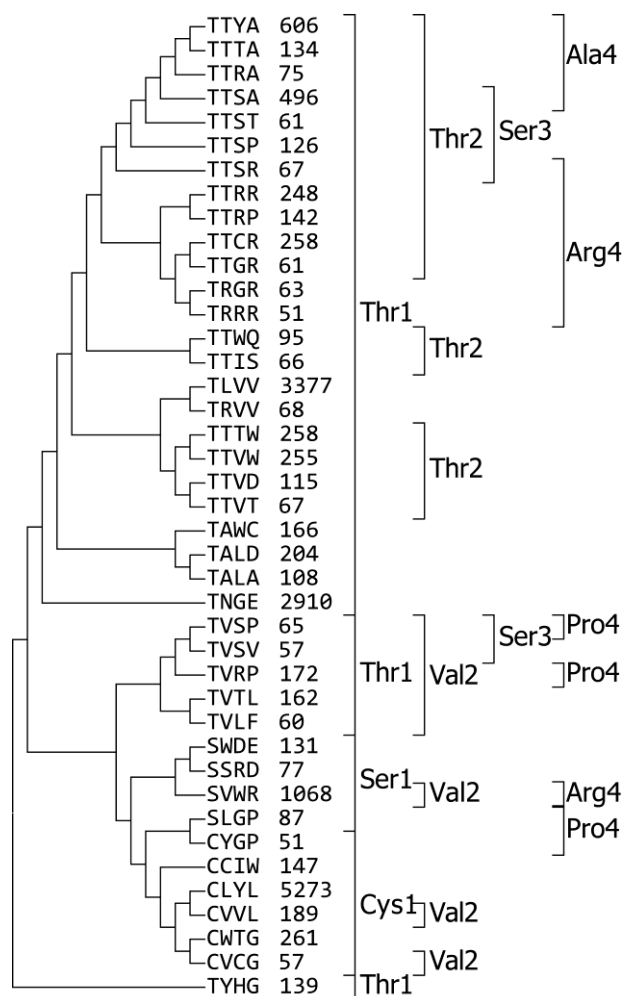


Figure 4.34. Phylogenetic tree for tetrapeptide sequences. The phylogenetic tree was generated using the Maximum Likelihood statistical method. Read numbers are attached to the peptide sequences. Brackets designate groups of sequences with conserved amino acids in the indicated position.

A sequence motif unique to tetrapeptides was difficult to detect. However, the combination of bacterial A β ₄₂-EGFP fluorescence data for the truncated variants of A β C5-116 (**Figure 4.13**) and the discovery of cyclo-TXXR sequences in the sorted pool, hinted towards a smaller bioactive variant of Family IA. Furthermore, the population of tetrapeptides also contained distinct sequences that required individual assessment. Interestingly, the most frequently occurring sequence (A β C4-1 or cyclo-CLYL) was rather unique among tetrapeptides. A β C4-1 was shown to enhance bacterial A β ₄₂-EGFP fluorescence. Another mostly unique tetrapeptide A β C4-2 (cyclo-TLVV) also enhanced

bacterial A β ₄₂-EGFP fluorescence (**Figure 4.35**). Notably, these sequences showed a greater enhancement of bacterial A β ₄₂-EGFP fluorescence compared to the truncated variants of A β C5-116.

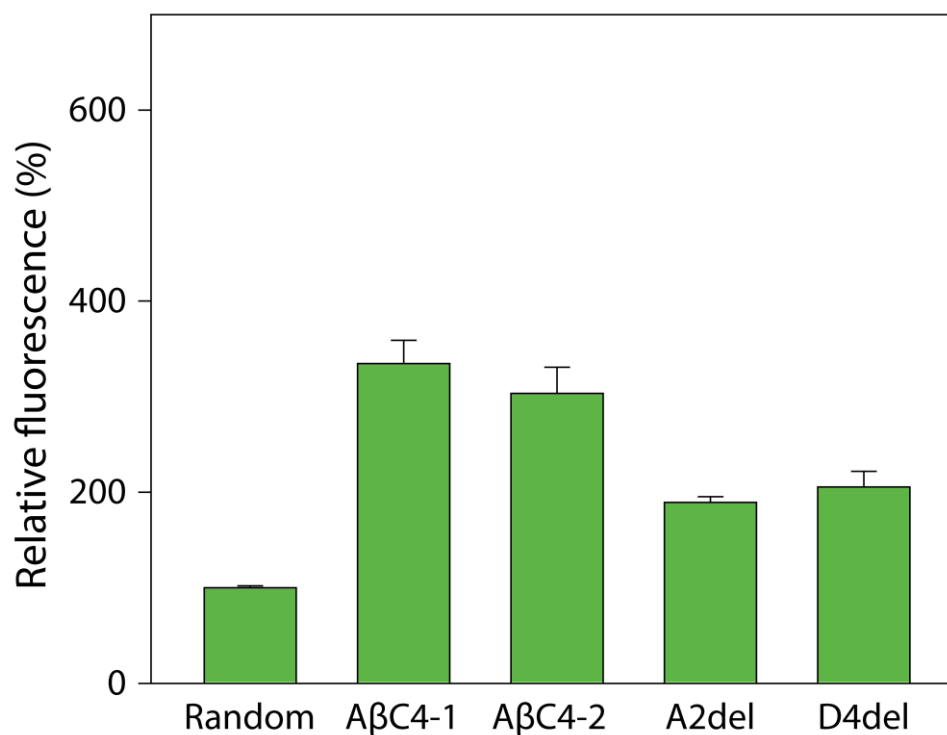


Figure 4.35. Bacterial A β ₄₂-EGFP fluorescence for tetrapeptides identified in the NGS-derived peptide pool. Fluorescence of *E. coli* BL21(DE3) cells co-expressing A β ₄₂-EGFP along with specific cyclic tetrapeptides. Measurements of fluorescence were compared to the “Random” peptide sample, which was arbitrarily set at 100%. Mean values \pm sd are reported ($n=1$ independent experiments, performed in triplicate).

4.3. Remarks on structure-activity analysis

The bacterial system described in chapter 2 displayed a number of advantages, as already discussed. In order to achieve hit deconvolution of the large number of hits generated by this bacterial system it was necessary to implement high-throughput protocols. Deep sequencing is an NGS protocol that nicely meets this requirement, as it can easily be applied in the analysis of the sorted population with minimal manipulation of the sorted sample. The advantages introduced by this protocol include the

simultaneous recognition of all sequences present in the sorted population, as well as the ability to perform a more thorough structure-activity analysis for the selected peptides.

Structure-activity analysis was performed using the deep sequencing results in conjunction with bacterial A β ₄₂-EGFP fluorescence experiments involving peptides (or variants of peptides) identified in the sorted pool. This analysis was essential in the characterization of the validated peptide hits A β C5-34 and A β C5-116. In particular, it revealed that A β C5-34 was intolerant to residue substitutions, verifying its rather unique sequence among the selected A β -targeting cyclic peptides. On the contrary, the A β C5-116 peptide was shown to be a much more flexible sequence. Importantly, deep sequencing enabled the characterization of the broader family in which A β C5-116 belongs, revealing characteristic preferences for specific residues and uncovering the chemical properties inherent of this family. Likewise, several more peptide families were discovered within the sorted pool and subsequently analyzed.

The identification of distinct peptide families was indicative of the bacterial system's ability to select hits with varying chemistries, while the discovery of a unique hit, A β C5-34 (cyclo-SASPT), indicated that additional hits with unique backbones might exist in the sorted peptide pool. Indeed, several sequences did not match any prevalent motifs, while appearing frequently within the sorted pool.

Altogether, the structure-activity analysis served to highlight the bacterial system's ability to isolate distinct hits as well as distinct families of hits, from a highly diverse library of cyclic peptides. The information acquired from this analysis was not only essential in the classification of the derived hits, but also provided useful chemical information that will potentially aid in the derivatization of discovered hits. Alternatively, this information could provide the basis for the design of secondary peptide libraries with altered or additional properties.

Chapter 5. Discussion: remarks and perspectives

The work presented in this thesis describes the development of an integrated bacterial system that allows easy and expedited discovery of cyclic peptides with the ability to interfere with the aggregation of A β and modulate its aggregation-associated cytotoxicity. This approach in hit identification for A β displays a number of advantageous properties, discussed below.

The first and most significant property of the described bacterial system is that it allows straightforward screening of large compound libraries with expanded chemical and structural diversities. The library that was utilized in this work contained $>10^7$ unique cyclic peptides, making it a highly diverse library compared to commercially available chemical libraries whose diversity does not usually surpass 10^5 members. Furthermore, even larger libraries can be utilized with this bacterial system, as indicated by work performed at the laboratory of Enzyme & Synthetic Biotechnology of the National Hellenic Research Foundation, which has investigated cyclic peptide libraries with diversities of more than 2×10^8 members (Delivoria *et al.*, in preparation). Additionally, the host used to express these libraries, i.e. *E. coli*, exhibits very high transformation efficiencies, allowing the production of libraries with up to 10^{10} unique cyclic peptides³³⁸. This diversity is a considerable enhancement over the diversity on offer by yeast-based expression systems (10^5 - 10^8)^{262, 338}. The associated benefit of increasing the diversity of screened libraries is that in this way, the probability of the probability of identifying molecules with the desired properties against the target-protein rises significantly. Furthermore, if the biological nature of the produced library is taken into account, the chances of hit identification against a target protein are further improved²⁸⁸. Moreover, *E. coli* is a very flexible host organism, with regard to the type of biosynthetic libraries it can generate. Existing protocols allow the generation of libraries of side-chain-to-backbone cyclized peptides³³⁹, bicyclic peptides³⁴⁰, lasso peptides³⁴¹, cyclotides^{342, 343} and more³⁴⁴, enabling the screening of biological libraries of varying structures and sizes. Finally, the ability to introduce non-natural amino acids in biosynthetic protocols

by using suppressor tRNA technologies also enhances the capabilities of this approach, by introducing new chemistries in the produced libraries^{345, 346}.

The second important property of the presented bacterial system is that the biosynthesized libraries are screened via a FACS-based screening assay. In this way, this system displays a significant advantage over typical microplate and solid media-based assays, since the screening is performed in a much higher-throughput manner and in significantly shorter times. In this case, biosynthesis and screening of the investigated library are performed as part of a single-step *in vivo* process. This combination displays a comparative advantage over mRNA or phage display methods, which rely on affinity-based selection of hits: the detection of hits by the presented system is based on the ability of the tested molecule to induce a target-protein-linked phenotype. Therefore this system selects hits with bioactivity against the target's misfolding/aggregation²⁹⁶, rather than detect hits based on their binding affinity.

The third property introduced by this integrated bacterial system is especially useful for the identification of effectors for proteins with insufficiently characterized misfolding and aggregation pathways³⁴⁷, such as A β . The presented technology does not require knowledge of oligomerization/aggregation states or the production and purification of specific intermediates, as it screens for molecules that interfere directly with the misfolding of the target protein, or the basic intermolecular interactions that develop in the target's aggregation. This property is partially owed to the bacterial screen, which is known to produce inhibitors for the early stages of A β aggregation²⁶⁰. However, other proteins, whose association to disease has been traced to the loss of native conformation rather than to an inherent potential for aggregation (A β), also stand to benefit from the technology presented here, since many such proteins tend to aggregate because of their acquired misfolding. Thus, misfolding-prone proteins such as variants of p53 and SOD1 associated with various forms of cancer and familial ALS, respectively, constitute suitable targets for the presented bacterial system, as inhibition of their aggregation could also lead to the stabilization of their functionality³⁴⁸.

³⁴⁹. Indeed, published work from the laboratory of Enzyme & Synthetic Biotechnology of the National Hellenic Research Foundation has yielded a cyclic peptide that rescues the misfolding of the familial ALS-linked mutant SOD1(A4V)³²⁴. Furthermore, ongoing work from the same laboratory has yielded cyclic peptide rescuers for the misfolding of the human p53 variants, as well as forms of polyglutaminated huntingtin, linked to Huntington's disease. Consequently, the integrated bacterial system presented herein has been shown to be a flexible approach for the identification of rescuers of disease-associated protein misfolding.

As a result of screening a combined cyclic peptide library of $>10^7$ members, this bacterial system was able to identify hundreds of cyclic peptides with bioactivity against the aggregation of A β . Two of those peptides, A β C5-34 and A β C5-116, were further investigated in a variety of biochemical, biophysical and biological assays, showing an interference with typical A β aggregation and causing the formation of atypical A β aggregates with reduced binding affinity for the neuronal surface and decreased neurotoxicity. The effects of A β C5-34 and A β C5-116 were also verified against the aggregation and cytotoxicity of A β in transgenic *Caenorhabditis elegans* models of AD. Moreover, site-directed mutagenesis and next-generation sequencing analysis of the complete pool of hits derived from screening, allowed the characterization of structure-activity relationships and the classification of hits into distinct families of bioactive cyclic peptides.

5.1. Perspective and future work

5.1.1. Cyclic peptides as anti-aggregation agents

The defining feature of the presented integrated bacterial system was the use of a highly diverse cyclic peptide library as the source of anti-aggregation hits. Peptide therapeutics development has greatly expanded in the last decade, owing to their appealing pharmacological properties. A review by Craik

et al. stated that peptides posed an alternative to the “rule of five³⁵⁰” class of therapeutics, as they usually display molecular weights >500 Da. Furthermore, peptides generally show great specificity for their respective protein-target, decreasing the chances of off-target side effects. This property is attributed to the long evolutionary selection of complementary conformations, that allows very specific binding of the target²⁹². Importantly, this principle of evolutionary selection can be applied in selection assays like the one presented in this thesis.

Cyclic peptides in particular, are a proven class of pharmaceuticals^{295, 296}. They tend to assume less flexible conformations, leading to energetically favourable target binding, while they are thermodynamically more stable and less susceptible to proteolysis, compared to their linear analogues. From 2006 to 2015, nine cyclic peptide drugs managed to enter the market, while numerous are undergoing clinical trials²⁹⁵. With regard to the cyclic peptides used in this thesis, these contained 4-6 amino acids (mean M.W.≈660 Da), placing them at the lower end of the size-scale of existing peptide pharmaceuticals²⁹². As such, they were intended to display the advantageous characteristics of peptides, while retaining a small-molecule-like behaviour. Despite lacking the structural complexity of larger peptides or antibodies³⁵¹⁻³⁵³, hits derived from the utilized libraries have minimal target-binding surfaces, which should allow easy derivatization into lead compounds^{296, 354}. For instance, such molecules are relatively easy to modify with the aim of improving specific pharmacological characteristics like the balance of lipophilicity/hydrophilicity and inherent peptide aggregation, as necessary.

5.1.2. Further study and development of AβC5-34 and AβC5-116

The novel peptides AβC5-34 and AβC5-116 are two promising candidates for the development of anti-amyloid drugs for AD, which led to a patent application (PCT/US18/33784). These peptides have been

tested in *in vitro* experiments revealing that they interfere with the aggregation of A β , causing the formation of atypical aggregates with reduced neurotoxicity.

Future *in vitro* studies on A β C5-34 and A β C5-116, should aim to illuminate certain properties that are necessary in the development of an AD treatment. For instance, the lead candidate should be able to cross into the brain, in order to gain access to accumulating A β . To this end, various static or dynamic *in vitro* models of the blood brain barrier can be used^{355, 356}. Subsequently, it would be necessary to study the manner in which the two peptides are metabolized. Isolated hepatocytes and hepatic microsomes have long been utilized in drug discovery, since they provide a way to estimate the metabolic clearance of a substance^{357, 358}. Hepatocytes offer the most biologically relevant option for this task, since they contain the enzymes that a substance will encounter during hepatic metabolism, as well as the membrane transporter proteins that will likely interact with the substance³⁵⁹. As such, primary cultures of hepatocytes could provide an approximation of the *in vivo* metabolic clearance for the two peptides^{357, 358}. Furthermore, the ability of A β C5-34 and A β C5-116 to bind plasma proteins should be estimated, in order to determine the bound and unbound fractions of the peptides in plasma³⁶⁰⁻³⁶², and future *in vitro* studies of A β C5-34 and A β C5-116 would also have to include a general evaluation of peptide-induced cytotoxicity³⁶³.

The pharmacokinetics of the A β C5-34 and A β C5-116 peptides would also have to be studied *in vivo*. The most common animal models used for AD drug candidate evaluation are usually mice. Administration of the peptides to a wild-type animal should enable the determination of their metabolic clearance, their plasma concentrations and their half-lives. Additionally, it should also allow the evaluation of the peptides' ability to localize in the brain, therefore establishing a measure of bioavailability. In a similar manner, administration of the peptides to wild-type animals should also allow the observation of the toxic effect of the peptides on key organs such as the liver and brain. However, the evaluation of anti-aggregation and toxicity-suppressing effects of the peptides would require a transgenic animal. A commonly used animal for this purpose is the APP/PS1 double

transgenic mouse that co-expresses five familial AD mutations (5×FAD mice). These animals overexpress A β ₄₂ leading to massive and rapid accumulation in the brain. Accordingly, this transgenic model of AD would allow evaluation of the peptide effect through memory and behavioural tests, observation of neuronal loss and immunohistochemical methods^{364, 365}.

Naturally, the additional biochemical and biological studies of A β C5-34 and A β C5-116 could generate valuable information needed for any necessary modifications of the peptide, that would improve their potency and efficacy against the toxic effects associated with the aggregation of A β . This kind of pre-clinical development studies are currently underway in the Laboratory of Enzyme & Synthetic Biotechnology at NHRF.

5.1.3. Evaluation of additional selected cyclic peptides

Ongoing research in the lab of Enzyme & Synthetic Biotechnology at NHRF has already assessed peptides from other pentapeptide families, showing positive *in vitro* results, while more selected peptides are pending evaluation. The list of 605 selected peptides contains numerous unique peptides and peptide families that warrant further investigation, and it should allow the discovery of many more unique therapeutic candidates. The categorization of cyclic peptides into families has simplified this task.

Moreover, research from the lab of Enzyme & Synthetic Biotechnology at NHRF has already assessed peptide hits obtained by screening a heptapeptide library, using the same bacterial system. Two of those hits have been successfully studied and evaluated in *in vitro* assays and in two *C. elegans* models of AD. Overall, it is expected that the assessment of additional selected peptides derived from this work, as well as peptides resulting from screening alternative libraries with even higher diversities (i.e.

heptapeptides), will reveal a multitude of candidates for the development of anti-amyloid therapeutics against AD.

5.1.4. Generation of secondary peptide libraries

Alternatively, the sequence motifs that were discovered through this work could be used in the generation of additional SICLOPPS libraries. For instance, it would be possible to generate libraries of peptides with extended backbones, which incorporate one or more bioactive sequence motifs. Peptides contained within these libraries would not resemble small molecules anymore, rather, any hits derived from such libraries would be expected to exhibit enhanced binding affinities due to larger contact surfaces²⁹⁶. Moreover, these hits would be likely to display multiple points of contact with A β . A secondary benefit to those libraries would be that the longer and more flexible peptide backbones, could impose a greater steric hindrance to A β intermolecular interactions³⁶⁶. In the case of A β , hits derived from such libraries could lead to enhanced disruption of aggregation³⁶⁷.

5.2. Conclusion

Overall, the approach presented in this thesis presents a simple and high-throughput means of investigating and exploiting the large chemical space occupied by this class of relatively underexplored biomolecules³⁶⁸. Furthermore, the search for an A β -related treatment remains at the epicenter of AD research, while A β accumulation is reported to commence as early as 20 years prior to dementia³⁶⁹⁻³⁷¹. As a result, the necessity for a disease-modifying treatment that targets A β is as pressing as ever. Based on the results discussed throughout this thesis, the presented integrated bacterial technology is anticipated to provide a much-needed aid in the development of therapeutic candidates against the neurotoxic effects associated with the accumulation of A β .

Chapter 6. Methods and materials

6.1. Chemicals, reagents, kits and media

All chemical reagents and media were purchased from Sigma-Aldrich (USA), unless otherwise stated. Isopropyl- β -D-thiogalactoside (IPTG) was purchased from MP Biomedicals (Germany).

Synthetic human amyloid peptides A β ₄₀ and A β ₄₂ were purchased from Eurogentec, Belgium (>95% pure). The synthetic cyclic peptides A β C5-34 (cyclo-SASPT) and A β C5-116 (cyclo-TAFDR) were purchased from Genscript (USA). Stock solutions of the synthetic cyclic peptides were created as follows: 32.5 mM in water for A β C5-34, 10 mM in 40% DMSO for A β C5-116.

All DNA-processing enzymes were purchased from New England Biolabs (USA) apart from alkaline phosphatase FastAP, which was purchased from ThermoFisher Scientific (USA).

Recombinant plasmids were purified using NucleoSpin Plasmid from Macherey-Nagel (Germany) or Plasmid Midi kits from Qiagen (Germany). PCR products and DNA extracted from agarose gels were purified using Nucleospin Gel and PCR Clean-up kits from Macherey-Nagel (Germany), respectively.

6.2. Bacterial cell transformation

The following *E. coli* strains were used throughout the work presented herein for various transformations, DNA preparations and protein expression:

- BL21(DE3), Novagen, USA: *E. coli* str. B F⁻ *ompT gal dcm lon hsdS_B(r_B⁻m_B⁻)* λ (DE3 [*lacI lacUV5-T7p07 ind1 sam7 nin5*]) [*malB*⁺]_{K-12}(λ ^S)
- DH5a: F⁻ *endA1 glnV44 thi-1 recA1 relA1 gyrA96 deoR nupG purB20 ϕ 80dlacZ Δ M15 Δ (lacZYA-argF)U169, hsdR17(r_K⁻m_K⁺), λ ⁻*

- MC1061, acquired from G. Georgiou, The University of Texas at Austin, USA: str. K-12 F⁻ λ⁻ Δ(*ara-leu*)7697 [*araD139*]_{B/r} Δ(*codB-lacI*)3 *galK16 galE15* e14⁻ *mcrA0 relA1 rpsL150*(Str^R) *spoT1 mcrB1 hsdR2*(r⁻m⁺)
- MC4100A, acquired from G. Georgiou, The University of Texas at Austin, USA: F⁻ [*araD139*]_{B/r} Δ(*argF-lac*)169* &lambda⁻ e14⁻ flhD5301 Δ(*fruK-yeiR*)725 (*fruA25*)‡ *relA1 rpsL150*(strR) *rbsR22* Δ(*fimB-fimE*)632(::IS1) *deoC1*
- NEB10-beta, New England Biolabs, USA: Δ(*ara-leu*) 7697 *araD139 fhuA* Δ*lacX74 galK16 galE15* e14⁻ φ80*dlacZΔM15 recA1 relA1 endA1 nupG rpsL* (Str^R) *rph spoT1* Δ(*mrr-hsdRMS-mcrBC*)

Electrically or chemically competent *E. coli* strains were used for transformations. Usually, 50-100 μL of competent cells rested on ice for 15-30 minutes, with 1-4 μL of purified plasmid DNA. For chemically competent strains, cells were heat-shocked at 42 °C for 45-50 sec. Electrically competent strains were electroporated using a Micropulser™ Electroporator (Bio-Rad, USA). Strains were promptly complemented with 500-1000 μL of LB broth. Transformed cells were incubated at 37 °C, 220 rpm for 50 minutes and were then plated on LB agar plates containing the required antibiotics. Colonies were formed overnight (14-16 h) at 37 °C and colonies were picked and used as necessary.

6.3. Library construction

Nine combinatorial cyclic peptide sub-libraries were initially created³²⁴, and combined prior to transformation and production.

Table 6.1: Plasmid vector libraries encoding tetra-, penta-, and hexapeptides. Vectors constructed by D.C. Delivoria at NHRF.

Peptide length	Sub-library	Plasmid vector	DNA
4	cyclo-CysX ₁ X ₂ X ₃	pSICLOPPS-CysX ₁ X ₂ X ₃	Cys(NNS) ₁ -(NNS) ₃
	cyclo-SerX ₁ X ₂ X ₃	pSICLOPPS-SerX ₁ X ₂ X ₃	Ser(NNS) ₁ -(NNS) ₃
	cyclo-ThrX ₁ X ₂ X ₃	pSICLOPPS-ThrX ₁ X ₂ X ₃	Thr(NNS) ₁ -(NNS) ₃

Peptide length	Sub-library	Plasmid vector	DNA
5	cyclo-CysX ₁ X ₂ X ₃ X ₄	pSICLOPPS-CysX ₁ X ₂ X ₃ X ₄	Cys(NNS) ₁ -(NNS) ₄
	cyclo-SerX ₁ X ₂ X ₃ X ₄	pSICLOPPS-SerX ₁ X ₂ X ₃ X ₄	Ser(NNS) ₁ -(NNS) ₄
	cyclo-ThrX ₁ X ₂ X ₃ X ₄	pSICLOPPS-ThrX ₁ X ₂ X ₃ X ₄	Thr(NNS) ₁ -(NNS) ₄
6	cyclo-CysX ₁ X ₂ X ₃ X ₄ X ₅	pSICLOPPS-CysX ₁ X ₂ X ₃ X ₄ X ₅	Cys(NNS) ₁ -(NNS) ₅
	cyclo-SerX ₁ X ₂ X ₃ X ₄ X ₅	pSICLOPPS-SerX ₁ X ₂ X ₃ X ₄ X ₅	Ser(NNS) ₁ -(NNS) ₅
	cyclo-ThrX ₁ X ₂ X ₃ X ₄ X ₅	pSICLOPPS-ThrX ₁ X ₂ X ₃ X ₄ X ₅	Thr(NNS) ₁ -(NNS) ₅

N: A, T, G, C and S: G, C,
as in Tavassoli & Benkovic, 2007³¹⁰

The generated plasmid vectors express libraries of fusion proteins consisting of four moieties. (i) A C-terminal splicing domain of the *Synechocystis* sp PCC6803 DnaE intein (I_C), (ii) a tetra-, penta-, or hexapeptide sequence, followed by (iii) the N-terminal splicing domain of the *Ssp* DnaE intein (I_N), and (iv) a chitin-binding domain (CBD). Protein expression for these fusions is controlled by the L(+)-arabinose-inducible P_{BAD} promoter. For the construction of the nine sub-libraries presented in **Table 6.1**, the degenerate 5'-primers GS032, GS033, GS034, GS072, GS073, GS074, GS075, GS076, GS077 were individually paired to the 3'-primer GS035. The plasmid vector pSICLOPPS (kind gift by Professor S. Benkovic, generated according to the protocol described in Tavassoli *et al.*, 2007³¹⁰) served as a template for PCR amplifications. In order to encode the necessary nucleophilic Cys, Ser, and Thr^{301,302}, the respective codons TGC, AGC, and ACC were chosen, as they represent the most frequently occurring codons encoding these amino acids in *E. coli*. The randomized amino acids (X) were encoded by NNS codons as shown in **Table 6.1**. In order to eliminate mismatches, a second PCR amplification was conducted for each construct. In this case, the product of the first PCR amplification served as template. For the second round of PCR amplification, the 5'-primers GS069, GS070 and GS071 corresponding to the Cys, Ser and Thr sub-libraries respectively, were individually paired to the 3'-primer GS035. The PCR products of the second amplification were purified and then digested with the restriction enzymes BglI and HindIII, (5 h at 37 °C). Subsequently, the digested products were ligated to the pSICLOPPSKanR vector, which had been previously digested with BglI and HindIII, and dephosphorylated. The insert/vector ratios used for these ligation reactions were optimized at a 12:1

insert/vector molar ratio (0.35, 0.7 and 3.5 μg of pSICLOPPSKanR vector, for each of the tetra-, penta- and hexapeptide sub-libraries, respectively). The ligation reaction took place at 16 °C, for 4 hours. Following ligation, the DNA was purified with spin-columns (Macherey-Nagel, Germany) and transformed into an electrically competent MC1061 strain. The transformed cells were plated on LB agar plates containing 25 $\mu\text{g}/\text{mL}$ chloramphenicol and were incubated at 37 °C for 14-16 h. Serial dilutions of the transformations were used to determine the total number of transformants for each sub-library and as a result, the combined library composed of the pSICLOPPS-NuX₁X₂X₃-X₅ plasmid vectors in nearly 31,240,000 independent transformants. All primers used in the construction of the libraries are presented in **Supplementary Table 1**.

6.4. Expression vector construction

The pSICLOPPS plasmid vector was a gift from Prof. Stephen Benkovic (University of Pennsylvania, USA). The plasmid vectors pA β_{42} -EGFP and pA β_{42} (F19S;L34P)-EGFP were gifts from Prof. Michael H. Hecht (Princeton University, USA). The pET28a(+) plasmid vector was purchased from Novagen, USA.

The pETSOD1(A4V)-EGFP plasmid vector was constructed by S. Panoutsou (NHRF)³²⁴ and the pETp53(Y220C)-EGFP plasmid vector was constructed by D.C. Delivoria (NHRF)³²⁴. The pETA β_{42} vector, used in the production of recombinant Met1-A β_{42} was constructed by lifting the A β_{42} sequence with PCR, using the forward primer DD004 and reverse primer IM060. Subsequently, the PCR product was digested with NcoI and XbaI and inserted into the similarly digested pET28(a)+ (Addgene, USA) commercial vector.

In order to construct the pSICLOPPS vectors responsible for encoding the variants of the two selected cyclic peptides, A β C5-34 and A β C5-116, the auxiliary pSICLOPPSKanR vector was used. The previously constructed pSICLOPPSKanR contained the aminoglycoside 3'-phosphotransferase-encoding gene

(kanamycin resistance), replacing the intein fusion starting from the 3' end of the I_c splicing domain. To construct the vectors pSICLOPPS-AβC5-34(S1C), pSICLOPPS-AβC5-34(S1T), pSICLOPPS-AβC5-34(S3A), pSICLOPPS-AβC5-34(P4A) and pSICLOPPS-AβC5-34(T5A), PCR mutagenesis was performed on the pSICLOPPS-AβC5-34 template using the 5'-primers IM033, IM034, IM036, IM037 and IM038, individually paired to the 3'-primer GS035. PCR was followed by BglI and HindIII digestion and ligation of the respective products to the BglI/HindIII digested pSICLOPPSKanR vector.

The vectors: pSICLOPPS-AβC5-116(T1C), pSICLOPPS-AβC5-116(T1S), pSICLOPPS-AβC5-116(F3A), pSICLOPPS-AβC5-116(D4A), pSICLOPPS-AβC5-116(R5A), pSICLOPPS-AβC5-116(A2F), pSICLOPPS-AβC5-116(A2S), pSICLOPPS-AβC5-116(A2P), pSICLOPPS-AβC5-116(A2T), pSICLOPPS-AβC5-116(A2Y), pSICLOPPS-AβC5-116(A2H), pSICLOPPS-AβC5-116(A2K), pSICLOPPS-AβC5-116(A2E), pSICLOPPS-AβC5-116(A2W), pSICLOPPS-AβC5-116(A2R), pSICLOPPS-AβC5-116(A2del), pSICLOPPS-AβC5-116(F3del) and pSICLOPPS-AβC5-116(D4del) were generated by using the pSICLOPPS-AβC5-116 as PCR template. PCR mutagenesis was performed with the 5'-primers IM027, IM028, IM030, IM031, IM032, IM043, IM044, IM045, IM046, IM047, IM048, IM049, IM050, IM051, IM052, IM039, IM040 and IM041, paired individually to the 3'-primer GS035. Digestion of the generated product with BglI and HindIII, was followed by ligation to the similarly digested pSICLOPPSKanR.

pSICLOPPS-AβC5-325, pSICLOPPS-AβC5-359, pSICLOPPS-AβC5-413, pSICLOPPS-AβC5-479 were generated by PCR amplification using the template pSICLOPPS-Random1 and the forward primers IM077, IM078, IM080 and IM081, respectively, along with the reverse primer GS035, digestion with BglI and HindIII and ligation into similarly digested pSICLOPPSKanR.

To construct the vectors: pSICLOPPS(H24L;F26A)-AβC5-3, pSICLOPPS(H24L;F26A)-AβC5-2, pSICLOPPS(H24L;F26A)-AβC5-17, pSICLOPPS(H24L;F26A)-AβC6-1, pSICLOPPS-(H24L;F26A)-AβC5-34, pSICLOPPS(H24L;F26A)-AβC5-26, pSICLOPPS(H24L;F26A)-AβC5-21, pSICLOPPS(H24L;F26A)-AβC5-116, pSICLOPPS-(H24L;F26A)-Random1 and pSICLOPPS(H24L;F26A)-Random2, the respective pSICLOPPS

peptide-encoding vectors were used. Specifically, the pSICLOPPS vectors encoding: TTVDR (A β C5-3), TTYAR (A β C5-2), TTTAR (A β C5-17), TPVWFD (A β C6-1), TAWCR (A β C5-27), TTWCR (A β C5-21), TAFDR (A β C5-116), random cyclic peptide 1 (undetermined sequence), random cyclic peptide 2 (undetermined sequence), were digested with BglI and HindIII. Ligation of the digested inserts to the BglI and HindIII digested pSICLOPPS(H24L;F26A)KanR vector afforded the series of vectors mentioned above, by replacing the KanR sequence. The auxiliary pSICLOPPS(H24L;F26A)KanR vector had been generated beforehand. Using the pSICLOPPS vector as template and the GS037 and DD015 primers, the c-terminal splicing domain of the Ssp DnaE intein I_C was mutated by PCR amplification. The PCR product underwent digestion with NcoI and BglI and was ligated to a similarly digested pSICLOPPSKanR.

All PCR and digestion products were purified with spin-columns (Macherey-Nagel, Germany) prior to any further processing. All primers used in the construction of plasmid vectors are presented in **Supplementary Table 1**. The complete list of plasmid vectors used in this work are presented in **Supplementary Table 2**.

6.5. Cyclic peptide library screening.

Chemically competent *E. coli* BL21(DE3) cells (Novagen, USA) were transformed with the pETA β_{42} -EGFP expression vector, which encodes the A β_{42} -EGFP fusion under the control of the T7 bacteriophage promoter. The transformed cells were then treated to become chemically competent and were co-transformed with the combined pSICLOPPS-NuX₁X₂X₃-X₅ vector library (**Table 6.2**). This transformation yielded approximately 10⁸ transformants that carried both the pETA β_{42} -EGFP vector and one of the pSICLOPPS-NuX₁X₂X₃-X₅ vectors. The transformants were pooled together and grown at 37 °C under shaking in Luria-Bertani (LB) broth. The broth contained 0.002% L-arabinose in order to induce cyclic peptide production. When OD₆₀₀ reached 0.5, 0.1 mM isopropyl- β -D-thiogalactoside

(IPTG) was added in order to induce expression of the A β ₄₂-EGFP reporter fusion. The culture was allowed to grow in the same conditions for two hours before recording the fluorescence (FITC-H - 530/30 nm) of 10,000 cells with a Becton-Dickinson FACS Aria system (BD Biosciences, USA). Following fluorescence measurement, $\sim 10^8$ cells were gated on a side-scatter (SSC-H) versus forward-scatter (FSC-H) plot so as to eliminate events that were caused by non-cellular particles, and subjected to cell sorting using the BD FACS Aria sorter and the FACSDiva software (BD Biosciences, USA). Sorting isolated the part of the bacterial population that exhibited the top 1-3% fluorescence. The isolated cells were used to obtain fresh cultures and were then used for two additional rounds of sorting, under the same conditions set for the initial sorting round. The bacterial population isolated after the second sorting round contained the sorted bacterial library and was enriched in high-fluorescence clones.

6.6. Isolation of individual bacterial clones from the sorted peptide library

The sorted population contained approximately 10^4 bacterial cells. These cells were grown at 37 °C under shaking and the culture was then used for plasmid purification. 5 μ L of this plasmid preparation were used to transform fresh chemically competent *E. coli* BL21(DE3) cells. The cells were plated on duplicate LB agar plates containing chloramphenicol 40 μ g/mL. Colonies were grown for 16 hours at 37 °C. Both plates contained more than 10^4 colonies. One plate was used to pool the colonies together and inoculate fresh LB broth containing chloramphenicol 40 μ g/mL. This was used for medium-sized plasmid preparation (Qiagen, Germany). The second plate was used to lift 30 non-overlapping colonies, which were used to inoculate fresh LB broth containing kanamycin 50 μ g/mL and chloramphenicol 40 μ g/mL. The cultures were grown at 37 °C under shaking overnight and were used for plasmid purification (Macherey-Nagel, Germany). These plasmid preparations were expected to contain one or both of the pETA β ₄₂-EGFP and pSICLOPPS-NuX₁X₂X₃-X₅ vectors. In order to isolate only the peptide encoding vector, the plasmid preparations were used to transform electrically competent

E. coli BL21(DE3) cells which were then plated on LB agar plates containing only 40 µg/mL chloramphenicol, as this would allow preferential selection of transformants that did not carry the pETA β_{42} -EGFP vector. Colonies were grown at 37 °C for 16 hours. To eliminate the possibility of carrying over a pETA β_{42} -EGFP plasmid, single colonies were lifted and used for duplicate liquid LB cultures, containing either kanamycin 50 µg/mL or chloramphenicol 40 µg/mL. Cultures were grown at 37 °C under shaking, for 16 hours. Chloramphenicol-containing cultures, whose kanamycin-containing duplicate did not grow, indicated the presence of the pSICLOPPS-NuX₁X₂X₃-X₅ vector and the absence of the pETA β_{42} -EGFP vector. The process was repeated for duplicate cultures that showed growth for both antibiotics. Eventually, 10 clones were isolated from the sorted sub-library of pSICLOPPS-NuX₁X₂X₃-X₅.

6.7. Bacterial cell fluorescence

Electro-competent BL21(DE3) cells (50 µL) were transformed with 2 µL of pETA β_{42} -EGFP and 2 µL of one of the isolated pSICLOPPS-NuX₁X₂X₃-X₅ vectors and plated on kanamycin 50 µg/mL and chloramphenicol 40 µg/mL containing LB agar plates. Colonies were grown for 16 hours at 37 °C. Individual colonies were picked the following day and used to inoculate 5 mL LB cultures, which were then left to grow for 16 hours at 37 °C under shaking conditions. Out of these cultures, 50 µL were used to inoculate 5 mL of fresh LB that contained kanamycin 50 µg/mL, chloramphenicol 40 µg/mL and arabinose 0.02% in order to induce peptide production. The cultures were placed at 37 °C under shaking and left until the OD₆₀₀ reached 0.4-0.5 units. At that point, IPTG 0.1 mM was added to the cultures to induce A β_{42} -EGFP production. The cultures were grown for two more hours before measuring their OD₆₀₀ and lifting the equivalent of 1 mL of culture with OD₆₀₀=1. Following a centrifugation at 6,000 g for 2 minutes, the cell pellets were resuspended in 100 µL PBS and transferred into a 96-well FLUOTRAC 200 plate (Greiner Bio One, Austria). In order to perform

fluorescence measurements a TECAN Safire II-Basic plate reader (Tecan, Austria) was used with the excitation set at 488 nm and emission at 510 nm.

6.8. Protein electrophoresis and western blot analysis

Bacterial cell pellets corresponding to 1 mL culture with $OD_{600}=1$ were collected as described above and were resuspended in 200 μ L PBS. Cell lysis was performed with twice-repeated sonication for 10 seconds at 4 °C, with samples remaining on ice between repetitions. The cell lysate was split in two equal volumes and 100 μ L of it were centrifuged at 13,000 x g for 10 min. The resulting supernatant represented the soluble lysate and the pellet (insoluble lysate) was re-suspended in 100 μ L PBS.

Denaturing SDS-PAGE: Samples were boiled for 5 min in the presence of a loading solution that contained SDS and mercaptoethanol. 10 μ L of each sample were loaded onto 12% or 15% gels that contained SDS. Gels were run at 120-180 V.

Native-PAGE: Samples were unboiled and the loading solution did not contain any denaturing agents. 10 μ L of each sample were loaded onto SDS-free 10% gels. Gels were run at ~4 °C using low voltages (80-100 V)

In-gel fluorescence: Semi-denaturing SDS-PAGE or native-PAGE were used for visualization of in-gel fluorescence under UV light. Pictures were acquired with a ChemiDoc-It² Imaging System, using a green fluorescence emission filter attached to the camera lens and an excitation filter attached to the UV lamp (UVP, UK). Exposure times ranged from 3-5 sec.

Western blots: The antibodies used in western blots presented in this thesis were a mouse anti-A β (6E10) (Covance, USA) at 1:2,000 dilution, a mouse anti-CBD (New England Biolabs, USA) at 1:25,000 or 1:100,000 dilution, and a HRP-conjugated goat anti-mouse antibody (Bio-Rad, USA) at 1:4,000.

Following gel electrophoresis, protein bands were transferred onto polyvinylidene fluoride (PVDF) membranes (Merck, Germany) using a semi-dry blotter (Thermo Fisher, USA). Transfer time was 50 minutes at 12 V. Membranes were blocked with TBS-tween buffer containing 5% dried, non-fat milk, for 1 h at room temperature. Afterwards, the membranes were washed thrice with TBS-tween and were then incubated with the appropriate antibody for 1h at room temperature. Additional washing TBS-tween was repeated three times and was followed by the addition of a secondary antibody. Antibodies were diluted as necessary in TBS-tween containing 0.5% non-fat dried milk. After a 1h incubation with the secondary antibody, protein bands were visualized using a ChemiDoc-It² Imaging System (UVP, UK).

6.9. Preparation of synthetic A β ₄₀ and A β ₄₂ stocks and solutions

Synthetic A β ₄₀ and A β ₄₂ were carefully dissolved in double-deionized water to a final concentration of 100 μ M, avoiding agitation and introduction of air in the solution. PBS (10 mM, pH 7.33) was then added to these solutions, to a final A β concentration of 50 μ M.

6.10. Recombinant production and purification of A β ₄₂

BL21(DE3) competent cells were transformed with the pETA β 42 plasmid vector that encoded Met-A β ₄₂. Single colonies were picked and used to inoculate LB liquid medium cultures, which were incubated for 16 hours at 37 °C (220 rpm). These were used to inoculate fresh LB of varying volumes, with a 1/100 inoculant/medium ratio. The new cultures were incubated at 37 °C (220 rpm) until they reached OD₆₀₀= 0.6-0.8, when protein production was induced with 0.01 mM IPTG. Cultures were incubated for 2-3 more hours at 37 °C (220 rpm). Subsequently, cultures were placed in ice-water for 20-30 minutes and were frequently shaken. Pellets were then collected in sterile tubes upon gentle

centrifugation (2,500 g, 4 °C, 15 min) and stored at -20 °C. In order to purify the produced Met-A β ₄₂ the pellets from 400 mL culture were thawed and resuspended in 50 mL phosphate buffer (phosphate 20 mM, EDTA 200 μ M, pH 8.0) and sonicated at low intensity, for 4 \times 30 sec cycles with 30 sec pauses between each cycle. The supernatant and pellet were separated after centrifugation at 18,000 g for 15 min, at 4 °C. Then, the pellet was resuspended in 8 M urea (solubilized in phosphate buffer 20 mM, EDTA 200 μ M, pH 8.0). All steps thus far, were performed on ice. Following a final sonication of the resolubilized pellet (8 M urea buffer, as above) the whole volume was used for anion exchange chromatography. Specifically, 5 mL of DEAE sephadex A-25 beads (Sigma-Aldrich, USA) was used to pack a benchtop column. Following equilibration with the same phosphate buffer used in the initial solubilization of the cell pellets, the sonicated mixture was carefully introduced in the column. Flow was gravity assisted and was not allowed to exceed 1 mL/min. The flow-through was then passed through the column a second time prior to sequential washes with 25 mL phosphate buffer and 25 mL phosphate buffer with 25 mM NaCl. Elution was performed with phosphate buffer containing increasing quantities of NaCl (50, 100, 150, 200 and 300 mM). Elution volumes were 5 mL per fraction and the protein was usually detected in the second and third fraction by western blot (anti-A β , 6E10). Elution fractions were then frozen or used in the next step of purification. Fractions eluted from DEAE purification were thawed and combined in a solution that contained Guanidine-HCl 5 M. The total volume was concentrated to \leq 0.8 mL using a spin-concentrator with a 3 KDa molecular weight cut-off, and injected in a 1 mL loading loop. The sample was purified using a size exclusion Superdex 75 10/300 column (GE Healthcare, USA), attached to an ÄKTA FPLC system (GE healthcare, USA). Flow was set at 0.5 mL/min and eluate volumes at 1 mL. Protein was detected with western blots (anti-A β , 6E10) and protein concentrations ranged from \sim 10 μ M to \sim 50 μ M (calculated using UV₂₈₀ absorption and an extinction coefficient of 1490 (M \times cm)⁻¹, at 280 nm. If necessary, eluates were combined, concentrated and purified for a second time, provided that the concentration was adequately high.

6.11. Circular dichroism measurements

The required amount of synthetic cyclic peptides were added to 50 μM $\text{A}\beta_{40}$ or $\text{A}\beta_{42}$ solutions in order to achieve the target molar ratio (cyclic peptide: $\text{A}\beta$). The developing structural changes of $\text{A}\beta_{40}$ or $\text{A}\beta_{42}$ were monitored for 30 days, while the samples were kept under quiescent conditions at 33 °C, in 1 mm path-length quartz cuvettes. CD spectra were recorded from 190 to 260 nm wavelengths, using a JASCO J-715 spectropolarimeter (Jasco Co., Japan). Each reported spectrum is the average of three scans at a rate of 100 $\text{nm}\cdot\text{min}^{-1}$ and a resolution of 0.5 nm. Three independent experiments were performed for each condition.

6.12. Thioflavin T staining of $\text{A}\beta$ fibril formations

For $\text{A}\beta_{40}$ or $\text{A}\beta_{42}$ fibril staining with thioflavin T (ThT), 100 μL of the 30-day aged 50 μM $\text{A}\beta$ solutions that had been prepared for CD experiments, were diluted with PBS 10 mM, pH 7.33, to a final volume of 200 μL and 25 μM of $\text{A}\beta$. The ThT stock solution was prepared by dissolving ThT in PBS 10 mM, pH 7.3. For amyloid-like fibril staining, 2.5 μL from a stock solution of ThT in PBS (10 mM, pH 7.33) were added to the $\text{A}\beta$ solutions, to a final ThT concentration of 5 μM . Following careful agitation by pipetting, the fluorescence emitted by fibril binding ThT was measured using a HITACHI F-2500 (Japan) spectrofluorometer. Excitation filters were set at 440 nm and emission at 480 nm.

6.13. Transmission electron microscopy

TEM analysis was performed on the 30-day aged 50 μM samples of $\text{A}\beta_{42}$ that were used for CD (50 μM $\text{A}\beta_{42}$ with or without 100 μM of the selected peptides). The samples were mixed by pipetting to create a homogeneous mixture and 2 μL were lifted and placed in a carbon-coated film on 200-mesh copper

grids (Agar Scientific, UK) for 5 minutes³⁷². Following adsorption the grids were washed with deionized water and were subsequently negatively stained for 5 min by applying a 2 μ l drop of newly prepared 1% (w/v) uranyl acetate in Milli-Q grade water. Any excess fluid was removed and grids were washed again with deionized water, before air-drying them. The recorded images were produced with a FEI CM20 electron microscope (FEI, USA) using a Gatan GIF200 imaging filter (Gatan, USA) combined with a Peltier-cooled slow-scan CCD camera.

6.14. Dynamic light scattering

The distribution of recombinant A β ₄₂ particles was recorded with a Zetasizer NanoZS90 (Malvern, UK) instrument. Following a 30 sec equilibration at 37 °C, DLS spectra were recorded as the average of six distinct runs per sample, each consisting of 10 sequential measurements of 30 sec. Spectra showed the particle signal intensity plotted against its hydrodynamic radius.

6.15. Neuronal cell cultures

Media and agents for primary neuronal cell cultures were purchased from Thermo Fisher Scientific (USA). After dissection from postnatal day 1 female pups of C57BL/6 mice, the hippocampus was incubated with 0.25% trypsin at 37 °C for 15 minutes, then rinsed with 10 mL of Hibernate solution containing heat-inactivated fetal bovine serum (FBS) 10% (v/v). Cells were grown in Neurobasal-A medium containing 2% B-27 supplement, 0.5 mM Gluta-MAX and 1% penicillin/streptomycin at 37 °C and 5% CO₂. Half the medium volume was replaced twice per week. Finally, neuronal hippocampal cells were plated at a density of $\sim 2 \times 10^4$ per well in 96-well plates for MTT assays and 5×10^5 per well

in 24-well plates for induced cell death assays. Measurements were taken after seven days of incubation, based on which, the cell viability of primary hippocampal neurons was calculated³⁷³.

U87MG cells (human glioblastoma-astrocytoma, epithelial-like cell line) were kindly provided by Dr. Maria Paravatou-Petsotas, Radiobiology Laboratory, Institute of Nuclear & Radiological Sciences & Technology, Energy & Safety, NCSR "Demokritos". Microscope observation and DAPI staining of the cell cultures indicated no contamination by mycoplasma. Growth media and chemical agents were purchased from Biochrom AG (Germany) and PAA Laboratories (USA). U87MG cells were grown in Dulbecco's modified Eagle medium (DMEM), supplemented with 10% fetal bovine serum (FBS), 2.5 mM L-glutamine, 1% penicillin/streptomycin at 37 °C and 5% CO₂. Cells were plated in 96-well plates at a density of 2×10⁴ cells per well and incubated at 37 °C for 24 h to allow cells to attach. The medium was then removed, and the *cells were* incubated in serum-free medium for 24 h, prior to Aβ treatment.

Cortical neurons were prepared from E16 C57BL/6 mouse embryos^{374, 375} and were seeded onto poly-D-lysine-coated glass coverslips (12 mm in diameter) in 24-well plates at a density of 1.5 × 10⁵ per well. The cultures were grown for 5-7 d in Neurobasal medium containing B-27 supplement, at 37 °C and 5% CO₂ prior to plating. Half the medium volume was replaced twice a week. Prior to cell viability measurements, primary mouse cortical neurons were treated with Aβ oligomers produced by the 7PA2 cell line, which is known to produce oligomeric Aβ species (kindly provided by Prof. Dominic Walsh, Brigham & Women's Hospital, USA). The 7PA2 cells were initially generated upon transfection of CHO cells with the human APP gene that carried the V717P mutation that is associated with familial AD and leads to increased Aβ production³³⁰. The treatment of primary mouse cortical neurons involved exposure to Aβ oligomers derived from 7PA2 cells cultured with or without AβC5-34, AβC5-116 or SOD1C5-4 (10 μM)³³⁰.

6.16. Cell viability measurements

Solutions of synthetic A β ₄₀ or A β ₄₂ (10 μ M) in PBS were incubated at 37 °C (3 d for A β ₄₀ and 1 d for A β ₄₂) with or without synthetic cyclic peptides (at 1:1 and 2:1 peptide/A β ratios). The solutions were subsequently diluted with fresh medium to a 1 μ M final A β concentration, and transferred into wells. Cells were exposed to the A β solutions for 24 h and 100 μ L of a 0.5 mg/mL stock solution of MTT in Neurobasal-A was added in wells containing primary hippocampal neuron cultures. The cells were then incubated for 3 h at 37 °C. Likewise, 100 μ L of a 1 mg/mL stock solution of MTT in DMEM complete medium was added to U87MG-containing wells, before incubating for 4 h at 37 °C. Removal of the medium was followed by the dilution of the cells in DMSO and the relative formazan concentration was calculated by measuring the absorbance at 540 nm (Tecan, Austria). Results were expressed in relation to the absorbance of the untreated cells (set at 100%). Three independent experiments were performed with six replicate wells for each condition. Cell death was also examined by phase-contrast microscopy (Carl Zeiss, Axiovert 25 CFL, Germany) using the same solutions as above. In each run, the effect of solutions of plain synthetic peptides and plain A β ₄₀ or A β ₄₂ was independently checked to serve as internal control. The viability measurement of primary cortical neurons was performed in a similar manner (n=6 independent experiments, with three replicates each).

6.17. Immunocytochemistry

Antibodies used in immunocytochemistry assays included anti-A β 6E10 (Covance, USA), a rabbit polyclonal C-terminal anti-APP [R1(57)] antibody (a kind gift from Dr. Pankj Mehta, Institute for Basic Research in Developmental Disabilities, Staten Island, New York) as primary antibodies (1:1000 in PBS, 10% normal goat serum (NGS), 1% Triton X-100). An Alexa Fluor 488-labeled anti-mouse IgG (ThermoFisher Scientific, USA) and a Cy3-labeled anti-rabbit IgG (ThermoFisher Scientific, USA) were

used as secondary antibodies (1:2,500 in PBS, 1% NGS, 1% Triton X-100). Treatment of primary mouse cortical neurons with vehicle, 1 μM $\text{A}\beta_{40}$, 1 μM $\text{A}\beta_{40}$ + 1 μM $\text{A}\beta\text{C5-34}$ or 1 μM $\text{A}\beta_{40}$ + 1 μM $\text{A}\beta\text{C5-116}$ was performed for 1 h at 37 °C. $\text{A}\beta_{40}$ had been pre-aggregated for 3 d at 37 °C with or without $\text{A}\beta\text{C5-34}$ or $\text{A}\beta\text{C5-116}$. Cell-free cover slips served in the assessment of non-specific binding of $\text{A}\beta$ to the glass. Following treatment, neurons were labeled with the mouse monoclonal anti- $\text{A}\beta$ antibody 6E10, to detect $\text{A}\beta$ binding on the cell surface. Additionally, neurons were also labeled with the rabbit polyclonal C-terminal anti-APP antibody R1(57), in order to distinguish $\text{A}\beta$ -specific labeling from full-length APP labeling. Following the removal of the medium, cells were rinsed twice with PBS and fixed in 4% paraformaldehyde (in PBS, pH 7.4) for 15 min at room temperature. Cells were then washed three times with PBS and blocking of non-specific binding was achieved with 10% normal goat serum (NGS), 0.1% Triton X-100 (in PBS) for 60 min at RT. Next, neurons were labeled by overnight incubation at 4 °C with 6E10 and R1(57) diluted in PBS, 10% NGS, 0.1% Triton X-100, twice. Rinsing with PBS was repeated three times, preceding a 1 h incubation at room temperature, with Alexa Fluor 488-labeled anti-mouse IgG and Cy3-labeled anti-rabbit IgG secondary antibodies (1:2,500) diluted in PBS, 1% NGS, 1% Triton X-100. After washing, DAPI (1:1000) was used to stain cell nuclei for 5 min and cells were used to mount coverslips. Finally, a Axio observer Z1 fluorescence microscope (Zeiss, Germany) and ZEN 2012 software was used to acquire images of neurons using a 40 \times objective lens. 15-20 photos were taken per experimental condition. Image acquisition settings for each fluorophore, were identical for all samples in an experiment.

6.18. *In vivo* assays in *C. elegans*.

The utilized *C. elegans* strains listed below, were maintained at 16 °C:

- CL2179:dvIs179 [myo-3p::GFP::3'UTR(long)+rol-6(su1006)] (<http://www.cgc.cbs.umn.edu/strain.php?id=26134>)

- CL2331: *dvl*s37 [*myo-3p::GFP::Aβ(3-42)+rol-6(su1006)*] (<http://www.cgc.cbs.umn.edu/strain.php?id=26135>);
- CL2006: *dvl*s2 [*Punc-54::Aβ(3-42); pRF4 (rol-6 (su1006))*] (<http://www.cgc.cbs.umn.edu/strain.php?id=7660>)³⁷⁶
- CL4176: *smg-1(cc546) I; dvl*s27 [*myo-3::Aβ(1-42)-let 3'UTR(pAF29); pRF4 (rol-6(su1006))*] (<http://www.cgc.cbs.umn.edu/strain.php?id=7663>)³⁷⁷

6.18.1. Treatment of *C. elegans* nematodes with selected cyclic peptides

Synthetic AβC5-34 and AβC5-116 treatment: Stock solutions of AβC5-34 and AβC5-116 peptides were created following the addition of DMSO and storage at -20 °C. The necessary volume of stock solution or DMSO (control) was added onto an *E. coli* OP50 bacterial lawn, to achieve the indicated cyclic peptide concentration.

Biosynthetically produced cyclic peptide treatment: *C. elegans* nematodes were provided with 200 μl of *E. coli* OP50 cultures that carried pSICLOPPS-AβC5-34, pSICLOPPS-AβC5-116 or pSICLOPPS-Random1. *E. coli* OP50 cells were transformed with the pSICLOPPS plasmids responsible for encoding AβC5-34 or AβC5-116 and protein production was induced. Following growth to an OD₆₀₀≈0.7, the bacteria were used to feed the worms. Nematodes were spread onto standard nematode growth medium (NGM) plates, containing 25 μg/mL chloramphenicol and 0.002% L-arabinose for the induction of cyclic peptide production.

In both cases, synchronized offspring were randomly distributed to treatment plates to avoid systematic differences in egg-lay batches. Treatment and control plates were handled, scored and assayed in parallel.

6.18.2. Analysis of cyclic-peptide-treated *C. elegans* nematodes

Paralysis assay: 90-120 synchronized CL2006 L4 larvae per condition were transferred onto NGM plates that contained live OP50 bacteria producing A β C5-34, A β C5-116 or Random1, and maintained at 20 °C. Likewise, 150-300 synchronized CL4176 animals per condition were transferred onto NGM plates that contained synthetic A β C5-34, A β C5-116 or 0.26% DMSO and maintained at 16 °C for 48 h prior to transgene induction via a temperature increase to 25 °C.

Paralysis: Scoring was commenced in the 1st day of adulthood for CL2006, and 24 h after the temperature up-shift for CL4176 nematodes. Paralysis was determined upon failure of nematodes to move their half end-body when prodded. Dead animals were excluded from scoring. Plates were numbered by an independent person and were subsequently scored. The evaluation of differences between paralysis curves and the estimation of *P* values for all independent data utilized the log-rank (Mantel–Cox) test. The letter “*n*” in the presented paralysis figures refers to the number of animals that paralyzed, divided by the total number of animals that were used, which includes paralyzed, deceased and censored animals.

Dot blot Analysis: CL4176 animals were allowed to lay eggs for 3 h on NGM plates that contained A β C5-34, A β C5-116 synthetic peptides or 0.26% DMSO (control). A temperature up-shift was used to induce paralysis and the progeny were again exposed to A β C5-34, A β C5-116 synthetic peptides or 0.26% DMSO, until 50% of the control population developed paralysis. Subsequently, the nematodes were collected and boiled in non-reducing Laemmli buffer. Dot blot analysis was performed using 1-5 μ g of protein lysates, and spotting it onto 0.2 μ m nitrocellulose membranes (Bio-Rad, USA). The membranes had been previously soaked in TBS and pre-heated at 80 °C. Immuno-staining of the blots was performed with two anti-A β antibodies: i) the 6E10 antibody, used to determine total A β aggregation/oligomerization and ii) the AB9234 antibody (Merck Millipore, Germany), which is specific for oligomeric A β species. Loading control was provided by the use of Actin, which was detected using

the anti-actin antibody sc-1615 (Santa Cruz, Germany), and used for normalization of the detected oligomeric and aggregated proteins.

Confocal microscopy analysis: In order to measure deposits of A β ₃₋₄₂, synchronized (L4 larval stage) CL2331 and CL2179 (control strain) nematodes were exposed to 10 μ M A β C5-34, 5 μ M A β C5-116 or 0.26% DMSO and subsequently grown at 20 °C, allowing for aggregation to occur. The nematodes were collected at the 2nd day of adulthood and subsequently mounted on 2% agarose pads on glass slides. Subsequently, they were anesthetized with 10 mM levamisole and observed at room temperature using a confocal laser-scanning microscope (Leica TCS SPE, Leica Lasertechnik GmbH, Germany). Three independent experiments was conducted, using \geq 20 animals per condition. Images of complete worms and of the posterior nematode area were acquired using 10 x 0.45 and 20 x 0.70 numerical apertures respectively.

6.19. High-throughput sequencing analysis

The plasmid vector library that was generated after FACS-enabled selection was digested with NcoI and BsrGI, resulting in a ~250 bp product that contained the variable peptide-encoding region and parts of the intein fusion. Following DNA extraction and purification, the digestion product underwent high-throughput sequencing analysis at the Genomics core facility of the Biomedical Sciences Research Center “Alexander Fleming” (Athens, Greece), using an Ion Torrent PROTON (Life Technologies, USA) high-throughput sequencing platform. Sequences with mismatches in the regions flanking the variable peptide-encoding region were excluded from the dataset and only the 12-, 15- or 18-bp-long peptide-encoding sequences were used for further analysis. All sequences containing stop-codons were excluded from the analysis. Sequence translation was performed using Bioedit (www.mbio.ncsu.edu/BioEdit/bioedit.html). Sequence alignments and phylogenetic trees were generated with the MEGA 7 software package (www.megasoftware.net).

6.20. Statistical analyses

Statistical analyses were performed with Prism (GraphPad Software Inc, USA) and Microsoft Office 2016 Excel (Microsoft, USA). Mean values were compared using unpaired t-tests to determine p-values. Population sizes of nematode experiments were determined based on previous laboratory experience (Lab of Dr. Niki Chondrogianni) and based on relevant literature as well. Accordingly, sufficient numbers of nematodes were used for the presented experiments. Finally, no samples or worms were excluded from the reported analyses, while the statistical analysis of next generation sequencing results included the complete dataset, as provided by the Genomics core facility of the Biomedical Sciences Research Center “Alexander Fleming” (Athens, Greece).

Appendix

Supplementary Tables

Supplementary Table 1: DNA primers used in this work.

Name	Primer sequence (5'-3')	Use
GS032	GGAATTCG <u>CCAATGGGGCGATCGCC</u> CACAA TTGC(NNS) ₃ TGCTTAAGTTTTGGC	Degenerate 5' primer for the construction of the CysX ₁ X ₂ X ₃ sub-library. BglI site is underlined.
GS033	GGAATTCG <u>CCAATGGGGCGATCGCC</u> CACAA TAGC(NNS) ₃ TGCTTAAGTTTTGGC	Degenerate 5' primer for the construction of the SerX ₁ X ₂ X ₃ sub-library. BglI site is underlined.
GS034	GGAATTCG <u>CCAATGGGGCGATCGCC</u> CACAA TACC(NNS) ₃ TGCTTAAGTTTTGGC	Degenerate 5' primer for the construction of the ThrX ₁ X ₂ X ₃ sub-library. BglI site is underlined.
GS072	GGAATTCG <u>CCAATGGGGCGATCGCC</u> CACAA TTGC(NNS) ₄ TGCTTAAGTTTTGGC	Degenerate 5' primer for the construction of the CysX ₁ X ₂ X ₃ X ₄ sub-library. BglI site is underlined.
GS073	GGAATTCG <u>CCAATGGGGCGATCGCC</u> CACAA TAGC(NNS) ₄ TGCTTAAGTTTTGGC	Degenerate 5' primer for the construction of the SerX ₁ X ₂ X ₃ X ₄ sub-library. BglI site is underlined.
GS074	GGAATTCG <u>CCAATGGGGCGATCGCC</u> CACAA TACC(NNS) ₄ TGCTTAAGTTTTGGC	Degenerate 5' primer for the construction of the ThrX ₁ X ₂ X ₃ X ₄ sub-library. BglI site is underlined.
GS075	GGAATTCG <u>CCAATGGGGCGATCGCC</u> CACAA TTGC(NNS) ₅ TGCTTAAGTTTTGGC	Degenerate 5' primer for the construction of the CysX ₁ X ₂ X ₃ X ₄ X ₅ sub-library. BglI site is underlined.
GS076	GGAATTCG <u>CCAATGGGGCGATCGCC</u> CACAA TAGC(NNS) ₅ TGCTTAAGTTTTGGC	Degenerate 5' primer for the construction of the SerX ₁ X ₂ X ₃ X ₄ X ₅ sub-library. BglI site is underlined.
GS077	GGAATTCG <u>CCAATGGGGCGATCGCC</u> CACAA TACC(NNS) ₅ TGCTTAAGTTTTGGC	Degenerate 5' primer for the construction of the ThrX ₁ X ₂ X ₃ X ₄ X ₅ sub-library. BglI site is underlined.
GS035	AAAAAA <u>AAGCTTTCATTGAAGCTGCC</u> CACAA GG	3' primer annealing to CBD. HindIII site is underlined.
GS069	AAAAAA <u>GCCAATGGGGCGATCGCC</u> CACAA TTGC	5' zipper primer for the construction of the Cys sub-libraries. BglI site is underlined.
GS070	AAAAAA <u>GCCAATGGGGCGATCGCC</u> CACAA TAGC	5' zipper primer for the construction of the Ser sub-libraries. BglI site is underlined.
GS071	AAAAAA <u>GCCAATGGGGCGATCGCC</u> CACAA TACC	5' zipper primer for the construction of the Thr sub-libraries. BglI site is underlined.
DD004	AAAAA <u>CCATGGATGCGGAATTCGCC</u> ATG	5' primer for the construction of pETA β ₄₂ . NcoI site is underlined.
IM060	CGCT <u>CTAGATTACGCAATCACCACGCC</u> CGC CGC	3' primer for the construction of pETA β ₄₂ . XbaI site is underlined.
IM033	CTAG <u>CCAATGGGGCGATCGCC</u> CACAATg _c GCCTCGCCGACGTGCTTAAGTTTTGGC	5' primer for the construction of pSICLOPPS-A β C5-34(S1C). Lower case indicates modification. BglI site is underlined.
IM034	CTAG <u>CCAATGGGGCGATCGCC</u> CACAATacc GCCTCGCCGACGTGCTTAAGTTTTGGC	5' primer for the construction of pSICLOPPS-A β C5-34(S1T). Lower case indicates modification. BglI site is underlined.
IM036	CTAG <u>CCAATGGGGCGATCGCC</u> CACAATAGC GCCg _c gCCGACGTGCTTAAGTTTTGGC	5' primer for the construction of pSICLOPPS-A β C5-34(S3A). Lower case indicates modification. BglI site is underlined.
IM037	CTAG <u>CCAATGGGGCGATCGCC</u> CACAATAGC GCCTCGg _c gACGTGCTTAAGTTTTGGC	5' primer for the construction of pSICLOPPS-A β C5-34(P4A). Lower case indicates modification. BglI site is underlined.

Name	Primer sequence (5'-3')	Use
IM038	CTAGCCAATGGGGCGATCGCCACAATAGC GCCTCGCCGgcgTGCTTAAGTTTTGGC	5' primer for the construction of pSICLOPPS-AβC5-34(T5A). Lower case indicates modification. BglI site is underlined.
IM027	CTGCTAGCCAATGGGGCGATCGCCACAAT tgcGCGTTCGACCGGTGCTTAAGTTTTGGC	5' primer for the construction of pSICLOPPS-AβC5-116(T1C). Lower case indicates modification. BglI site is underlined.
IM028	CTGCTAGCCAATGGGGCGATCGCCACAAT agcGCGTTCGACCGGTGCTTAAGTTTTGGC	5' primer for the construction of pSICLOPPS-AβC5-116(T1S). Lower case indicates modification. BglI site is underlined.
IM030	CTGCTAGCCAATGGGGCGATCGCCACAAT ACCGCGgcgGACCGGTGCTTAAGTTTTGGC	5' primer for the construction of pSICLOPPS-AβC5-116(F3A). Lower case indicates modification. BglI site is underlined.
IM031	CTGCTAGCCAATGGGGCGATCGCCACAAT ACCGCGTTCgcgGCGGTGCTTAAGTTTTGGC	5' primer for the construction of pSICLOPPS-AβC5-116(D4A). Lower case indicates modification. BglI site is underlined.
IM032	CTGCTAGCCAATGGGGCGATCGCCACAAT ACCGCGTTCGACgcgTGCTTAAGTTTTGGC	5' primer for the construction of pSICLOPPS-AβC5-116(R5A). Lower case indicates modification. BglI site is underlined.
IM043	CTAGCCAATGGGGCGATCGCCACAATACC tttTTCGACCGGTGCTTAAGTTTTGGC	5' primer for the construction of pSICLOPPS-AβC5-116(A2F). Lower case indicates modification. BglI site is underlined.
IM044	CTAGCCAATGGGGCGATCGCCACAATACC agcTTCGACCGGTGCTTAAGTTTTGGC	5' primer for the construction of pSICLOPPS-AβC5-116(A2S). Lower case indicates modification. BglI site is underlined.
IM045	CTAGCCAATGGGGCGATCGCCACAATACC ccgTTCGACCGGTGCTTAAGTTTTGGC	5' primer for the construction of pSICLOPPS-AβC5-116(A2P). Lower case indicates modification. BglI site is underlined.
IM046	CTAGCCAATGGGGCGATCGCCACAATACC accTTCGACCGGTGCTTAAGTTTTGGC	5' primer for the construction of pSICLOPPS-AβC5-116(A2T). Lower case indicates modification. BglI site is underlined.
IM047	CTAGCCAATGGGGCGATCGCCACAATACC tatTTCGACCGGTGCTTAAGTTTTGGC	5' primer for the construction of pSICLOPPS-AβC5-116(A2Y). Lower case indicates modification. BglI site is underlined.
IM048	CTAGCCAATGGGGCGATCGCCACAATACC catTTCGACCGGTGCTTAAGTTTTGGC	5' primer for the construction of pSICLOPPS-AβC5-116(A2H). Lower case indicates modification. BglI site is underlined.
IM049	CTAGCCAATGGGGCGATCGCCACAATACC aaaTTCGACCGGTGCTTAAGTTTTGGC	5' primer for the construction of pSICLOPPS-AβC5-116(A2K). Lower case indicates modification. BglI site is underlined.
IM050	CTAGCCAATGGGGCGATCGCCACAATACC gaaTTCGACCGGTGCTTAAGTTTTGGC	5' primer for the construction of pSICLOPPS-AβC5-116(A2E). Lower case indicates modification. BglI site is underlined.
IM051	CTAGCCAATGGGGCGATCGCCACAATACC tggTTCGACCGGTGCTTAAGTTTTGGC	5' primer for the construction of pSICLOPPS-AβC5-116(A2W). Lower case indicates modification. BglI site is underlined.
IM052	CTAGCCAATGGGGCGATCGCCACAATACC cgtTTCGACCGGTGCTTAAGTTTTGGC	5' primer for the construction of pSICLOPPS-AβC5-116(A2R). Lower case indicates modification. BglI site is underlined.
IM039	CTAGCCAATGGGGCGATCGCCACAATACC TTCGACCGGTGCTTAAGTTTTGGC	5' primer for the construction of pSICLOPPS-AβC5-116(A2del). BglI site is underlined.
IM040	CTAGCCAATGGGGCGATCGCCACAATACC GCGGACCGGTGCTTAAGTTTTGGC	5' primer for the construction of pSICLOPPS-AβC5-116 (F3del). BglI site is underlined.
IM041	CTAGCCAATGGGGCGATCGCCACAATACC GCGTTCGCGGTGCTTAAGTTTTGGC	5' primer for the construction of pSICLOPPS-AβC5-116(D4del). BglI site is underlined.

Name	Primer sequence (5'-3')	Use
IM077	CTAGCCAATGGGGCGATCGCCACAATaccg ccaccgtgcgtTGCTTAAGTTTTGGCACCGAAA TTTTAACCG	5' primer for the construction of pSICLOPPS-AβC5-479. Lower case indicates peptide DNA sequence. BglI site is underlined.
IM078	CTAGCCAATGGGGCGATCGCCACAATaccg cgatgtgcgtTGCTTAAGTTTTGGCACCGAAA TTTTAACCG	5' primer for the construction of pSICLOPPS-AβC5-359. Lower case indicates peptide DNA sequence. BglI site is underlined.
IM080	CTAGCCAATGGGGCGATCGCCACAATaccg tgtggattcgtTGCTTAAGTTTTGGCACCGAAA TTTTAACCG	5' primer for the construction of pSICLOPPS-AβC5-325. Lower case indicates peptide DNA sequence. BglI site is underlined.
IM081	CTAGCCAATGGGGCGATCGCCACAATaccg gccatgcgtTGCTTAAGTTTTGGCACCGAAA TTTTAACCG	5' primer for the construction of pSICLOPPS-AβC5-413. Lower case indicates peptide DNA sequence. BglI site is underlined.
GS037	CTATAACTATGGCTGGAATG	5' primer annealing to the pSICLOPPS backbone, before the 5'-end of the C-terminal domain of the Ssp DnaE intein.
DD015	TTTTTGGCCCATGGGCTAGCAGcATTAA GGTCTTGGGGAAGACCAATATC	3' primer for the H24L/F26A mutagenesis of the C-terminal domain of the Ssp DnaE intein (I _C). Lower case indicates modification. BglI site is underlined.

Supplementary Table 2. Plasmid vectors used in this work.

Plasmid	Encoded Protein	Marker	Origin of replication	Source
pETAβ ₄₂ -EGFP	Aβ ₄₂ -GFP	Kan ^R	ColE1	Prof. M. H. Hecht
pETAβ ₄₂	Met1-Aβ ₄₂	Kan ^R	ColE1	This work
pETAβ ₄₂ (F19S;L34P)-EGFP	Aβ ₄₂ (F19S;L34P)-GFP	Kan ^R	ColE1	Prof. M. H. Hecht
pETSOD1(A4V)-EGFP	SOD1(A4V)-GFP	Kan ^R	ColE1	S. Panoutsou
pETp53(Y220C)-EGFP	p53C(Y220C)-GFP	Kan ^R	ColE1	D.C. Delivoria
pSICLOPPS	I _C -SGGYLPPL-IN-CBD	Cm ^R	ACYC	Prof. S. Benkovic
pSICLOPPS-CysX ₁ X ₂ X ₃	I _C -CysX ₁ X ₂ X ₃ -IN-CBD sub-library	Cm ^R	ACYC	D.C. Delivoria
pSICLOPPS-SerX ₁ X ₂ X ₃	I _C -SerX ₁ X ₂ X ₃ -IN-CBD sub-library	Cm ^R	ACYC	D.C. Delivoria
pSICLOPPS-ThrX ₁ X ₂ X ₃	I _C -ThrX ₁ X ₂ X ₃ -IN-CBD sub-library	Cm ^R	ACYC	D.C. Delivoria
pSICLOPPS-CysX ₁ X ₂ X ₃ X ₄	I _C -CysX ₁ X ₂ X ₃ X ₄ -IN-CBD sub-library	Cm ^R	ACYC	D.C. Delivoria
pSICLOPPS-SerX ₁ X ₂ X ₃ X ₄	I _C -SerX ₁ X ₂ X ₃ X ₄ -IN-CBD sub-library	Cm ^R	ACYC	D.C. Delivoria
pSICLOPPS-ThrX ₁ X ₂ X ₃ X ₄	I _C -ThrX ₁ X ₂ X ₃ X ₄ -IN-CBD sub-library	Cm ^R	ACYC	D.C. Delivoria
pSICLOPPS-CysX ₁ X ₂ X ₃ X ₄ X ₅	I _C -CysX ₁ X ₂ X ₃ X ₄ X ₅ -IN-CBD sub-library	Cm ^R	ACYC	D.C. Delivoria
pSICLOPPS-SerX ₁ X ₂ X ₃ X ₄ X ₅	I _C -SerX ₁ X ₂ X ₃ X ₄ X ₅ -IN-CBD sub-library	Cm ^R	ACYC	D.C. Delivoria
pSICLOPPS-ThrX ₁ X ₂ X ₃ X ₄ X ₅	I _C -ThrX ₁ X ₂ X ₃ X ₄ X ₅ -IN-CBD sub-library	Cm ^R	ACYC	D.C. Delivoria
pSICLOPPS-AβC5-34	I _C -SASPT-IN-CBD	Cm ^R	ACYC	This work
pSICLOPPS-AβC5-34(S1C)	I _C -CASPT-IN-CBD	Cm ^R	ACYC	This work
pSICLOPPS-AβC5-34(S1T)	I _C -TASPT-IN-CBD	Cm ^R	ACYC	This work
pSICLOPPS-AβC5-34(S3A)	I _C -SAAPT-IN-CBD	Cm ^R	ACYC	This work
pSICLOPPS-AβC5-34(P4A)	I _C -SASAT-IN-CBD	Cm ^R	ACYC	This work
pSICLOPPS-AβC5-34(T5A)	I _C -SASPA-IN-CBD	Cm ^R	ACYC	This work

Plasmid	Encoded Protein	Marker	Origin of replication	Source
pSICLOPPS-AβC5-116	Ic-TAFDR-IN-CBD	Cm ^R	ACYC	This work
pSICLOPPS-AβC5-116(T1C)	Ic-CAFDR-IN-CBD	Cm ^R	ACYC	This work
pSICLOPPS-AβC5-116(T1S)	Ic-SAFDR-IN-CBD	Cm ^R	ACYC	This work
pSICLOPPS-AβC5-116(F3A)	Ic-TAADR-IN-CBD	Cm ^R	ACYC	This work
pSICLOPPS-AβC5-116(D4A)	Ic-TAFAR-IN-CBD	Cm ^R	ACYC	This work
pSICLOPPS-AβC5-116(R5A)	Ic-TAFDA-IN-CBD	Cm ^R	ACYC	This work
pSICLOPPS-AβC5-116(A2F)	Ic-TFFDR-IN-CBD	Cm ^R	ACYC	This work
pSICLOPPS-AβC5-116(A2W)	Ic-TWFDL-IN-CBD	Cm ^R	ACYC	This work
pSICLOPPS-AβC5-116(A2Y)	Ic-TYFDR-IN-CBD	Cm ^R	ACYC	This work
pSICLOPPS-AβC5-116(A2S)	Ic-TSFDR-IN-CBD	Cm ^R	ACYC	This work
pSICLOPPS-AβC5-116(A2T)	Ic-TTFDR-IN-CBD	Cm ^R	ACYC	This work
pSICLOPPS-AβC5-116(A2E)	Ic-TEFDR-IN-CBD	Cm ^R	ACYC	This work
pSICLOPPS-AβC5-116(A2R)	Ic-TRFDR-IN-CBD	Cm ^R	ACYC	This work
pSICLOPPS-AβC5-116(A2H)	Ic-THFDR-IN-CBD	Cm ^R	ACYC	This work
pSICLOPPS-AβC5-116(A2K)	Ic-TKFDR-IN-CBD	Cm ^R	ACYC	This work
pSICLOPPS-AβC5-116(A2P)	Ic-TPFDR-IN-CBD	Cm ^R	ACYC	This work
pSICLOPPS-AβC5-116(delA2)	Ic-TFDR-IN-CBD	Cm ^R	ACYC	This work
pSICLOPPS-AβC5-116(delF3)	Ic-TADR-IN-CBD	Cm ^R	ACYC	This work
pSICLOPPS-AβC5-116(delD4)	Ic-TAFR-IN-CBD	Cm ^R	ACYC	This work
pSICLOPPS-AβC5-325	Ic-TVWIR-IN-CBD	Cm ^R	ACYC	This work
pSICLOPPS-AβC5-359	Ic-TAMWR-IN-CBD	Cm ^R	ACYC	This work
pSICLOPPS-AβC5-413	Ic-TSHAR-IN-CBD	Cm ^R	ACYC	This work
pSICLOPPS-AβC5-479	Ic-TTTVR-IN-CBD	Cm ^R	ACYC	This work
pSICLOPPS-Random1	Ic-unknown peptide sequence1-IN-CBD	Cm ^R	ACYC	This work
pSICLOPPS-Random2	Ic-unknown peptide sequence2-IN-CBD	Cm ^R	ACYC	This work
pSICLOPPS(H24L;F26A)-AβC5-2	Ic(H24L;F26A)-TTYAR-IN-CBD	Cm ^R	ACYC	This work
pSICLOPPS(H24L;F26A)-AβC5-3	Ic(H24L;F26A)-TTVDR-IN-CBD	Cm ^R	ACYC	This work
pSICLOPPS(H24L;F26A)-AβC5-17	Ic(H24L;F26A)-TTTAR-IN-CBD	Cm ^R	ACYC	This work
pSICLOPPS(H24L;F26A)-AβC5-21	Ic(H24L;F26A)-TTWCR-IN-CBD	Cm ^R	ACYC	This work
pSICLOPPS(H24L;F26A)-AβC5-26	Ic(H24L;F26A)-TAWCR-IN-CBD	Cm ^R	ACYC	This work
pSICLOPPS(H24L;F26A)-AβC5-34	Ic(H24L;F26A)-SASPT-IN-CBD	Cm ^R	ACYC	This work
pSICLOPPS(H24L;F26A)-AβC5-116	Ic(H24L;F26A)-TAFDR-IN-CBD	Cm ^R	ACYC	This work
pSICLOPPS(H24L;F26A)-AβC6-1	Ic(H24L;F26A)-TPVWFD-IN-CBD	Cm ^R	ACYC	This work
pSICLOPPS(H24L;F26A)-Random1	Ic(H24L;F26A)-unknown peptide sequence1-IN-CBD	Cm ^R	ACYC	This work
pSICLOPPS(H24L;F26A)-Random2	Ic(H24L;F26A)- unknown peptide sequence2-IN-CBD	Cm ^R	ACYC	This work

Supplementary Table 3. Pentapeptide sequences identified in the sorted peptide pool more than 50 times.

Number	Peptide name	Amino acid sequence	Number of reads	Reads/Total pentapeptide reads (%)	Reads/Total peptide reads (%)
1	AβC5-1	T V E W L	466,621	11.493	10.299

Number	Peptide name	Amino acid sequence					Number of reads	Reads/Total pentapeptide reads (%)	Reads/Total peptide reads (%)
2	AβC5-2	T	T	Y	A	R	304,753	7.506	6.727
3	AβC5-3	T	T	V	D	R	214,461	5.282	4.734
4	AβC5-4	T	I	E	W	L	194,190	4.783	4.286
5	AβC5-5	T	T	T	W	R	175,510	4.323	3.874
6	AβC5-6	T	I	E	F	L	172,228	4.242	3.801
7	AβC5-7	T	T	L	H	R	134,018	3.301	2.958
8	AβC5-8	T	T	F	A	R	96,700	2.382	2.134
9	AβC5-9	T	V	L	D	R	89,669	2.209	1.979
10	AβC5-10	T	L	T	S	I	87,879	2.165	1.940
11	AβC5-11	T	V	N	R	L	73,050	1.799	1.612
12	AβC5-12	T	T	W	A	R	65,929	1.624	1.455
13	AβC5-13	T	A	L	D	R	62,792	1.547	1.386
14	AβC5-14	C	T	I	N	R	47,860	1.179	1.056
15	AβC5-15	T	A	N	V	R	47,855	1.179	1.056
16	AβC5-16	S	T	V	W	R	45,726	1.126	1.009
17	AβC5-17	T	T	T	A	R	40,135	0.989	0.886
18	AβC5-18	T	T	I	A	R	37,150	0.915	0.820
19	AβC5-19	T	V	W	D	R	37,091	0.914	0.819
20	AβC5-20	T	T	I	S	R	37,044	0.912	0.818
21	AβC5-21	T	T	W	C	R	36,295	0.894	0.801
22	AβC5-22	T	V	L	W	R	35,820	0.882	0.791
23	AβC5-23	C	L	T	F	I	33,203	0.818	0.733
24	AβC5-24	C	T	W	M	R	29,133	0.718	0.643
25	AβC5-25	T	T	L	A	R	28,989	0.714	0.640
26	AβC5-26	T	A	W	C	R	28,391	0.699	0.627
27	AβC5-27	T	T	S	A	R	28,188	0.694	0.622
28	AβC5-28	S	T	R	W	R	27,745	0.683	0.612
29	AβC5-29	T	T	L	E	R	27,514	0.678	0.607
30	AβC5-30	T	S	T	A	R	27,456	0.676	0.606
31	AβC5-31	C	T	F	A	R	26,842	0.661	0.592
32	AβC5-32	T	V	E	L	L	26,779	0.660	0.591
33	AβC5-33	C	T	T	W	R	26,068	0.642	0.575
34	AβC5-34	S	A	S	P	T	25,673	0.632	0.567
35	AβC5-35	T	V	R	D	R	25,428	0.626	0.561
36	AβC5-36	S	T	V	A	V	25,297	0.623	0.558
37	AβC5-37	C	T	S	V	R	24,446	0.602	0.540
38	AβC5-38	C	T	V	A	R	23,694	0.584	0.523
39	AβC5-39	S	V	L	W	R	23,559	0.580	0.520
40	AβC5-40	C	S	L	W	R	23,446	0.578	0.518
41	AβC5-41	T	G	W	A	R	21,784	0.537	0.481
42	AβC5-42	T	T	N	R	L	21,738	0.535	0.480
43	AβC5-43	C	T	F	M	R	21,532	0.530	0.475
44	AβC5-44	T	A	W	A	R	20,807	0.512	0.459
45	AβC5-45	T	T	W	V	R	20,798	0.512	0.459
46	AβC5-46	T	L	L	W	R	19,957	0.492	0.440
47	AβC5-47	T	T	I	D	R	19,735	0.486	0.436
48	AβC5-48	C	A	E	V	R	19,588	0.482	0.432
49	AβC5-49	C	V	S	W	R	19,498	0.480	0.430
50	AβC5-50	T	A	L	A	R	19,433	0.479	0.429

Number	Peptide name	Amino acid sequence					Number of reads	Reads/Total pentapeptide reads (%)	Reads/Total peptide reads (%)
51	AβC5-51	T	S	V	D	R	19,249	0.474	0.425
52	AβC5-52	C	V	E	W	R	19,212	0.473	0.424
53	AβC5-53	T	T	V	W	R	18,669	0.460	0.412
54	AβC5-54	T	I	E	C	L	17,758	0.437	0.392
55	AβC5-55	C	V	I	V	Q	17,597	0.433	0.388
56	AβC5-56	T	L	D	H	L	16,964	0.418	0.374
57	AβC5-57	T	V	V	V	Q	16,255	0.400	0.359
58	AβC5-58	T	I	E	W	V	16,172	0.398	0.357
59	AβC5-59	C	T	D	T	I	16,157	0.398	0.357
60	AβC5-60	T	I	E	L	L	15,931	0.392	0.352
61	AβC5-61	T	W	F	E	L	15,763	0.388	0.348
62	AβC5-62	C	L	V	L	Q	15,318	0.377	0.338
63	AβC5-63	S	V	I	W	R	15,107	0.372	0.333
64	AβC5-64	C	T	Y	C	R	14,817	0.365	0.327
65	AβC5-65	C	T	Y	T	I	14,804	0.365	0.327
66	AβC5-66	T	T	H	W	R	14,304	0.352	0.316
67	AβC5-67	T	A	R	D	R	14,213	0.350	0.314
68	AβC5-68	T	V	D	Y	L	13,544	0.334	0.299
69	AβC5-69	S	T	I	D	L	13,519	0.333	0.298
70	AβC5-70	C	L	D	Q	L	13,481	0.332	0.298
71	AβC5-71	T	L	D	A	L	13,390	0.330	0.296
72	AβC5-72	C	V	T	W	R	13,345	0.329	0.295
73	AβC5-73	T	T	R	D	R	12,894	0.318	0.285
74	AβC5-74	C	L	E	F	I	12,664	0.312	0.280
75	AβC5-75	T	L	V	S	I	12,511	0.308	0.276
76	AβC5-76	C	T	L	W	R	12,138	0.299	0.268
77	AβC5-77	C	T	M	C	I	11,870	0.292	0.262
78	AβC5-78	C	T	W	E	R	11,484	0.283	0.253
79	AβC5-79	C	L	V	V	Q	10,918	0.269	0.241
80	AβC5-80	T	S	V	H	R	10,181	0.251	0.225
81	AβC5-81	C	T	V	M	I	9,813	0.242	0.217
82	AβC5-82	T	A	V	W	R	9,781	0.241	0.216
83	AβC5-83	T	T	G	C	R	9,362	0.231	0.207
84	AβC5-84	C	T	F	H	R	9,276	0.228	0.205
85	AβC5-85	T	V	V	V	F	8,874	0.219	0.196
86	AβC5-86	C	V	L	H	R	8,343	0.205	0.184
87	AβC5-87	S	A	L	Y	V	8,336	0.205	0.184
88	AβC5-88	C	V	W	W	R	8,074	0.199	0.178
89	AβC5-89	T	A	T	D	R	7,984	0.197	0.176
90	AβC5-90	C	T	A	Y	M	7,971	0.196	0.176
91	AβC5-91	T	L	V	T	V	7,939	0.196	0.175
92	AβC5-92	C	V	W	V	R	7,844	0.193	0.173
93	AβC5-93	C	V	S	H	R	7,742	0.191	0.171
94	AβC5-94	T	V	L	F	R	7,442	0.183	0.164
95	AβC5-95	C	T	V	W	V	7,262	0.179	0.160
96	AβC5-96	S	S	V	W	R	7,185	0.177	0.159
97	AβC5-97	C	V	R	V	R	7,147	0.176	0.158
98	AβC5-98	C	T	M	W	R	7,120	0.175	0.157
99	AβC5-99	S	A	V	H	R	6,481	0.160	0.143

Number	Peptide name	Amino acid sequence	Number of reads	Reads/Total pentapeptide reads (%)	Reads/Total peptide reads (%)
100	AβC5-100	C I F W R	6,400	0.158	0.141
101	AβC5-101	T T E T I	6,290	0.155	0.139
102	AβC5-102	T T Y N R	6,067	0.149	0.134
103	AβC5-103	C I I L N	6,062	0.149	0.134
104	AβC5-104	T T T A V	5,517	0.136	0.122
105	AβC5-105	T V R W R	5,450	0.134	0.120
106	AβC5-106	C T T A R	5,430	0.134	0.120
107	AβC5-107	C P M H L	5,161	0.127	0.114
108	AβC5-108	T V W V Q	5,058	0.125	0.112
109	AβC5-109	S C V W R	4,998	0.123	0.110
110	AβC5-110	C A W A R	4,990	0.123	0.110
111	AβC5-111	C T C W V	4,963	0.122	0.110
112	AβC5-112	C S W M R	4,903	0.121	0.108
113	AβC5-113	S V V W R	4,900	0.121	0.108
114	AβC5-114	C T W H L	4,808	0.118	0.106
115	AβC5-115	S L V T V	4,677	0.115	0.103
116	AβC5-116	T A F D R	4,243	0.105	0.094
117	AβC5-117	T T R C R	4,237	0.104	0.094
118	AβC5-118	T T F W R	4,216	0.104	0.093
119	AβC5-119	T T V T L	4,204	0.104	0.093
120	AβC5-120	S T L H R	4,167	0.103	0.092
121	AβC5-121	T I K D R	3,970	0.098	0.088
122	AβC5-122	C A T A R	3,420	0.084	0.075
123	AβC5-123	T T V H R	3,371	0.083	0.074
124	AβC5-124	C T T M R	3,129	0.077	0.069
125	AβC5-125	C T W V R	3,050	0.075	0.067
126	AβC5-126	T T L L R	3,016	0.074	0.067
127	AβC5-127	S L V T A	2,907	0.072	0.064
128	AβC5-128	C T S A R	2,885	0.071	0.064
129	AβC5-129	T T L F R	2,630	0.065	0.058
130	AβC5-130	T A Y H R	2,594	0.064	0.057
131	AβC5-131	T W P V L	2,534	0.062	0.056
132	AβC5-132	C F I L Q	2,361	0.058	0.052
133	AβC5-133	C T R M R	2,170	0.053	0.048
134	AβC5-134	T P L W L	2,133	0.053	0.047
135	AβC5-135	T L L T L	2,120	0.052	0.047
136	AβC5-136	T A L H R	2,026	0.050	0.045
137	AβC5-137	C T M W V	2,009	0.049	0.044
138	AβC5-138	C T W L R	1,951	0.048	0.043
139	AβC5-139	T T S P R	1,904	0.047	0.042
140	AβC5-140	S S W A R	1,852	0.046	0.041
141	AβC5-141	S C W C R	1,754	0.043	0.039
142	AβC5-142	T Y P D L	1,748	0.043	0.039
143	AβC5-143	C S T W R	1,683	0.041	0.037
144	AβC5-144	T W P D L	1,682	0.041	0.037
145	AβC5-145	S A V W R	1,637	0.040	0.036
146	AβC5-146	T T W S R	1,612	0.040	0.036
147	AβC5-147	T A M H R	1,611	0.040	0.036
148	AβC5-148	C I E A V	1,600	0.039	0.035

Number	Peptide name	Amino acid sequence	Number of reads	Reads/Total pentapeptide reads (%)	Reads/Total peptide reads (%)
149	AβC5-149	C A V H R	1,576	0.039	0.035
150	AβC5-150	C C I A F	1,507	0.037	0.033
151	AβC5-151	C P Q H I	1,498	0.037	0.033
152	AβC5-152	T A N R L	1,452	0.036	0.032
153	AβC5-153	C I D W M	1,345	0.033	0.030
154	AβC5-154	S A V S L	1,309	0.032	0.029
155	AβC5-155	T S L D R	1,251	0.031	0.028
156	AβC5-156	S V Y W I	1,196	0.029	0.026
157	AβC5-157	S V L T I	1,180	0.029	0.026
158	AβC5-158	T T G A R	1,172	0.029	0.026
159	AβC5-159	C T C H R	1,143	0.028	0.025
160	AβC5-160	C V V W R	1,133	0.028	0.025
161	AβC5-161	T T V T I	1,126	0.028	0.025
162	AβC5-162	T S V W R	1,094	0.027	0.024
163	AβC5-163	T T W T V	1,071	0.026	0.024
164	AβC5-164	C M V V F	1,058	0.026	0.023
165	AβC5-165	T I T T L	1,053	0.026	0.023
166	AβC5-166	C T M A R	1,022	0.025	0.023
167	AβC5-167	C T I H R	1,012	0.025	0.022
168	AβC5-168	S T I N R	1,000	0.025	0.022
169	AβC5-169	C V I L Q	1,000	0.025	0.022
170	AβC5-170	C A M H I	966	0.024	0.021
171	AβC5-171	C A Q W R	960	0.024	0.021
172	AβC5-172	C W S A Q	960	0.024	0.021
173	AβC5-173	T T H A R	953	0.023	0.021
174	AβC5-174	S T L W L	948	0.023	0.021
175	AβC5-175	T C V T V	946	0.023	0.021
176	AβC5-176	T A G W R	945	0.023	0.021
177	AβC5-177	T A T A R	925	0.023	0.020
178	AβC5-178	C A V V Q	891	0.022	0.020
179	AβC5-179	T I V V F	884	0.022	0.020
180	AβC5-180	T I D F L	869	0.021	0.019
181	AβC5-181	C C M W R	846	0.021	0.019
182	AβC5-182	T L S H L	831	0.020	0.018
183	AβC5-183	C T I R R	830	0.020	0.018
184	AβC5-184	T V L A R	818	0.020	0.018
185	AβC5-185	T T F N R	800	0.020	0.018
186	AβC5-186	C C A W R	786	0.019	0.017
187	AβC5-187	C A R A R	773	0.019	0.017
188	AβC5-188	T G M R R	768	0.019	0.017
189	AβC5-189	T T V A R	757	0.019	0.017
190	AβC5-190	C L R T L	739	0.018	0.016
191	AβC5-191	T T V T V	736	0.018	0.016
192	AβC5-192	T G L A R	720	0.018	0.016
193	AβC5-193	T T T E V	709	0.017	0.016
194	AβC5-194	S A F F R	703	0.017	0.016
195	AβC5-195	C T C W N	683	0.017	0.015
196	AβC5-196	C S V F I	682	0.017	0.015
197	AβC5-197	T I D V V	680	0.017	0.015

Number	Peptide name	Amino acid sequence					Number of reads	Reads/Total pentapeptide reads (%)	Reads/Total peptide reads (%)
198	AβC5-198	T	S	W	C	R	679	0.017	0.015
199	AβC5-199	S	A	M	W	R	678	0.017	0.015
200	AβC5-200	C	I	D	W	I	661	0.016	0.015
201	AβC5-201	T	L	A	F	I	657	0.016	0.015
202	AβC5-202	C	T	M	M	R	648	0.016	0.014
203	AβC5-203	C	G	Y	P	V	648	0.016	0.014
204	AβC5-204	S	V	W	C	R	640	0.016	0.014
205	AβC5-205	C	I	G	W	I	616	0.015	0.014
206	AβC5-206	S	A	W	W	R	609	0.015	0.013
207	AβC5-207	C	I	G	W	R	605	0.015	0.013
208	AβC5-208	C	V	E	W	V	582	0.014	0.013
209	AβC5-209	T	T	R	A	R	580	0.014	0.013
210	AβC5-210	C	V	L	L	R	576	0.014	0.013
211	AβC5-211	T	P	E	T	L	565	0.014	0.012
212	AβC5-212	T	I	A	W	L	538	0.013	0.012
213	AβC5-213	C	V	K	F	R	532	0.013	0.012
214	AβC5-214	T	V	M	T	V	524	0.013	0.012
215	AβC5-215	T	T	P	W	R	524	0.013	0.012
216	AβC5-216	C	P	T	S	I	516	0.013	0.011
217	AβC5-217	S	L	V	T	L	509	0.013	0.011
218	AβC5-218	T	V	L	H	R	497	0.012	0.011
219	AβC5-219	C	V	E	L	L	489	0.012	0.011
220	AβC5-220	T	L	T	A	L	482	0.012	0.011
221	AβC5-221	T	I	E	S	L	480	0.012	0.011
222	AβC5-222	T	L	M	T	V	474	0.012	0.010
223	AβC5-223	T	G	L	D	R	464	0.011	0.010
224	AβC5-224	C	V	L	H	I	461	0.011	0.010
225	AβC5-225	C	T	Y	A	L	455	0.011	0.010
226	AβC5-226	T	T	W	T	G	454	0.011	0.010
227	AβC5-227	C	A	A	V	R	451	0.011	0.010
228	AβC5-228	C	S	L	H	I	447	0.011	0.010
229	AβC5-229	C	A	L	V	R	444	0.011	0.010
230	AβC5-230	T	T	S	D	R	442	0.011	0.010
231	AβC5-231	T	V	E	R	L	440	0.011	0.010
232	AβC5-232	T	V	T	H	I	429	0.011	0.009
233	AβC5-233	T	V	V	T	L	426	0.010	0.009
234	AβC5-234	C	C	R	V	R	425	0.010	0.009
235	AβC5-235	C	T	E	F	L	419	0.010	0.009
236	AβC5-236	T	P	T	T	L	408	0.010	0.009
237	AβC5-237	T	L	V	T	L	394	0.010	0.009
238	AβC5-238	C	T	C	W	L	386	0.010	0.009
239	AβC5-239	T	T	M	H	R	384	0.009	0.008
240	AβC5-240	C	S	W	I	R	382	0.009	0.008
241	AβC5-241	C	T	W	T	R	376	0.009	0.008
242	AβC5-242	T	T	S	T	R	376	0.009	0.008
243	AβC5-243	C	G	V	L	P	370	0.009	0.008
244	AβC5-244	T	T	R	V	R	366	0.009	0.008
245	AβC5-245	T	T	R	F	R	364	0.009	0.008
246	AβC5-246	S	A	L	W	R	357	0.009	0.008

Number	Peptide name	Amino acid sequence					Number of reads	Reads/Total pentapeptide reads (%)	Reads/Total peptide reads (%)
247	AβC5-247	C	T	L	Y	V	356	0.009	0.008
248	AβC5-248	T	T	T	H	R	339	0.008	0.007
249	AβC5-249	C	C	V	T	L	336	0.008	0.007
250	AβC5-250	T	H	A	W	R	334	0.008	0.007
251	AβC5-251	T	T	Y	A	G	332	0.008	0.007
252	AβC5-252	T	V	I	W	R	331	0.008	0.007
253	AβC5-253	T	T	W	F	R	327	0.008	0.007
254	AβC5-254	S	T	L	V	R	326	0.008	0.007
255	AβC5-255	T	T	S	R	R	325	0.008	0.007
256	AβC5-256	C	I	N	T	L	308	0.008	0.007
257	AβC5-257	T	I	E	W	S	301	0.007	0.007
258	AβC5-258	T	T	S	C	R	301	0.007	0.007
259	AβC5-259	C	V	L	V	R	297	0.007	0.007
260	AβC5-260	T	T	W	T	R	295	0.007	0.007
261	AβC5-261	C	L	S	T	L	290	0.007	0.006
262	AβC5-262	T	T	S	S	R	286	0.007	0.006
263	AβC5-263	T	H	L	A	R	284	0.007	0.006
264	AβC5-264	T	S	G	A	R	282	0.007	0.006
265	AβC5-265	T	I	D	V	L	278	0.007	0.006
266	AβC5-266	T	T	L	R	R	274	0.007	0.006
267	AβC5-267	S	V	T	T	V	272	0.007	0.006
268	AβC5-268	C	V	W	A	R	272	0.007	0.006
269	AβC5-269	C	T	T	C	R	270	0.007	0.006
270	AβC5-270	T	A	T	W	R	266	0.007	0.006
271	AβC5-271	T	L	G	W	L	260	0.006	0.006
272	AβC5-272	T	C	M	W	R	254	0.006	0.006
273	AβC5-273	T	L	P	W	L	252	0.006	0.006
274	AβC5-274	T	W	L	E	L	250	0.006	0.006
275	AβC5-275	T	A	H	V	R	249	0.006	0.005
276	AβC5-276	T	S	W	A	R	249	0.006	0.005
277	AβC5-277	C	S	T	V	R	247	0.006	0.005
278	AβC5-278	T	T	W	L	R	241	0.006	0.005
279	AβC5-279	T	A	F	W	V	240	0.006	0.005
280	AβC5-280	C	T	A	A	R	240	0.006	0.005
281	AβC5-281	T	A	E	W	L	238	0.006	0.005
282	AβC5-282	T	R	L	V	E	237	0.006	0.005
283	AβC5-283	T	V	D	A	V	233	0.006	0.005
284	AβC5-284	C	T	V	T	L	230	0.006	0.005
285	AβC5-285	C	A	M	T	I	226	0.006	0.005
286	AβC5-286	C	V	P	S	I	226	0.006	0.005
287	AβC5-287	C	T	T	L	I	221	0.005	0.005
288	AβC5-288	C	T	T	V	R	220	0.005	0.005
289	AβC5-289	T	I	G	N	L	218	0.005	0.005
290	AβC5-290	T	T	W	R	G	213	0.005	0.005
291	AβC5-291	T	T	L	D	R	213	0.005	0.005
292	AβC5-292	C	V	E	W	L	212	0.005	0.005
293	AβC5-293	T	S	S	G	L	211	0.005	0.005
294	AβC5-294	T	V	E	W	P	207	0.005	0.005
295	AβC5-295	T	V	E	S	L	207	0.005	0.005

Number	Peptide name	Amino acid sequence					Number of reads	Reads/Total pentapeptide reads (%)	Reads/Total peptide reads (%)
296	AβC5-296	T	T	P	H	R	207	0.005	0.005
297	AβC5-297	T	C	V	T	I	204	0.005	0.005
298	AβC5-298	T	T	R	G	R	201	0.005	0.004
299	AβC5-299	T	T	V	G	R	200	0.005	0.004
300	AβC5-300	T	R	R	V	V	197	0.005	0.004
301	AβC5-301	T	T	T	R	R	191	0.005	0.004
302	AβC5-302	T	Q	Q	S	Q	185	0.005	0.004
303	AβC5-303	T	I	D	V	S	184	0.005	0.004
304	AβC5-304	T	S	I	N	R	182	0.004	0.004
305	AβC5-305	T	T	A	D	R	181	0.004	0.004
306	AβC5-306	T	H	R	V	L	179	0.004	0.004
307	AβC5-307	C	S	W	A	R	175	0.004	0.004
308	AβC5-308	C	S	E	Y	V	173	0.004	0.004
309	AβC5-309	T	L	E	W	L	172	0.004	0.004
310	AβC5-310	T	V	G	W	L	170	0.004	0.004
311	AβC5-311	T	I	T	F	L	167	0.004	0.004
312	AβC5-312	T	T	S	P	G	167	0.004	0.004
313	AβC5-313	T	L	C	T	I	162	0.004	0.004
314	AβC5-314	T	I	D	C	L	160	0.004	0.004
315	AβC5-315	T	T	S	E	R	158	0.004	0.003
316	AβC5-316	T	T	C	A	R	157	0.004	0.003
317	AβC5-317	T	T	A	W	R	156	0.004	0.003
318	AβC5-318	S	L	D	L	I	155	0.004	0.003
319	AβC5-319	T	G	R	G	G	153	0.004	0.003
320	AβC5-320	T	T	V	E	R	150	0.004	0.003
321	AβC5-321	T	T	T	F	R	148	0.004	0.003
322	AβC5-322	T	I	E	F	P	147	0.004	0.003
323	AβC5-323	T	A	V	D	R	147	0.004	0.003
324	AβC5-324	C	P	C	Y	L	146	0.004	0.003
325	AβC5-325	T	V	W	I	R	144	0.004	0.003
326	AβC5-326	C	A	A	W	R	143	0.004	0.003
327	AβC5-327	T	I	E	F	V	141	0.003	0.003
328	AβC5-328	T	R	P	V	E	141	0.003	0.003
329	AβC5-329	T	T	V	R	R	141	0.003	0.003
330	AβC5-330	C	I	M	T	I	139	0.003	0.003
331	AβC5-331	T	W	S	G	L	139	0.003	0.003
332	AβC5-332	T	V	G	V	L	138	0.003	0.003
333	AβC5-333	T	H	V	R	R	137	0.003	0.003
334	AβC5-334	C	T	P	Y	R	135	0.003	0.003
335	AβC5-335	T	R	R	V	L	134	0.003	0.003
336	AβC5-336	S	H	L	A	R	133	0.003	0.003
337	AβC5-337	C	T	V	V	R	132	0.003	0.003
338	AβC5-338	T	I	D	W	L	132	0.003	0.003
339	AβC5-339	T	P	A	S	I	130	0.003	0.003
340	AβC5-340	C	V	I	V	R	126	0.003	0.003
341	AβC5-341	T	I	V	A	W	125	0.003	0.003
342	AβC5-342	T	C	I	V	F	125	0.003	0.003
343	AβC5-343	T	N	L	D	R	125	0.003	0.003
344	AβC5-344	T	T	P	G	R	125	0.003	0.003

Number	Peptide name	Amino acid sequence					Number of reads	Reads/Total pentapeptide reads (%)	Reads/Total peptide reads (%)
345	AβC5-345	T	V	E	W	S	124	0.003	0.003
346	AβC5-346	S	H	L	D	R	123	0.003	0.003
347	AβC5-347	C	M	T	H	L	122	0.003	0.003
348	AβC5-348	T	T	L	T	R	119	0.003	0.003
349	AβC5-349	S	M	L	S	D	118	0.003	0.003
350	AβC5-350	T	H	R	V	V	118	0.003	0.003
351	AβC5-351	T	V	E	W	F	117	0.003	0.003
352	AβC5-352	T	V	G	V	V	117	0.003	0.003
353	AβC5-353	T	I	D	W	W	115	0.003	0.003
354	AβC5-354	T	I	E	S	V	115	0.003	0.003
355	AβC5-355	T	A	T	V	R	115	0.003	0.003
356	AβC5-356	C	V	R	I	R	114	0.003	0.003
357	AβC5-357	C	C	T	W	R	113	0.003	0.002
358	AβC5-358	T	R	E	W	L	111	0.003	0.002
359	AβC5-359	T	A	M	W	R	110	0.003	0.002
360	AβC5-360	C	P	C	H	L	109	0.003	0.002
361	AβC5-361	T	T	K	W	R	108	0.003	0.002
362	AβC5-362	T	T	W	D	R	107	0.003	0.002
363	AβC5-363	S	T	I	V	R	106	0.003	0.002
364	AβC5-364	T	T	M	A	R	106	0.003	0.002
365	AβC5-365	T	T	G	G	R	106	0.003	0.002
366	AβC5-366	T	L	P	S	I	105	0.003	0.002
367	AβC5-367	T	T	M	V	R	105	0.003	0.002
368	AβC5-368	S	T	C	G	G	104	0.003	0.002
369	AβC5-369	S	N	L	W	R	104	0.003	0.002
370	AβC5-370	T	V	S	G	C	104	0.003	0.002
371	AβC5-371	C	T	M	F	L	101	0.002	0.002
372	AβC5-372	C	L	T	L	I	100	0.002	0.002
373	AβC5-373	C	H	F	V	T	98	0.002	0.002
374	AβC5-374	C	A	L	W	R	97	0.002	0.002
375	AβC5-375	T	N	L	A	R	97	0.002	0.002
376	AβC5-376	T	I	E	R	L	96	0.002	0.002
377	AβC5-377	T	I	R	D	R	96	0.002	0.002
378	AβC5-378	T	T	T	G	R	96	0.002	0.002
379	AβC5-379	T	R	L	G	R	95	0.002	0.002
380	AβC5-380	T	P	T	S	I	94	0.002	0.002
381	AβC5-381	T	T	H	T	R	93	0.002	0.002
382	AβC5-382	C	T	T	T	I	92	0.002	0.002
383	AβC5-383	T	T	I	T	R	92	0.002	0.002
384	AβC5-384	T	T	Y	T	R	90	0.002	0.002
385	AβC5-385	T	T	L	Y	R	90	0.002	0.002
386	AβC5-386	S	P	M	H	L	89	0.002	0.002
387	AβC5-387	S	H	S	P	T	89	0.002	0.002
388	AβC5-388	S	H	L	H	R	89	0.002	0.002
389	AβC5-389	T	H	L	D	R	89	0.002	0.002
390	AβC5-390	S	V	C	W	I	88	0.002	0.002
391	AβC5-391	T	L	L	I	R	88	0.002	0.002
392	AβC5-392	T	T	C	D	R	87	0.002	0.002
393	AβC5-393	T	T	G	R	R	87	0.002	0.002

Number	Peptide name	Amino acid sequence					Number of reads	Reads/Total pentapeptide reads (%)	Reads/Total peptide reads (%)
394	AβC5-394	T	T	V	S	R	86	0.002	0.002
395	AβC5-395	T	T	Q	H	R	85	0.002	0.002
396	AβC5-396	T	T	T	P	R	84	0.002	0.002
397	AβC5-397	S	V	E	V	L	83	0.002	0.002
398	AβC5-398	T	P	C	G	G	82	0.002	0.002
399	AβC5-399	T	P	R	P	T	82	0.002	0.002
400	AβC5-400	T	A	F	A	R	82	0.002	0.002
401	AβC5-401	C	I	V	V	Q	81	0.002	0.002
402	AβC5-402	C	L	T	S	I	81	0.002	0.002
403	AβC5-403	S	T	L	W	R	81	0.002	0.002
404	AβC5-404	S	T	D	F	I	80	0.002	0.002
405	AβC5-405	T	T	S	H	R	78	0.002	0.002
406	AβC5-406	T	I	E	V	S	77	0.002	0.002
407	AβC5-407	S	P	R	L	Q	76	0.002	0.002
408	AβC5-408	S	T	W	V	R	76	0.002	0.002
409	AβC5-409	T	R	Q	S	Q	76	0.002	0.002
410	AβC5-410	T	V	L	G	R	76	0.002	0.002
411	AβC5-411	T	V	E	F	L	75	0.002	0.002
412	AβC5-412	T	T	Q	R	R	75	0.002	0.002
413	AβC5-413	T	T	V	T	G	74	0.002	0.002
414	AβC5-414	T	S	H	A	R	74	0.002	0.002
415	AβC5-415	T	T	T	C	R	74	0.002	0.002
416	AβC5-416	T	H	E	A	V	73	0.002	0.002
417	AβC5-417	T	S	S	P	A	73	0.002	0.002
418	AβC5-418	T	T	V	G	P	73	0.002	0.002
419	AβC5-419	T	H	E	F	L	73	0.002	0.002
420	AβC5-420	T	V	S	R	L	73	0.002	0.002
421	AβC5-421	S	A	S	P	A	72	0.002	0.002
422	AβC5-422	T	M	E	W	L	72	0.002	0.002
423	AβC5-423	T	A	W	R	R	72	0.002	0.002
424	AβC5-424	T	V	C	S	I	71	0.002	0.002
425	AβC5-425	T	N	Q	F	L	70	0.002	0.002
426	AβC5-426	T	T	T	A	Q	70	0.002	0.002
427	AβC5-427	T	T	S	T	G	70	0.002	0.002
428	AβC5-428	T	T	C	G	R	69	0.002	0.002
429	AβC5-429	C	T	T	Y	R	68	0.002	0.002
430	AβC5-430	T	T	T	D	I	68	0.002	0.002
431	AβC5-431	S	P	E	A	I	68	0.002	0.002
432	AβC5-432	T	T	T	W	W	67	0.002	0.001
433	AβC5-433	T	V	D	V	V	66	0.002	0.001
434	AβC5-434	T	T	S	G	R	65	0.002	0.001
435	AβC5-435	S	R	R	V	R	64	0.002	0.001
436	AβC5-436	T	V	N	A	L	64	0.002	0.001
437	AβC5-437	C	T	L	W	N	62	0.002	0.001
438	AβC5-438	T	T	T	S	R	62	0.002	0.001
439	AβC5-439	T	I	G	V	V	61	0.002	0.001
440	AβC5-440	T	A	T	G	R	61	0.002	0.001
441	AβC5-441	T	A	W	D	R	61	0.002	0.001
442	AβC5-442	T	I	E	L	S	60	0.001	0.001

Number	Peptide name	Amino acid sequence				Number of reads	Reads/Total pentapeptide reads (%)	Reads/Total peptide reads (%)	
443	AβC5-443	T	H	A	V	T	60	0.001	0.001
444	AβC5-444	T	T	H	H	R	60	0.001	0.001
445	AβC5-445	T	T	L	H	W	59	0.001	0.001
446	AβC5-446	T	T	T	L	E	59	0.001	0.001
447	AβC5-447	S	P	R	R	Q	59	0.001	0.001
448	AβC5-448	T	A	Y	A	R	58	0.001	0.001
449	AβC5-449	T	A	N	A	R	58	0.001	0.001
450	AβC5-450	T	T	A	A	P	58	0.001	0.001
451	AβC5-451	T	R	D	V	R	58	0.001	0.001
452	AβC5-452	T	H	V	D	R	58	0.001	0.001
453	AβC5-453	T	Q	L	S	Q	57	0.001	0.001
454	AβC5-454	T	L	F	W	R	57	0.001	0.001
455	AβC5-455	T	P	D	A	I	56	0.001	0.001
456	AβC5-456	T	T	T	W	G	56	0.001	0.001
457	AβC5-457	C	F	D	L	I	55	0.001	0.001
458	AβC5-458	T	H	R	L	L	55	0.001	0.001
459	AβC5-459	T	T	A	A	R	55	0.001	0.001
460	AβC5-460	T	T	E	F	L	54	0.001	0.001
461	AβC5-461	C	T	F	T	R	54	0.001	0.001
462	AβC5-462	T	S	E	W	L	54	0.001	0.001
463	AβC5-463	T	V	V	D	R	54	0.001	0.001
464	AβC5-464	T	T	P	A	R	54	0.001	0.001
465	AβC5-465	T	T	R	G	P	53	0.001	0.001
466	AβC5-466	S	S	L	W	R	53	0.001	0.001
467	AβC5-467	T	Q	P	V	T	53	0.001	0.001
468	AβC5-468	C	V	M	V	R	53	0.001	0.001
469	AβC5-469	T	T	I	G	R	53	0.001	0.001
470	AβC5-470	T	T	Y	A	K	52	0.001	0.001
471	AβC5-471	C	T	P	W	R	51	0.001	0.001
472	AβC5-472	T	M	Y	A	R	51	0.001	0.001
473	AβC5-473	T	H	V	A	R	51	0.001	0.001
474	AβC5-474	T	T	W	P	R	51	0.001	0.001
475	AβC5-475	T	T	V	D	P	51	0.001	0.001
476	AβC5-476	T	T	G	D	R	51	0.001	0.001
477	AβC5-477	C	G	A	W	R	50	0.001	0.001
478	AβC5-478	C	V	T	F	R	50	0.001	0.001
479	AβC5-479	T	Q	S	A	Q	50	0.001	0.001
480	AβC5-480	T	T	T	V	R	50	0.001	0.001
481	AβC5-481	C	L	D	L	I	50	0.001	0.001
482	AβC5-482	T	V	F	G	R	50	0.001	0.001
483	AβC5-483	T	R	V	G	R	50	0.001	0.001
		Sum				4,059,907	100.000	89.611	

Supplementary Table 4. Hexapeptide sequences identified in the sorted peptide pool more than 50 times.

Number	Peptide name	Amino acid sequence						Number of reads	Reads/Total hexapeptide reads (%)	Reads/Total peptide reads (%)
1	AβC6-1	T	P	V	W	F	D	131,935	29.151	2.912
2	AβC6-2	T	P	A	W	F	D	111,132	24.555	2.453
3	AβC6-3	T	V	T	S	V	L	44,094	9.743	0.973
4	AβC6-4	T	L	E	F	F	D	27,057	5.978	0.597
5	AβC6-5	T	Y	T	T	T	I	21,522	4.755	0.475
6	AβC6-6	T	V	T	W	F	D	17,100	3.778	0.377
7	AβC6-7	S	W	V	W	C	R	15,349	3.391	0.339
8	AβC6-8	T	L	L	I	R	W	13,135	2.902	0.290
9	AβC6-9	T	L	T	T	I	I	12,182	2.692	0.269
10	AβC6-10	T	L	K	W	L	N	11,016	2.434	0.243
11	AβC6-11	S	M	L	S	D	C	5,696	1.259	0.126
12	AβC6-12	T	S	L	V	A	L	5,598	1.237	0.124
13	AβC6-13	S	R	V	S	V	V	3,810	0.842	0.084
14	AβC6-14	T	T	T	T	V	V	3,708	0.819	0.082
15	AβC6-15	C	I	S	V	R	L	3,371	0.745	0.074
16	AβC6-16	T	I	V	R	Q	A	2,321	0.513	0.051
17	AβC6-17	C	V	I	V	R	T	1,708	0.377	0.038
18	AβC6-18	S	V	T	S	L	V	1,492	0.330	0.033
19	AβC6-19	T	P	T	T	V	L	1,428	0.316	0.032
20	AβC6-20	T	V	V	R	E	V	1,263	0.279	0.028
21	AβC6-21	T	K	E	Y	F	D	1,231	0.272	0.027
22	AβC6-22	T	L	T	T	L	V	1,028	0.227	0.023
23	AβC6-23	T	V	A	F	S	T	905	0.200	0.020
24	AβC6-24	S	G	L	C	E	L	766	0.169	0.017
25	AβC6-25	S	I	V	S	L	V	766	0.169	0.017
26	AβC6-26	T	L	H	W	F	E	647	0.143	0.014
27	AβC6-27	T	C	S	W	F	D	623	0.138	0.014
28	AβC6-28	T	L	E	Y	F	M	556	0.123	0.012
29	AβC6-29	T	I	A	Q	F	L	528	0.117	0.012
30	AβC6-30	T	F	T	S	L	L	528	0.117	0.012
31	AβC6-31	T	I	T	S	E	I	487	0.108	0.011
32	AβC6-32	T	L	C	W	L	N	455	0.101	0.010
33	AβC6-33	T	L	L	V	R	K	405	0.089	0.009
34	AβC6-34	C	V	A	Q	R	K	405	0.089	0.009
35	AβC6-35	C	G	V	A	E	S	384	0.085	0.008
36	AβC6-36	T	P	I	V	F	D	384	0.085	0.008
37	AβC6-37	T	L	W	V	F	D	355	0.078	0.008
38	AβC6-38	T	G	T	T	L	V	351	0.078	0.008
39	AβC6-39	T	V	T	E	V	L	319	0.070	0.007
40	AβC6-40	T	P	L	W	F	N	316	0.070	0.007

Number	Peptide name	Amino acid sequence						Number of reads	Reads/Total hexapeptide reads (%)	Reads/Total peptide reads (%)
41	AβC6-41	T	S	V	E	Y	E	307	0.068	0.007
42	AβC6-42	T	L	G	W	L	D	307	0.068	0.007
43	AβC6-43	T	M	M	F	S	T	297	0.066	0.007
44	AβC6-44	T	P	P	W	F	D	289	0.064	0.006
45	AβC6-45	T	V	T	N	V	V	276	0.061	0.006
46	AβC6-46	T	P	C	W	F	D	252	0.056	0.006
47	AβC6-47	T	L	S	W	Y	D	239	0.053	0.005
48	AβC6-48	T	P	V	L	V	D	236	0.052	0.005
49	AβC6-49	T	L	E	Y	L	W	233	0.051	0.005
50	AβC6-50	T	I	F	W	F	D	227	0.050	0.005
51	AβC6-51	T	W	Q	W	C	K	226	0.050	0.005
52	AβC6-52	S	L	R	G	R	G	219	0.048	0.005
53	AβC6-53	T	P	A	L	V	D	208	0.046	0.005
54	AβC6-54	C	R	T	T	V	V	204	0.045	0.005
55	AβC6-55	T	P	G	W	F	D	180	0.040	0.004
56	AβC6-56	T	I	T	S	I	I	180	0.040	0.004
57	AβC6-57	T	L	S	V	F	D	176	0.039	0.004
58	AβC6-58	T	P	G	L	V	D	142	0.031	0.003
59	AβC6-59	T	L	S	W	F	N	141	0.031	0.003
60	AβC6-60	T	P	V	V	E	I	137	0.030	0.003
61	AβC6-61	T	P	G	L	V	R	122	0.027	0.003
62	AβC6-62	T	L	T	T	L	I	121	0.027	0.003
63	AβC6-63	T	L	D	F	F	D	114	0.025	0.003
64	AβC6-64	T	R	A	S	V	L	109	0.024	0.002
65	AβC6-65	T	P	S	W	F	D	105	0.023	0.002
66	AβC6-66	T	M	Q	V	S	V	105	0.023	0.002
67	AβC6-67	C	G	T	G	G	E	103	0.023	0.002
68	AβC6-68	T	P	A	L	F	D	101	0.022	0.002
69	AβC6-69	T	P	A	W	S	D	86	0.019	0.002
70	AβC6-70	T	M	F	V	S	V	86	0.019	0.002
71	AβC6-71	T	V	S	Q	V	V	82	0.018	0.002
72	AβC6-72	T	T	W	S	R	L	72	0.016	0.002
73	AβC6-73	S	P	S	S	V	Q	70	0.015	0.002
74	AβC6-74	T	R	E	S	V	L	69	0.015	0.002
75	AβC6-75	T	A	T	S	V	L	68	0.015	0.002
76	AβC6-76	T	H	G	Q	S	Q	66	0.015	0.001
77	AβC6-77	T	V	T	S	V	P	60	0.013	0.001
78	AβC6-78	T	P	A	R	F	D	55	0.012	0.001
79	AβC6-79	T	P	A	W	L	D	55	0.012	0.001
80	AβC6-80	T	P	V	W	L	D	55	0.012	0.001

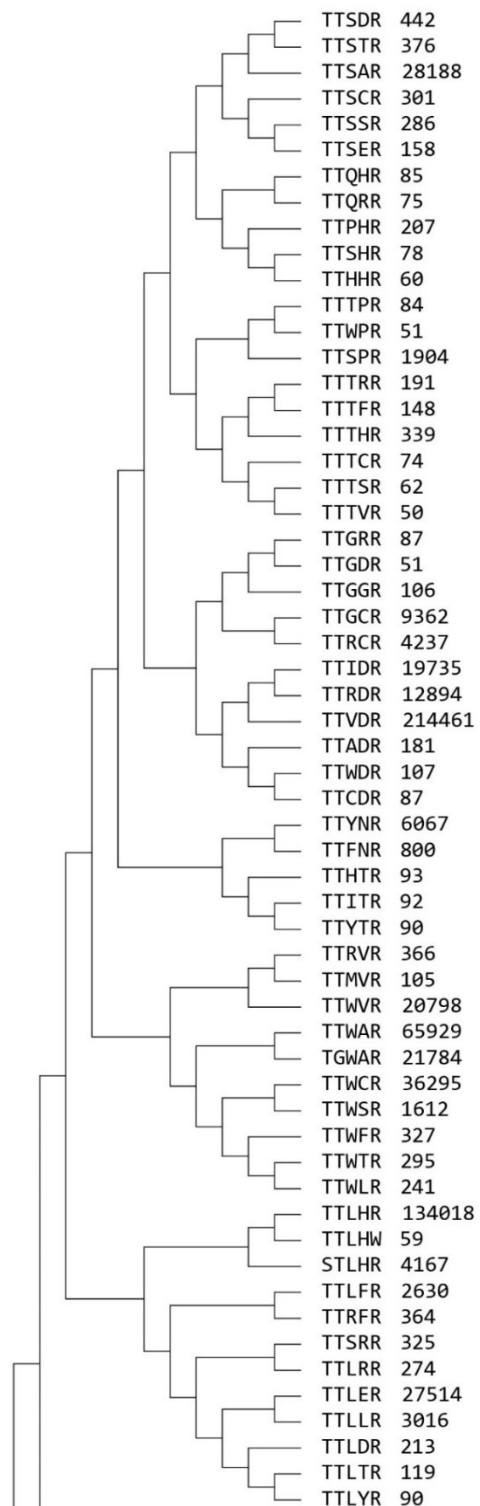
Number	Peptide name	Amino acid sequence						Number of reads	Reads/Total hexapeptide reads (%)	Reads/Total peptide reads (%)
81	AβC6-81	S	P	G	P	S	Q	51	0.011	0.001
		Sum						452,587	100.000	9.990

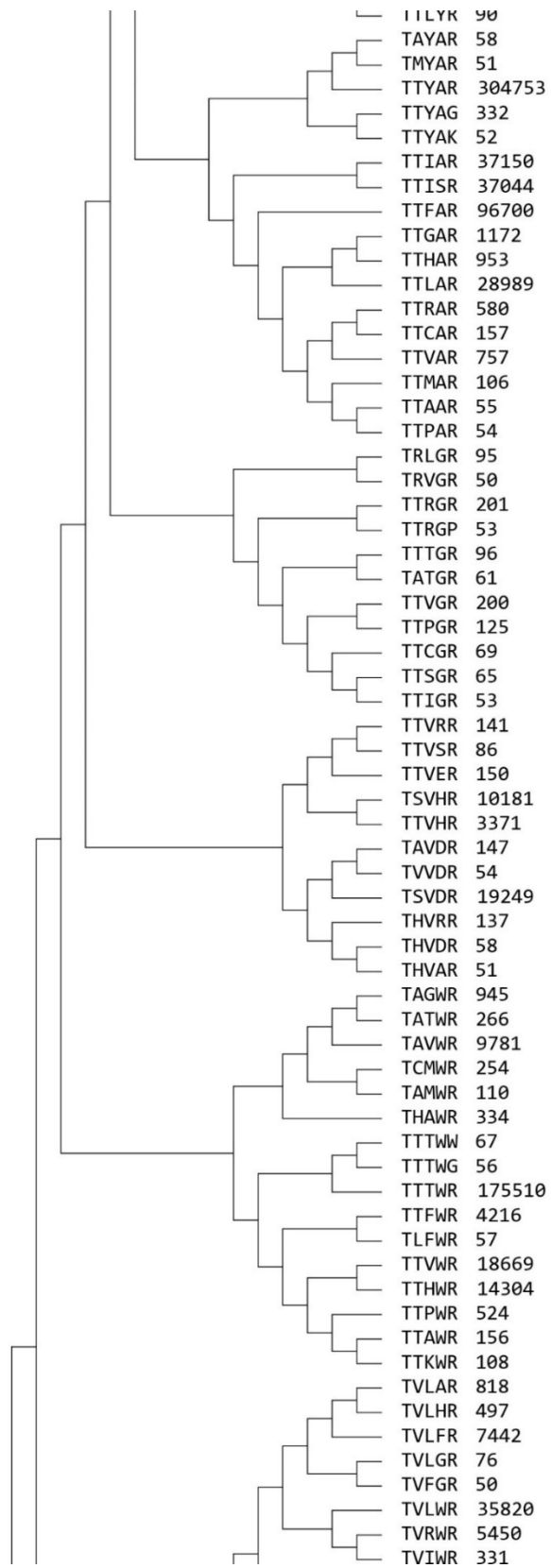
Supplementary Table 5. Tetrapeptide sequences identified in the sorted peptide pool more than 50 times.

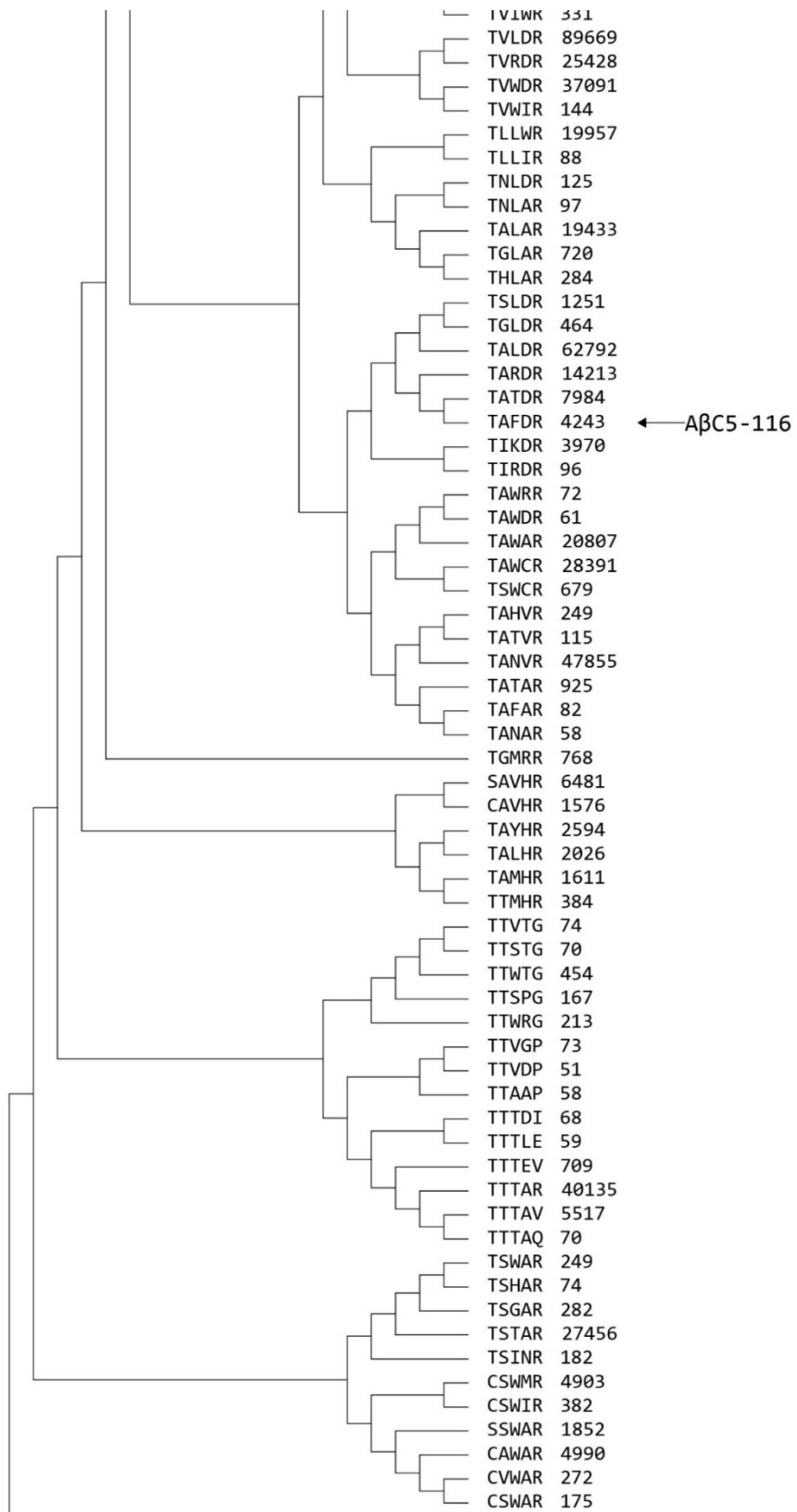
Number	Peptide name	Amino acid sequence				Number of reads	Reads/Total tetrapeptide reads (%)	Reads/Total peptide reads (%)
1	AβC4-1	C	L	Y	L	5,273	29.176	0.00116
2	AβC4-2	T	L	V	V	3,377	18.685	0.00075
3	AβC4-3	T	N	G	E	2,910	16.101	0.00064
4	AβC4-4	S	V	W	R	1,068	5.909	0.00024
5	AβC4-5	T	T	Y	A	606	3.353	0.00013
6	AβC4-6	T	T	S	A	496	2.744	0.00011
7	AβC4-7	C	W	T	G	261	1.444	0.00006
8	AβC4-8	T	T	T	W	258	1.428	0.00006
9	AβC4-9	T	T	C	R	258	1.428	0.00006
10	AβC4-10	T	T	V	W	255	1.411	0.00006
11	AβC4-11	T	T	R	R	248	1.372	0.00005
12	AβC4-12	T	A	L	D	204	1.129	0.00005
13	AβC4-13	C	V	V	L	189	1.046	0.00004
14	AβC4-14	T	V	R	P	172	0.952	0.00004
15	AβC4-15	T	A	W	C	166	0.918	0.00004
16	AβC4-16	T	V	T	L	162	0.896	0.00004
17	AβC4-17	C	C	I	W	147	0.813	0.00003
18	AβC4-18	T	T	R	P	142	0.786	0.00003
19	AβC4-19	T	Y	H	G	139	0.769	0.00003
20	AβC4-20	T	T	T	A	134	0.741	0.00003
21	AβC4-21	S	W	D	E	131	0.725	0.00003
22	AβC4-22	T	T	S	P	126	0.697	0.00003
23	AβC4-23	T	T	V	D	115	0.636	0.00003
24	AβC4-24	T	A	L	A	108	0.598	0.00002
25	AβC4-25	T	T	W	Q	95	0.526	0.00002
26	AβC4-26	S	L	G	P	87	0.481	0.00002
27	AβC4-27	S	S	R	D	77	0.426	0.00002
28	AβC4-28	T	T	R	A	75	0.415	0.00002
29	AβC4-29	T	R	V	V	68	0.376	0.00002
30	AβC4-30	T	T	V	T	67	0.371	0.00001
31	AβC4-31	T	T	S	R	67	0.371	0.00001
32	AβC4-32	T	T	I	S	66	0.365	0.00001

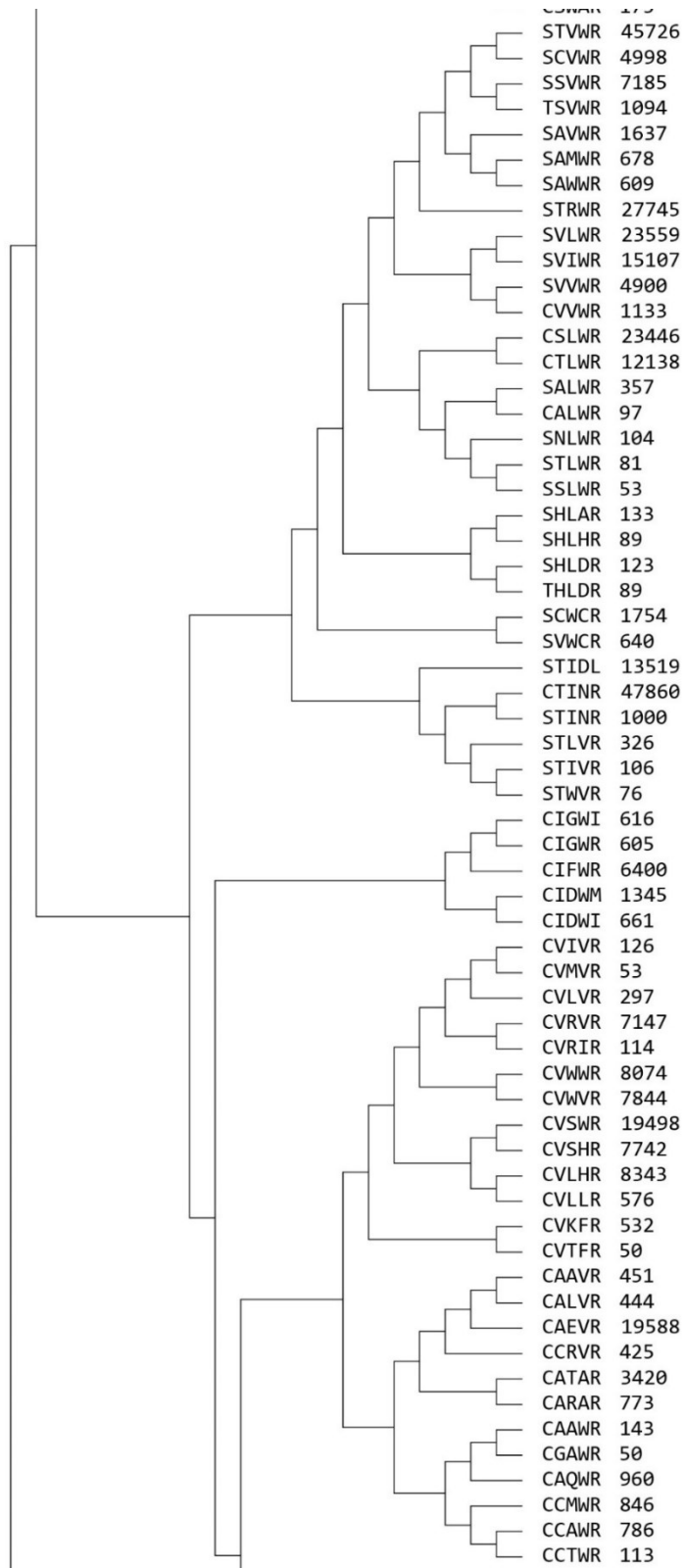
Number	Peptide name	Amino acid sequence				Number of reads	Reads/Total tetrapeptide reads (%)	Reads/Total peptide reads (%)
33	AβC4-33	T	V	S	P	65	0.360	0.00001
34	AβC4-34	T	R	G	R	63	0.349	0.00001
35	AβC4-35	T	T	G	R	61	0.338	0.00001
36	AβC4-36	T	T	S	T	61	0.338	0.00001
37	AβC4-37	T	V	L	F	60	0.332	0.00001
38	AβC4-38	C	V	C	G	57	0.315	0.00001
39	AβC4-39	T	V	S	V	57	0.315	0.00001
40	AβC4-40	C	Y	G	P	51	0.282	0.00001
41	AβC4-41	T	R	R	R	51	0.282	0.00001
		Sum				18,073	100.000	0.00399

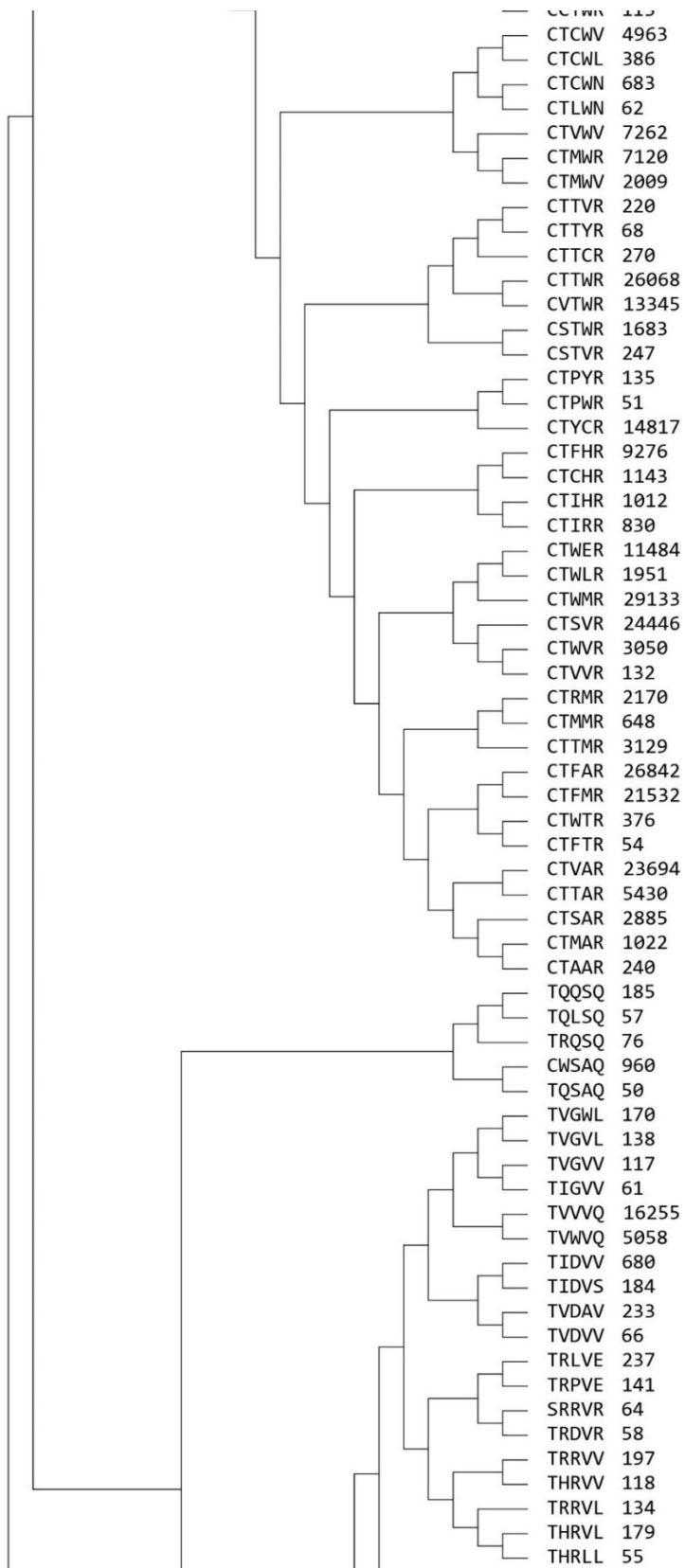
Supplementary Figures

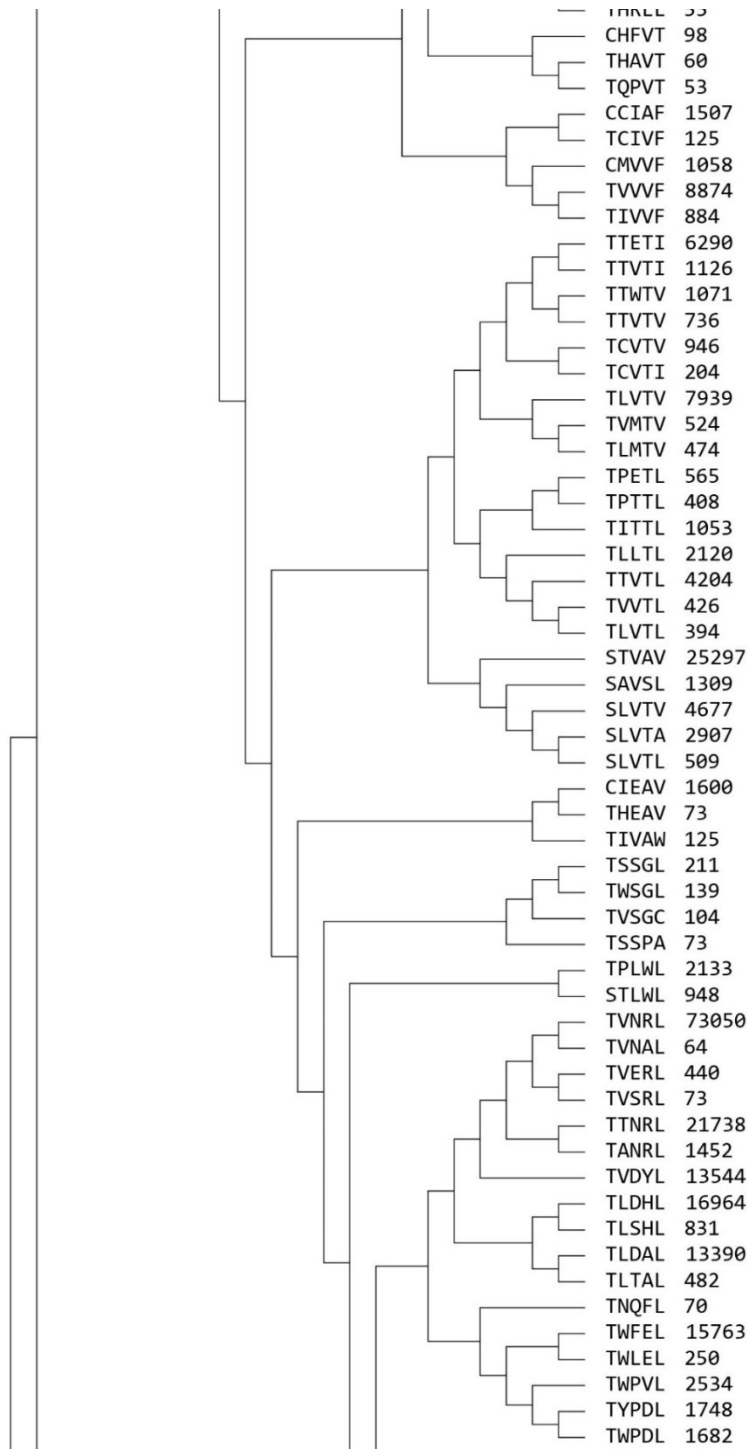


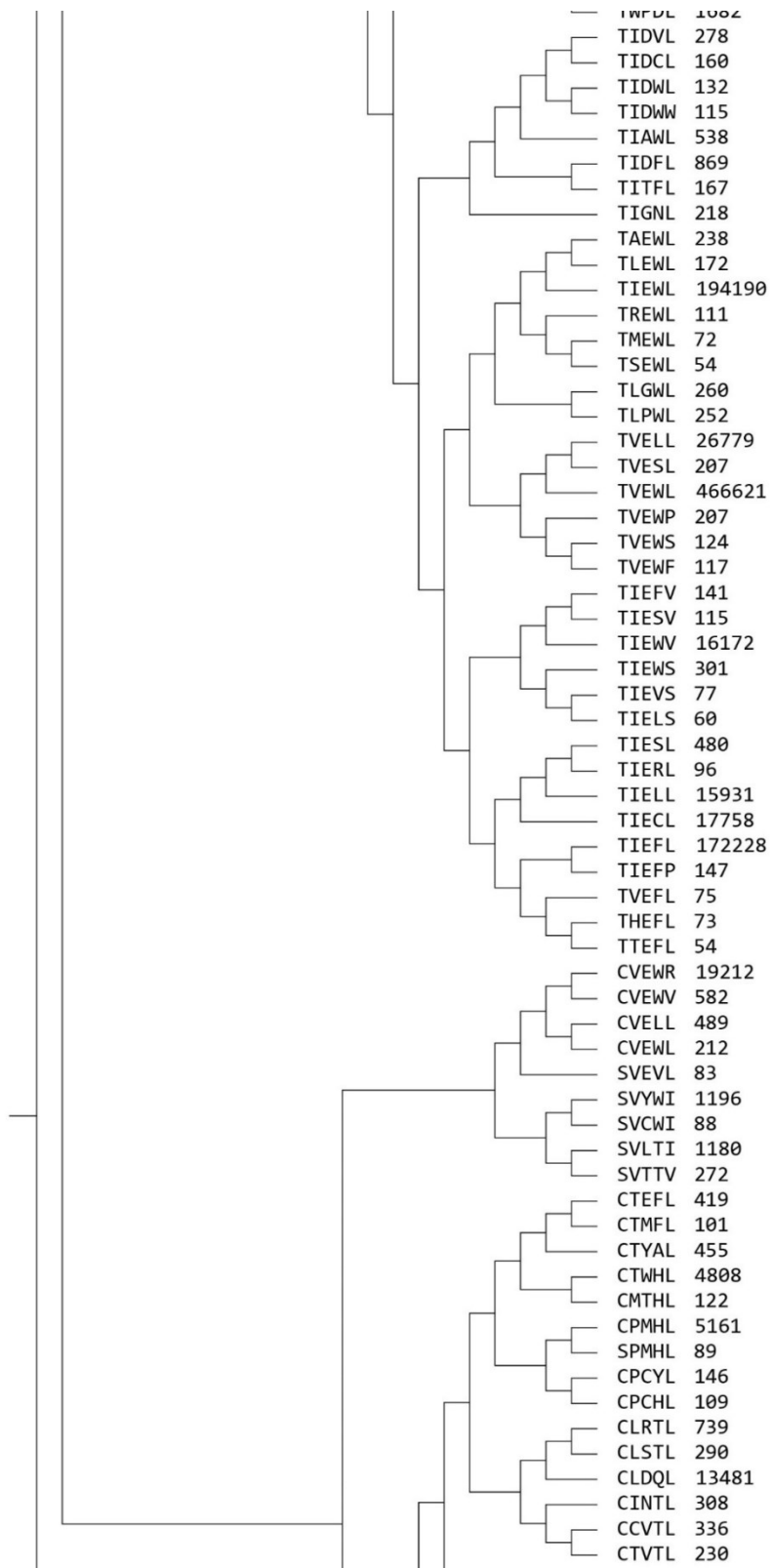


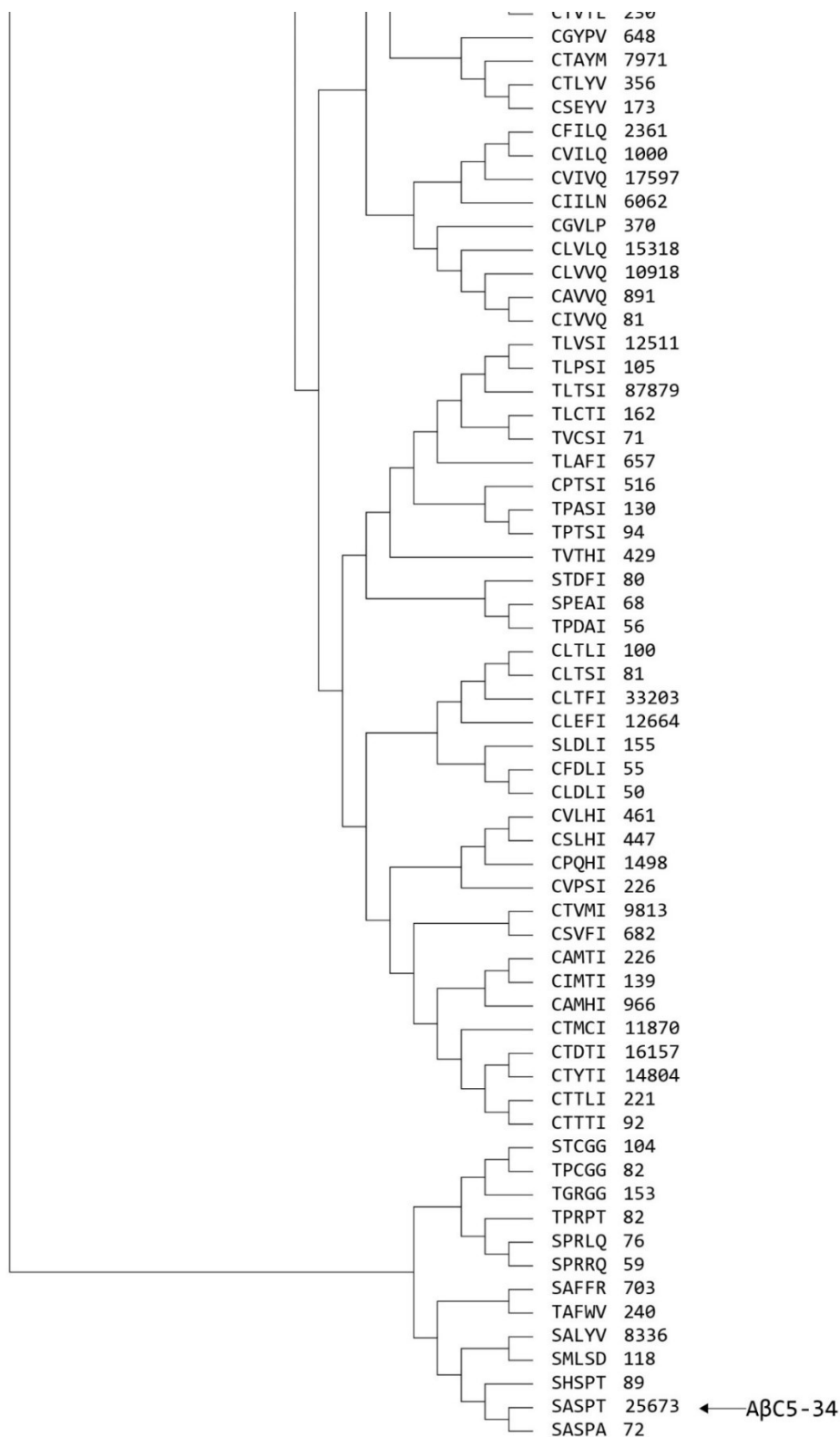




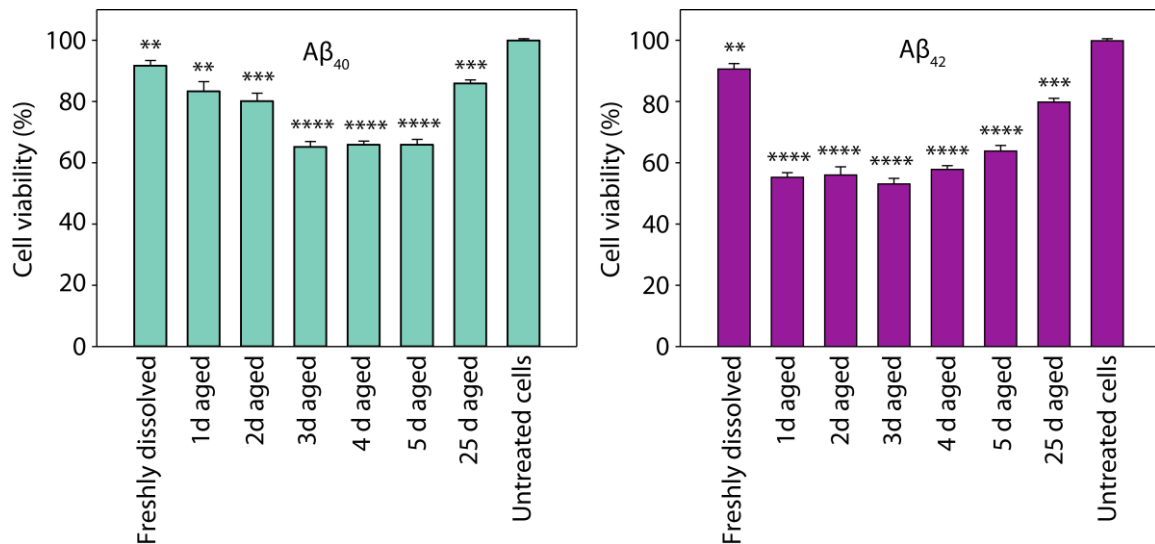








Supplementary Figure 1. Complete phylogenetic tree for pentapeptides identified in the sorted peptide pool. Tree was generated using the UPGMA statistical method with a p-distance substitution model. Read numbers are attached next to the peptide sequences. Brackets designate possible pentapeptide families.



Supplementary Figure 2. Effect of pre-incubation time on the cytotoxicity of Aβ solutions. Cell viability of primary hippocampal neurons, as determined by the MTT assay. Cells were treated for 24 h at 37 °C with 1 μM solutions of Aβ₄₀ or Aβ₄₂, pre-incubated for 0–25 d. An MTT stock solution in Neurobasal-A complete medium was added in each well to a final concentration of 0.5 mg/mL and was subsequently incubated for 3 h at 37 °C. Results are expressed as the percentage of MTT reduction, assuming that the absorbance of control (untreated) cells was 100%. Mean values ± s.e.m. of three independent experiments (*n*=3) with six replicate wells for each condition are reported. Statistical significances of the differences in the levels of viability between cells untreated and treated with Aβ or between cells treated with Aβ in the presence and absence of the selected cyclic peptides are also reported. ***P* ≤ 0.01, ****P* ≤ 0.001; *****P* ≤ 0.0001. MTT assays were performed in the lab of Dr. Maria Pelecanou at National Center for Scientific Research “Demokritos”.

References

1. Anfinsen, C.B., Haber, E., Sela, M. & White, F. The kinetics of formation of native ribonuclease during oxidation of the reduced polypeptide chain. *Proc. Natl. Acad. Sci. USA* **47**, 1309-1314 (1961).
2. Karplus, M. Behind the folding funnel diagram. *Nat. Chem. Biol.* **7**, 401-404 (2011).
3. Wetlaufer, D.B. Nucleation, rapid folding, and globular intrachain regions in proteins. *Proc. Natl. Acad. Sci. USA* **70**, 697-701 (1973).
4. Fersht, A. A guide to enzyme catalysis and protein folding. *Struct. Mech. Prot. Sci.*, 508-539 (1999).
5. Karplus, M. & Weaver, D.L. Protein-folding dynamics. *Nature* **260**, 404-406 (1976).
6. Harrison, S.C. & Durbin, R. Is there a single pathway for the folding of a polypeptide chain? *Proc. Natl. Acad. Sci. USA* **82**, 4028-4030 (1985).
7. Zwanzig, R., Szabo, A. & Bagchi, B. Levinthal's paradox. *Proc. Natl. Acad. Sci. USA* **89**, 20-22 (1992).
8. Bryngelson, J.D., Onuchic, J.N., Socci, N.D. & Wolynes, P.G. Funnels, pathways, and the energy landscape of protein folding: A synthesis. *Prot. Struct. Funct. Bioinf.* **21**, 167-195 (1995).
9. Onuchic, J.N., Wolynes, P.G., Luthey-Schulten, Z. & Socci, N.D. Toward an outline of the topography of a realistic protein-folding funnel. *Proc. Natl. Acad. Sci. USA* **92**, 3626-3630 (1995).
10. Wolynes, P.G., Onuchic, J.N. & Thirumalai, D. Navigating the folding routes. *Science* **267**, 1619-1620 (1995).
11. Dill, K.A. & Chan, H.S. From Levinthal to pathways to funnels. *Nat. Struct. Mol. Biol.* **4**, 10-19 (1997).
12. Frauenfelder, H., Sligar, S. & Wolynes, P. The energy landscapes and motions of proteins. *Science* **254**, 1598-1603 (1991).
13. Ferreira, D.U., Hegler, J.A., Komives, E.A. & Wolynes, P.G. Localizing frustration in native proteins and protein assemblies. *Proc. Natl. Acad. Sci. USA* **104**, 19819-19824 (2007).
14. Chiti, F. & Dobson, C.M. Protein misfolding, functional amyloid, and human disease. *Annu. Rev. Biochem.* **75**, 333-366 (2006).
15. Baldwin, A.J. *et al.* Metastability of native proteins and the phenomenon of amyloid formation. *J. Am. Chem. Soc.* **133**, 14160-14163 (2011).
16. Knowles, T.P., Vendruscolo, M. & Dobson, C.M. The amyloid state and its association with protein misfolding diseases. *Nat. Rev. Mol. Cell Biol.* **15**, 384-396 (2014).
17. Rauscher, S., Baud, S., Miao, M., Keeley, F.W. & Pomes, R. Proline and glycine control protein self-organization into elastomeric or amyloid fibrils. *Structure* **14**, 1667-1676 (2006).
18. Eisenberg, D.S. & Sawaya, M.R. Implications for Alzheimer's disease of an atomic resolution structure of amyloid- β (1-42) fibrils. *Proc. Natl. Acad. Sci. USA* **113**, 9398-9400 (2016).
19. Colvin, M.T. *et al.* High resolution structural characterization of A β 42 amyloid fibrils by magic angle spinning NMR. *J. Am. Chem. Soc.* **137**, 7509-7518 (2015).
20. Walti, M.A., Orts, J., Vogeli, B., Campioni, S. & Riek, R. Solution NMR studies of recombinant A β (1-42): from the presence of a micellar entity to residual beta-sheet structure in the soluble species. *ChemBiochem* **16**, 659-669 (2015).
21. Haass, C. & Selkoe, D.J. Soluble protein oligomers in neurodegeneration: lessons from the Alzheimer's amyloid beta-peptide. *Nat. Rev. Mol. Cell Biol.* **8**, 101-112 (2007).
22. Finder, V.H. & Glockshuber, R. Amyloid- β Aggregation. *Neurodegen. Dis.* **4**, 13-27 (2007).
23. Dobson, C.M. Protein-misfolding diseases: Getting out of shape. *Nature* **418**, 729-730 (2002).
24. Monsellier, E. & Chiti, F. Prevention of amyloid-like aggregation as a driving force of protein evolution. *EMBO Rep.* **8**, 737-742 (2007).
25. Watters, A.L. *et al.* The highly cooperative folding of small naturally occurring proteins is likely the result of natural selection. *Cell* **128**, 613-624 (2007).

26. Sipe, J.D. *et al.* Nomenclature 2014: Amyloid fibril proteins and clinical classification of the amyloidosis. *Amyloid* **21**, 221-224 (2014).
27. Fitzpatrick, A.W.P. *et al.* Atomic structure and hierarchical assembly of a cross- β amyloid fibril. *Proc. Natl. Acad. Sci. USA* **110**, 5468-5473 (2013).
28. Colvin, M.T. *et al.* Atomic Resolution Structure of Monomorphic A β 42 Amyloid Fibrils. *J. Am. Chem. Soc.* **138**, 9663-9674 (2016).
29. Wälti, M.A. *et al.* Atomic-resolution structure of a disease-relevant A β (1–42) amyloid fibril. *Proc. Natl. Acad. Sci. USA* **113**, E4976-E4984 (2016).
30. Nonaka, T., Watanabe, S.T., Iwatsubo, T. & Hasegawa, M. Seeded Aggregation and Toxicity of α -Synuclein and Tau Cellular Models of Neurodegenerative Diseases. *J. Biol. Chem.* **285**, 34885-34898 (2010).
31. Wogulis, M. *et al.* Nucleation-Dependent Polymerization Is an Essential Component of Amyloid-Mediated Neuronal Cell Death. *J. Neurosci.* **25**, 1071-1080 (2005).
32. Jahn, T.R. & Radford, S.E. Folding versus aggregation: polypeptide conformations on competing pathways. *Arch. Biochem. Biophys.* **469**, 100-117 (2008).
33. Crowther, D.C. *et al.* Intraneuronal A β , non-amyloid aggregates and neurodegeneration in a Drosophila model of Alzheimer's disease. *Neuroscience* **132**, 123-135 (2005).
34. Rousseau, F., Schymkowitz, J. & Serrano, L. Protein aggregation and amyloidosis: confusion of the kinds? *Curr. Opin. Struct. Biol.* **16**, 118-126 (2006).
35. Olesen, J. *et al.* The economic cost of brain disorders in Europe. *Eur. J. Neurol.* **19**, 155-162 (2012).
36. Werner, P., Savva, G.M., Maidment, I., Thyrian, J.R. & Fox, C. in *Mental Health and Older People: A Guide for Primary Care Practitioners.* (eds. C.A. Chew-Graham & M. Ray) 197-209 (Springer International Publishing, Cham; 2016).
37. Aldwin, C.M., Igarashi, H., Gilmer, D.F. & Levenson, M.R. Health, illness, and optimal aging: Biological and psychosocial perspectives. (Springer Publishing Company, 2017).
38. Hippus, H. & Neundörfer, G. The discovery of Alzheimer's disease. *Dialogues Clin. Neurosci.* **5**, 101-108 (2003).
39. Wimo, A. *et al.* The worldwide costs of dementia 2015 and comparisons with 2010. *Alzheimer's & Dementia* **13**, 1-7 (2017).
40. Alzheimer's, A. 2015 Alzheimer's disease facts and figures. *Alzheimer's & dementia: J. Alzheimer's Assoc.* **11**, 332 (2015).
41. Rogers, S.L. & Friedhoff, L.T. The efficacy and safety of donepezil in patients with Alzheimer's disease: results of a US multicentre, randomized, double-blind, placebo-controlled trial. *Dement. Geriatr. Cogn. Dis.* **7**, 293-303 (1996).
42. Raskind, M.A., Peskind, E., Wessel, T., Yuan, W. & Group, G.U.-S. Galantamine in AD A 6-month randomized, placebo-controlled trial with a 6-month extension. *Neurology* **54**, 2261-2268 (2000).
43. Tariot, P.N. *et al.* A 5-month, randomized, placebo-controlled trial of galantamine in AD. *Neurology* **54**, 2269-2276 (2000).
44. Birks, J., Grimley Evans, J., Iakovidou, V. & Tsolaki, M. Rivastigmine for Alzheimer's disease. *Cochrane Database Syst. Rev.* **4** (2000).
45. Farlow, M., Anand, R., Messina Jr, J., Hartman, R. & Veach, J. A 52-week study of the efficacy of rivastigmine in patients with mild to moderately severe Alzheimer's disease. *Eur. Neurol.* **44**, 236-241 (2000).
46. Tariot, P.N. *et al.* Memantine treatment in patients with moderate to severe Alzheimer disease already receiving donepezil: a randomized controlled trial. *JAMA* **291**, 317-324 (2004).
47. Reisberg, B. *et al.* Memantine in moderate-to-severe Alzheimer's disease. *N. Engl. J. Med.* **348**, 1333-1341 (2003).
48. Howard, R. *et al.* Donepezil and Memantine for Moderate-to-Severe Alzheimer's Disease. *N. Engl. J. Med.* **366**, 893-903 (2012).

49. Zanetti, O., Solerte, S. & Cantoni, F. Life expectancy in Alzheimer's disease (AD). *Arch. Gerontol. Geriatr.* **49**, 237-243 (2009).
50. Galimberti, D. & Scarpini, E. Disease-modifying treatments for Alzheimer's disease. *Ther. Adv. Neurol. Dis.* **4**, 203-216 (2011).
51. Glenner, G.G. & Wong, C.W. Alzheimer's disease and Down's syndrome: sharing of a unique cerebrovascular amyloid fibril protein. *Biochem. Biophys. Res. Commun.* **122**, 1131-1135 (1984).
52. Rumble, B. *et al.* Amyloid A4 protein and its precursor in Down's syndrome and Alzheimer's disease. *N. Engl. J. Med.* **320**, 1446-1452 (1989).
53. Kang, J. *et al.* The precursor of Alzheimer's disease amyloid A4 protein resembles a cell-surface receptor. *Nature* **325**, 733-736 (1987).
54. Robakis, N.K., Ramakrishna, N., Wolfe, G. & Wisniewski, H.M. Molecular cloning and characterization of a cDNA encoding the cerebrovascular and the neuritic plaque amyloid peptides. *Proc. Natl. Acad. Sci. USA* **84**, 4190-4194 (1987).
55. Masters, C.L. *et al.* Amyloid plaque core protein in Alzheimer disease and Down syndrome. *Proc. Natl. Acad. Sci. USA* **82**, 4245-4249 (1985).
56. Mann, D. *et al.* An analysis of the morphology of senile plaques in Down's syndrome patients of different ages using immunocytochemical and lectin histochemical techniques. *Neuropathol. Appl. Neurobiol.* **15**, 317-329 (1989).
57. Kitaguchi, N., Takahashi, Y., Tokushima, Y., Shiojiri, S. & Ito, H. Novel precursor of Alzheimer's disease amyloid protein shows protease inhibitory activity. *Nature* **331**, 530-532 (1988).
58. Turner, P.R., O'Connor, K., Tate, W.P. & Abraham, W.C. Roles of amyloid precursor protein and its fragments in regulating neural activity, plasticity and memory. *Progr. Neurobiol.* **70**, 1-32 (2003).
59. Tanaka, S. *et al.* Tissue-specific expression of three types of β -protein precursor mRNA: Enhancement of protease inhibitor-harboring types in Alzheimer's disease brain. *Biochem. Biophys. Res. Commun.* **165**, 1406-1414 (1989).
60. Chua, L.-M., Lim, M.-L. & Wong, B.-S. The Kunitz-protease inhibitor domain in amyloid precursor protein reduces cellular mitochondrial enzymes expression and function. *Biochem. Biophys. Res. Commun.* **437**, 642-647 (2013).
61. Schubert, W. *et al.* Localization of Alzheimer β A4 amyloid precursor protein at central and peripheral synaptic sites. *Brain Res.* **563**, 184-194 (1991).
62. Shigematsu, K., McGeer, P.L. & McGeer, E.G. Localization of amyloid precursor protein in selective postsynaptic densities of rat cortical neurons. *Brain Res.* **592**, 353-357 (1992).
63. Caporaso, G.L. *et al.* Morphologic and biochemical analysis of the intracellular trafficking of the Alzheimer beta/A4 amyloid precursor protein. *J. Neurosci.* **14**, 3122-3138 (1994).
64. Esch, F. *et al.* Cleavage of amyloid beta peptide during constitutive processing of its precursor. *Science* **248**, 1122-1124 (1990).
65. Sisodia, S., Koo, E., Beyreuther, K., Unterbeck, A. & Price, D. Evidence that beta-amyloid protein in Alzheimer's disease is not derived by normal processing. *Science* **248**, 492-495 (1990).
66. Sisodia, S.S. Beta-amyloid precursor protein cleavage by a membrane-bound protease. *Proc. Natl. Acad. Sci. USA* **89**, 6075-6079 (1992).
67. Wirak, D. *et al.* Deposits of amyloid beta protein in the central nervous system of transgenic mice. *Science* **253**, 323-325 (1991).
68. Sandhu, F.A., Salim, M. & Zain, S.B. Expression of the human beta-amyloid protein of Alzheimer's disease specifically in the brains of transgenic mice. *J. Biol. Chem.* **266**, 21331-21334 (1991).
69. Golde, T., Estus, S., Younkin, L., Selkoe, D. & Younkin, S. Processing of the amyloid protein precursor to potentially amyloidogenic derivatives. *Science* **255**, 728-730 (1992).
70. Haass, C., Hung, A. & Selkoe, D. Processing of beta-amyloid precursor protein in microglia and astrocytes favors an internal localization over constitutive secretion. *J. Neurosci.* **11**, 3783-3793 (1991).

71. Nordstedt, C. *et al.* Alzheimer beta/A4 amyloid precursor protein in human brain: aging-associated increases in holoprotein and in a proteolytic fragment. *Proc. Natl. Acad. Sci. USA* **88**, 8910-8914 (1991).
72. Beyreuther, K. & Masters, C.L. Amyloid Precursor Protein (APP) and BZA4 Amyloid in the Etiology of Alzheimer's Disease: Precursor-Product Relationships in the Derangement of Neuronal Function. *Brain Pathol.* **1**, 241-251 (1991).
73. Hardy, J. & Allsop, D. Amyloid deposition as the central event in the aetiology of Alzheimer's disease. *Trends Pharmacol. Sci.* **12**, 383-388 (1991).
74. Selkoe, D.J. The molecular pathology of Alzheimer's disease. *Neuron* **6**, 487-498 (1991).
75. Hardy, J. & Higgins, G. Alzheimer's disease: the amyloid cascade hypothesis. *Science* **256**, 184-185 (1992).
76. Kowall, N.W., Beal, M.F., Busciglio, J., Duffy, L.K. & Yankner, B.A. An *in vivo* model for the neurodegenerative effects of beta amyloid and protection by substance P. *Proc. Natl. Acad. Sci. USA* **88**, 7247-7251 (1991).
77. Yankner, B. *et al.* Neurotoxicity of a fragment of the amyloid precursor associated with Alzheimer's disease. *Science* **245**, 417-420 (1989).
78. Koh, J.-y., Yang, L.L. & Cotman, C.W. β -Amyloid protein increases the vulnerability of cultured cortical neurons to excitotoxic damage. *Brain Res.* **533**, 315-320 (1990).
79. Mattson, M.P. *et al.* β -Amyloid peptides destabilize calcium homeostasis and render human cortical neurons vulnerable to excitotoxicity. *J. Neurosci.* **12**, 376-389 (1992).
80. Grundke-Iqbal, I. *et al.* Microtubule-associated protein tau. A component of Alzheimer paired helical filaments. *J. Biol. Chem.* **261**, 6084-6089 (1986).
81. Delacourte, A. & Defossez, A. Alzheimer's disease: Tau proteins, the promoting factors of microtubule assembly, are major components of paired helical filaments. *J. Neurol. Sci.* **76**, 173-186 (1986).
82. Baudier, J., Lee, S.H. & Cole, R.D. Separation of the different microtubule-associated tau protein species from bovine brain and their mode II phosphorylation by Ca²⁺/phospholipid-dependent protein kinase C. *J. Biol. Chem.* **262**, 17584-17590 (1987).
83. Hardy, J. & Selkoe, D.J. The amyloid hypothesis of Alzheimer's disease: progress and problems on the road to therapeutics. *Science* **297**, 353-356 (2002).
84. Selkoe, D.J. & Hardy, J. The amyloid hypothesis of Alzheimer's disease at 25 years. *EMBO Mol. Med.* **8**, 595-608 (2016).
85. Haass, C. *et al.* Amyloid β -peptide is produced by cultured cells during normal metabolism. *Nature* **359**, 322-325 (1992).
86. Seubert, P. *et al.* Isolation and quantification of soluble Alzheimer's [beta]-peptide from biological fluids. *Nature* **359**, 325-327 (1992).
87. Shoji, M. *et al.* Production of the Alzheimer amyloid β protein by normal proteolytic processing. *Science*, 126-129 (1992).
88. Goate, A. *et al.* Segregation of a missense mutation in the amyloid precursor protein gene with familial Alzheimer's disease. *Nature* **349**, 704-706 (1991).
89. Chartier-Harlin, M.-C. *et al.* Early-onset Alzheimer's disease caused by mutations at codon 717 of the β -amyloid precursor protein gene. *Nature* **353**, 844 (1991).
90. Murrell, J., Farlow, M., Ghetti, B. & Benson, M.D. A Mutation in the Amyloid Precursor Protein Associated with Hereditary Alzheimer's Disease. *Science* **254**, 97-99 (1991).
91. Hendriks, L. *et al.* Presenile dementia and cerebral haemorrhage linked to a mutation at codon 692 of the β -amyloid precursor protein gene. *Nat. Gen.* **1**, 218 (1992).
92. Moechars, D., Lorent, K., De Strooper, B., Dewachter, I. & Van Leuven, F. Expression in brain of amyloid precursor protein mutated in the alpha-secretase site causes disturbed behavior, neuronal degeneration and premature death in transgenic mice. *EMBO J.* **15**, 1265-1274 (1996).
93. Moechars, D. *et al.* Early Phenotypic Changes in Transgenic Mice That Overexpress Different Mutants of Amyloid Precursor Protein in Brain. *J. Biol. Chem.* **274**, 6483-6492 (1999).

94. Nelson, K.K., Schlondorff, J. & Blobel, C.P. Evidence for an interaction of the metalloprotease-disintegrin tumour necrosis factor alpha convertase (TACE) with mitotic arrest deficient 2 (MAD2), and of the metalloprotease-disintegrin MDC9 with a novel MAD2-related protein, MAD2beta. *Biochem J* **343**, 673-680 (1999).
95. KOIKE, H. *et al.* Membrane-anchored metalloprotease MDC9 has an α -secretase activity responsible for processing the amyloid precursor protein. *Biochem. J.* **343**, 371-375 (1999).
96. Lammich, S. *et al.* Constitutive and regulated α -secretase cleavage of Alzheimer's amyloid precursor protein by a disintegrin metalloprotease. *Proc. Natl. Acad. Sci. USA* **96**, 3922-3927 (1999).
97. Kojro, E., Gimpl, G., Lammich, S., März, W. & Fahrenholz, F. Low cholesterol stimulates the nonamyloidogenic pathway by its effect on the α -secretase ADAM 10. *Proc. Natl. Acad. Sci. USA* **98**, 5815-5820 (2001).
98. Buxbaum, J.D. *et al.* Evidence that tumor necrosis factor α converting enzyme is involved in regulated α -secretase cleavage of the Alzheimer amyloid protein precursor. *J. Biol. Chem.* **273**, 27765-27767 (1998).
99. SLACK, B.E. & SEAH, C.C. Constitutive shedding of the amyloid precursor protein ectodomain is up-regulated by tumour necrosis factor- α converting enzyme. *Biochem. J.* **357**, 787-794 (2001).
100. Kuhn, P.H. *et al.* ADAM10 is the physiologically relevant, constitutive α -secretase of the amyloid precursor protein in primary neurons. *EMBO J.* **29**, 3020-3032 (2010).
101. Jorissen, E. *et al.* The disintegrin/metalloproteinase ADAM10 is essential for the establishment of the brain cortex. *J. Neurosci.* **30**, 4833-4844 (2010).
102. Lichtenthaler, S.F. Alpha-secretase in Alzheimer's disease: molecular identity, regulation and therapeutic potential. *J. Neurochem.* **116**, 10-21 (2011).
103. Endres, K. & Fahrenholz, F. Regulation of alpha-secretase ADAM10 expression and activity. *Exp. Brain Res.* **217**, 343-352 (2012).
104. Hussain, I. *et al.* Identification of a Novel Aspartic Protease (Asp 2) as β -Secretase. *Mol. Cell. Neurosci.* **14**, 419-427 (1999).
105. Holsinger, R., McLean, C.A., Beyreuther, K., Masters, C.L. & Evin, G. Increased expression of the amyloid precursor β -secretase in Alzheimer's disease. *Ann. Neurol.* **51**, 783-786 (2002).
106. Cummings, J., Morstorf, T. & Lee, G. Alzheimer's drug-development pipeline: 2016. *Alzheimer's & Dement: Translational Research & Clinical Interventions* **2**, 222-232 (2016).
107. Cummings, J., Lee, G., Mortsdorf, T., Ritter, A. & Zhong, K. Alzheimer's disease drug development pipeline: 2017. *Alzheimer's & Dement: Translational Research & Clinical Interventions* **3**, 367-384 (2017).
108. Steiner, H., Fluhner, R. & Haass, C. Intramembrane Proteolysis by γ -Secretase. *The J. Biol. Chem.* **283**, 29627-29631 (2008).
109. Sastre, M. *et al.* Presenilin-dependent γ -secretase processing of β -amyloid precursor protein at a site corresponding to the S3 cleavage of Notch. *EMBO Rep.* **2**, 835-841 (2001).
110. Weidemann, A. *et al.* A Novel ϵ -Cleavage within the Transmembrane Domain of the Alzheimer Amyloid Precursor Protein Demonstrates Homology with Notch Processing. *Biochemistry* **41**, 2825-2835 (2002).
111. Qi-Takahara, Y. *et al.* Longer Forms of Amyloid β Protein: Implications for the Mechanism of Intramembrane Cleavage by γ -Secretase. *J. Neurosci.* **25**, 436-445 (2005).
112. Takami, M. *et al.* γ -Secretase: Successive Tripeptide and Tetrapeptide Release from the Transmembrane Domain of β -Carboxyl Terminal Fragment. *J. Neurosci.* **29**, 13042-13052 (2009).
113. Haass, C., Kaether, C., Thinakaran, G. & Sisodia, S. Trafficking and Proteolytic Processing of APP. *Cold Spring Harbor Perspect. Med.* **2**, a006270 (2012).
114. Schenk, D., Basi, G.S. & Pangalos, M.N. Treatment Strategies Targeting Amyloid β -Protein. *Cold Spring Harbor Perspect. Med.* **2**, a006387 (2012).

115. Prasher, V. *et al.* Molecular mapping of Alzheimer-type dementia in Down's syndrome. *Ann. Neurol.* **43**, 380-383 (1998).
116. Rovelet-Lecrux, A. *et al.* APP locus duplication causes autosomal dominant early-onset Alzheimer disease with cerebral amyloid angiopathy. *Nat. Gen.* **38**, 24 (2006).
117. Jonsson, T. *et al.* A mutation in APP protects against Alzheimer's disease and age-related cognitive decline. *Nature* **488**, 96-99 (2012).
118. Benilova, I. *et al.* The Alzheimer disease protective mutation A2T modulates kinetic and thermodynamic properties of amyloid- β (A β) aggregation. *J. Biol. Chem.* **289**, 30977-30989 (2014).
119. Maloney, J.A. *et al.* Molecular mechanisms of Alzheimer disease protection by the A673T allele of amyloid precursor protein. *J. Biol. Chem.* **289**, 30990-31000 (2014).
120. Zheng, X., Liu, D., Roychoudhuri, R., Teplow, D.B. & Bowers, M.T. Amyloid β -Protein Assembly: Differential Effects of the Protective A2T Mutation and Recessive A2V Familial Alzheimer's Disease Mutation. *ACS Chem. Neurosci.* **6**, 1732-1740 (2015).
121. Meisl, G. *et al.* Differences in nucleation behavior underlie the contrasting aggregation kinetics of the A β 40 and A β 42 peptides. *Proc. Natl. Acad. Sci. USA* **111**, 9384-9389 (2014).
122. Bitan, G. *et al.* Amyloid β -protein (A β) assembly: A β 40 and A β 42 oligomerize through distinct pathways. *Proc. Natl. Acad. Sci. USA* **100**, 330-335 (2003).
123. Kim, W. & Hecht, M.H. Sequence determinants of enhanced amyloidogenicity of Alzheimer A β 42 peptide relative to A β 40. *J. Biol. Chem.* **280**, 35069-35076 (2005).
124. Jarrett, J.T. & Lansbury, P.T. Seeding "one-dimensional crystallization" of amyloid: a pathogenic mechanism in Alzheimer's disease and scrapie? *Cell* **73**, 1055-1058 (1993).
125. Kirkitadze, M.D., Condrón, M.M. & Teplow, D.B. Identification and characterization of key kinetic intermediates in amyloid β -protein fibrillogenesis. *J. Mol. Biol.* **312**, 1103-1119 (2001).
126. Walsh, D.M., Hartley, D.M., Condrón, M.M., Selkoe, D.J. & Teplow, D.B. *In vitro* studies of amyloid β -protein fibril assembly and toxicity provide clues to the aetiology of Flemish variant (Ala692 \rightarrow Gly) Alzheimer's disease. *Biochem. J.* **355**, 869-877 (2001).
127. Mori, H., Takio, K., Ogawara, M. & Selkoe, D.J. Mass spectrometry of purified amyloid beta protein in Alzheimer's disease. *J. Biol. Chem.* **267**, 17082-17086 (1992).
128. Nussbaum, J.M. *et al.* Prion-like behaviour and tau-dependent cytotoxicity of pyroglutamylated amyloid-[bgr]. *Nature* **485**, 651-655 (2012).
129. Cole, N.B. *et al.* Lipid droplet binding and oligomerization properties of the Parkinson's disease protein alpha-synuclein. *J. Biol. Chem.* **277**, 6344-6352 (2002).
130. Saito, T. *et al.* Potent amyloidogenicity and pathogenicity of A [beta] 43. *Nat. Neurosci.* **14**, 1023-1032 (2011).
131. Puzzo, D. & Arancio, O. Fibrillar β -amyloid impairs the late phase of long term potentiation. *Curr. Alzheimer Res.* **3**, 179-183 (2006).
132. Yoshiike, Y., Akagi, T. & Takashima, A. Surface structure of amyloid- β fibrils contributes to cytotoxicity. *Biochemistry* **46**, 9805-9812 (2007).
133. Katzman, R. Alzheimer's Disease. *N. Engl. J. Med.* **314**, 964-973 (1986).
134. Terry, R.D. *et al.* Physical basis of cognitive alterations in Alzheimer's disease: synapse loss is the major correlate of cognitive impairment. *Ann. Neurol.* **30**, 572-580 (1991).
135. Dickson, D.W. *et al.* Correlations of synaptic and pathological markers with cognition of the elderly. *Neurobiol. Aging* **16**, 285-298 (1995).
136. Wang, J., Dickson, D.W., Trojanowski, J.Q. & Lee, V.M.-Y. The levels of soluble versus insoluble brain A β distinguish Alzheimer's disease from normal and pathologic aging. *Exp. Neurol.* **158**, 328-337 (1999).
137. McLean, C.A. *et al.* Soluble pool of A β amyloid as a determinant of severity of neurodegeneration in Alzheimer's disease. *Ann. Neurol.* **46**, 860-866 (1999).
138. Lue, L.-F. *et al.* Soluble amyloid β peptide concentration as a predictor of synaptic change in Alzheimer's disease. *Am. J. Pathol.* **155**, 853-862 (1999).

139. Shankar, G.M. *et al.* Amyloid β -protein dimers isolated directly from Alzheimer brains impair synaptic plasticity and memory. *Nat. Med.* **14**, 837 (2008).
140. Esparza, T.J. *et al.* Amyloid-beta oligomerization in Alzheimer dementia versus high-pathology controls. *Ann. Neurol.* **73**, 104-119 (2013).
141. Lesné, S. *et al.* A specific amyloid- β protein assembly in the brain impairs memory. *Nature* **440**, 352-357 (2006).
142. Reed, M.N. *et al.* Cognitive effects of cell-derived and synthetically derived A β oligomers. *Neurobiol. Aging* **32**, 1784-1794 (2011).
143. Klyubin, I. *et al.* Amyloid β protein dimer-containing human CSF disrupts synaptic plasticity: prevention by systemic passive immunization. *J. Neurosci.* **28**, 4231-4237 (2008).
144. Roher, A.E. *et al.* Morphology and toxicity of A β -(1-42) dimer derived from neuritic and vascular amyloid deposits of Alzheimer's disease. *J. Biol. Chem.* **271**, 20631-20635 (1996).
145. Shankar, G.M. *et al.* Biochemical and immunohistochemical analysis of an Alzheimer's disease mouse model reveals the presence of multiple cerebral A β assembly forms throughout life. *Neurobiol. Dis.* **36**, 293-302 (2009).
146. Oda, T. *et al.* Clusterin (apoJ) alters the aggregation of amyloid β -peptide (A β 1-42) and forms slowly sedimenting A β complexes that cause oxidative stress. *Exp. Neurol.* **136**, 22-31 (1995).
147. Lambert, M.P. *et al.* Diffusible, nonfibrillar ligands derived from A β 1-42 are potent central nervous system neurotoxins. *Proc. Natl. Acad. Sci. USA* **95**, 6448-6453 (1998).
148. Sokolov, Y. *et al.* Soluble amyloid oligomers increase bilayer conductance by altering dielectric structure. *J. Gen. Physiol.* **128**, 637-647 (2006).
149. Kim, H.-J. *et al.* Selective neuronal degeneration induced by soluble oligomeric amyloid beta protein. *FASEB J.* **17**, 118-120 (2003).
150. Hoshi, M. *et al.* Spherical aggregates of β -amyloid (amylospheroid) show high neurotoxicity and activate tau protein kinase I/glycogen synthase kinase-3 β . *Proc. Natl. Acad. Sci. USA* **100**, 6370-6375 (2003).
151. Harper, J.D., Wong, S.S., Lieber, C.M. & Lansbury, P.T. Observation of metastable A β amyloid protofibrils by atomic force microscopy. *Chem. Biol.* **4**, 119-125 (1997).
152. Yamamoto, N. *et al.* A Ganglioside-induced Toxic Soluble A β Assembly ITS ENHANCED FORMATION FROM A β BEARING THE ARCTIC MUTATION. *J. Biol. Chem.* **282**, 2646-2655 (2007).
153. Yanagisawa, K. Role of gangliosides in Alzheimer's disease. *BBA-Biomembranes* **1768**, 1943-1951 (2007).
154. Hong, S. *et al.* Soluble A β oligomers are rapidly sequestered from brain ISF *in vivo* and bind GM1 ganglioside on cellular membranes. *Neuron* **82**, 308-319 (2014).
155. Stege, G. *et al.* The molecular chaperone α B-crystallin enhances amyloid β neurotoxicity. *Biochem. Biophys. Res. Commun.* **262**, 152-156 (1999).
156. Chromy, B.A. *et al.* Self-assembly of A β 1-42 into globular neurotoxins. *Biochemistry* **42**, 12749-12760 (2003).
157. De Felice, F.G. *et al.* A β oligomers induce neuronal oxidative stress through an N-methyl-D-aspartate receptor-dependent mechanism that is blocked by the Alzheimer drug memantine. *J. Biol. Chem.* **282**, 11590-11601 (2007).
158. Shankar, G.M. *et al.* Natural oligomers of the Alzheimer amyloid- β protein induce reversible synapse loss by modulating an NMDA-type glutamate receptor-dependent signaling pathway. *J. Neurosci.* **27**, 2866-2875 (2007).
159. Townsend, M., Mehta, T. & Selkoe, D.J. Soluble A β inhibits specific signal transduction cascades common to the insulin receptor pathway. *J. Biol. Chem.* **282**, 33305-33312 (2007).
160. Zhao, W.-Q. *et al.* Amyloid beta oligomers induce impairment of neuronal insulin receptors. *FASEB J.* **22**, 246-260 (2008).
161. De Felice, F.G. *et al.* Protection of synapses against Alzheimer's-linked toxins: insulin signaling prevents the pathogenic binding of A β oligomers. *Proc. Natl. Acad. Sci. USA* **106**, 1971-1976 (2009).

162. Magdesian, M.H. *et al.* Amyloid- β binds to the extracellular cysteine-rich domain of Frizzled and inhibits Wnt/ β -catenin signaling. *J. Biol. Chem.* **283**, 9359-9368 (2008).
163. Almeida, C.G., Takahashi, R.H. & Gouras, G.K. β -Amyloid accumulation impairs multivesicular body sorting by inhibiting the ubiquitin-proteasome system. *J. Neurosci.* **26**, 4277-4288 (2006).
164. Yang, A.J., Chandswangbhuvana, D., Margol, L. & Glabe, C.G. Loss of endosomal/lysosomal membrane impermeability is an early event in amyloid A β 1-42 pathogenesis. *J. Neurosci.* **52**, 691-698 (1998).
165. Nimmrich, V. *et al.* Amyloid β oligomers (A β 1-42 globulomer) suppress spontaneous synaptic activity by inhibition of P/Q-type calcium currents. *J. Neurosci.* **28**, 788-797 (2008).
166. Kawahara, M. & Kuroda, Y. Molecular mechanism of neurodegeneration induced by Alzheimer's β -amyloid protein: channel formation and disruption of calcium homeostasis. *Brain Res. Bull.* **53**, 389-397 (2000).
167. Soto, C. Unfolding the role of protein misfolding in neurodegenerative diseases. *Nat. Rev. Neurosci.* **4**, 49-60 (2003).
168. Guo, Q. *et al.* Alzheimer's Presenilin Mutation Sensitizes Neural Cells to Apoptosis Induced by Trophic Factor Withdrawal and Amyloid β -Peptide: Involvement of Calcium and Oxyradicals. *J. Neurosci.* **17**, 4212-4222 (1997).
169. Russo, C. *et al.* Presenilin-1 mutations in Alzheimer's disease. *Nature* **405**, 531 (2000).
170. Lee, J.-H. *et al.* Lysosomal Proteolysis and Autophagy Require Presenilin 1 and Are Disrupted by Alzheimer-Related PS1 Mutations. *Cell* **141**, 1146-1158 (2010).
171. Citron, M. Mutation of the [beta]-amyloid precursor protein in familial Alzheimer's disease increases [beta]-protein production. *Nature* **360**, 672-674 (1992).
172. Tomita, T. *et al.* The presenilin 2 mutation (N141I) linked to familial Alzheimer disease (Volga German families) increases the secretion of amyloid β protein ending at the 42nd (or 43rd) residue. *Proc. Natl. Acad. Sci. USA* **94**, 2025-2030 (1997).
173. Jiang, Q. *et al.* ApoE Promotes the Proteolytic Degradation of A β . *Neuron* **58**, 681-693 (2008).
174. Tai, L.M. *et al.* Levels of Soluble Apolipoprotein E/Amyloid- β (A β) Complex Are Reduced and Oligomeric A β Increased with APOE4 and Alzheimer Disease in a Transgenic Mouse Model and Human Samples. *J. Biol. Chem.* **288**, 5914-5926 (2013).
175. Castellano, J.M. *et al.* Human apoE isoforms differentially regulate brain amyloid- β peptide clearance. *Sci. Transl. Med.* **3**, 89ra57-89ra57 (2011).
176. Fan, J., Donkin, J. & Wellington, C. Greasing the wheels of A β clearance in Alzheimer's disease: the role of lipids and apolipoprotein E. *BioFactors* **35**, 239-248 (2009).
177. Verghese, P.B. *et al.* ApoE influences amyloid-beta (A β) clearance despite minimal apoE/A β association in physiological conditions. *Proc. Natl. Acad. Sci. USA* **110**, E1807-1816 (2013).
178. Taylor, B.M. *et al.* Spontaneous aggregation and cytotoxicity of the β -amyloid A β 1-40: a kinetic model. *J. Prot. Chem.* **22**, 31-40 (2003).
179. Benseny-Cases, N., Cócera, M. & Cladera, J. Conversion of non-fibrillar β -sheet oligomers into amyloid fibrils in Alzheimer's disease amyloid peptide aggregation. *Biochem. Biophys. Res. Commun.* **361**, 916-921 (2007).
180. Bartolini, M. *et al.* Insight Into the Kinetic of Amyloid β (1-42) Peptide Self-Aggregation: Elucidation of Inhibitors' Mechanism of Action. *ChemBioChem* **8**, 2152-2161 (2007).
181. Walsh, D.M. *et al.* Amyloid β -protein fibrillogenesis Structure and biological activity of protofibrillar intermediates. *J. Biol. Chem.* **274**, 25945-25952 (1999).
182. Cerf, E. *et al.* Antiparallel β -sheet: a signature structure of the oligomeric amyloid β -peptide. *Biochem. J.* **421**, 415-423 (2009).
183. Klimov, D.K. & Thirumalai, D. Dissecting the assembly of A β 16-22 amyloid peptides into antiparallel β sheets. *Structure* **11**, 295-307 (2003).
184. Nerelius, C. *et al.* α -Helix targeting reduces amyloid- β peptide toxicity. *Proc. Natl. Acad. Sci. USA* **106**, 9191-9196 (2009).

185. Marcon, G. *et al.* Amyloid formation from HypF-N under conditions in which the protein is initially in its native state. *J. Mol. Biol.* **347**, 323-335 (2005).
186. Canet, D. *et al.* Mechanistic studies of the folding of human lysozyme and the origin of amyloidogenic behavior in its disease-related variants. *Biochemistry* **38**, 6419-6427 (1999).
187. Necula, M., Kaye, R., Milton, S. & Glabe, C.G. Small molecule inhibitors of aggregation indicate that amyloid β oligomerization and fibrillization pathways are independent and distinct. *J. Biol. Chem.* **282**, 10311-10324 (2007).
188. Bitan, G. *et al.* Amyloid β -protein (A β) assembly: A β 40 and A β 42 oligomerize through distinct pathways. *Proc. Natl. Acad. Sci. USA* **100**, 330-335 (2003).
189. Lazo, N.D., Grant, M.A., Condon, M.C., Rigby, A.C. & Teplow, D.B. On the nucleation of amyloid β -protein monomer folding. *Prot. Sci.* **14**, 1581-1596 (2005).
190. Chen, Y.-R. & Glabe, C.G. Distinct Early Folding and Aggregation Properties of Alzheimer Amyloid- β Peptides A β 40 and A β 42 STABLE TRIMER OR TETRAMER FORMATION BY A β 42. *J. Biol. Chem.* **281**, 24414-24422 (2006).
191. Bernstein, S.L. *et al.* Amyloid- β protein oligomerization and the importance of tetramers and dodecamers in the aetiology of Alzheimer's disease. *Nat. Chem.* **1**, 326-331 (2009).
192. Munter, L.-M. *et al.* GxxxG motifs within the amyloid precursor protein transmembrane sequence are critical for the etiology of A β 42. *EMBO J.* **26**, 1702-1712 (2007).
193. Barrett, P.J. *et al.* The Amyloid Precursor Protein has a Flexible Transmembrane Domain and Binds Cholesterol. *Science* **336**, 1168-1171 (2012).
194. Glabe, C.G. Structural Classification of Toxic Amyloid Oligomers. *J. Biol. Chem.* **283**, 29639-29643 (2008).
195. Nelson, R. *et al.* Structure of the cross- β spine of amyloid-like fibrils. *Nature* **435**, 773-778 (2005).
196. Sawaya, M.R. *et al.* Atomic structures of amyloid cross- β spines reveal varied steric zippers. *Nature* **447**, 453-457 (2007).
197. Gremer, L. *et al.* Fibril structure of amyloid- β (1-42) by cryo-electron microscopy. *Science* **358**, 116-119 (2017).
198. Fukumoto, H., Cheung, B.S., Hyman, B.T. & Irizarry, M.C. β -Secretase protein and activity are increased in the neocortex in Alzheimer disease. *Arch. Neurol.* **59**, 1381-1389 (2002).
199. Yang, L.-B. *et al.* Elevated β -secretase expression and enzymatic activity detected in sporadic Alzheimer disease. *Nat. Med.* **9**, 3-4 (2003).
200. Li, R. *et al.* Amyloid β peptide load is correlated with increased β -secretase activity in sporadic Alzheimer's disease patients. *Proc. Natl. Acad. Sci. USA* **101**, 3632-3637 (2004).
201. Mawuenyega, K.G. *et al.* Decreased clearance of CNS β -amyloid in Alzheimer's disease. *Science* **330**, 1774-1774 (2010).
202. Walsh, D.M. & Selkoe, D.J. A β oligomers—a decade of discovery. *J. Neurochem.* **101**, 1172-1184 (2007).
203. Kuchibhotla, K.V. *et al.* A β plaques lead to aberrant regulation of calcium homeostasis *in vivo* resulting in structural and functional disruption of neuronal networks. *Neuron* **59**, 214-225 (2008).
204. Meyer-Luehmann, M. *et al.* Rapid appearance and local toxicity of amyloid- β plaques in a mouse model of Alzheimer's disease. *Nature* **451**, 720 (2008).
205. Vassar, R. BACE1 inhibitor drugs in clinical trials for Alzheimer's disease. *Alzheimer's Res. Ther.* **6**, 89 (2014).
206. Yan, R. Stepping closer to treating Alzheimer's disease patients with BACE1 inhibitor drugs. *Transl. Neurodegen.* **5**, 13 (2016).
207. Yan, R. *et al.* Membrane-anchored aspartyl protease with Alzheimer's disease β -secretase activity. *Nature* **402**, 533-537 (1999).
208. Yan, R. Physiological Functions of the β -Site Amyloid Precursor Protein Cleaving Enzyme 1 and 2. *Front. Mol. Neurosci.* **10** (2017).

209. Kennedy, M.E. *et al.* The BACE1 inhibitor verubecestat (MK-8931) reduces CNS β -amyloid in animal models and in Alzheimer's disease patients. *Sci. Transl. Med.* **8**, 363ra150-363ra150 (2016).
210. Doody, R.S. *et al.* A phase 3 trial of semagacestat for treatment of Alzheimer's disease. *N. Engl. J. Med.* **369**, 341-350 (2013).
211. Wong, G.T. *et al.* Chronic treatment with the gamma-secretase inhibitor LY-411,575 inhibits beta-amyloid peptide production and alters lymphopoiesis and intestinal cell differentiation. *J. Biol. Chem.* **279**, 12876-12882 (2004).
212. Milano, J. *et al.* Modulation of notch processing by γ -secretase inhibitors causes intestinal goblet cell metaplasia and induction of genes known to specify gut secretory lineage differentiation. *Toxicol. Sci.* **82**, 341-358 (2004).
213. Wilcock, G.K. *et al.* Efficacy and safety of tarenflurbil in mild to moderate Alzheimer's disease: a randomised phase II trial. *Lancet Neurol.* **7**, 483-493 (2008).
214. Green, R.C. *et al.* Effect of tarenflurbil on cognitive decline and activities of daily living in patients with mild Alzheimer disease: a randomized controlled trial. *JAMA* **302**, 2557-2564 (2009).
215. Crump, C.J. *et al.* BMS-708,163 targets presenilin and lacks notch-sparing activity. *Biochemistry* **51**, 7209-7211 (2012).
216. Jia, Q., Deng, Y. & Qing, H. Potential Therapeutic Strategies for Alzheimer's Disease Targeting or Beyond β -Amyloid: Insights from Clinical Trials. *BioMed Res. Internatl.* **22** (2014).
217. Martone, R.L. *et al.* Begacestat (GSI-953): a novel, selective thiophene sulfonamide inhibitor of amyloid precursor protein gamma-secretase for the treatment of Alzheimer's disease. *J. Pharmacol. Exp. Ther.* **331**, 598-608 (2009).
218. Wang, J., Ho, L. & Pasinetti, G.M. The Development of NIC5-15, a natural anti-diabetic agent, in the treatment of Alzheimer's disease. *Alzheimer's & Dement* **1**, S62 (2005).
219. Ono, K., Hasegawa, K., Yamada, M. & Naiki, H. Nicotine breaks down preformed Alzheimer's β -amyloid fibrils *in vitro*. *Biol. Psychiatry* **52**, 880-886 (2002).
220. Ono, K. *et al.* Vitamin A exhibits potent anti-amyloidogenic and fibril-destabilizing effects *in vitro*. *Exp. Neurol.* **189**, 380-392 (2004).
221. Hong, H.S. *et al.* Inhibition of Alzheimer's amyloid toxicity with a tricyclic pyrone molecule *in vitro* and *in vivo*. *J. Neurochem.* **108**, 1097-1108 (2009).
222. Rodríguez-Rodríguez, C. *et al.* Design, selection, and characterization of thioflavin-based intercalation compounds with metal chelating properties for application in Alzheimer's disease. *J. Am. Chem. Soc.* **131**, 1436-1451 (2009).
223. Ehrnhoefer, D.E. *et al.* Green tea (-)-epigallocatechin-gallate modulates early events in huntingtin misfolding and reduces toxicity in Huntington's disease models. *Hum. Mol. Genet.* **15**, 2743-2751 (2006).
224. Ehrnhoefer, D.E. *et al.* EGCG redirects amyloidogenic polypeptides into unstructured, off-pathway oligomers. *Nat. Struct. Mol. Biol.* **15**, 558-566 (2008).
225. Bieschke, J. *et al.* EGCG remodels mature α -synuclein and amyloid- β fibrils and reduces cellular toxicity. *Proc. Natl. Acad. Sci. USA* **107**, 7710-7715 (2010).
226. Soto, C. *et al.* β -sheet breaker peptides inhibit fibrillogenesis in a rat brain model of amyloidosis: implications for Alzheimer's therapy. *Nat. Med.* **4**, 822-826 (1998).
227. Ma, K., Thomason, L.A. & McLaurin, J. scyllo-Inositol, Preclinical, and Clinical Data for Alzheimer's. *Current State of Alzheimer's Disease Research and Therapeutics* **64**, 177 (2012).
228. Gervais, F. *et al.* Targeting soluble A β peptide with Tramiprosate for the treatment of brain amyloidosis. *Neurobiol. Aging* **28**, 537-547 (2007).
229. Aisen, P.S. *et al.* Tramiprosate in mild-to-moderate Alzheimer's disease—a randomized, double-blind, placebo-controlled, multi-centre study (the Alphase Study). *Arch. Med. Sci.* **7**, 102 (2011).
230. Lee, D. *et al.* A guanidine-appended scyllo-inositol derivative AAD-66 enhances brain delivery and ameliorates Alzheimer's phenotypes. *Sci. Rep.* **7**, 14125 (2017).

231. Habchi, J. *et al.* An anticancer drug suppresses the primary nucleation reaction that initiates the production of the toxic A β 42 aggregates linked with Alzheimer's disease. *Sci. Adv.* **2**, e1501244 (2016).
232. Eckman, E. & Eckman, C. A β -degrading enzymes: modulators of Alzheimer's disease pathogenesis and targets for therapeutic intervention. *Biochem. Soc. Trans.* **33**, 1101-1105 (2005).
233. Deane, R., Wu, Z. & Zlokovic, B.V. RAGE (Yin) versus LRP (Yang) balance regulates Alzheimer amyloid β -peptide clearance through transport across the blood-brain barrier. *Stroke* **35**, 2628-2631 (2004).
234. Galasko, D. *et al.* Clinical trial of an inhibitor of RAGE-A β interactions in Alzheimer disease. *Neurology* **82**, 1536-1542 (2014).
235. Schenk, D. *et al.* Immunization with amyloid-[beta] attenuates Alzheimer-disease-like pathology in the PDAPP mouse. *Nature* **400**, 173-177 (1999).
236. Bard, F. *et al.* Peripherally administered antibodies against amyloid [beta]-peptide enter the central nervous system and reduce pathology in a mouse model of Alzheimer disease. *Nat. Med.* **6**, 916-919 (2000).
237. Doody, R.S. *et al.* Phase 3 trials of solanezumab for mild-to-moderate Alzheimer's disease. *N. Engl. J. Med.* **370**, 311-321 (2014).
238. Hake, A.M. *et al.* Tolerability, pharmacokinetics and pharmacodynamics in healthy subjects of single doses of ly2599666, an antigen-binding fragment (fab) of a monoclonal antibody that targets soluble monomer a β linked to a poly (ethylene glycol) molecule. *Alzheimer's & Dement: J. Alzheimer's Assoc.* **13**, P250 (2017).
239. Cummings, J. *et al.* A randomized, double-blind, placebo-controlled phase 2 study to evaluate the efficacy and safety of crenezumab in patients with mild to moderate Alzheimer's disease. *Alzheimer's & Dement: J. Alzheimer's Assoc.* **10**, P275 (2014).
240. Sevigny, J. *et al.* Aducanumab (BIIB037), an anti-amyloid beta monoclonal antibody, in patients with prodromal or mild Alzheimer's disease: Interim results of a randomized, double-blind, placebo-controlled, phase 1b study. *Alzheimer's & Dementia: J. Alzheimer's Assoc.* **11**, P277 (2015).
241. Sevigny, J. *et al.* The antibody aducanumab reduces A β plaques in Alzheimer's disease. *Nature* **537**, 50-56 (2016).
242. Viglietta, V. *et al.* Titration dosing of aducanumab: Results of a 12-month interim analysis from a randomized, double-blind, placebo-controlled Phase 1b study (PRIME) in patients with prodromal or mild Alzheimer's Disease (S7. 003). *Neurology* **88**, S7. 003 (2017).
243. Serrano-Pozo, A. *et al.* Beneficial effect of human anti-amyloid- β active immunization on neurite morphology and tau pathology. *Brain* **133**, 1312-1327 (2010).
244. Escott-Price, V. *et al.* Common polygenic variation enhances risk prediction for Alzheimer's disease. *Brain* **138**, 3673-3684 (2015).
245. Langbaum, J.B. *et al.* Ushering in the study and treatment of preclinical Alzheimer disease. *Nat. Rev. Neurol.* **9**, 371 (2013).
246. Bulawa, C.E. *et al.* Tafamidis, a potent and selective transthyretin kinetic stabilizer that inhibits the amyloid cascade. *Proc. Natl. Acad. Sci. USA* **109**, 9629-9634 (2012).
247. Coelho, T. *et al.* Mechanism of Action and Clinical Application of Tafamidis in Hereditary Transthyretin Amyloidosis. *Neurol. Ther.* **5**, 1-25 (2016).
248. Necula, M. *et al.* Methylene blue inhibits amyloid A β oligomerization by promoting fibrillization. *Biochemistry* **46**, 8850-8860 (2007).
249. Vlad, S.C., Miller, D.R., Kowall, N.W. & Felson, D.T. Protective effects of NSAIDs on the development of Alzheimer disease. *Neurology* **70**, 1672-1677 (2008).
250. McKee, A.C. *et al.* Ibuprofen reduces A β , hyperphosphorylated tau and memory deficits in Alzheimer mice. *Brain Res.* **1207**, 225-236 (2008).

251. Galasko, D.R. *et al.* Safety, tolerability, pharmacokinetics, and A β levels after short-term administration of R-flurbiprofen in healthy elderly individuals. *Alzheimer Dis. Assoc. Dis.* **21**, 292-299 (2007).
252. Geerts, H. Drug evaluation:(R)-flurbiprofen--an enantiomer of flurbiprofen for the treatment of Alzheimer's disease. *IDrugs* **10**, 121-133 (2007).
253. Gupta-Bansal, R., Frederickson, R.C.A. & Brunden, K.R. Proteoglycan-mediated Inhibition of A β Proteolysis: a potential cause of senile plaque accumulation. *J. Biol. Chem.* **270**, 18666-18671 (1995).
254. McLaurin, J., Golomb, R., Jurewicz, A., Antel, J.P. & Fraser, P.E. Inositol stereoisomers stabilize an oligomeric aggregate of Alzheimer amyloid β peptide and inhibit A β -induced toxicity. *J. Biol. Chem.* **275**, 18495-18502 (2000).
255. Josephson, K., Ricardo, A. & Szostak, J.W. mRNA display: from basic principles to macrocycle drug discovery. *Drug Discov. Today* **19**, 388-399 (2014).
256. Nagai, Y. *et al.* Inhibition of polyglutamine protein aggregation and cell death by novel peptides identified by phage display screening. *J. Biol. Chem.* **275**, 10437-10442 (2000).
257. Sunita, G.G., Masayuki, K. & Koichi, N. Strong Inhibition of Beta-Amyloid Peptide Aggregation Realized by Two-Steps Evolved Peptides. *Chem. Biol. Drug Des.* **85**, 356-368 (2015).
258. Thomas, v.G. *et al.* Reduction of Alzheimer's Disease Amyloid Plaque Load in Transgenic Mice by D3, a D-Enantiomeric Peptide Identified by Mirror Image Phage Display. *ChemMedChem* **3**, 1848-1852 (2008).
259. Saunders, J.C. *et al.* An *in vivo* platform for identifying inhibitors of protein aggregation. *Nat. Chem. Biol.* **12**, 94-101 (2016).
260. Kim, W. *et al.* A high-throughput screen for compounds that inhibit aggregation of the Alzheimer's peptide. *ACS Chem. Biol.* **1**, 461-469 (2006).
261. Zhang, X. *et al.* A potent small molecule inhibits polyglutamine aggregation in Huntington's disease neurons and suppresses neurodegeneration *in vivo*. *Proc. Natl. Acad. Sci. USA* **102**, 892-897 (2005).
262. Kritzer, J.A. *et al.* Rapid selection of cyclic peptides that reduce [alpha]-synuclein toxicity in yeast and animal models. *Nat. Chem. Biol.* **5**, 655-663 (2009).
263. Tardiff, D.F. & Lindquist, S. Phenotypic screens for compounds that target the cellular pathologies underlying Parkinson's disease. *Drug Discov. Today* **10**, e121-128 (2013).
264. Lee, L.L., Ha, H., Chang, Y.T. & DeLisa, M.P. Discovery of amyloid-beta aggregation inhibitors using an engineered assay for intracellular protein folding and solubility. *Prot. Sci.* **18**, 277-286 (2009).
265. Tardiff, D.F. *et al.* Yeast reveal a "druggable" Rsp5/Nedd4 network that ameliorates alpha-synuclein toxicity in neurons. *Science* **342**, 979-983 (2013).
266. Pouplana, S. *et al.* Thioflavin-S staining of bacterial inclusion bodies for the fast, simple, and inexpensive screening of amyloid aggregation inhibitors. *Curr. Med. Chem.* **21**, 1152-1159 (2014).
267. Espargaro, A., Medina, A., Di Pietro, O., Munoz-Torrero, D. & Sabate, R. Ultra rapid *in vivo* screening for anti-Alzheimer anti-amyloid drugs. *Sci. Rep.* **6**, 23349 (2016).
268. Espargaro, A. *et al.* Combined *in vitro* Cell-Based/*in Silico* Screening of Naturally Occurring Flavonoids and Phenolic Compounds as Potential Anti-Alzheimer Drugs. *J. Nat. Prod.* **80**, 278-289 (2017).
269. Wang, L., Maji, S.K., Sawaya, M.R., Eisenberg, D. & Riek, R. Bacterial Inclusion Bodies Contain Amyloid-Like Structure. *PLoS Biology* **6**, e195 (2008).
270. Navarro, S. & Ventura, S. Fluorescent dye ProteoStat to detect and discriminate intracellular amyloid-like aggregates in Escherichia coli. *Biotechnol J* **9**, 1259-1266 (2014).
271. Garai, K., Crick, S.L., Mustafi, S.M. & Frieden, C. Expression and purification of amyloid-beta peptides from Escherichia coli. *Prot. Expr. Purif.* **66**, 107-112 (2009).

272. Shimomura, O. Structure of the chromophore of Aequorea green fluorescent protein. *FEBS Let.* **104**, 220-222 (1979).
273. Prasher, D.C. Using GFP to see the light. *Trends Genet.* **11**, 320-323 (1995).
274. Yang, F., Moss, L.G. & Phillips, G.N. The molecular structure of green fluorescent protein. *Nat. Biotechnol.* **14**, 1246-1251 (1996).
275. Waldo, G.S., Standish, B.M., Berendzen, J. & Terwilliger, T.C. Rapid protein-folding assay using green fluorescent protein. *Nat. Biotechnol.* **17**, 691-695 (1999).
276. Wurth, C., Guimard, N.K. & Hecht, M.H. Mutations that reduce aggregation of the Alzheimer's A β 42 peptide: an unbiased search for the sequence determinants of A β amyloidogenesis. *J. Mol. Biol.* **319**, 1279-1290 (2002).
277. McKoy, A.F., Chen, J., Schupbach, T. & Hecht, M.H. A novel inhibitor of Amyloid β (A β) peptide aggregation from high throughput screening to efficacy in an animal model of Alzheimer disease. *J. Biol. Chem.* **287**, 38992-39000 (2012).
278. Baine, M. *et al.* Inhibition of A β 42 aggregation using peptides selected from combinatorial libraries. *J. Pept. Sci.* **15**, 499-503 (2009).
279. Nair, S., Traini, M., Dawes, I.W. & Perrone, G.G. Genome-wide analysis of *Saccharomyces cerevisiae* identifies cellular processes affecting intracellular aggregation of Alzheimer's amyloid-beta42: importance of lipid homeostasis. *Mol. Biol. Cell* **25**, 2235-2249 (2014).
280. Bozyczko-Coyne, D. *et al.* CEP-1347/KT-7515, an inhibitor of SAPK/JNK pathway activation, promotes survival and blocks multiple events associated with A β -induced cortical neuron apoptosis. *J. Neurochem.* **77**, 849-863 (2001).
281. Weiner, J.H. *et al.* A novel and ubiquitous system for membrane targeting and secretion of cofactor-containing proteins. *Cell* **93**, 93-101 (1998).
282. Palmer, T. & Berks, B.C. The twin-arginine translocation (Tat) protein export pathway. *Nat. Rev. Microbiol.* **10**, 483-496 (2012).
283. Waraho-Zhmeyev, D., Gkogka, L., Yu, T.Y. & DeLisa, M.P. A microbial sensor for discovering structural probes of protein misfolding and aggregation. *Prion* **7**, 151-156 (2013).
284. Fisher, A.C., Kim, W. & DeLisa, M.P. Genetic selection for protein solubility enabled by the folding quality control feature of the twin-arginine translocation pathway. *Prot. Sci.* **15**, 449-458 (2006).
285. Lee, L.L., Ha, H., Chang, Y.T. & DeLisa, M.P. Discovery of amyloid-beta aggregation inhibitors using an engineered assay for intracellular protein folding and solubility. *Prot. Sci.* **18**, 277-286 (2009).
286. Zlokarnik, G. *et al.* Quantitation of transcription and clonal selection of single living cells with beta-lactamase as reporter. *Science* **279**, 84-88 (1998).
287. Kostelidou, K., Matis, I. & Skretas, G. Microbial genetic screens for monitoring protein misfolding associated with neurodegeneration: tools for identifying disease-relevant genes and for screening synthetic and natural compound libraries for the discovery of potential therapeutics. *Curr. Pharm. Des.* (Ahead of Print) (2018).
288. Hert, J., Irwin, J.J., Laggner, C., Keiser, M.J. & Shoichet, B.K. Quantifying biogenic bias in screening libraries. *Nat. Chem. Biol.* **5**, 479-483 (2009).
289. Bohacek, R.S., McMartin, C. & Guida, W.C. The art and practice of structure-based drug design: A molecular modeling perspective. *Med. Res. Rev.* **16**, 3-50 (1996).
290. Neumann, H. & Neumann-Staubitz, P. Synthetic biology approaches in drug discovery and pharmaceutical biotechnology. *Appl. Microbiol. Biotechnol.* **87**, 75-86 (2010).
291. Lau, J.L. & Dunn, M.K. Therapeutic peptides: Historical perspectives, current development trends, and future directions. *Bioorg. Med. Chem.* **10**, 2700-2707 (2017).
292. Craik, D.J., Fairlie, D.P., Liras, S. & Price, D. The future of peptide-based drugs. *Chem. Biol. Drug Des.* **81**, 136-147 (2013).
293. Wang, L. & Schultz, P.G. Expanding the genetic code. *Angew. Chem. Int. Ed. Engl.* **44**, 34-66 (2005).

294. Heinis, C. & Winter, G. Encoded libraries of chemically modified peptides. *Curr. Opin. Chem. Biol.* **26**, 89-98 (2015).
295. Zorzi, A., Deyle, K. & Heinis, C. Cyclic peptide therapeutics: past, present and future. *Curr. Opin. Chem. Biol.* **38**, 24-29 (2017).
296. Tavassoli, A. SICLOPPS cyclic peptide libraries in drug discovery. *Curr. Opin. Chem. Biol.* **38**, 30-35 (2017).
297. Chen, S. *et al.* Bicyclic Peptide Ligands Pulled out of Cysteine-Rich Peptide Libraries. *J. Am. Chem. Soc.* **135**, 6562-6569 (2013).
298. Kawakami, T. *et al.* Diverse backbone-cyclized peptides via codon reprogramming. *Nat. Chem. Biol.* **5**, 888 (2009).
299. Bowers, A.A. Biochemical and biosynthetic preparation of natural product-like cyclic peptide libraries. *MedChemComm* **3**, 905-915 (2012).
300. Bernstein, S.L. *et al.* Amyloid β -Protein: Monomer Structure and Early Aggregation States of A β 42 and Its Pro19 Alloform. *J. Am. Chem. Soc.* **127**, 2075-2084 (2005).
301. Scott, C.P., Abel-Santos, E., Wall, M., Wahnou, D.C. & Benkovic, S.J. Production of cyclic peptides and proteins *in vivo*. *Proc. Natl. Acad. Sci. USA* **96**, 13638-13643 (1999).
302. Scott, C.P., Abel-Santos, E., Jones, A.D. & Benkovic, S.J. Structural requirements for the biosynthesis of backbone cyclic peptide libraries. *Chem. Biol.* **8**, 801-815 (2001).
303. Shah, N.H. & Muir, T.W. Inteins: Nature's Gift to Protein Chemists. *Chem. Sci.* **5**, 446-461 (2014).
304. Perler, F.B. *et al.* Protein splicing elements: inteins and exteins--a definition of terms and recommended nomenclature. *Nucl. Acids Res.* **22**, 1125 (1994).
305. Perler, F.B. InBase: the intein database. *Nucl. Acids Res.* **30**, 383-384 (2002).
306. Wu, H., Hu, Z. & Liu, X.-Q. Protein trans-splicing by a split intein encoded in a split DnaE gene of *Synechocystis* sp. PCC6803. *Proc. Natl. Acad. Sci. USA* **95**, 9226-9231 (1998).
307. Horswill, A.R., Savinov, S.N. & Benkovic, S.J. A systematic method for identifying small-molecule modulators of protein-protein interactions. *Proc. Natl. Acad. Sci. USA* **101**, 15591-15596 (2004).
308. Tavassoli, A. & Benkovic, S.J. Genetically selected cyclic-peptide inhibitors of AICAR transformylase homodimerization. *Angew. Chem. Int. Ed. Engl.* **44**, 2760-2763 (2005).
309. Naumann, T.A., Tavassoli, A. & Benkovic, S.J. Genetic selection of cyclic peptide Dam methyltransferase inhibitors. *Chembiochem* **9**, 194-197 (2008).
310. Tavassoli, A. & Benkovic, S.J. Split-intein mediated circular ligation used in the synthesis of cyclic peptide libraries in *E. coli*. *Nat. Prot.* **2**, 1126-1133 (2007).
311. Villar-Pique, A. *et al.* The effect of amyloidogenic peptides on bacterial aging correlates with their intrinsic aggregation propensity. *J. Mol. Biol.* **421**, 270-281 (2012).
312. de Groot, N.S. & Ventura, S. Effect of temperature on protein quality in bacterial inclusion bodies. *FEBS Lett.* **580**, 6471-6476 (2006).
313. Carrió, M., González-Montalbán, N., Vera, A., Villaverde, A. & Ventura, S. Amyloid-like properties of bacterial inclusion bodies. *J. Mol. Biol.* **347**, 1025-1037 (2005).
314. de Groot, N.S., Sabate, R. & Ventura, S. Amyloids in bacterial inclusion bodies. *Trends Biochem. Sci.* **34**, 408-416 (2009).
315. Dasari, M. *et al.* Bacterial Inclusion Bodies of Alzheimer's Disease β -Amyloid Peptides Can Be Employed To Study Native-Like Aggregation Intermediate States. *ChemBioChem* **12**, 407-423 (2011).
316. Rosa, M., Roberts, C.J. & Rodrigues, M.A. Connecting high-temperature and low-temperature protein stability and aggregation. *PLoS ONE* **12**, e0176748 (2017).
317. Evans, T.C. & Xu, M.-Q. Mechanistic and kinetic considerations of protein splicing. *Chem. Rev.* **102**, 4869-4884 (2002).
318. Ding, Y. *et al.* Crystal structure of a mini-intein reveals a conserved catalytic module involved in side chain cyclization of asparagine during protein splicing. *J. Biol. Chem.* **278**, 39133-39142 (2003).

319. Ghosh, I., Sun, L. & Xu, M.-Q. Zinc inhibition of protein trans-splicing and identification of regions essential for splicing and association of a split intein. *J. Biol. Chem.* **276**, 24051-24058 (2001).
320. Naumann, T.A., Savinov, S.N. & Benkovic, S.J. Engineering an affinity tag for genetically encoded cyclic peptides. *Biotechnol. Bioeng.* **92**, 820-830 (2005).
321. García-Fruitós, E. *et al.* Aggregation as bacterial inclusion bodies does not imply inactivation of enzymes and fluorescent proteins. *Microb. Cell Fact.* **4**, 27 (2005).
322. Gavrín, L.K., Denny, R.A. & Saiah, E. Small molecules that target protein misfolding. *J. Med. Chem.* **55**, 10823-10843 (2012).
323. Nilsberth, C. *et al.* The 'Arctic' APP mutation (E693G) causes Alzheimer's disease by enhanced A β protofibril formation. *Nat. Neurosci.* **4**, 887-893 (2001).
324. Matis, I. *et al.* An integrated bacterial system for the discovery of chemical rescuers of disease-associated protein misfolding. *Nat. Biomed. Eng.* **1**, 838-852 (2017).
325. Levine, H. Thioflavine T interaction with synthetic Alzheimer's disease β -amyloid peptides: Detection of amyloid aggregation in solution. *Prot. Sci.* **2**, 404-410 (1993).
326. Vandersteen, A. *et al.* Molecular plasticity regulates oligomerization and cytotoxicity of the multipolypeptide-length amyloid- β peptide pool. *J. Biol. Chem.* **287**, 36732-36743 (2012).
327. Ahmed, M. *et al.* Structural conversion of neurotoxic amyloid-beta(1-42) oligomers to fibrils. *Nat. Struct. Mol. Biol.* **17**, 561-567 (2010).
328. Walsh, D.M. *et al.* A facile method for expression and purification of the Alzheimer's disease-associated amyloid beta-peptide. *FEBS J.* **276**, 1266-1281 (2009).
329. Lomakin, A., Benedek, G.B. & Teplow, D.B. Monitoring protein assembly using quasielastic light scattering spectroscopy. *Methods Enzymol.* **309**, 429-459 (1999).
330. Walsh, D.M. *et al.* Naturally secreted oligomers of amyloid β protein potently inhibit hippocampal long-term potentiation *in vivo*. *Nature* **416**, 535-539 (2002).
331. Walsh, D.M., Klyubin, I., Fadeeva, J.V., Rowan, M.J. & Selkoe, D.J. Amyloid- β oligomers: their production, toxicity and therapeutic inhibition. *Biochem. Soc. Transact.* **30**, 552-557 (2002).
332. Jin, S. *et al.* Amyloid- β (1-42) Aggregation Initiates Its Cellular Uptake and Cytotoxicity. *J. Biol. Chem.* **291**, 19590-19606 (2016).
333. Mason, R.P. *et al.* Distribution and Fluidizing Action of Soluble and Aggregated Amyloid β -Peptide in Rat Synaptic Plasma Membranes. *J. Biol. Chem.* **274**, 18801-18807 (1999).
334. Alexander, A.G., Marfil, V. & Li, C. Use of *Caenorhabditis elegans* as a model to study Alzheimer's disease and other neurodegenerative diseases. *Front. Genet.* **5**, (2014).
335. Hall, T.A. *et al.* BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. In: *Nucleic acids symposium series*. [London]: Information Retrieval Ltd., c1979-c2000., 1999. p. 95-98.
336. Kumar, S., Stecher, G. & Tamura, K. MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. *Mol. Biol. Evol.* **33**, 1870-1874 (2016).
337. Smith, C.K., Withka, J.M. & Regan, L. A Thermodynamic Scale for the beta.-Sheet Forming Tendencies of the Amino Acids. *Biochemistry* **33**, 5510-5517 (1994).
338. Packer, M.S. & Liu, D.R. Methods for the directed evolution of proteins. *Nat. Rev. Genet.* **16**, 379 (2015).
339. Frost, J.R., Jacob, N.T., Papa, L.J., Owens, A.E. & Fasan, R. Ribosomal synthesis of macrocyclic peptides *in vitro* and *in vivo* mediated by genetically encoded aminothiol unnatural amino acids. *ACS Chem. Biol.* **10**, 1805-1816 (2015).
340. Bionda, N. & Fasan, R. Ribosomal synthesis of natural-product-like bicyclic peptides in *Escherichia coli*. *ChemBioChem* **16**, 2011-2016 (2015).
341. Maksimov, M.O., Pan, S.J. & James Link, A. Lasso peptides: structure, function, biosynthesis, and engineering. *Nat. Prod. Rep.* **29**, 996-1006 (2012).
342. Austin, J., Wang, W., Puttamadappa, S., Shekhtman, A. & Camarero, J.A. Biosynthesis and Biological Screening of a Genetically Encoded Library Based on the Cyclotide MCotI-I. *ChemBioChem* **10**, 2663-2670 (2009).

343. Craik, D.J. & Malik, U. Cyclotide biosynthesis. *Curr Opin Chem Biol* **17**, 546-554 (2013).
344. Arnison, P.G. *et al.* Ribosomally synthesized and post-translationally modified peptide natural products: overview and recommendations for a universal nomenclature. *Nat. Prod. Rep.* **30**, 108-160 (2013).
345. Liu, C.C. & Schultz, P.G. Adding new chemistries to the genetic code. *Annu Rev Biochem* **79**, 413-444 (2010).
346. Young, T.S. *et al.* Evolution of cyclic peptide protease inhibitors. *Proc. Natl. Acad. Sci. USA* **108**, 11052-11056 (2011).
347. Eisenberg, D.S. & Sawaya, M.R. Structural studies of amyloid proteins at the molecular level. *Annu. Rev. Biochem.* **86**, 69-95 (2017).
348. Wright, G.S., Antonyuk, S.V., Kershaw, N.M., Strange, R.W. & Samar Hasnain, S. Ligand binding and aggregation of pathogenic SOD1. *Nat. Commun.* **4**, 1758 (2013).
349. Soragni, A. *et al.* A designed inhibitor of p53 aggregation rescues p53 tumor suppression in ovarian carcinomas. *Cancer Cell* **29**, 90-103 (2016).
350. Lipinski, C.A., Lombardo, F., Dominy, B.W. & Feeney, P.J. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv. Drug Deliv. Rev.* **23**, 3-25 (1997).
351. Habicht, G. *et al.* Directed selection of a conformational antibody domain that prevents mature amyloid fibril formation by stabilizing Abeta protofibrils. *Proc. Natl. Acad. Sci. USA* **104**, 19232-19237 (2007).
352. Lafaye, P., Achour, I., England, P., Duyckaerts, C. & Rougeon, F. Single-domain antibodies recognize selectively small oligomeric forms of amyloid β , prevent A β -induced neurotoxicity and inhibit fibril formation. *Mol. Immunol.* **46**, 695-704 (2009).
353. Kasturirangan, S. *et al.* Nanobody specific for oligomeric beta-amyloid stabilizes nontoxic form. *Neurobiol. Aging* **33**, 1320-1328 (2012).
354. Spurr, I.B. *et al.* Targeting Tumour Proliferation with a Small-Molecule Inhibitor of AICAR Transformylase Homodimerization. *ChemBioChem* **13**, 1628-1634 (2012).
355. Cecchelli, R. *et al.* Modelling of the blood-brain barrier in drug discovery and development. *Nat. Rev. Drug Discov.* **6**, 650 (2007).
356. He, Y., Yao, Y., Tzirka, S.E. & Cao, Y. Cell-Culture Models of the Blood-Brain Barrier. *Stroke* **45**, 2514-2526 (2014).
357. Brian Houston, J. & Carlile, D.J. Prediction of hepatic clearance from microsomes, hepatocytes, and liver slices. *Drug Metabol. Rev.* **29**, 891-922 (1997).
358. Grime, K. & Riley, R. The impact of *in vitro* binding on *in vitro-in vivo* extrapolations, projections of metabolic clearance and clinical drug-drug interactions. *Curr. Drug Metabol.* **7**, 251-264 (2006).
359. Ayrton, A. & Morgan, P. Role of transport proteins in drug absorption, distribution and excretion. *Xenobiotica* **31**, 469-497 (2001).
360. Trainor, G.L. The importance of plasma protein binding in drug discovery. *Expert Opin. Drug Discov.* **2**, 51-64 (2007).
361. Smith, D.A., Di, L. & Kerns, E.H. The effect of plasma protein binding on *in vivo* efficacy: misconceptions in drug discovery. *Nat. Rev. Drug Discov.* **9**, 929 (2010).
362. Bohnert, T. & Gan, L.-S. Plasma protein binding: From discovery to development. *J. Pharm. Sci.* **102**, 2953-2994 (2013).
363. Kepp, O., Galluzzi, L., Lipinski, M., Yuan, J. & Kroemer, G. Cell death assays for drug discovery. *Nat. Rev. Drug Discov.* **10**, 221 (2011).
364. Oakley, H. *et al.* Intraneuronal β -Amyloid Aggregates, Neurodegeneration, and Neuron Loss in Transgenic Mice with Five Familial Alzheimer's Disease Mutations: Potential Factors in Amyloid Plaque Formation. *J. Neurosci.* **26**, 10129-10140 (2006).

365. Eimer, W.A. & Vassar, R. Neuron loss in the 5XFAD mouse model of Alzheimer's disease correlates with intraneuronal A β (42) accumulation and Caspase-3 activation. *Mol. Neurodegener.* **8**, 2-2 (2013).
366. Nie, Q., Du, X.G. & Geng, M.Y. Small molecule inhibitors of amyloid beta peptide aggregation as a potential therapeutic strategy for Alzheimer's disease. *Acta Pharmacol. Sin.* **32**, 545-551 (2011).
367. Cheng, P.-N., Spencer, R., Woods, R.J., Glabe, C.G. & Nowick, J.S. Heterodivalent Linked Macrocyclic β -Sheets with Enhanced Activity against A β Aggregation: Two Sites Are Better Than One. *J. Am. Chem. Soc.* **134**, 14179-14184 (2012).
368. Driggers, E.M., Hale, S.P., Lee, J. & Terrett, N.K. The exploration of macrocycles for drug discovery—an underexploited structural class. *Nat. Rev. Drug Discov.* **7**, 608-624 (2008).
369. Sperling, R.A., Karlawish, J. & Johnson, K.A. Preclinical Alzheimer disease—the challenges ahead. *Nat. Rev. Neurol.* **9**, 54 (2012).
370. Sperling, R., Mormino, E. & Johnson, K. The evolution of preclinical Alzheimer's disease: Implications for prevention trials. *Neuron* **84**, 608-622 (2014).
371. Dubois, B. *et al.* Preclinical Alzheimer's disease: Definition, natural history, and diagnostic criteria. *Alzheimer's Dement.* **12**, 292-323 (2016).
372. Broersen, K. *et al.* A standardized and biocompatible preparation of aggregate-free amyloid beta peptide for biophysical and biological studies of Alzheimer's disease. *Prot. Eng. Des. Sel.* **24**, 743-750 (2011).
373. Brewer, G.J. & Torricelli, J.R. Isolation and culture of adult neurons and neurospheres. *Nat. Protoc.* **2**, 1490-1498 (2007).
374. Friedman, W. *et al.* Differential actions of neurotrophins in the locus coeruleus and basal forebrain. *Exp. Neurol.* **119**, 72-78 (1993).
375. Dietrich, P., Rideout, H.J., Wang, Q. & Stefanis, L. Lack of p53 delays apoptosis, but increases ubiquitinated inclusions, in proteasomal inhibitor-treated cultured cortical neurons. *Mol. Cell. Neurosci.* **24**, 430-441 (2003).
376. Link, C.D. Expression of human beta-amyloid peptide in transgenic *Caenorhabditis elegans*. *Proc. Natl. Acad. Sci.* **92**, 9368-9372 (1995).
377. Drake, J., Link, C.D. & Butterfield, D.A. Oxidative stress precedes fibrillar deposition of Alzheimer's disease amyloid β -peptide (1-42) in a transgenic *Caenorhabditis elegans* model. *Neurobiol. Aging* **24**, 415-420 (2003).