

**MOLECULAR EVOLUTION OF POTENTIALLY
THERAPEUTIC COMPOUNDS AGAINST PROTEIN
MISFOLDING DISEASES**

DOCTORAL THESIS

By

Dafni Chrysanthi Delivoria



Academic Supervisor: Prof. Fragiskos Kolisis

National Technical University of Athens

School of Chemical Engineering, Biotechnology Laboratory

2019

**ΧΡΗΣΗ ΜΕΘΟΔΩΝ ΜΟΡΙΑΚΗΣ ΕΞΕΛΙΞΗΣ ΓΙΑ
ΤΗΝ ΑΝΑΚΑΛΥΨΗ ΕΝ ΔΥΝΑΜΕΙ ΘΕΡΑΠΕΥΤΙΚΩΝ
ΕΝΩΣΕΩΝ ΚΑΤΑ ΑΣΘΕΝΕΙΩΝ ΠΟΥ ΣΧΕΤΙΖΟΝΤΑΙ
ΜΕ ΤΗΝ ΠΡΟΒΛΗΜΑΤΙΚΗ ΑΝΑΔΙΠΛΩΣΗ
ΠΡΩΤΕΪΝΩΝ**

ΔΙΔΑΚΤΟΡΙΚΗ ΔΙΑΤΡΙΒΗ

Δάφνη Χρυσάνθη Δεληβοριά



Ακαδημαϊκός Επιβλέπων: Καθ. Φραγκίσκος Κολίσης

Εθνικό Μετσόβιο Πολυτεχνείο

Σχολή Χημικών Μηχανικών, Εργαστήριο Βιοτεχνολογίας

2019

To my loving family

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 γνωμών του συγγραφέα.

(Ν. 5343/1392, Άρθρο 202)

Supervising Committee (Τριμελής Συμβουλευτική Επιτροπή)

Professor Emeritus, National Technical University of Athens (Supervisor)	Fragiskos Kolisis
Professor, National Technical University of Athens	Dimitrios Kekos
Research Associate Professor, National Hellenic Research Foundation	Georgios Skretas

Examination Committee (Επταμελής Εξεταστική Επιτροπή)

Professor Emeritus, National Technical University of Athens (Supervisor)	Fragiskos Kolisis
Professor, National Technical University of Athens	Dimitrios Kekos
Research Associate Professor, National Hellenic Research Foundation	Georgios Skretas
Associate Professor, University of Athens	Dimitris Hatzinikolaou
Assistant Professor, National Technical University of Athens	Evangelos Topakas
Professor, University of Athens	Spiros Efthimiopoulos
Professor, Universitat Autònoma de Barcelona	Salvador Ventura

Acknowledgements

The work described in this doctoral thesis was performed in the laboratory of Enzyme and Synthetic Biotechnology of the Institute of Biology, Medicinal Chemistry and Biotechnology at the National Hellenic Research Foundation (NHRF) and in collaboration with the Biotechnology laboratory of the School of Chemical Engineering at the National Technical University of Athens (NTUA). Financial support was provided by the following projects: (i) “STHENOS-b” (MIS 5002398) which is funded by the Operational Programme “Competitiveness, Entrepreneurship and Innovation” (NSRF 2014-2020) and co-financed by Greece and the EU (European Regional Development Fund), (ii) “NEUROTHERAPY” in the framework of the research grant “Aristeia” and (iii) “STHENOS”, the last two financed by the Hellenic General Secretariat of Research and Technology (GSRT) and the National Strategic Reference Framework (NSRF). This work was also supported by a short-term scientific mission Grant from COST Action BM1405.

First, I would like to express my sincere gratitude to my supervisor at NHRF, Research Associate Professor Georgios Skretas and my academic supervisor at NTUA, Professor Fragiskos Kolisis, for trusting me with this highly interesting and ambitious project and for their supervision and advice during my doctoral thesis. Especially, I am extremely grateful to Dr. Skretas whose everyday mentorship, guidance, dedication and encouragement have been a source of inspiration and significantly impacted on my academic development throughout these years.

Also, I would like to extend my appreciation to Professor Dimitris Kekos from the NTUA, Associate Professor Dimitris Hatzinikolaou from the University of Athens (UoA), Assistant Professor Evangelos Topakas from the NTUA, Professor Spiros Efthimiopoulos from

the UoA and Professor Salvador Ventura from the University of Barcelona for the honour of being in my supervising and examination committees.

I am forever obliged to my former colleague, Dimitra Gialama, who introduced me in the field of biotechnology and trained me in a wide range of molecular biology techniques when I first started at NHRF. Moreover, I would like to express my heartfelt gratitude to former and current members of our laboratory in NHRF and specifically Dimitra Gialama, Dimitra Zarafeta, Myrto Michou, Stamatia Bellou, Ilias Matis, Stefania Panoutsou, Zacharoula Linardaki, Maria Giannakou, Elena Vassilopoulou and Spyridoula Krikoni for their everyday support, advice and friendship. Thanks to all my former and current colleagues, this experience has truly been unforgettable. Stamatia is also gratefully acknowledged for her invaluable assistance with the evaluation of the sorted bacterial population against p53(Y220C), Zacharoula and Dr. Vassiliki Pletsas for performing and supervising the cancer cell line experiments respectively and Ilias for cloning the mutated variants of the selected peptides against A β 42.

I would also like to sincerely thank Nikoletta Papaevgeniou and Dr. Niki Chondrogianni for performing and supervising respectively the *C. elegans* paralysis experiments, Eleni Siapi for performing the mass spectrometry experiments, as well as Elena Taki and Professor Vassilis Zoumpourlis for performing and supervising respectively the stem cell experiments. Furthermore, I am truly thankful to Dr. Evangelia Chrysina and Dr. Spyros Zografos for hosting me in their laboratory at NHRF to perform the FPLC experiments and Dr. Christina Drakou, Panagiota Liggri and Dr. Katerina Tsitsanou for their assistance and support. Professor Aristotelis Xenakis and the members of his team at NHRF are also greatly appreciated for their daily collaboration and friendly working environment. Professor Alexander Pintzas is especially acknowledged for facilitating the DSF experiments and for his support throughout these years.

Moreover, I am particularly grateful to Professor George Georgiou for accepting me as a visiting student in his laboratory at the University of Texas at Austin as well as to the members of his laboratory: Joe Taft, Carl Denard, Giulia Agnello, Constantine Chrysostomou and Sebastian Schatzle, for introducing me in flow cytometric cell sorting, for their scientific advice, and most importantly for their support and friendship during my stay in Austin.

Likewise, I would like to express my appreciation to Professor Michele Vendruscolo for hosting me at the Centre for Misfolding Diseases of the Department of Chemistry of the University of Cambridge, where I was trained in *in vitro* aggregation techniques and *C. elegans* models of AD. My warmest gratitude goes to Dr. Johnny Habchi and Dr. Michele Perni for their invaluable supervision during my stay in Cambridge, Sean Chia and Sam Casford for introducing me *in vitro* and *in vivo* techniques respectively, Rodrigo Lessa Cataldi for his help with TEM measurements and Benedetta Mannini for her help with DLS experiments. Johnny, Michele, Sean, Sam, Rodrigo, Benedetta as well as Sunehra Sarwat, and Giulia Vecchi are also sincerely appreciated for their hospitality and friendship during my stay in the UK.

Lastly, this endeavour would have been impossible without the love and support from my friends and family, whom I am truly fortunate to have in my life.

Περίληψη

Η προβληματική αναδίπλωση πρωτεϊνών και η επικείμενη συσσωμάτωσή τους αποτελεί κοινό μοριακό χαρακτηριστικό μιας πληθώρας ανθρώπινων ασθενειών που ονομάζονται ασθένειες προβληματικής αναδίπλωσης πρωτεϊνών (protein misfolding diseases, PMDs). Στις PMDs περιλαμβάνονται ασθένειες με ποικίλες παθολογίες και συμπτώματα, όπως η νόσος Alzheimer, η νόσος Parkinson, η κυστική ίνωση και ο καρκίνος. Ανεξάρτητα με την παθογένεσή τους, η συντριπτική πλειοψηφία των PMDs είναι επί της παρούσης ανίατη και επιβάλλει ένα τεράστιο κοινωνικο-οικονομικό αντίκτυπο στην ανθρωπότητα. Συνεπώς, η ανακάλυψη μορίων με την ικανότητα να επιδιορθώνουν την αναδίπλωσή των προβληματικών αυτών πρωτεϊνών αποτελεί μία ευοίωνη προσέγγιση για την ανάπτυξη νέων φαρμάκων κατά των PMDs.

Για το σκοπό αυτό, στην παρούσα διδακτορική διατριβή αναφέρεται η ανάπτυξη μίας καινοτόμου βακτηριακής πλατφόρμας για την ανάκαλυψη εν δυνάμει θεραπευτικών ενώσεων κατά ενός ευρέος φάσματος PMDs. Σε αυτό το σύστημα, βακτηριακά κύτταρα *Escherichia coli* τροποποιήθηκαν γενετικά ώστε να εκτελούν δύο ταυτόχρονες διεργασίες: (i) να παράγουν συνδυαστικές βιβλιοθήκες με πάνω από 200 εκατομμύρια διαφορετικά κυκλικά ολιγοπεπτίδια και (ii) να επιτρέπουν την ταυτοποίηση των βιοδραστικών ενώσεων με την ικανότητα να επιδιορθώνουν την προβληματική αναδίπλωση των πρωτεϊνών-στόχων ή/και να αναστέλλουν την παθολογική συσσωμάτωσή τους, χρησιμοποιώντας ένα απλό σύστημα γενετικής επιλογής. Το σύστημα αυτό στηρίζεται στη σύνδεση της προβληματικής αναδίπλωσης της πρωτεΐνης-στόχου με την εμφάνιση ενός φθορίζοντα φαινότυπου, με αποτέλεσμα να επιτρέπεται ο εντοπισμός και η απομόνωση των βιοδραστικών ενώσεων χρησιμοποιώντας κυτταρομετρία ροής υψηλής απόδοσης. Με τον τρόπο αυτό, μειώνεται σημαντικά ο χρόνος καθώς και το κόστος της ανακάλυψης πιθανών θεραπευτικών ενώσεων κατά των PMDs.

Η εν λόγω μεθοδολογία χρησιμοποιήθηκε κατά τεσσάρων μη-συναφών πρωτεϊνικών στόχων: (i) μίας καρκινογόνου παραλλαγής της ανθρώπινης ογκοκατασταλτικής πρωτεΐνης p53 που περιέχει την αντικατάσταση της τυροσίνης στη θέση 220 από κυστεΐνη (p53(Y220C)), (ii) του αμυλοειδούς β πεπτιδίου (Aβ42) που έχει συσχετισθεί ευρέως με τη νόσο Alzheimer, (iii) μίας μεταλλαγμένης μορφής της δισμουτάσης του υπεροξειδίου 1 που περιέχει την αντικατάσταση της αλανίνης στη θέση 4 από βαλίνη (SOD1(A4V)) και η οποία έχει συσχετισθεί με δριμύτατες μορφές της νόσου του κινητικού νευρώνα και (iv) με μία παθολόγνα μορφή της πρωτεΐνης huntingtin που περιέχει 97 κατάλοιπα γλουταμίνης (HTT-97Q), και η οποία έχει συσχετισθεί με την νόσο του Huntington. Χρησιμοποιώντας το εν λόγω σύστημα ήταν δυνατή η απομόνωση τεσσάρων διακριτών βακτηριακών πληθυσμών που παράγουν κυκλικά πεπτίδια με την ικανότητα να επιδιορθώνουν την προβληματική αναδίπλωση ή/και να αναστέλλουν την παθολογική συσσωμάτωσή των πρωτεϊνών-στόχων. Τα επιλεγμένα κυκλικά πεπτίδια κατά της p53(Y220C) και της Aβ42 εξετάστηκαν περαιτέρω *in vitro* και *in vivo* με σκοπό την επιβεβαίωση της βιοδραστικότητάς τους.

Συγκεκριμένα, στην περίπτωση της p53(Y220C), εφαρμογή της παραπάνω βιοτεχνολογικής μεθοδολογίας οδήγησε στην ανακάλυψη έξι ενώσεων με πιθανή βιοδραστικότητα. Οι ενώσεις αυτές εξετάστηκαν περαιτέρω ως προς την ικανότητά τους να αυξάνουν τη θερμοδυναμική σταθερότητα της πρωτεΐνης-στόχου, να αναστέλλουν τη συσσωμάτωσή της και να αποκαθιστούν τη φυσιολογική λειτουργία της σε μία καρκινική σειρά ανθρώπινου μελανώματος που φέρει την Y220C μετάλλαξη. Στην πλειοψηφία τους, οι επιλεγμένες ενώσεις βρέθηκαν ικανές να επιβραδύνουν τον ανεξέλεγκτο πολλαπλασιασμό των καρκινικών κυττάρων, με μία εξ' αυτών να παρουσιάζει αξιόλογη αντικαρκινική επίδραση. Επιπλέον, η ένωση αυτή βρέθηκε να αυξάνει σημαντικά τη θερμοδυναμική σταθερότητα της p53(Y220C) μεταλλαγής, ενώ δύο εκ των υπόλοιπων ενώσεων βρέθηκαν ικανές να αναστέλλουν τη συσσωμάτωση της πρωτεΐνης, χωρίς να επηρεάζουν τη θερμοδυναμική της

σταθερότητα. Τα αποτελέσματα αυτά υποδηλώνουν ότι το βακτηριακό σύστημα επιτρέπει την ανακάλυψη εν δυνάμει θεραπευτικών ενώσεων κατά της p53(Y220C) μεταλλαγής με πιθανή δράση μέσω διαφορετικών μοριακών μηχανισμών.

Στην περίπτωση της Αβ42, η εφαρμογή του βακτηριακού συστήματος επέτρεψε τον εντοπισμό παραπάνω από 400 κυκλικών πεπτιδίων με πιθανή δράση κατά της συσσωμάτωσης της Αβ42, τα οποία ομαδοποιήθηκαν σε 20 διακριτές οικογένειες βάσει της πεπτιδικής τους ακολουθίας. Στη συνέχεια, δύο αντιπροσωπευτικά πεπτίδια των δύο επικρατούντων οικογενειών συντέθηκαν οργανικά και εξετάστηκαν περαιτέρω *in vitro* και *in vivo*. Οι μελέτες αυτές ανέδειξαν την ικανότητα και των δύο επιλεγμένων κυκλικών πεπτιδίων να αναστέλλουν τη συσσωμάτωση της Αβ42 με μεγάλη αποτελεσματικότητα. Επίσης, τα δύο πεπτίδια βρέθηκαν να επηρεάζουν σε διαφορετικό βαθμό την πρωτογενή και δευτερογενή πυρηνογένεση της διαδικασίας συσσωμάτωσης, υποδηλώνοντας ότι μόρια από διαφορετικές οικογένειες μπορεί να δρουν μέσω διαφορετικών μηχανισμών. Κατόπιν, οι δύο επιλεγμένες κυκλικές ενώσεις μελετήθηκαν σε διαγονιδιακά στελέχη του νηματώδη σκώληκα *Caenorhabditis elegans* που λειτουργούν ως μοντέλα της νόσου Alzheimer και βρέθηκαν ικανά να προστατεύουν τα ζώα από την κυτταροτοξική συσσωμάτωση της Αβ42 καθώς οδήγησαν σε μείωση των εναποθέσεων της πρωτεΐνης στα κύτταρα των σκωλήκων, αύξηση της κινητικότητάς τους και μείωση του ρυθμού παράλυσής τους. Τέλος, εφαρμόζοντας αλληλούχιση νέας γενιάς και στοχευμένη μεταλλαξιγένεση επιτεύχθει η γρήγορη ταυτοποίηση των σχέσεων δομής-δραστικότητας των κυκλικών πεπτιδίων και τα απαραίτητα χαρακτηριστικά για μέγιστη δραστηριότητα.

Συνολικά, η βακτηριακή πλατφόρμα που παρουσιάζεται στην παρούσα εργασία αποτελεί μία ιδιαιτέρως ευπροσάρμοστη τεχνολογία που επιτρέπει την γρήγορη και οικονομική αξιολόγηση παραπάνω από 200 εκατομμυρίων διαφορετικών κυκλικών πεπτιδίων με σκοπό την ανακάλυψη και τον χαρακτηρισμό ενώσεων με την ικανότητα να επιδιορθώνουν

την προβληματική αναδίπλωση ή/και την συσσωμάτωση των πρωτεϊνών-στόχων. Εξ' όσων γνωρίζουμε, το γεγονός αυτό την καθιστά το πρώτο σύστημα σάρωσης μορίων που έχει αναφερθεί μέχρι σήμερα στη βιβλιογραφία, που επιτρέπει την αξιολόγηση ενός πρωτοφανούς αριθμού ενώσεων, με κριτήριο τη λειτουργικότητά τους έναντι της παθογενούς πρωτεϊνικής αναδίπλωσης και το οποίο έχει πιθανή εφαρμογή σε ένα ευρύ φάσμα PMDs.

Abstract

Protein misfolding and aggregation are defining features of a wide range of human conditions which have been collectively termed protein misfolding diseases (PMDs). These include disorders with diverse pathologies and symptoms, such as Alzheimer's disease, Parkinson's disease, cystic fibrosis and cancer. Whichever their pathogenesis however, the vast majority of PMDs remain to date incurable and impose a very high socio-economic burden on humanity.

To address this unmet medical need, in this thesis we report the development of a novel integrated bacterial platform for the discovery of potential therapeutics against a wide range of PMDs. In this system, *Escherichia coli* cells are genetically engineered in order to perform two simultaneous tasks: (i) produce combinatorial libraries of more than 200 million drug-like, head-to-tail cyclic oligopeptides using protein-splicing technology and (ii) enable the identification of the bioactive cyclic peptides that correct the problematic folding and/or inhibit the aggregation of disease-associated misfolding-prone proteins (MisPs) using a genetic assay that links the folding of the target MisP with a fluorescent phenotype. In this way, the bioactive cyclic peptide hits can be identified in an ultrahigh-throughput manner using flow cytometric cell sorting, thus significantly decreasing the overall cost, time and complexity of early drug discovery for PMDs.

We utilized the developed platform against four unrelated targets: (i) a carcinogenic variant of the tumour suppressor protein p53, which contains a tyrosine to cysteine substitution at position 220 (p53(Y220C)), (ii) the 42 residue form of the amyloid β peptide (A β 42), widely associated with Alzheimer's disease, (iii) the Cu/Zn superoxide dismutase 1 containing an alanine to valine substitution at position 4 (SOD1(A4V)), which is linked with amyotrophic lateral sclerosis and (iv) a pathogenic variant of huntingtin containing a 97-glutamine

expansion (HTT-97Q), which is associated with Huntington's disease. This process resulted in the isolation of four distinct bacterial populations that produce cyclic peptides with the ability to rescue protein misfolding and aggregation of the respective target MisP. The identified hits against p53(Y220C) and A β 42 were further tested *in vitro* and *in vivo* in order to verify their bioactivity.

Specifically, using this system, we were able to identify six putative rescuers of p53(Y220C) misfolding, which were further evaluated for their ability to increase the thermodynamic stability of the target protein, inhibit its aggregation and restore its physiological function in mammalian cell lines. The majority of the selected hits were found to affect cancer cell growth, with one of them exhibiting a pronounced anti-cancer effect. Furthermore, *in vitro* evaluation of the selected hits revealed that the latter cyclic peptide was able to significantly stabilize p53(Y220C), while two other cyclic peptides were able to interfere with the aggregation of p53(Y220C), without affecting its thermodynamic stability. This indicates that the bacterial platform enables the identification of bioactive cyclic peptides that rescue the pathogenic misfolding of p53(Y220C) by different mechanisms.

Moreover, using this system we were able to identify more than 400 putative modulators of A β 42 aggregation, which we divided into 20 distinct clusters sharing similar sequence characteristics. Two representative members of the two most dominant clusters were chemically synthesized and evaluated *in vitro* and *in vivo*. Both selected macrocycles were found to potently inhibit the aggregation of A β 42 *in vitro* at sub-stoichiometric ratios, by interfering with both the primary and secondary nucleation steps of A β 42 aggregation, albeit to a different extent. This suggests that members from different clusters may ameliorate the pathogenesis of A β 42 by diverse inhibitory mechanisms. Furthermore, when tested in two *Caenorhabditis elegans* models of Alzheimer's disease both cyclic peptides were able to protect from the A β 42 aggregation associated cytotoxicity, by decreasing the A β 42 deposits

found in the worms' body wall muscle cells, increasing their locomotion and delaying their paralysis. Finally, a combination of deep sequencing and site-directed mutagenesis analyses allowed the rapid definition of structure-activity relationships and consensus motifs required for optimal bioactivity among each selected cluster.

Overall, the herein described approach represents a highly adaptable ultrahigh-throughput strategy that enables the facile and cost-effective investigation of more than 200 million different molecules and the discovery and characterization of potent rescuers of pathogenic protein misfolding and/or aggregation. To our knowledge, this constitutes the largest functional screen of drug-like molecular entities described to date, with potential applicability against a broad range of PMDs.

Thesis Publications

This thesis is based on the work contained in the following papers:

1. Delivoria, D.C., Chia S., Habchi, J., Perni, M., Papaevgeniou, N., Matis, I., Chondrogianni, N., Dobson, C.M., Vendruscolo, M., Skretas, G. 2019. Bacterial biosynthesis and direct functional screening of molecular libraries with expanded diversities for discovering inhibitors of pathogenic protein aggregation. *submitted*.
2. Matis, I., Delivoria, D.C., Mavroidi, B., Papaevgeniou, N., Panoutsou, S., Bellou, S., Papavasileiou, K.D., Linardaki, Z.I., Stavropoulou, A.V., Vekrellis, K., Boukos, N., Kolisis, F.N., Gonos, E.S., Margarity, M., Papadopoulos, M.G., Efthimiopoulos, S., Pelecanou, M., Chondrogianni, N. & Skretas, G. 2017. An integrated bacterial system for the discovery of chemical rescuers of disease-associated protein misfolding. *Nature Biomedical Engineering*, 1(10), 838-852. doi: 10.1038/s41551-017-0144-3

Publications in preparation at the time of writing:

1. Delivoria, D.C., Papaevgeniou, N., Matis, I., Chondrogianni, Skretas, G. A bacterial discovery platform uncovers different families of peptide macrocycles against A β aggregation and neurotoxicity.

Papers not included in the thesis:

1. Gialama, D., Kostelidou, K., Michou, M., Delivoria, D.C., Kolisis, F.N., Skretas, G. 2017. Development of *Escherichia coli* strains that withstand membrane protein-induced toxicity and achieve high-level recombinant membrane protein production. *ACS Synthetic Biology*. 6(2):284-300. doi: 10.1021/acssynbio.6b00174.

2. Gialama, D., Delivoria, D.C., Michou, M., Giannakopoulou, A., Skretas, G. 2017. Functional requirements for DjlA- and RraA-mediated enhancement of recombinant membrane protein production in the engineered *Escherichia coli* strains SuptoxD and SuptoxR. *Journal of Molecular Biology*, 429(12): 1800-1816. doi: 10.1016/j.jmb.2017.05.003.

Table of Contents

Supervising Committee (Τριμελής Συμβουλευτική Επιτροπή)	7
Examination Committee (Επταμελής Εξεταστική Επιτροπή)	7
Acknowledgements.....	ix
Περίληψη	xiii
Abstract.....	xvii
Thesis Publications	xxi
Table of Contents.....	xxiii
List of Figures	xxx
List of Tables	xxxvii
List of Abbreviations	xxxix
Chapter 1 - Protein misfolding and its association with human diseases	1
1.1. Protein folding	1
1.2. <i>In vivo</i> protein folding and homeostasis	4
1.3. Protein Misfolding Diseases	9
1.3.1. Mechanisms of PMDs	10
1.3.2. PMDs with respect to gain and loss of function.....	19
1.3.3. p53-related cancer as a PMD	21
1.3.4. Alzheimer's disease and the amyloid- β peptide.....	24
1.4. Therapeutic strategies against PMDs.....	27
1.4.1. Manipulating the proteostasis network.....	28

1.4.2. Enhancing aggregate clearance by immunization	29
1.4.3. Targeting the production of MisPs	29
1.4.4. Stabilizing the native state of MisPs.....	31
1.4.5. Inhibiting the toxic aggregation of MisPs	32
1.5. Thesis aims and outline.....	36
Chapter 2 – Construction and biosynthesis of a combinatorial cyclic peptide library with expanded diversity	41
2.1. Peptides in drug discovery	41
2.2. Methods for the preparation of cyclic peptide libraries	42
2.3. Construction of a combinatorial cyclic peptide library with expanded diversity using the SICLOPPS technology.....	49
2.4. Quality assessment of the constructed combinatorial library of random cyclic oligopeptides.....	53
2.5. Discussion.....	61
Chapter 3 – Development of a generalized genetic screen for monitoring protein misfolding and aggregation.....	63
3.1. Screens for the identification of aggregation inhibitors as potential therapeutics against PMDs	63
3.1.1. <i>In silico</i> screening assays for the identification of protein aggregation inhibitors.....	63
3.1.2 <i>In vitro</i> screening assays for the identification of protein aggregation inhibitors ...	65
3.1.3. Cell based screens for the identification of protein aggregation inhibitors	67

3.2. Development of a generalized genetic screen for monitoring protein aggregation <i>in vivo</i>	70
3.2.1. Monitoring the aggregation of p53 oncogenic variants	71
3.2.2. Monitoring the aggregation of SOD1 variants	74
3.2.3. Monitoring the aggregation of polyQ-HTT _{ex1} variants	78
3.2.4. Optimization of the MisP-GFP assay	80
3.3. Discussion	84
Chapter 4 - Development of an integrated bacterial system for the discovery of potential PMD therapeutics	85
4.1. Targeting cancer - Identification of p53C(Y220C) folding rescuers using the generated ultrahigh-throughput biotechnology platform	87
4.2. Targeting AD – Identification of Aβ42 aggregation inhibitors using the generated ultrahigh-throughput biotechnology platform	100
4.3. Targeting ALS – Identification of SOD1(A4V) folding rescuers using the generated ultrahigh-throughput biotechnology platform	107
4.4. Targeting HD – Identification of HTT _{ex1} -97Q aggregation inhibitors using the generated ultrahigh-throughput biotechnology platform	109
4.5. Discussion	111
Chapter 5 – Evaluation of the selected cyclic peptides on their ability to rescue the misfolding of p53(Y220C) <i>in vitro</i> and restore its pro-apoptotic function in cancer cell lines	113
5.1. Biosynthesis and isolation of selected cyclic peptides from bacterial cells	113

5.2. Evaluation of the selected cyclic peptides' ability to promote cell death of a cancer cell line expressing the p53(Y220C) mutation	116
5.3. Biosynthesis and purification of T-p53C and T-p53C(Y220C) for use in <i>in vitro</i> experiments	121
5.4. Evaluation of the selected cyclic peptides' ability to increase the thermodynamic stability of T-p53C(Y220C) <i>in vitro</i>	122
5.5. Evaluation of the selected cyclic peptides' ability to inhibit T-p53C(Y220C) aggregation <i>in vitro</i>	128
5.6. Discussion	131
Chapter 6 – Evaluation of the selected cyclic peptides on their ability to affect Aβ42 aggregation and its associated neurotoxicity <i>in vitro</i> and <i>in vivo</i>	133
6.1. High-throughput analysis of the isolated hits	133
6.2. <i>In vitro</i> evaluation of A β C7-1 and A β C7-14	140
6.3. <i>In vivo</i> evaluation of A β C7-1 and A β C7-14	146
6.4. Structure-activity relationships of A β C7-1 and A β C7-14	152
6.5. Discussion	157
Chapter 7 – Conclusion	159
Chapter 8 – Future perspectives	165
Chapter 9 - Materials and Methods	169
Reagents and chemicals	169
Construction of the combinatorial cyclic peptide library	169
Plasmid constructions	171

Plasmids involved in cyclic peptide production	171
Plasmids involved in MisP production	173
Cyclic peptide library screening for the identification of p53C(Y220C) folding rescuers....	176
FACS sorting utilizing a two-plasmid system comprising of pETp53C(Y220C)-GFP and the combined pSICLOPPS-NuX ₁ X ₂ X ₃ -X ₅ vector library.....	176
FACS sorting sequential integration of two two-plasmid systems comprising of pETT- p53C(Y220C)-GFP or pASKT-p53C(Y220C)-GFP and the combined pSICLOPPS- NuX ₁ X ₂ X ₃ -X ₅ vector library	177
FACS sorting utilizing a three-plasmid system comprising of pCDFT-p53C(Y220C)-BFP, pETA β 42-GFP and the pSICLOPPS-NuX ₁ X ₂ X ₃ -X ₆ vector libraries.....	177
Cyclic peptide library screening for the identification of A β 42 aggregation inhibitors	178
Cyclic peptide library screening for the identification of SOD1(A4V) folding rescuers	179
Cyclic peptide library screening for the identification of HTT _{ex1} -97Q aggregation inhibitors	180
Protein/cyclic peptide production in liquid cultures.	180
Bacterial cell fluorescence	181
In-gel fluorescence and western blot analyses	182
High-throughput sequencing analysis	183
Peptide sequence similarity analysis and clustering	184
Cyclic peptide on-column purification	184
Chitin affinity chromatography	184
Immobilized metal affinity chromatography	185

T-p53C and T-p53(Y220C) overexpression and purification by IMAC and SEC	186
Preparation of A β 42 samples	187
Differential Scanning Fluorimetry (DSF)	187
T-p53C and T-p53C(Y220C) aggregation kinetics experiments.....	188
A β 42 aggregation kinetics experiments.....	188
Transmission Electron Microscopy	189
Dot-blot assay	189
Cell viability experiments	190
<i>C. elegans</i> motility assay	190
<i>C. elegans</i> paralysis assay.....	192
Statistical analyses	193
References	195
Appendices	231
Appendix A – Graphical representation of fitting the data obtained from DSF experiments involving T-p53C(Y220C), to the Boltzmann equation using the OriginPro software.	231
Appendix B – Sequences and frequency of appearance of the A β -targeting heptapeptide sequences appearing at least twenty times within the sorted population, as determined by high-throughput sequencing of the enriched library after the 7th round of sorting.	234
Appendix C – Sequences and frequency of appearance of the A β -targeting Cluster I heptapeptide sequences, as determined by high-throughput sequencing of the enriched library after the 7th round of sorting.	245

Appendix D – Sequences and frequency of appearance of the A β -targeting Cluster II heptapeptide sequences as determined by high-throughput sequencing of the enriched library after the 7th round of sorting.248

Appendix E – Plasmids used herein.....249

Appendix F – Primers used herein.....254

List of Figures

Figure 1.1. Schematic representation of a complex energy landscape	2
Figure 1.2. Schematic representation of a combined energy landscape for both protein folding and aggregation.....	3
Figure 1.3. Schematic of the possible conformational states that can be adopted by a nascent protein and the in-between state transitions.....	5
Figure 1.4. The proteostasis network (PN)	6
Figure 1.5. Protein folding in the ER.....	7
Figure 1.6. Schematic of the amyloid aggregation mechanism.....	16
Figure 1.7. Functions of p53 in response to stress signals.....	21
Figure 1.8. Schematic of p53's domain structure indicating the relative frequency of oncogenic mutations for each position.....	22
Figure 1.9. Ribbon diagram of the DNA-binding domain of p53 bound to a target DNA sequence.....	22
Figure 1.10. Schematic of APP processing to release the disease-associated A β peptide	25
Figure 1.11. Schematic of brain tissue from normal patients and AD patients	25
Figure 1.12. Schematic representation of the energy landscape of A β aggregation depicting the relationship between the size of A β aggregated species to their associated toxicity.....	26
Figure 2.1. Display technologies for the production and high-throughput screening of cyclic peptides	46
Figure 2.2. Schematic of protein splicing	48
Figure 2.3. SICLOPPS mechanism standard pathway and side reactions.....	51
Figure 2.4. Schematic of the generation of the cyclic-NuX ₁ X ₂ X ₃ -X ₆ peptide libraries via the SICLOPPS technology.....	53

Figure 2.5. Indicative western blot analysis using an anti-CBD antibody of fourteen randomly selected individual clones from the constructed cyclo-NuX ₁ X ₂ X ₃ X ₄ X ₅ hexapeptide sub-library, demonstrating that individual clones can exhibit variable levels of expression	54
Figure 2.6. Heat map representation of the amino acid distribution at each position of the constructed cyclo-CysX ₁ X ₂ X ₃ -X ₆ (top), cyclo-SerX ₁ X ₂ X ₃ -X ₆ (middle) and cyclo-ThrX ₁ X ₂ X ₃ -X ₆ (bottom) sub-libraries, as demonstrated by the deep sequencing analysis results.....	59
Figure 3.1. Schematic of the MisP-GFP genetic system for monitoring MisP folding and aggregation.....	71
Figure 3.2. Monitoring the aggregation of p53 oncogenic variants.....	73
Figure 3.3. Solubility analysis of <i>E. coli</i> BL21(DE3) cells overexpressing p53C-GFP fusions	74
Figure 3.4. Monitoring the folding and misfolding of SOD1 variants using the MisP-GFP assay	76
Figure 3.5. Solubility analysis of SOD1 variants overexpressed in <i>E. coli</i> Origami2(DE3)...	77
Figure 3.6. Monitoring the aggregation of polyQ-HTT _{ex1} variants using the MisP-GFP assay	79
Figure 3.7. Effect of different optimization parameters on the bacterial fluorescence of <i>E. coli</i> BL21(DE3) cells producing p53C-GFP fusions	81
Figure 3.8. Effect of varying incubation temperatures and IPTG concentrations on the bacterial fluorescence of <i>E. coli</i> Origami2(DE3) cells overexpressing SOD1-GFP fusions from the pET28 vector.....	82
Figure 3.9. Effect of different fluorescent protein partners for monitoring protein aggregation in <i>E. coli</i> Tuner(DE3) cells	83
Figure 4.1. Schematic of the utilized bacterial platform for discovering inhibitors of protein aggregation.....	86

Figure 4.2. Schematic of the utilized two-plasmid system and the gating strategy for FACS sorting for the identification of rescuers of MisP-GFP misfolding and aggregation.....	88
Figure 4.3. Histograms of BL21(DE3) cells co-expressing p53C(Y220C) and the initial cyclo-NuX ₁ X ₂ X ₃ -X ₅ oligopeptide library (grey) or the enriched library after the second round of sorting (green).....	88
Figure 4.4. First attempt of FACS sorting for the identification of p53C(Y220C) folding rescuers	90
Figure 4.5. Schematic of the utilized two-plasmid system and the gating strategy utilized in our second attempt of FACS sorting for the identification of T-p53C(Y220C) folding rescuers .	91
Figure 4.6. Histograms of BL21(DE3) cells co-expressing T-p53C(Y220C) and the initial cyclo-NuX ₁ X ₂ X ₃ -X ₅ oligopeptide library or the enriched libraries after three, four or five rounds of sorting	93
Figure 4.7. Second attempt of FACS sorting for the identification of p53C(Y220C) folding rescuers	93
Figure 4.8. Relative fluorescence of <i>E. coli</i> Tuner(DE3) cells overexpressing T-p53C-BFP and T-p53(Y220C)-BFP from the pCDFT-p53C-BFP and pCDFT-p53(Y220C)-BFP vectors	94
Figure 4.9. Schematic of the utilized three-plasmid system and the gating strategy for FACS sorting for the identification of p53C(Y220C) folding rescuers.....	95
Figure 4.10. FACS sorting of <i>E. coli</i> Tuner(DE3) cells overexpressing T-p53(Y220C)-BFP, A β 42-GFP and the cyclo-NuX ₁ X ₂ X ₃ -X ₆ libraries	96
Figure 4.11. Identification of p53C(Y220C) folding rescuers via FACS sorting.....	98
Figure 4.12. Solubility analysis of <i>E. coli</i> Tuner(DE3) cells overexpressing T-p53C(Y220C)-GFP and the five individual cyclic peptide sequences.....	99
Figure 4.13. FACS sorting of <i>E. coli</i> Tuner(DE3) cells overexpressing A β 42-GFP and the combined cyclic heptapeptide library	101

Figure 4.14. Identification of A β 42 aggregation inhibitors via FACS sorting	103
Figure 4.15. Solubility analysis of <i>E. coli</i> Tuner(DE3) cells overexpressing A β 42-GFP and the ten individual cyclic peptide sequences	104
Figure 4.16. Emission spectra of <i>E. coli</i> Tuner(DE3) cells overexpressing A β 42 along with four of the selected cyclic heptapeptide sequences tested in (B) and stained with ThS.....	105
Figure 4.17. FACS sorting for the identification of SOD1(A4V) folding rescuers.....	108
Figure 4.18. FACS sorting for the identification of HTT _{ex1-97Q} aggregation inhibitors.....	110
Figure 5.1. Cyclic peptide production and purification via on-column protein splicing using chitin affinity chromatography	114
Figure 5.2. Cyclic peptide production and purification via on-column protein splicing using IMAC	115
Figure 5.3. Full scan spectrum (ESI-MS) of the sample containing the free cyclic peptide after on-column intein-mediated circular ligation.....	116
Figure 5.4. Viability of melanoma WM164 (p53(Y220C)) cells after treatment for 48 h with 1% or 10% v/v of the selected cyclic peptides, as determined by the MTT assay	117
Figure 5.5. Full scan spectrum (ESI-MS) of the synthesized p53C4-16 by solid-phase chemical synthesis.....	118
Figure 5.6. Viability of melanoma WM164 (p53(Y220C)) cells after treatment for 48 h with different concentration of p53C4-16, as determined by the MTT assay in (A) the absence or (B) the presence of 330 μ M (100 μ g/ml) cisplatin.....	119
Figure 5.7. Viability of human mesenchymal cells derived from Wharton's Jelly after treatment for 3 h with different concentration of p53C4-16, as determined by the MTT assay.....	120
Figure 5.8. SDS-PAGE of samples collected during protein purification of T-p53C(Y220C)-His ₆ by Ni-IMAC after overexpression in <i>E. coli</i> Tuner(DE3).	121

Figure 5.9. Size-exclusion chromatography (SEC) of T-p53C-His ₆ (top) and T-p53C(Y220C)-His ₆ (bottom) after the IMAC purification shown in Figure 5.8	122
Figure 5.10. Characteristic diagram of fluorescence intensity vs. temperature of a representative globular protein in the presence of SYPRO orange	123
Figure 5.11. Thermal denaturation analysis of T-p53C and T-p53C(Y220C) using DSF	124
Figure 5.12. Effect of p53C4-16 to the thermal stability of T-p53C(Y220C) using DSF.....	125
Figure 5.13. Effect of p53C5-18 to the thermal stability of T-p53C(Y220C) using DSF.....	127
Figure 5.14. Effect of p53C7-10 to the thermal stability of T-p53C(Y220C) using DSF.....	127
Figure 5.15. Kinetic studies of T-p53C and T-p53C(Y220C) in the absence and presence of the selected peptides	130
Figure 6.1. Low frequency cyclic heptapeptides from the sorted pool affect the aggregation of A β 42-GFP	134
Figure 6.2. Sequence analysis of the selected heptapeptides	135
Figure 6.3. Schematic of the linear representations (circular permutants) of a cyclic heptapeptide.	137
Figure 6.4. Network visualization of all clusters identified using the Girvan-Newman algorithm of the Gephi software.....	137
Figure 6.5. Visualization of the main clusters formed by the selected cyclic heptapeptides according to their sequence similarities, as in Figure 6.4	139
Figure 6.6. Chemical structures of the selected cyclic heptapeptides A β C7-1 and A β C7-14.	140
Figure 6.7. Kinetic profiles of the aggregation of 2 μ M A β 42 in the absence and presence of (A) A β C7-1 or (B) A β C7-14, at different molar ratios.....	141
Figure 6.8. Normalised $t_{1/2}$, t_{lag} and t_{growth} values of the kinetic aggregation reactions from (Figure 6.7) in the presence of (A) A β C7-1 or (B) A β C7-14	143

Figure 6.9. Representative TEM images of 2 μ M A β 42 fibrils in the absence (left) and presence (right) of either 10 μ M A β C7-1 (top) or 4 μ M A β C7-14 (bottom).....	145
Figure 6.10. Time course of the aggregation of 2 μ M A β 42 in the presence and absence of 0.5 μ M A β C7-14 using a dot blot assay and the A β 42 fibril-specific antibody OC.....	145
Figure 6.11. A β C7-1 and A β C7-14 increase the motility of A β worms.....	147
Figure 6.12. A β C7-1 and A β C7-14 increase the speed of movement of A β worms.....	148
Figure 6.13. Total fitness of individual A β and wild-type worms in the absence and presence of A β C7-1 and A β C7-14 at 7 d of adulthood	149
Figure 6.14. Imaging of A β and wild-type worms in the absence and presence of A β C7-1 and A β C7-14 at 7 d of adulthood	150
Figure 6.15. Paralysis curves of <i>C. elegans</i> CL4176 expressing human A β 42 and treated with 10 μ M of A β C7-1 and A β C7-14.....	151
Figure 6.16. Relative fluorescence of <i>E. coli</i> Tuner(DE3) cells overexpressing A β 42-GFP and A β C7-1 or the indicated variants thereof as measured by flow cytometry.....	153
Figure 6.17. Relative fluorescence of <i>E. coli</i> Tuner(DE3) cells overexpressing A β 42-GFP and A β C7-14 or the indicated variants thereof as measured by flow cytometry.....	153
Figure 6.18. Solubility analysis of <i>E. coli</i> Tuner(DE3) cells overexpressing A β 42-GFP and the two selected cyclic peptide sequences along with the indicated variants	154
Figure 6.19. Heat map representation of the amino acid distribution at each position of the peptide sequences corresponding to Cluster I (top) and Cluster II (bottom), as demonstrated by the deep sequencing analysis results.....	156
Figure A. 1. Graphical representation of the fluorescence data acquired by DSF for T-p53C and T-p53C(Y220C) as in Figure 5.11, after fitting to the Boltzmann equation using OriginPro.	231

Figure A. 2. Graphical representation of the fluorescence data acquired by DSF for T-p53C(Y220C) in the presence and absence of p53C4-16 as in Figure 5.12, after fitting to the Boltzmann equation using OriginPro.....232

Figure A. 3. Graphical representation of the fluorescence data acquired by DSF for T-p53C(Y220C) in the presence and absence of p53C5-18 and p53C7-10 as in Figure 5.13 and Figure 5.14, after fitting to the Boltzmann equation using OriginPro.233

List of Tables

Table 1.1. List of proteins forming functional amyloids in microbes and their physiologic role	18
Table 1.2. List of polypeptides forming functional amyloids in humans and their physiologic role	19
Table 1.3. Examples of protein misfolding diseases and their associated MisPs	20
Table 2.1. Theoretical diversity of the constructed combinatorial cyclo-NuX ₁ X ₂ X ₃ -X ₆ oligopeptide library.....	50
Table 2.2. Quality assessment of the constructed library via molecular biology techniques ..	55
Table 2.3. Sequencing results of the peptide-encoding regions of 23 randomly selected clones from the constructed pSICLOPPS-NuX ₁ X ₂ X ₃ , pSICLOPPS-NuX ₁ X ₂ X ₃ X ₄ , and pSICLOPPS-NuX ₁ X ₂ X ₃ X ₄ X ₅ vector sub-libraries.....	56
Table 2.4. Deep sequencing analysis of the peptide-encoding regions of ~3.7 million clones from the constructed pSICLOPPS-NuX ₁ X ₂ X ₃ -X ₆ library	58
Table 2.5. Top 10 most frequent sequences of the cyclo-CysX ₁ X ₂ X ₃ X ₄ X ₅ and cyclo-SerX ₁ X ₂ X ₃ X ₄ X ₅ sub-libraries as demonstrated by the deep sequencing analysis	60
Table 3.1. Statistics for fALS patients with the herein studied SOD1 mutations.....	75
Table 4.1. Cyclic oligopeptide sequences encoded by the selected bacterial clones exhibiting enhanced T-p53C(Y220C)-GFP fluorescence	99
Table 4.2. Cyclic heptapeptide sequences encoded by the selected bacterial clones exhibiting enhanced Aβ ₄₂ -GFP fluorescence.....	106
Table 5.1. Calculated T _m values of T-p53C and T-p53C(Y220C) after fitting to the Boltzmann equation.....	124
Table 5.2. Calculated T _m values of T-p53C(Y220C) in the absence or presence of increasing concentration of p53C4-16, after fitting to the Boltzmann equation.....	126

Table 5.3. Calculated T_m values of T-p53C(Y220C) in the absence or presence of increasing concentration of p53C5-18 or p53C7-10, after fitting to the Boltzmann equation.....	128
Table 6.1. Enrichment (blue) and depletion (red) of the 20 amino acids in each position of the heptapeptide sequences. Values represent the log ₂ fold change of the amino acid distribution of peptides from the selected pool compared to the initial library.	136
Table 6.2. Distribution of the heptapeptide sequences in the different clusters identified....	138
Table A. 1. Plasmids that encode SICLOPPS libraries and individual members.....	249
Table A. 2. Plasmids that encode MisPs.....	252
Table A. 3. Primers used for the construction of pSICLOPPS libraries and individual members	254
Table A. 4. Primers used for the construction of plasmids encoding MisPs	257

List of Abbreviations

α -syn	α -synuclein
AD	Alzheimer's disease
AL	Light chain amyloidosis
ALS	Amyotrophic later sclerosis
APP	Amyloid precursor protein
A β	Amyloid- β peptide
BACE-1	β -secretase
BFP	Blue fluorescent protein
bRo5	Beyond the rule of 5 space
CBD	Chitin-binding domain
CFTR	Cystic fibrosis transmembrane conductance regulator
CFTR(Δ F508)	CFTR variant with a deletion of the phenylalanine residue at position 508
DSF	Differential scanning fluorimetry
ER	Endoplasmic reticulum
ERAD	endoplasmic reticulum-associated degradation
FACS	Fluorescence activated cell sorting
fALS	Familial ALS
FSC	Forward-scatter
GFP	Green fluorescent protein
HD	Huntington's disease
Hsp	Heat-shock protein
HTT	Huntingtin
HTT _{ex1}	First exon of the huntingtin gene
IAPP	Islet amyloid polypeptide
I _C	C-terminal domain of the split Ssp DnaE intein
IMAC	Immobilized metal affinity chromatography
I _N	N-terminal domain of the split Ssp DnaE intein
IPTG	Isopropyl- β -D-thiogalactoside
LB	Luria-Bertani
LC	Immunoglobulin light chains

MisPs	Misfolded proteins
NGM	Nematode growth medium
Ni-NTA	Nickel resin
NMR	Nuclear magnetic resonance
NPMDs	Neurodegenerative PMDs
OD600	Optical density at 600 nm
p53C	DNA-binding core domain of p53, amino acids 94-312
p53C(Y220C)	p53C containing the substitution of tyrosine at position 220 by cysteine
PBS	Phosphate-buffered saline
PD	Parkinson's disease
PMDs	Protein misfolding diseases
PN	Proteostasis network
polyQ-HTT	Polyglutamine expansions of huntingtin
PS1 and PS2	Presenilin 1 and 2
RFP	Red fluorescent protein
SDS-PAGE	Sodium dodecyl sulfate polyacrylamide gel electrophoresis
SEC	Size-exclusion chromatography
SICLOPPS	Split-intein circular ligation of peptides and proteins
SOD1	Cu/Zn superoxide dismutase 1
SOD1(A4V)	SOD1 containing the substitution of alanine at position 4 by valine
SSC	Side-scatter
Ssp DnaE	<i>Synechocystis sp PCC6803 DnaE intein</i>
TBST	Tris-buffered saline containing 0.1% Tween-20
TEM	Transmission electron microscopy
Tet	Tetracycline
ThS	Thioflavin S
ThT	Thioflavin T
T _m	Melting temperature
T-p53C	Highly stabilized variant of p53C containing the four point mutations M133L, V203A, N239Y and N268D
TTR	Transthyretin

TTR-FAP	Transthyretin familial amyloid polyneuropathy
UPR	Unfolded protein response
UPS	Ubiquitin-proteasome system
wt	Wild type

Three and one letter abbreviations for the 20 naturally occurring amino acids:

Alanine	Ala	A
Arginine	Arg	R
Asparagine	Asn	N
Aspartic acid	Asp	D
Cysteine	Cys	C
Glutamic acid	Glu	E
Glutamine	Gln	Q
Glycine	Gly	G
Histidine	His	H
Isoleucine	Ile	I
Leucine	Leu	L
Lysine	Lys	K
Methionine	Met	M
Phenylalanine	Phe	F
Proline	Pro	P
Serine	Ser	S
Threonine	Thr	T
Tryptophan	Trp	W
Tyrosine	Tyr	Y
Valine	Val	V

Chapter 1 - Protein misfolding and its association with human diseases

1.1. Protein folding

Proteins are the most essential molecules in living systems, playing crucial roles in every biological process [1]. In order to function, most proteins need to fold into a specific three-dimensional structure characterized by good stability under physiological conditions. Protein folding is typically a highly efficient but complex process, whose elucidation has been of fundamental interest for modern science. Pioneering work from Anfinsen and co-workers in the 1960s, suggested that proteins can fold reversibly into their lowest free-energy conformation, known as the native state, and all the necessary information for achieving this is encoded in the protein's amino acid sequence [2]. Shortly afterwards, Levinthal indicated that due to the vast number of possible configurations, the time needed for a protein to randomly find its native state is exorbitant (e.g. 10^{52} years for a 100-residue protein), whereas proteins usually fold in a timeframe of milliseconds to seconds [3]. He later proposed that folding occurs through a predefined and sequential pathway, thus confining the protein's conformational space and folding time (sequential model) [4]. Since then, many folding mechanism models have been proposed, such as the nucleation-growth mechanism [5-7], the diffusion-collision model [8,9], the framework model [10], the jigsaw-puzzle model [11], the hydrophobic collapse model [12-14] and the nucleation-condensation model [15,16]. However, none of these models were able to resolve Levinthal's paradox.

In the late 20th century, Sali, Shakhnovich and Karplus proposed a new model known as the "new view" [17], providing a definitive solution to the Levinthal paradox. This model,

replaces the idea of a pathway containing a predetermined sequence of steps, with a funnel describing parallel events. It proposes that proteins fold in a diffusion-like process [18-21], in which an ensemble of intermediate conformations navigate the free-energy landscape, all finding their way to the same thermodynamically favored native state (Figure 1.1). The energy landscape of each protein is encoded by its amino acid sequence and has naturally evolved to afford rapid and efficient folding [22].

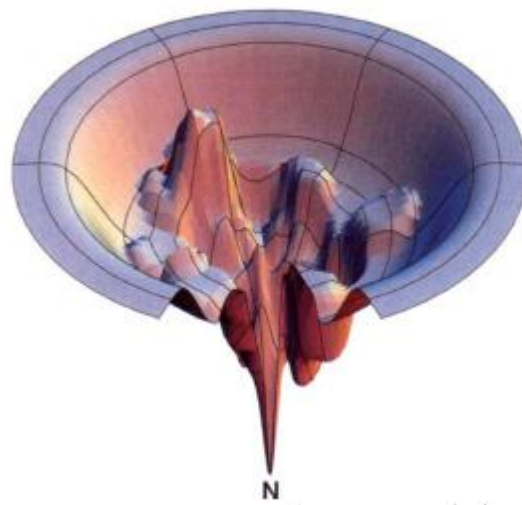


Figure 1.1. Schematic representation of a complex energy landscape. The vertical direction of the funnel represents the internal free energy and the horizontal plane represents the conformation coordinates. Each conformation is represented by a point on the energy surface. N is the native conformation. Hills represent high energy conformations (e.g. unfavourable ϕ ψ angles); valleys correspond to conformations more favourable than others nearby [23].

However, more recent studies have shown that several proteins self-associate, forming insoluble and usually inactive aggregated states that are more thermodynamically favored compared to their native state. In these cases, the latter is merely a local minimum in the energy landscape, kinetically trapped from the aggregated global minimum [24,25]. Therefore, under physiological conditions, proteins adopt their native form only as a metastable state, due to the high kinetic barriers associated with aggregation.

The aggregation process can be incorporated in a protein's folding funnel diagram that shows all its possible conformational states (Figure 1.2). The shape of the aggregation funnel depends on the propensity of a protein to aggregate as well as the structure of the formed aggregates. For example, proteins such as α -synuclein (α -syn), that do not form stable native states but create highly structured aggregates have a very deep and sharp aggregation funnel [26].

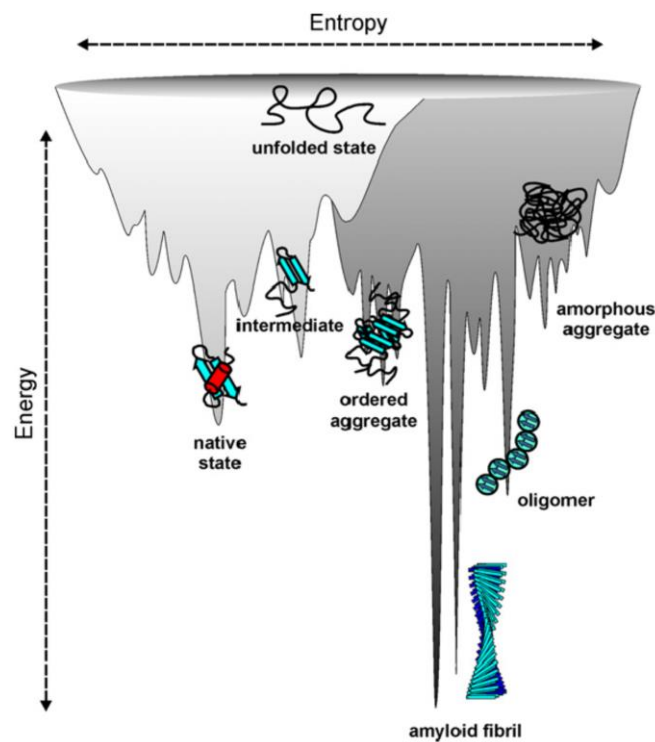


Figure 1.2. Schematic representation of a combined energy landscape for both protein folding and aggregation [27].

1.2. *In vivo* protein folding and homeostasis

While the fundamental principles of protein folding are universal, *in vivo* protein folding is governed by many additional factors. Firstly, the cellular environment is exceptionally crowded, consisting of ~400 g/l of macromolecules, thus increasing their encounter rate and favoring aggregation [28]. Furthermore, a variety of proteins are only marginally stable at physiological conditions, being at constant risk of misfolding, while others, even in their native-state, can self-assemble, eventually forming native-like toxic aggregates [29,30]. Moreover, approximately 15-30% of mammalian proteins are comprised of intrinsically disordered proteins - i.e. proteins that partially or entirely lack a unique 3D structure - that have been associated with various protein aggregation diseases [31]. Finally, the high turnover of proteins in the cell, which requires full or partial unfolding, further increases the opportunities for inappropriate intra-molecular interactions leading to misfolding and eventually aggregation [32]. The various conformational states that a polypeptide chain can assume after its biosynthesis and the in-between transformations of the different states are summarized in Figure 1.3.

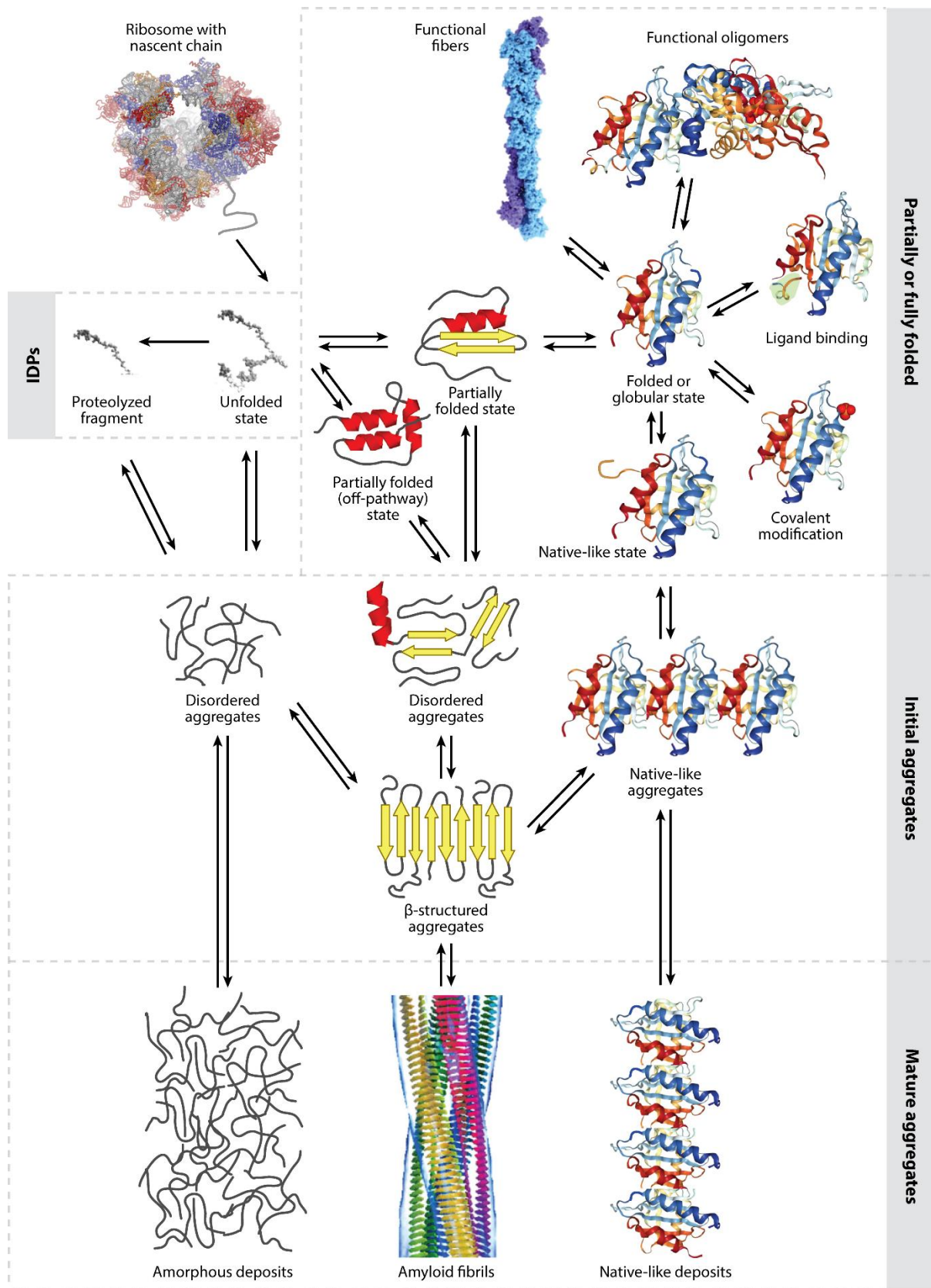


Figure 1.3. Schematic of the possible conformational states that can be adopted by a nascent protein and the in-between state transitions. Protein aggregation can result in amorphous, amyloid or native-like deposits, all of which have been associated with disease states [30].

In order to assist proper *in vivo* protein folding and to counteract the intrinsic propensity of proteins to form thermodynamically stable aggregates, biological systems have evolved a quality control system that assists protein folding and prevents protein aggregation. This system is referred to as the “proteostasis network” and consists of molecular chaperones and co-chaperones as well as protein degradation processes, such as the ubiquitin-proteasome system (UPS), autophagy as well as endoplasmic reticulum-associated degradation (ERAD) (Figure 1.4 and Figure 1.5) [22,33]. The first discovery of a chaperone system resulted from studies of the heat-shock response [34,35]. It was found that cultured cells, as well as whole organisms, respond to conditions of increased stress, such as elevated temperature, by increasing the concentration of a number of proteins, termed heat-shock proteins (Hsps), which are categorized by their size in kDa (e.g. Hsp40, Hsp60, etc.).

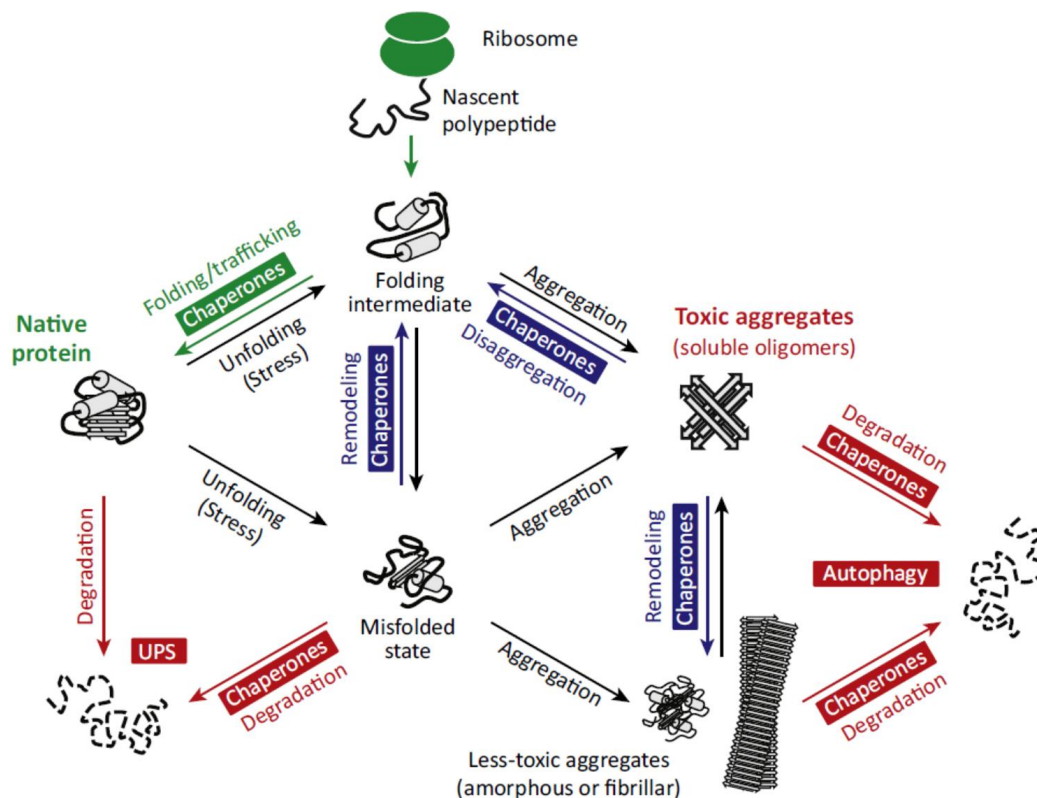


Figure 1.4. The proteostasis network (PN). Protein homeostasis is maintained by assuring correct protein folding and elimination of toxic aggregates. This is achieved by three mechanisms of the PN involving chaperones: proper folding and trafficking (green), conformational maintenance (blue) and degradation by the ubiquitin-proteasome system or autophagy (red) [33].

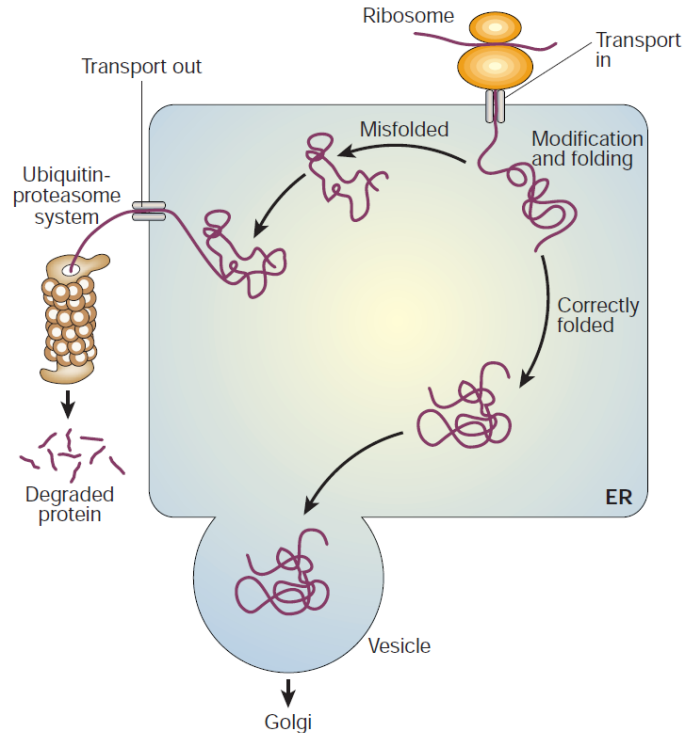


Figure 1.5. Protein folding in the ER. Nascent polypeptides enter the ER where they acquire their native state assisted by molecular chaperones and folding catalysts. Correctly folded proteins are released from the ER, while misfolded proteins are targeted for ubiquitination and proteasomal degradation [22].

To date, more than 330 human molecular chaperones have been identified and have been categorized into nine chaperone families: Hsp40, Hsp60, Hsp70, Hsp90, Hsp104, small Hsps (sHsps), prefoldin, tetratricopeptide repeat (TRP)-domain-containing and chaperones involved in the “unfolded protein response” of the ER and mitochondria (UPR^{ER} and UPR^{mt} respectively) [36]. These act at different stages of the folding process [37], i.e. they may interact with the nascent polypeptide chain as it emerges from the ribosome in order to promote its folding, they may assist the formation of protein complexes [38], they may inhibit protein aggregation of misfolded proteins and finally they may convert toxic oligomers into higher order assemblies in order to alleviate their toxicity (Figure 1.4) [39]. Furthermore, chaperones also participate in protein degradation via the UPS or autophagy, by either aiding the normal catabolism of native proteins after they have fulfilled their functionality or by disassembling

aggregated species (Figure 1.4) [33]. Finally, chaperones contribute to the trafficking of proteins to their correct location in the cell, and also prevent the secretion of misfolded proteins by the Golgi apparatus [22]. Interestingly, recent studies have also revealed that molecular chaperones, such as clusterin, are secreted in the extracellular space, where they interact with aggregation-prone protein species and inhibit their propagation, thus offering an additional line of defense against harmful protein aggregation [40]. Importantly, in all cases, molecular chaperones act by guiding proteins to their expected fate rather than by accelerating the individual steps of the folding process [22]. The latter is performed by folding catalysts, such as peptidyl-prolyl isomerases, that accelerate the cis/trans isomerization of peptide bonds N-terminal to proline residues [41], and oxidoreductases and disulphide isomerases that catalyse the formation and reorganization of disulphide bonds, respectively [42].

The importance of the proteostasis network regarding cellular fitness has been well-documented as its decline and failure have been repeatedly associated with human diseases [33]. Indeed, mutations in parkin, a regulator of protein degradation in the UPS, have been directly associated with Parkinson's disease [43], while defects in the lysosomal degradation of various biomolecules, such as proteins, lipids and glycans, have been associated with over 70 different human disorders collectively termed as lysosomal storage diseases [44]. It is believed that a number of factors, such as aging, genetic mutations, obesity and oxidative stress promote protein aggregation and result in the collapse of the proteostasis network [30]. Aging in particular is widely considered to be the leading risk factor for protein aggregation as an increasing body of evidence links aging with proteostasis network failure. Specifically, in aged human brain tissues approximately 30% of chaperones are downregulated [36], while others are sequestered by the accumulated protein aggregates, thus decreasing their cellular levels [45]. Furthermore, the activity of both the UPS and the autophagy machinery has been found to be severely affected by aging, as a result of the decrease of the levels of active proteasome

complexes and autophagy-related proteins respectively, amongst other factors [46,47]. It is therefore not surprising that a number of late-onset diseases, such as Alzheimer's disease (AD) and Parkinson's disease (PD) have been associated with protein aggregation and the collapse of the proteostasis network.

1.3. Protein Misfolding Diseases

The phenomenon of protein misfolding and aggregation is a defining feature of numerous human diseases that have been collectively termed protein misfolding diseases (PMDs), proteinopathies or conformational diseases [30,44,48-50]. These include conditions of very high socioeconomic impact and unmet medical need [30]. Importantly, although for some PMDs symptomatic therapies do exist, there are no approved disease modifying therapies that can prevent, delay or reverse the progression of the disease for the vast majority of these conditions [51-53]. Notable exceptions are tafamidis and migalastat, two small-molecule pharmacological chaperones that rescue the misfolding of transthyretin, associated with transthyretin familial amyloid polyneuropathy (TTR-FAP), and α -galactosidase, associated with Fabry disease, respectively [54,55], as well as the pharmacological chaperones ivacaftor, lumacaftor and tezacaftor, combinations of which have been approved for treatment of cystic fibrosis [56-60].

PMDs include diseases with diverse pathologies, ranging from neurodegeneration disorders to cancer. These can be (i) familial, such as Huntington's disease (HD), (ii) sporadic, such as most cases of PD, (iii) iatrogenic, such as dialysis-related amyloidosis and (iv) infectious, such as Creutzfeldt-Jacob disease [30]. All PMDs however, irrespective of their pathologies, have been associated with the aberrant folding and/or aggregation of one or more misfolded proteins (MisPs). The nature of the associated MisP varies greatly [30]: MisPs can

be globular, such as mutants of the tumour suppressor protein p53 associated with certain forms of cancer, or intrinsically disordered, such as the amyloid- β peptide ($A\beta$) associated with AD. Furthermore, MisPs can be localized in (i) the nucleus, such as p53, (ii) the cytosol, such as α -syn associated with PD, (iii) it can be membrane-embedded, such as the cystic fibrosis transmembrane conductance regulator (CFTR) associated with cystic fibrosis or even (iv) extracellularly secreted, such as $A\beta$. Understanding more on each MisP contribution's to its respective disease is fundamental for the development of successful therapeutics.

1.3.1. Mechanisms of PMDs

As mentioned earlier, the decline of the proteostasis network can result in the accumulation of aggregated proteins and the onset of human diseases, but this is by no means the only mechanism by which PMDs emerge. Indeed, PMDs can be classified into five categories according to their mechanisms of pathogenesis [61]:

1.3.1.1. Enhancement of the proteostasis network leading to improper protein degradation

Although the proteostasis network is essential for cellular fitness, a hyperactivity of the degradation systems can sometimes result in human disease. This occurs through the degradation of mutated but still fairly functional proteins, whose total deficiency induces the manifestation of a more severe disease. A typical example is cystic fibrosis, a serious disease caused by mutations in CFTR, the most common mutation being the deletion of the phenylalanine residue at position 508 ($\Delta F508$) [62]. In this case, CFTR($\Delta F508$) is misfolded and degraded by the proteasome before reaching the plasma membrane, where it normally functions as a chloride channel. Studies performed by Balch and co-workers [63] as well as Cyr and co-workers [64] have demonstrated that inhibition of proteasomal degradation leads

to a somewhat stable and partially functional CFTR(Δ F508). This indicates that although mutated, the protein is still able to reach its proper location, i.e. the plasma membrane, and therefore, a therapeutic strategy targeting the activity of the proteostasis network, would result in increased cellular levels of CFTR(Δ F508) that could be valuable for individuals carrying this mutation [65].

Lysosomal storage disorders, such as Fabry disease and Gaucher disease, are also associated with enhanced protein degradation [44]. Specifically, in Gaucher disease, mutations in β -glucocerebrosidase affect the enzyme's ability to degrade its substrate, glucosylceramide, which results in the latter's intracellular accumulation and induces of a variety of symptoms, such as bone damage, enlarged liver and spleen and hematologic abnormalities in different degrees [66]. Studies by Ron and Horowitz [67] have revealed that this variability is strongly correlated to the degree of ER retention and proteasomal degradation, as the mutated protein can still remain partially functional. Indeed, even individuals with the same genotype can exhibit disease heterogeneity due to different extents of ER retention and degradation.

1.3.1.2. Mutations leading to dominant-negative effects

An alternative mechanism by which PMDs can manifest is through the antagonism of a mutated protein with its wild-type form leading to a dominant-negative phenotype in heterozygote individuals. A classic example of this category is the tumour suppressor protein p53, a homotetrameric transcription factor that protects the genome from carcinogenesis [68]. Mutant forms of p53 have been associated with approximately 50% of human cancer cases, as a loss of the protein's normal function leads to uncontrolled cell proliferation [68]. Moreover, mutant p53 is still able to associate with its wild-type form leading to the formation of heterotetramers that similarly do not function properly [69]. In addition, although the cellular levels of p53 are strictly regulated by the proteostasis network, mutant p53 is sometimes

degraded to a smaller extent leading to a prevalence of the mutant form over the wild-type and, thus, decreasing the probability of formation of a functional homotetrameric protein [70]. Furthermore, mutant p53 can also co-aggregate with its wild type form as well as with its paralogs p63 and p73, leading to intracellular accumulation of aggregated species, that are targeted by the proteostasis network [71]. This results in a deficiency of transcription factors that control cell growth and apoptosis and therefore the progression of the disease [71].

Another example of a dominant-negative PMD is epidermolysis bullosa simplex, an inherited connective tissue disorder caused by mutant forms of the keratin proteins KRT5 and KRT14 [72]. Normally, multiple keratin molecules form filaments that provide structure to the skin's epidermis. On the other hand, mutated keratin is misfolded and creates intracellular aggregates [73]. In heterozygote individuals these keratin filaments are constructed by both wild-type and mutant forms; this impairs their overall function and causes cells to rupture when subjected to mechanical stress [74].

1.3.1.3. Aberrant localization of aggregation-prone proteins

After proper folding, proteins are directed to their appropriate location in order to fulfil their normal function. Mutations in a number of proteins associated with PMDs have been found to destabilize their structure and inhibit their intracellular trafficking. Improper localization of misfolded proteins can also result in dual toxicity as the protein will be absent from its proper location and therefore will be unable to function normally but furthermore, will accumulate at an improper location resulting in protein aggregation and possibly exhibiting a dominant-negative phenotype.

One example of this category is the mutation of α 1-antitrypsin, a serine protease inhibitor (serpin). Normally, α 1-antitrypsin is produced in hepatocytes and secreted into the bloodstream, where it inhibits the elastolytic destruction of lung tissue. Mutations in this

protein can therefore cause liver damage, due to increased protein aggregation, as well as emphysema due to uncontrolled elastase activity [75].

A different PMD associated with improper localization is synpolydactyly, a congenital limb malformation disorder caused by mutations in the transcription factor Hoxd13 [76]. Under physiological conditions, Hoxd13 is sequestered in the nucleus where it regulates the transcription of a variety of genes involved in the development of limbs. Mutant forms of Hoxd13 containing poly-alanine expansions are trafficked from the nucleus into the cytoplasm where they form large amorphous aggregates, thus inhibiting their normal function and leading to the manifestation of the disease [77]. Interestingly, mutant Hoxd13 is co-aggregated with its wild-type form resulting in a dominant-negative phenotype [77].

Most PMDs associated with misfolded membrane proteins fall into this category. For example, retinitis pigmentosa, a group of retinal degenerative diseases characterized by progressive vision loss, has been associated with over 120 point mutations in rhodopsin, a G-protein-coupled receptor (GPCR) composed of the apoprotein opsin and the 11-cis-retinal chromophore [78]. Class II mutations in rhodopsin lead to a misfolding of the receptor and an inability to incorporate into the plasma membrane where it normally functions [78]. In addition, these mutants form intracellular aggregates that impair cell fitness, eventually leading to apoptosis [78,79].

Another example is Charcot-Marie-Tooth disease (CMTD), associated with mutations in the gene encoding the peripheral myelin protein 22 (PMP22). PMP22 is normally synthesized in Schwann cells and trafficked to the plasma membrane, where it plays a crucial role in the development and maintenance of myelin [80]. In individuals with CMTD type 1E, point mutations in PMP22 lead to failure of protein trafficking from the ER to the plasma membrane and thus cause protein accumulation in the ER-Golgi intermediate compartment

[80]. Furthermore, mutant PMP22 may co-aggregate with wild-type PMP22, exhibiting a dominant-negative toxic effect [81]. Both mechanisms result in defective myelin stability, with mild to severe symptoms of neuropathy.

Similarly, a rare neurological disorder called episodic ataxia type-2 is associated with mutations in the voltage-gated calcium channel (Cav) [82]. The most common mutations appear in the Cav2.1. subunit affecting protein folding and resulting in ER sequestration and proteasomal degradation. In addition, mutant Cav2.1. subunits bind to their wild-type counterparts and induce their degradation, thus decreasing overall channel activity even in heterozygotes [83].

1.3.1.4. Alterations on the protein's physiological structure leading to a novel toxic function

Another mechanism associated with PMD pathogenesis involves the acquisition of protein conformations that deviate from their proper 3D structure and lead to a novel pathogenic function and cellular toxicity. A representative example is sickle-cell anemia, the first known PMD, which is characterized by decreased red blood cell elasticity that causes extreme pain, cell damage and anemia [84,85]. In this disease, a point mutation in hemoglobin of red blood cells exposes a hydrophobic region onto the proteins surface ultimately resulting in protein aggregation [84,85]. Interestingly, the disease manifests only in homozygote individuals indicating that hemoglobin's aggregation does not operate via a dominant-negative mechanism.

Another example is apolipoprotein E, a lipid transport molecule that exists in three common variants known as E2, E3 and E4. The ApoE4 variant represents a major risk factor for AD and corresponds to a more rigid protein structure stabilized by a salt bridge that is missing in the wild-type ApoE3 [86]. As ApoE4 is not as flexible, its binding to lipid molecules

is affected resulting in disruption of mitochondrial function [87] and reduction of neurite outgrowth [88]. Furthermore, ApoE4 has been associated with enhanced levels of the A β 42 peptide, whose aggregation has been strongly implicated in AD [89].

A noteworthy class of proteins with aberrant folding that leads to disease pathogenesis are oncogenic proteins, which are mutated in a diversity of cancers leading in uncontrolled cell proliferation. A representative example of this class is the non-receptor tyrosine kinase SRC. Mutant SRC is highly unstable compared to its wild-type form and thus activates the Hsp90 chaperone to assist its folding and proper membrane localization [90]. In this manner, Hsp90 ultimately increases the mutant SRC activity and leads to cancer development.

Importantly, a specific PMD can occur through different mechanisms according to the nature of the protein's mutation. For example, in the case of retinitis pigmentosa, Class IV mutations result in correct overall protein folding and localization but constitute the protein unable to associate with its chromophore substrate [78]. This structural alteration eventually leads in decreased protein stability and tissue degeneration.

1.3.1.5. Amyloid accumulation

Probably the most studied mechanism of PMD pathogenesis is associated with the formation of intracellular or extracellular amyloid deposits. These assemblies are characterized by a highly ordered cross- β architecture consisted of parallel or antiparallel β -sheets whose constituent β -strands are perpendicular to the fibril axis [30,91,92]. Notably, this is a common feature of amyloid fibrils formed by proteins with different primary sequences, indicating that the overall structure is held together by interactions of the polypeptide backbone, common in all proteins [91]. On the contrary, side chain variations affect spacing of the β -strands, and kinetics of the amyloid formation mechanism amongst others [93].

The formation of an amyloid fibril is a multistep process involving three macroscopic stages: a lag phase, a growth phase and the final equilibration phase. These are characterized by four microscopic stages, i.e. primary nucleation, secondary nucleation, elongation and fragmentation involving a variety of protein species ranging from oligomers to higher order aggregates (Figure 1.6) [94].

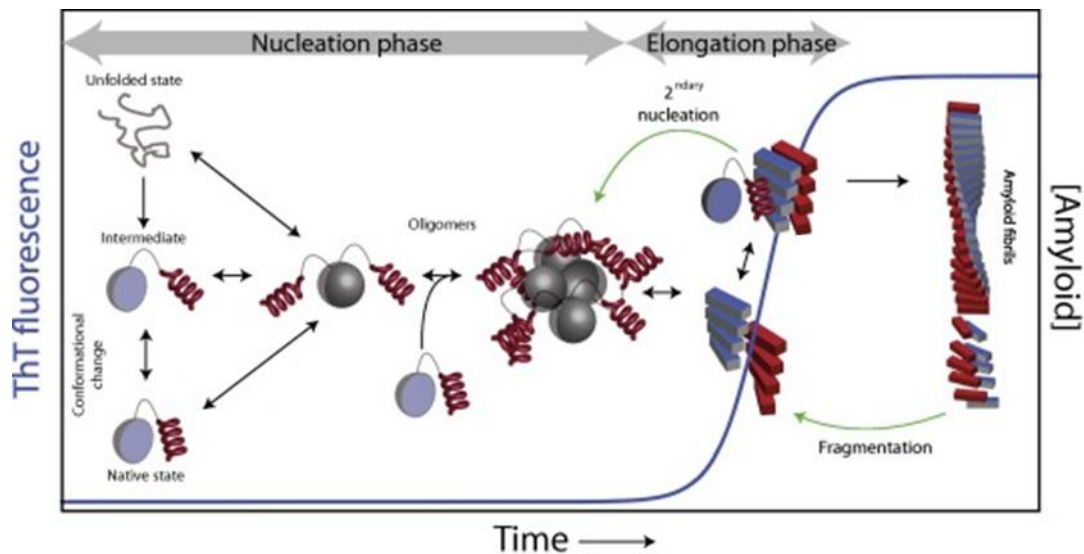


Figure 1.6. Schematic of the amyloid aggregation mechanism. The aggregation process initiates from a monomeric unfolded, partially folded or correctly folded precursor that self-associates forming various oligomeric species, prefibrillar assemblies and ultimately amyloid fibrils. Black and green arrows represent primary and secondary pathways respectively. ThT fluorescence is commonly used to follow the aggregation reaction. Adapted from [95].

The formation of amyloid fibrils has been associated with over 50 PMDs with very different pathologies, ranging from neurodegenerative disorders to cataract [30]. These have been collectively termed amyloidoses and can be either localized or systemic, i.e. the amyloid deposits may appear only in a specific tissue/organ, or in multiple organs respectively.

Neurodegenerative PMDs (NPMDs) have been extensively studied in recent times, due to the dramatic increase of their incidence rate and the financial burden that they impose to the health care system, attributed mainly to the increased aging populations [96]. Each of these

NPMDs, although exhibiting diverse pathologies, they are all associated with the intracellular or extracellular formation of amyloid aggregates of different proteins, which can be intrinsically disordered, aberrantly overexpressed, or mutated [30]. For example, AD has been linked with the aggregation A β , PD with α -syn and HD with polyglutamine expansions of huntingtin (polyQ-HTT). Interestingly, some aggregation-prone proteins have been associated with multiple NPMDs suggesting that their misfolding may not be a cause but an effect of disease-related cell stress, however, resulting in further increased tissue damage and faster progression of the disease [97].

Two representative examples of systemic amyloidosis are light chain amyloidosis (AL), associated with mutations in the immunoglobulin light chains (LC) that increase their aggregation propensity [98] and TTR-FAP, associated with destabilizing mutations in the tetrameric transporter transthyretin [99]. In both cases, amyloid accumulation appears in various organs and results in extensive tissue damage and organ failure.

Interestingly, although in some diseases, such as AL amyloidosis and cataract, toxicity is caused by fibril accumulation [98,100,101], in others, such as TTR-FAP, AD and PD, amyloid deposits are believed to be a protective cellular mechanism that sequesters the truly toxic oligomeric, or in TTR-FAP also monomeric, species mitigating their pathology [39,102]. However, even in these cases, the amyloid fibrils may be involved in the transmission of protein aggregation between cells, contributing in disease progression [103-106].

Notably, the formation of amyloid fibrils is not always associated with a pathogenic state. Indeed in several living systems, including bacteria, yeasts and humans, a number of endogenous proteins are converted into amyloid fibrils that fulfil a specific function. Proteins that form functional amyloids in microbes and humans are summarized in Table 1.1 and Table 1.2 [30,107].

Table 1.1. List of proteins forming functional amyloids in microbes and their physiologic role [107].

Protein name	Species	Physiological role
Curli	<i>Escherichia coli</i> , <i>Salmonella enterica</i>	Biofilm formation and interaction with host
FapC	<i>Pseudomonas</i> species	Biofilm formation
TasA	<i>Bacillus subtilis</i>	Biofilm formation
MTP	<i>Mycobacterium tuberculosis</i>	Pili formation, Cell adhesion
Als	<i>Candida albicans</i>	Cell adhesion
Chaplin	<i>Streptomyces coelicolor</i>	Formation of hydrophobic layer that enables aerial growth and differentiation
Hydrophobin	Fungi	Coat formation
Microcin E492	<i>Klebsiella pneumoniae</i>	Sequestration of toxic species
HpaG	<i>Xanthomonas</i> species	Virulence factor
Sup35p	<i>Saccharomyces cerevisiae</i>	Regulation of translation termination

Table 1.2. List of polypeptides forming functional amyloids in humans and their physiologic role [30].

Peptide or protein name	Physiological role
Intralumenal domain of melanocyte protein PMEL	Production of fibrous striations inside melanosomes, upon which melanin granules form
Various peptide hormones in pituitary secretory granules	Natural storage in pituitary secretory granules
Receptor-interacting serine/threonine-protein kinase 1/3 (RIP1/RIP3)	Mediation of the tumor necrosis factor–induced programmed cell death
Fragments of prostatic acid phosphatase and semenogelins	Unknown for humans, utilized by the HIV virus for infection

1.3.2. PMDs with respect to gain and loss of function

The aforementioned mechanisms involving PMDs can have two distinct impacts on the associated proteins: (i) loss of physiological function or (ii) gain of toxic function. In loss-of-function PMDs, the associated protein is misfolded and unable to fulfil its physiologic role thus resulting in the progression of disease. Examples of this category include cystic fibrosis, phenylketonuria and lysosomal storage disorders, such as Gaucher disease and Fabry disease [108,109]. In gain-of-toxic-function diseases, the affected MisPs form toxic aggregated species that impair cell fitness and result in a pathological phenotype. These diseases include sickle-cell anemia, type 2 diabetes and neurodegenerative diseases, such as AD and amyotrophic later sclerosis (ALS) [30]. Importantly, some proteins may exhibit both a loss-of-function and a gain-of-toxic-function phenotype. As mentioned earlier, mutation in both p53 and α 1-antitrypsin inhibit the proteins' physiological functions but also lead to the formation of intracellular aggregates contributing to the diseases' progression [71,75]. Examples of PMDs with their associated protein and underlying mechanism are presented in the following table.

Table 1.3. Examples of protein misfolding diseases and their associated MisPs [30,110].

Associated diseases	Peptide or protein name
<i>Gain-of-function diseases</i>	
Alzheimer disease	Amyloid- β peptide ($A\beta$), tau protein
Parkinson disease	α -synuclein (α -syn), tau protein
Transmissible spongiform encephalopathy	Prion protein (PrP)
Huntington disease	poly-glutaminated huntingtin exon 1 (polyQ-HTT)
Light-chain amyloidosis	Fragments of immunoglobulin light chains
Heavy-chain amyloidosis (mainly renal)	Fragments of immunoglobulin heavy chains
AA amyloidosis	Full or N-term fragments of serum amyloid A protein
Type 2 diabetes	Islet amyloid polypeptide (IAPP)
Amyotrophic lateral sclerosis	Cu/Zn superoxide dismutase 1 (SOD1)
Sickle cell anemia	Hemoglobin
Familial amyloidotic polyneuropathy	Transthyretin (TTR)
Dialysis-related amyloidosis	β 2-microglobulin (β 2-m)
Fibrinogen amyloidosis (mainly renal)	Fragments of fibrinogen α -chain
Spinocerebellar ataxia 1	Ataxin-1
<i>Loss-of-function diseases</i>	
Cystic fibrosis	Cystic fibrosis transmembrane conductance regulator
Gaucher's disease	Glucocerebrosidase
Fabry disease	α -galactosidase
Hypogonadotropic hypogonadism	Gonadotropin-releasing hormone
Nephrogenic diabetes insipidus	Vasopressin receptor 2
Retinitis pigmentosa	Rhodopsin
<i>Both Loss- and Gain-of-function diseases</i>	
α 1-Antitrypsin deficiency	α 1-Antitrypsin
Cancer	Transcription factor p53

1.3.3. p53-related cancer as a PMD

The tumor suppressor protein p53 is a transcription factor with a leading role in protecting cells from carcinogenesis, thus often called the “guardian of the genome” [68]. Upon cellular stress, such as DNA damage or hypoxia, p53 is activated through a cascade of events, resulting in the expression of target genes that are involved in cell-cycle arrest, DNA repair, and if damage is too extensive, ultimately apoptosis (Figure 1.7) [111].

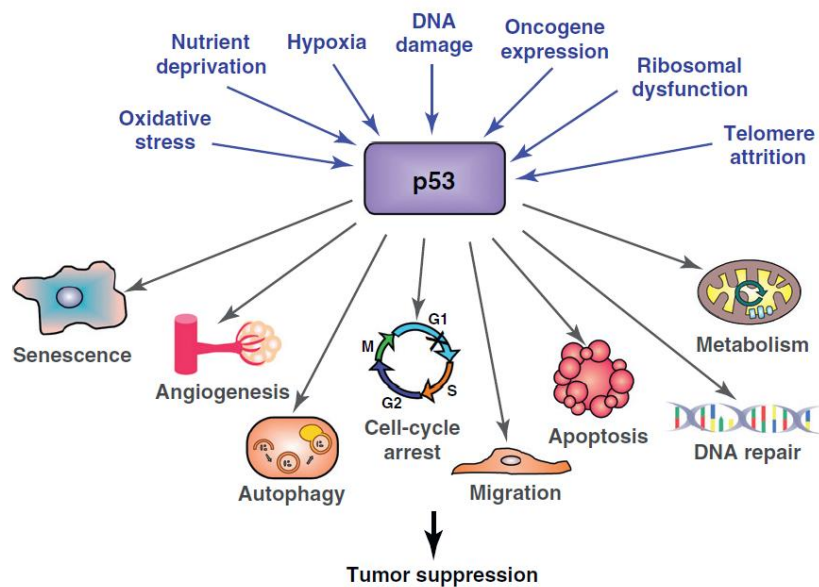


Figure 1.7. Functions of p53 in response to stress signals [111].

p53 is a homotetrameric protein, whose sequence consists of an intrinsically disordered N-terminal transactivation domain, a proline-rich region, a structured DNA-binding core domain, a tetramerization domain and an intrinsically disordered C-terminal regulatory domain (Figure 1.8) [112]. The DNA-binding core domain consists of a β -sandwich scaffold and a DNA-binding surface, which is formed by two loops (L2 and L3) tethered by a zinc atom, and a loop-sheet-helix motif (Figure 1.9) [112]. Besides this structured region, p53's high level of intrinsic disorder provides a dual function: (i) enables the interaction with different proteins, thus facilitating its diverse biological functions and (ii) lowers the protein's stability and half-

life, thus hindering its uncontrollable function that would otherwise lead to inappropriate apoptosis and senescence [113-115]. In order to further avoid hyperactivity of p53, its cellular protein levels are strictly controlled by the ubiquitin ligases MDM2 and MDMX, which act by targeting p53 for proteasomal degradation and also binding to its transactivation domain and inhibiting p53's function [116].

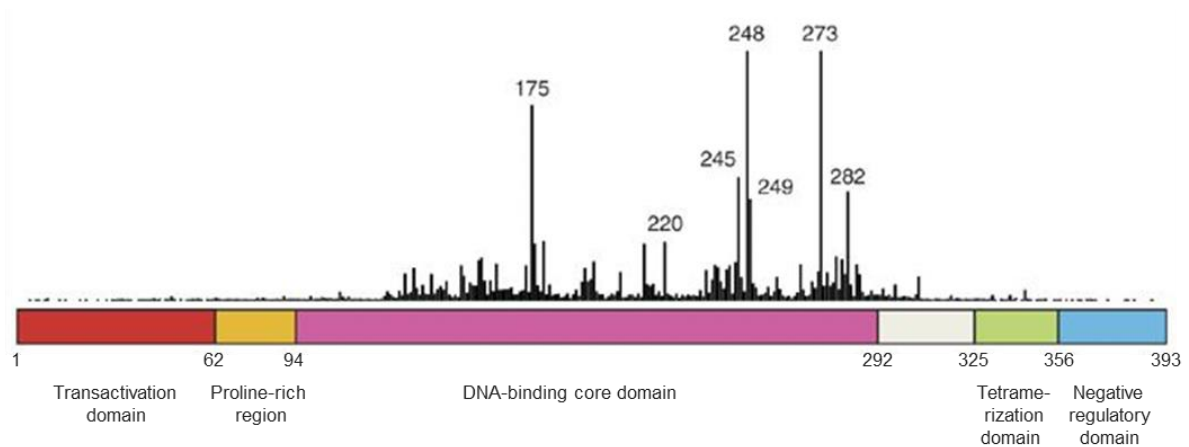


Figure 1.8. Schematic of p53's domain structure indicating the relative frequency of oncogenic mutations for each position [112].

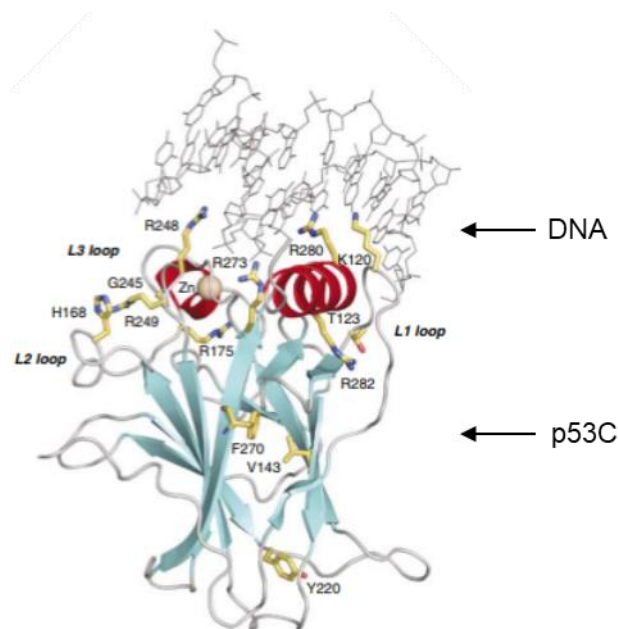


Figure 1.9. Ribbon diagram of the DNA-binding domain of p53 bound to a target DNA sequence. Mutations at positions at or near the DNA-binding site, such as R273 and R280, are called contact mutations while those at positions located at the periphery of p53C, such as V143, Y220 and F270, are called structural mutations [112].

The importance of p53's guardianship is highlighted by the fact that in almost every cancer p53 is inactivated either by mutations in the *TP53* gene or by deregulation of its signaling pathways [117]. Notably, while in most cases *TP53* mutations are acquired, in a rare hereditary syndrome called Li-Fraumeni, germ-line mutations in *TP53* result in the development of various types of cancer starting at a young age [118]. The vast majority of *TP53* mutations appear at the protein's DNA-binding core domain (p53C) as illustrated in Figure 1.8. These can be divided into two categories: contact mutations, which appear at or near the DNA-binding domain of p53 and directly abolish DNA-p53 binding, and structural mutations, which appear at the periphery of the core domain and create local conformational changes to p53 (Figure 1.9) [119]. Due to the marginal stability of p53, these changes can result in the denaturation and, therefore, inactivation of p53 at body temperature, whereas the protein remains active at lower temperatures [119]. Interestingly, second-site mutations that increase p53's thermodynamic stability, have also been found capable of rescuing the misfolding of structural mutants of p53 and restoring their proper function [120].

The most common structural mutant of p53 is the substitution of tyrosine at position 220 with cysteine p53(Y220C), appearing in ~ 100,000 new cancer cases each year [112]. This substitution results in the development of a crevice at the periphery of p53's core domain that destabilizes it thermodynamically by ~ 4 kcal/mol, making more than 80% of the protein unfolded at body temperature [121]. Therefore, protein misfolding results in a loss-of-function phenotype and uncontrolled cell proliferation. At the same time, p53(Y220C) has been found to acquire a gain-of-toxic-function phenotype, as it interact with p63 and p73 isoforms inhibiting their normal function [122] and also to accumulate in tumour cells due to insufficient degradation by MDM2 [123]. All together, these features make p53(Y220C) a highly attractive target for the discovery of rescuers of protein misfolding.

1.3.4. Alzheimer's disease and the amyloid- β peptide

Alzheimer's disease (AD) is a progressive neurodegenerative disorder named after the German physician who first described the disease in 1907 [124], with clinical symptoms that include progressive cognitive impairment, changes in behaviour and, ultimately, death [125]. Currently, it is the most common cause of dementia affecting almost 50 million people worldwide, a number that is predicted to rise to over 130 million by 2050 [126,127]. Importantly, AD remains to date incurable, despite the enormous efforts for the development of prevention and therapeutic strategies, and represents a substantial economic burden on our society.

Age is the strongest risk factor for AD, with the majority of cases manifesting after 65 years of age (late onset AD), however, in ~ 5% of AD cases, the symptoms can occur before the age of 60 (early onset AD) and in some cases as early as 30 years of age [127,128]. Late onset AD is considered of a sporadic nature, resulting by a combination of environmental and genetic factors, including, but not limited to, the aforementioned ApoE4 risk factor [129]. On the contrary, early onset AD is caused by autosomal dominant mutations appearing in three distinct genes, namely the *APP* gene, which encodes the amyloid precursor protein (APP), and the *PSEN1* and *PSEN2* genes, which encode the proteins presenilin 1 and 2 (PS1 and PS2) [129]. Importantly, all three proteins are involved in the production of A β , supporting its involvement in the pathology of AD (Figure 1.10) [130].

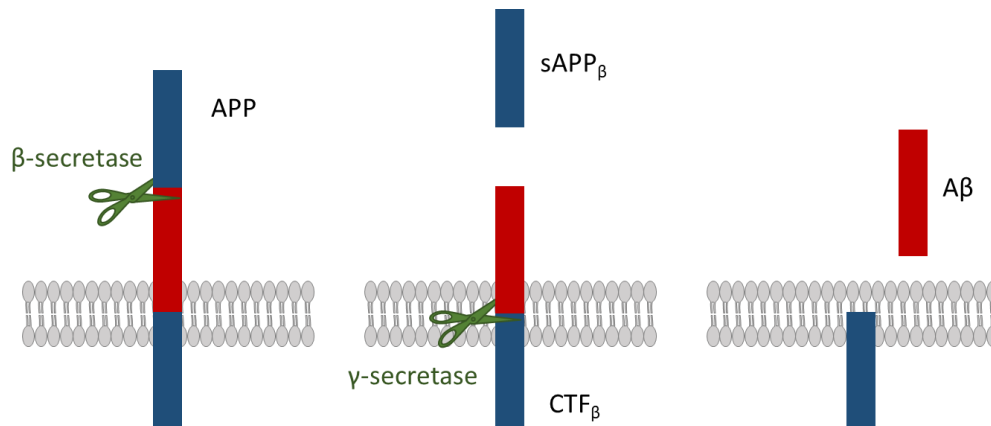


Figure 1.10. Schematic of APP processing to release the disease-associated A β peptide. APP is first cleaved by β -secretase BACE-1 producing the sAPP β fragment and the C terminus CTF β . The latter is then processed by the γ -secretase complex, which includes the PS1 and PS2 subunits. This process results in the release of the disease-associated A β peptide. Adapted from [131].

Pathologically, both late and early onset AD are characterized by the presence of extracellular amyloid deposits, consisting mainly of A β , as well as intracellular neurofibrillary tangles, consisting of hyper-phosphorylated forms of the tau protein (Figure 1.11) [132].

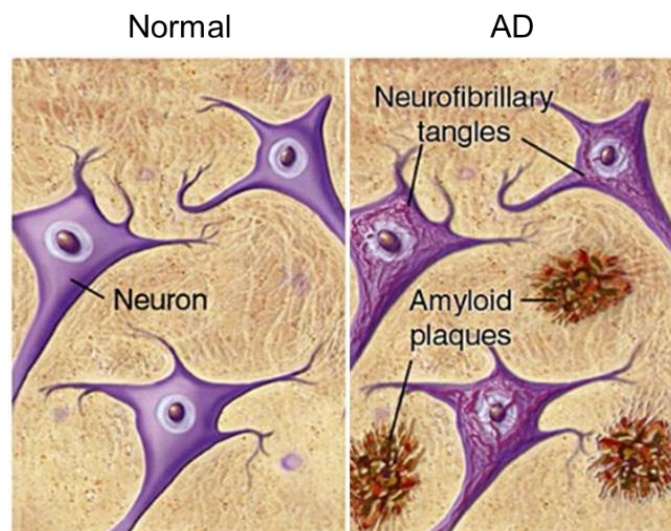


Figure 1.11. Schematic of brain tissue from normal patients and AD patients [133].

Until recently, there was a general consensus that the formation of these A β amyloid plaques was responsible for the disease pathogenesis, giving rise to the so-called amyloid hypothesis [134]. However, as amyloid plaque density was found to correlate poorly with the severity of the disease, research was directed towards other species involved in this aggregation process. This ultimately gave rise to the “toxic A β oligomer” hypothesis, which stipulates that soluble A β oligomeric species, ranging from dimers to larger soluble pre-fibrillar assemblies, are actually responsible for the neurotoxicity in AD [135]. Indeed, it is now widely believed that the amyloid plaques are probably the consequence of a protective cellular mechanism that sequesters the toxic oligomeric A β species, while the occasional adverse effects associated with A β fibrils can be attributed to their fragmentation into smaller species with enhanced toxicity [39,136]. To corroborate this, accumulating amount of evidence suggests that the observed toxicity is inversely correlated to the molecular size of the A β aggregated species (Figure 1.12) [137].

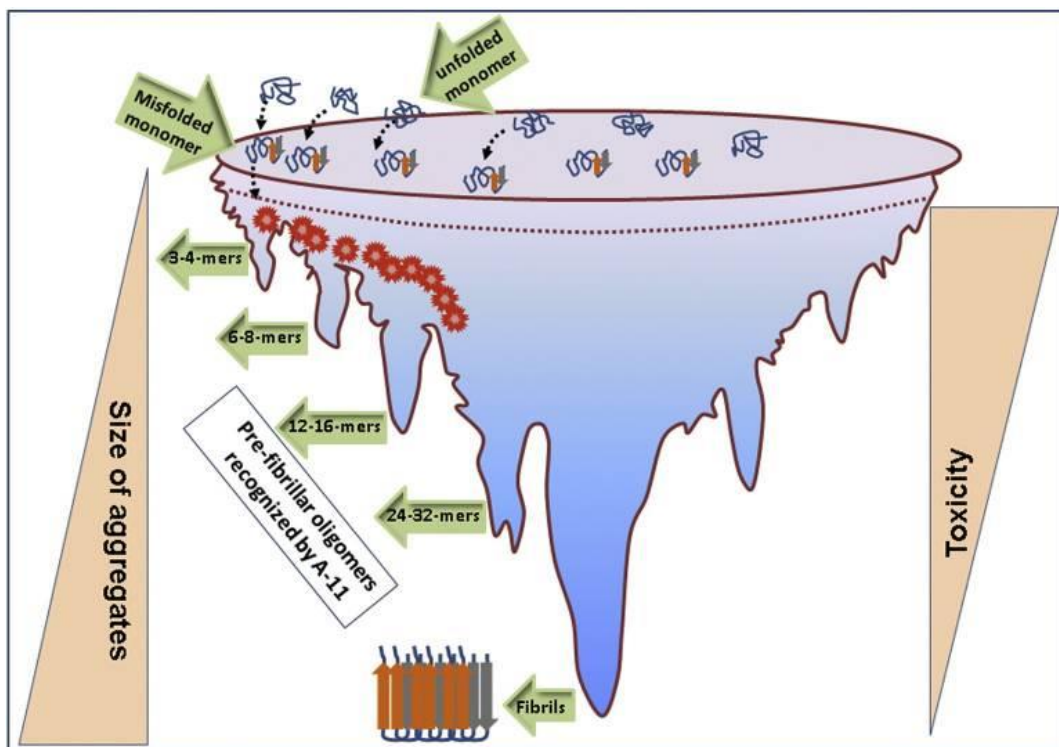


Figure 1.12. Schematic representation of the energy landscape of A β aggregation depicting the relationship between the size of A β aggregated species to their associated toxicity [137].

The mechanism by which the A β oligomeric species exert their toxicity is largely unknown. However, recent studies support that they interact with a wide range of receptors causing their malfunction, and also affect the cell membrane permeability, thus impairing calcium homeostasis [136,138]. Moreover, A β oligomers have also been associated with the degradation of the UPS, which could lead to the intracellular accumulation and subsequent aggregation of other proteins, such as tau, resulting in increased cellular toxicity and neuronal death [136,138,139]. Interestingly, recent studies suggest that both oligomers and mature fibrils are involved in the neuronal transmission of protein aggregation by a prion-like mechanism, thus contributing to disease progression [103,136].

Overall, while the mechanisms that underlie AD are not entirely clear, there is strong evidence supporting that A β plays a key role in AD pathogenesis and thus, targeting the aggregation of A β represents a highly promising therapeutic approach.

1.4. Therapeutic strategies against PMDs

As PMDs include notorious and mostly incurable disease with very high socio-economic impact, there is an unmet medical need to develop effective therapeutics against these diseases. While all PMDs are associated with protein aggregation, they have very different pathologies and diverse symptoms. For this reason, the approved treatments that are currently used vary greatly, ranging from surgery, such as liver transplantation to treat TTR-FAP [140] and fibrinogen amyloidosis [141] or cataract surgery [142], to substrate reduction therapy and enzyme replacement therapy to treat lysosomal storage disorders [143,144]. In the following section the most common approaches that are currently followed for the development of new and effective therapeutics are presented.

1.4.1. Manipulating the proteostasis network

As mentioned earlier, cells have evolved a robust quality control system comprising of molecular chaperones and degradation processes, in order to protect them from aggregation. Exploiting this system by either enhancing the function of molecular chaperones or stimulating aggregate degradation could be an effective strategy for the treatment of different PMDs. Indeed, activation of the heat-shock response has been used as a therapeutic approach against various diseases such as AD, PD, HD and cancer. For example, activation of the heat shock factor 1 (HSF1) by celastrol has exhibited cytoprotective effects from polyglutamine toxicity associated with HD [145]. Likewise, overexpression of Hsp70 has been associated with decreased toxicity and neurodegeneration in models of HD and PD [146,147]. Notably, the small molecule Hsp70 activator arimoclomol is currently in Phase 3 trials for the treatment of ALS.

On the other hand, inhibition of Hsp90 has been utilized as a therapeutic approach against various PMDs. For example, the Hsp90 inhibitor NVP-AUY992 has been found to promote degradation of polyQ-HTT [148] and 17-AAG has been found to reduce the levels of wild-type and mutant LRRK2, which has been associated with late-onset PD [149], as well as several p53 mutants [150]. Notably, ganetespib, another Hsp90 inhibitor, is currently under Phase I and Phase II trials for the treatment of various types of cancer such as ovarian, breast and lung [151]

Based on similar principles, activation of the cellular degradation processes can be promising for the treatment of PMDs. For example, the immunosuppressant rapamycin has been found to induce autophagy and enhance clearance of α -syn, tau and polyglutamine- or polyalanine-expanded proteins, thus mitigating their associated toxicity [152,153]. Likewise, activation of UPR-associated transcription factors, has been found to selectively reduce

secretion of destabilized TTR mutants and amyloidogenic LC, resulting in their subsequent degradation and decreased aggregate formation [154,155].

1.4.2. Enhancing aggregate clearance by immunization

Another approach to enhance clearance of protein aggregated species is by utilizing antibodies, which bind to such species and promote their degradation by antibody-mediated phagocytosis. This approach has been widely used against AD with many antibodies currently in clinical trials. For example, antibodies aducanumab and gantenerumab specifically bind to aggregated forms of A β and enhance their clearance resulting in increased cognitive function [156,157]. Similarly, the antibody crenezumab binds to A β oligomers and higher order aggregates and antibody BAN2401 selectively recognised A β protofibrils, both preventing plaque formation and disease progression [158,159]. On the same note, antibodies against tau, such as BIIB076, ABBV-8E12 and RO7105705 have also been studied for the treatment of AD and are currently in clinical trials [97].

Besides AD, immunization has been considered a potential therapeutic approach for other PMDs as well. For example, two antibodies, BIIB054 and RO7046015 have been developed against aggregated forms of α -syn and are both currently in Phase 2 trials for the treatment of PD [160,161]. Similarly, the α -miSOD1 antibody, currently in pre-clinical development for ALS treatment, selectively recognises misfolded variants of the Cu/Zn superoxide dismutase 1 (SOD1), reducing its aggregation and its associated neurodegeneration and [162].

1.4.3. Targeting the production of MisPs

One of the most obvious approaches towards the prevention of the aggregation of a specific disease –associated MisP, is the selective decrease of its cellular levels. Indeed this

approach has been employed against many PMDs. In the case of Alzheimer's disease, the disease-associated A β peptide is produced by proteolytic cleavage of the β -secretase (BACE-1) and a γ -secretase (Figure 1.10). Inhibitors of both enzymes have been extensively studied, however results so far have been unsatisfactory [97]. Specifically, semagacestat, the only γ -secretase inhibitor that has reached Phase 3 clinical trials, was withdrawn due to adverse reactions that led to increased incidence of skin cancer and infections [163], while next-generation γ -secretase inhibitors, such as avagacestat, that were developed in an attempt to alleviate these toxic effects, were not able to achieve the intended goals and the related trials have since been halted [164]. On the other hand, first-generation BACE-1 inhibitors had unfavourable physicochemical properties, while later frontrunners, such as verubecestat, were withdrawn from clinical trials due to insufficient efficacy [97,165]. However, other BACE-1 inhibitors are still being evaluated in clinical trials in the hope of positive effects, also taking into account that innovative clinical trial design, including patients in the early and asymptomatic disease stages, may play an important role in the trials' results.

Notably, proteolysis has been involved in the production of other aggregation-prone proteins, besides A β . Some examples include calpain cleavage and caspase-6 processing of huntingtin [166,167], caspase-mediated cleavage of the N-terminal segment of polyQ expanded ataxin-3 [168,169], proteasome-mediated C-terminal cleavage of α -syn [170,171], caspase-mediated C-terminal processing of tau [172,173] and sequential cleavage of gelsolin by furin and MTI-matrix metalloproteinase [174]. However, selective targeting of the corresponding proteases with minimal side effects, remains a challenge for drug development.

Another post-translational modification that results in the production of aggregation-prone protein species is phosphorylation. Indeed hyper-phosphorylation of tau has been linked with many tauopathies, such as AD and frontotemporal dementia with parkinsonism-17, implicating the GSK3 and CDK5 kinases among others [175,176]. Interestingly, GSK3

inhibition by lithium or the small molecule AR-A014418, resulted in decreased tau phosphorylation and reduced neurodegeneration in mice carrying the disease –associated variant P301L [177,178].

Another approach to reduce the levels of a PMD-associated proteins is using gene silencing by RNA interference (RNAi) or antisense drug technology. In 2018, two such molecules, the RNAi therapeutic Onpatro and the antisense Tegsedi were approved for the treatment of TTR-amyloidosis [179], while three antisense drugs, IONIS-HTT_{RX}, IONIS-SOD1_{RX} and IONIS-MAPT_{RX}, which target the production of HTT, SOD1 and tau, respectively, are currently in clinical trials with very encouraging preliminary results [97]. Importantly, the recently developed CRISPR-Cas9 technology has also been used for gene editing of proteins associated with PMDs, such as polyQ-HTT [180] and α 1-antitrypsin [181], and is believed to hold great promise for next-generation PMD therapeutics [182].

1.4.4. Stabilizing the native state of MisPs

Another approach to targeting PMDs is the stabilization of the native state of globular MisPs. One very successful example of this category is tafamidis, an approved small-molecule drug for the treatment of TTR-FAP, which functions by stabilizing mutated transthyretin to its native homotetrameric form, preventing its misfolding and aggregation and therefore, delaying disease progression [54]. Similarly, lumacaftor and tezacaftor, which have recently been approved for the treatment of cystic fibrosis, increase the stability of the CFTR(Δ F508) mutation and therefore enhance its trafficking to the cellular membrane where it normally functions [183,184].

Other protein targets that could benefit from native state stabilization are destabilized mutants of p53, such as the aforementioned p53(Y220C) structural variant. Indeed, Fersht and co-workers have already identified two p53(Y220C) stabilizers, PhiKan083 and PK7088, that

bind to the mutated protein and restore its apoptotic function [185,186]. However, there is no published work involving clinical studies of these molecules.

Similarly, point mutations in SOD1, which is associated with ALS [187], have been found to increase the homodimer's dissociation rate, leading to misfolding, protein aggregation and onset of the disease. In 2005, Lansbury and co-workers identified various small-molecule stabilizers of SOD1 mutants [188], however, these results were later disproven and attributed to experimental error [189]. Nevertheless, Samar Hasnain and co-workers have recently identified another SOD1 stabilizer, ebselen, which rescues the misfolding of mutant SOD1 and alleviates the associated toxicity [190].

1.4.5. Inhibiting the toxic aggregation of MisPs

The identification of aggregation inhibitors is probably the most widely used therapeutic approach. Indeed an extensive range of compounds have been identified that operate either by inhibiting the different stages of the protein aggregation mechanism or by directing the protein aggregation to “off-pathway” routes. These can be classified into two categories: small chemical compounds and peptides.

1.4.5.1. Small chemical compound inhibitors

Small molecules are generally preferred by researchers and pharmaceutical companies due to their drug-like properties, such as improved pharmacokinetics and blood-brain barrier permeability, and therefore it is not surprising that they have been widely studied as PMD therapeutics. Indeed a wide range of small-molecule synthetic and natural compounds have been identified as inhibitors of disease-associated protein aggregation, with several of them exhibiting a broad activity against PMDs of different pathologies. For example, curcumin has been found to inhibit aggregation of A β , α -syn and the islet amyloid polypeptide (IAPP) [191],

while rosmarinic acid inhibits aggregation of A β , tau and crystallin among others [192-194]. Likewise, Congo Red has been found to inhibit fibrillogenesis of A β , α -syn, polyQ-HTT and protein kinase C γ , whose aggregation has been associated with spinocerebellar ataxia type 14 [195-197]. Furthermore, compounds from the catechol chemical class have been shown to inhibit the aggregation of α -syn, A β and LC [198-200].

Moreover, epigallocatechin-3-gallate (EGCG) has been shown to inhibit the associated toxicity of various MisPs, such as polyQ-HTT, α -syn, A β , tau, LC, IAPP and several prions, by binding to oligomeric or higher order aggregates and re-directing their aggregation off-pathway, thereby inhibiting the formation of toxic aggregated species [201]. Similarly, resveratrol inhibits mutant p53 aggregation and directs the A β aggregation off-pathway by assisting the production of non-toxic oligomers and fibrils [202,203]. Interestingly, both compounds are currently in clinical trials with EGCG examined as a treatment of AD [204] and resveratrol for type 2 diabetes, Friedreich's ataxia, HD and AD among others [205-208].

Other noteworthy aggregation inhibitors include myricetin, which inhibits tau and α -syn fibrillization [209,210], baicalein and rifampicin, both inhibiting the aggregation of α -syn [210,211], maysin and bexarotene, which inhibit A β oligomer formation [212,213], and NPT200-11 and LMTX, which inhibit the aggregation of α -syn and tau, respectively, and are currently in clinical trials for the treatment of the associated diseases [214,215]. Although many small molecules have been identified as potent inhibitors, their lack of efficacy in clinical trials, increased toxicity, lack of specificity and in most cases unclear mechanism of action, render them unsuccessful drug candidates [194].

1.4.5.2. Peptide inhibitors

Another strategy for identifying aggregation inhibitors is the rational design of short peptides that target the protein region associated with protein misfolding and the protein-

protein interactions that result in protein aggregation. These peptides usually include a complementary to the MisP segment and a disruptive feature which hinders protein aggregation, such as non-natural amino acids, β -sheet breakers, charged residues or bulky groups [216]. For example, the peptides KLVFFKKKK and KLVFFEEEE, which comprise of the A β ₁₆₋₂₀ sequence and a short tail of charged amino acids, have been found to inhibit A β aggregation and its associated toxicity [217]. The same motif of A β has been utilized by other research teams by incorporating bulky groups such as cholic acid [218], N-methyl amino acids [219], β -sheet breakers [220,221] and D-amino acids [222,223]. Besides A β ₁₆₋₂₀, other amyloidogenic motifs, such as A β ₂₅₋₃₅, the GxMxG motif and C-terminal A β fragments, have also been utilized successfully for the rational design of A β aggregation inhibitors [224-226].

Similarly, peptide inhibitors of the aggregation of other disease-associated MisPs have also been identified. For example, Eisenberg and co-workers have identified ReACp53, a 17-residue p53 aggregation inhibitor, which comprises of the p53₂₅₂₋₂₅₈ aggregation-prone region LTIITLE and a poly-arginine tag that creates steric hindrance and inhibits adhering of other p53 molecules [227]. Likewise, Engel and co-workers have identified several short peptides complementary to the SOD1₁₀₆₋₁₁₁ aggregation-prone sequence that were able to inhibit protein aggregation and affect aggregate morphology [228]. Based on similar principles, Argon and co-workers have designed TAT-TISS, an inhibitor of immunoglobulin light chain aggregation associated with AL [229]; Fraser and co-workers have designed two hexapeptides that inhibit IAPP aggregation associated with type 2 diabetes [230]; Wiman and co-workers have identified a peptide that corresponds to the C-terminal p53₃₆₁₋₃₈₂ section and restores the apoptotic function of p53 mutants in cancer cell lines [231]; and López de la Paz and co-workers have identified several tau aggregation inhibitors containing D-amino acids [223].

Besides rational design, combinatorial library screening has also been used for the identification of potent peptide inhibitors against PMDs. For example, Moffet and co-workers

have identified three short peptide inhibitors that mimic aggregation-prone regions of A β , using a high-throughput bacterial screening assay [232], while Kamijo and co-workers identified eight peptide A β aggregation inhibitors that were rich in arginine residues, using phage display [233].

In conclusion, peptide aggregation inhibitors have many advantages compared to small molecule inhibitors as they are highly specific and exhibit limited toxicity. However, their poor pharmacokinetic properties, such as rapid degradation and low cell membrane permeation, render them problematic for drug development [234]. Nevertheless, recent advances in peptide chemistry, including peptide design, synthesis, formulation and delivery, have revived interest in peptide therapeutics, especially for use against difficult targets, such as protein-protein interactions and protein aggregation.

1.5. Thesis aims and outline

As discussed in the previous section, a large number of human diseases, named protein misfolding diseases (PMDs), arise from a common molecular mechanism, i.e. the inability of a specific protein to adopt its proper 3D conformation [235]. These include diseases with varying pathologies, such as AD, PD, cystic fibrosis and cancer [30,44,48-50]. Importantly, while some of these conditions are rather rare, such as cystic fibrosis [236], others have a significant socioeconomic impact, such as AD, which due to the increasing number of patients worldwide has been recently referred to as a “21st century plague” [237].

Despite the many advances in the understanding of the biology and mechanisms that underlie PMDs, most of them remain incurable and new disease-modifying agents are in urgent demand. The delay in drug development against PMDs can be attributed to a number of factors, including limited chemical diversity of the test compounds and time-consuming screening methodologies. Indeed, the diversity of the frequently tested small-molecule libraries is usually not higher than 10^5 - 10^6 [238]. This number is significantly limited compared to the $\sim 10^{60}$ theoretically possible low-molecular-weight structures [239]. Furthermore, even when larger chemical libraries are available, the drug discovery process is hampered by the inability of most screening methodologies to identify bioactive hits in a high-throughput manner. To corroborate this, the most commonly used functional screening assays utilize multi-well formats and therefore become very laborious and impractical for libraries with more than 10^6 - 10^7 members.

As part of an alternative strategy, some drug discovery programs rely on ultrahigh-throughput screening technologies, such as phage display, mRNA display and DNA-encoded library technology, which enable the generation of vast libraries containing up to 10^{13} test compounds and their subsequent screening for binding affinity to an immobilized target [240]. Indeed, these approaches have found great success in many cases and especially against

notoriously difficult targets, such as protein-protein interactions [241,242]. However, their inability to select for bioactivity rather than mere binding, ultimately increases the cost, complexity and time of the hit identification process significantly [241]. Therefore, in order to accelerate early drug discovery, ultrahigh-throughput methodologies that allow the deeper investigation of chemical space and the identification of hits with the desirable bioactivity are urgently required.

To that end, we have developed a synthetic biotechnology platform that enables the biosynthesis of molecular libraries with greatly expanded diversities and their simultaneous functional screening for the facile identification of potent aggregation inhibitors. Specifically, combinatorial libraries of macrocyclic peptides with expanded diversities are biosynthesized in the cytoplasm of *Escherichia coli* cells and simultaneously screened for their ability to correct the problematic folding of disease-associated MisPs, using an ultrahigh-throughput screening assay that links the aggregation of MisP with a selectable phenotype. We used this system to generate a library of ~200 million peptide macrocycles and to perform simultaneous functional screening for the discovery of putative therapeutic compounds against two major human diseases, cancer and AD, which although exhibit different pathologies, they have both been extensively associated with protein misfolding.

This thesis is organised in nine chapters. Chapter 1 presented the theoretical background related to protein folding and misfolding, the mechanisms by which the latter can lead to human disease and the current therapeutic strategies against such diseases. Emphasis was given in the two targeted PMDs, i.e. cancer and AD.

Chapter 2 describes the genetic engineering of *E. coli* cells in order to produce combinatorial libraries of more than 200 million drug-like head-to-tail cyclic peptides. This is

performed by utilizing protein splicing elements called inteins, which are widely used in protein engineering and biotechnology.

Chapter 3 describes the development of a generalized genetic assay for monitoring protein misfolding *in vivo*. This assay is based on monitoring the aggregation and misfolding of a MisP of interest by measuring the fluorescence levels of *E. coli* cells that produce recombinant fusions of the MisP with a fluorescence protein. The generality of this assay was demonstrated by monitoring the aggregation of four unrelated proteins, i.e. p53, A β , SOD1 and polyQ-HTT, and variants thereof that have been associated with cancer, AD, ALS and HD respectively.

Chapter 4 describes the combination of the constructed library with the generalized genetic assay in order to identify macrocyclic rescuers of protein misfolding and/or aggregation in an ultrahigh-throughput manner. This was achieved by introducing fluorescence activated cell sorting (FACS) into the system, in order to enable rapid screening of expanded libraries of cyclic peptides and selection of hits that enhance bacterial fluorescence due to an aggregation inhibitory activity. This system was successfully applied against four protein targets: p53(Y220C), A β 42, SOD1(A4V) and HTT_{ex1-97Q}, resulting in the selection of numerous bioactive or potentially bioactive cyclic peptides.

Chapter 5 describes the evaluation of the selected hits against p53(Y220C) *in vitro* and in cancer cell lines carrying this mutation. Notably, the majority of the selected hits were found to promote cancer cell death, with one cyclic peptide exhibiting a significant anti-cancer effect. Moreover, *in vitro* assessment of the selected hits revealed that the latter macrocycle was able to thermodynamically stabilize p53(Y220C), while two other cyclic peptides were able to interfere with the aggregation of p53(Y220C) without affecting its thermodynamic stability.

Chapter 6 describes the evaluation of the selected hits against A β aggregation, which resulted in the identification of >400 putative inhibitors, forming distinct clusters with different sequence characteristics. Representative members of the two most dominant clusters were tested *in vitro* and in established *Caenorhabditis elegans* models of AD and demonstrated their aggregation inhibitory and neuroprotective function. Furthermore, a combination of deep sequencing and site-directed mutagenesis analyses demonstrated how this system can accelerate the determination of structure-activity relationships (SAR) and define consensus motifs required for high bioactivity in the discovered molecules.

Finally, in Chapters 7 and 8, the results presented in the previous chapters and the future perspectives are discussed and in Chapter 9 the methods used throughout this thesis are presented.

In conclusion, the work presented in this thesis describes the development of a highly adaptable system that enables the discovery and characterization of potent aggregation inhibitors in a fully unbiased manner. Importantly, this system allows the facile investigation of more than 200 million different molecules and thus comprises the largest, to our knowledge, *in vivo* high-throughput functional screen against protein aggregation that has been reported to date.

Chapter 2 – Construction and biosynthesis of a combinatorial cyclic peptide library with expanded diversity

2.1. Peptides in drug discovery

In recent years, there has been a revival of interest in peptide-based molecules in drug discovery and development. Specifically, during the past two decades, 35 new peptide drugs targeting metabolic diseases, cancer, cardiovascular diseases, infectious diseases and neurological disorders amongst others, have entered the market, while over 150 are currently in active development [243].

Overall, peptides possess several favourable characteristics, thus bridging a significant gap between small molecule and biological therapeutics. Foremost, the high chemical and structural diversity of peptides render them extremely versatile, thus increasing the chances of interaction with pharmaceutical targets [244]. Furthermore, their functional groups can be altered chemically, producing molecules with improved properties. Moreover, peptides are often characterized by high specificity and binding affinity to their targets, due to their extensive interaction surface area, as well as low *in vivo* toxicity due to their nontoxic amino acid metabolites [240]. Finally, the small size of peptides enables the interaction with intracellular drug targets that are hardly accessible by larger biologics [245]. The last two characteristics make peptides ideal candidates against notorious targets, such as protein-protein

interactions, as they provide larger interaction surfaces compared to small-molecules and enhanced accessibility compared to protein biologics [246].

However, linear peptides exhibit some characteristics that are unattractive in drug development, i.e. high protease susceptibility, making them highly unstable under physiological conditions, and poor oral bioavailability and cell membrane permeability compared to small-molecules, making them less suitable against intracellular targets [234]. One of the most promising approaches to overcome these obstacles is cyclization of the linear peptide. Protecting the amino and carboxyl termini of the peptides by cyclization dramatically increases their stability *in vivo* as it reduces proteolytic degradation by both exopeptidases and endopeptidases [247]. Furthermore, in some cases the restricted conformation of cyclic peptides and the absence of charged termini have been found to increase membrane permeability, although it has been also shown, that there are some exceptions to this rule [248-251]. Finally, the binding affinity of cyclic peptides is typically highly improved compared to their more flexible linear analogues as the propensity for β -turn formation is increased and as cyclization enhances the rigidity of their structure resulting in lower entropic penalty upon binding to their target and formation of favoured side-chain conformations that facilitate the molecular recognition process [244,252].

2.2. Methods for the preparation of cyclic peptide libraries

The increased interest in cyclic peptides during the past years has encouraged the development of various methods for the production of combinatorial libraries of this type of molecules, divided into two main categories: (i) synthetic methods, which include chemical synthesis in solution and solid-phase, and (ii) genetic methods, which include cellular display

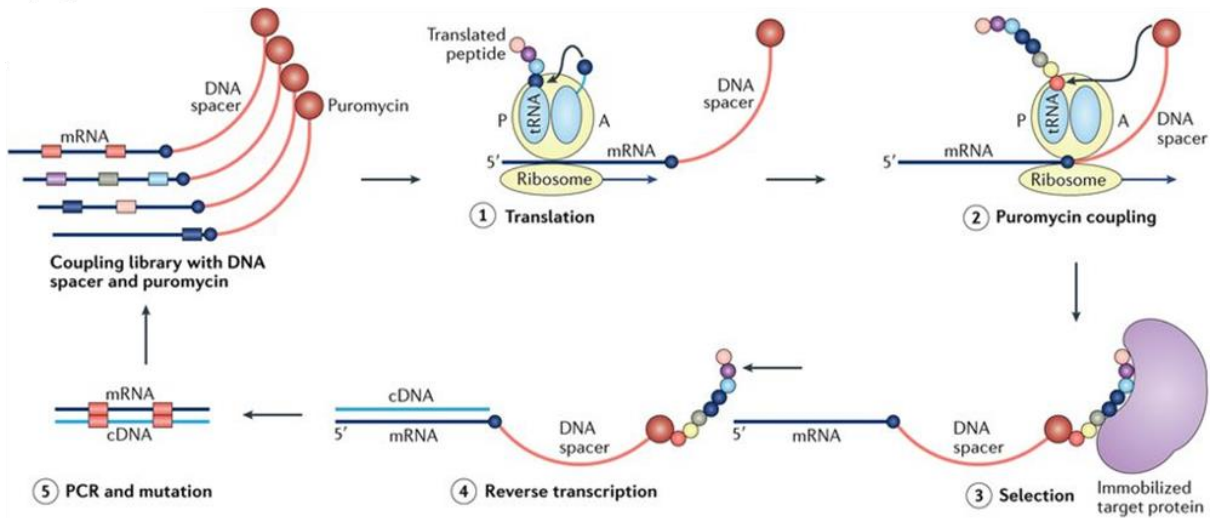
technologies, such as phage, yeast or bacterial display, non-cellular display technologies, such as mRNA and ribosome display, and intein-based techniques.

In general, synthetic approaches have many drawbacks compared to genetic methods, with the two most important being that the preparation of large cyclic peptide libraries is very laborious, and when these are synthesized in pooled formats hit deconvolution is highly challenging [253]. However, a very promising synthetic approach has recently emerged, where DNA-encoded peptide libraries can be synthesized on solid-phase support using split-and-pool synthesis [254]. Using this method libraries of trillions of members have been created and screened against various targets, while hit identification can be easily performed through DNA sequencing [255-257].

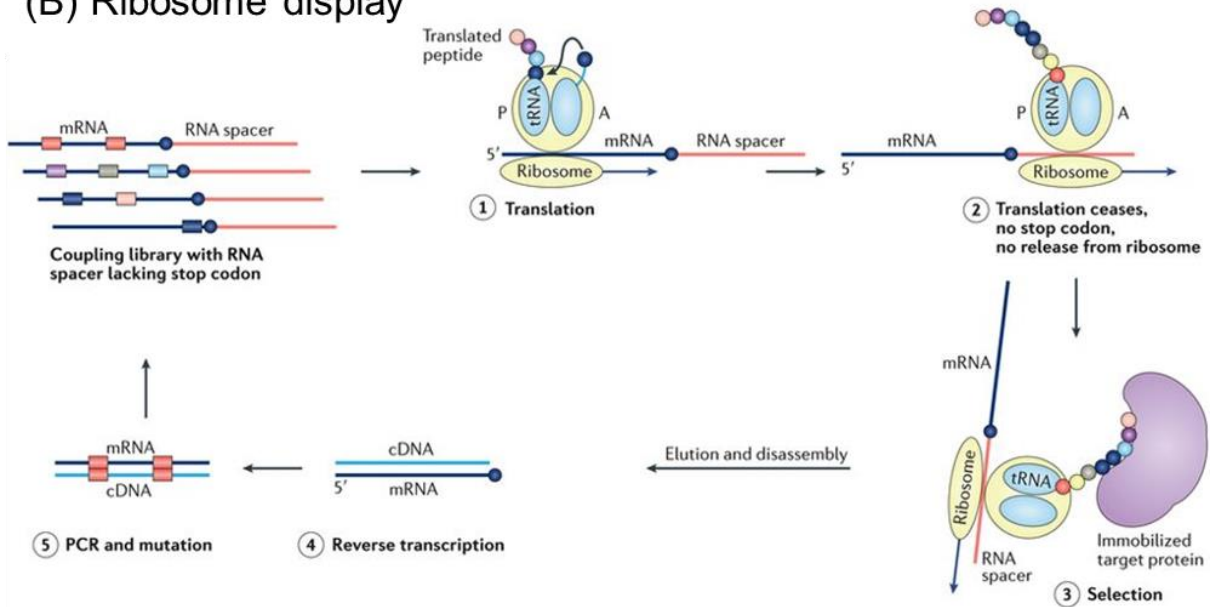
In both cellular and non-cellular display technologies the same principle applies: a library of peptides is first displayed onto the surface of biological entities, such as bacteriophages, yeasts, bacteria, mRNAs or ribosomes, then cyclized either via disulphide bridges between Cys residues, or via chemical linkers, and finally screened for binding to an protein target [258-262]. The top hits can be easily selected through multiple cycles of screening and identified through simple sequencing of the respective DNA sequence (Figure 2.1) [258-262]. In mRNA and ribosome display technologies, the use of flexizymes, i.e. flexible tRNA acylation ribozymes, has also enabled the incorporation of a broad range of artificial amino acids into the translated peptide sequence, which can then be cyclized via cross-linking [263]. These have been successfully utilized by the RaPID (Random nonstandard peptide integrated discovery) system, a combination of non-standard peptide translation and *in vitro* mRNA display, which has proven highly effective against multiple targets [264]. Notably, phage and mRNA display technologies have also enabled the production and screening of

bicyclic peptides, that contain either two independent loops or loops knotted in a θ conformation, by using carefully selected chemical linkers and click chemistry [265,266].

(A) mRNA display



(B) Ribosome display



(C) Phage display

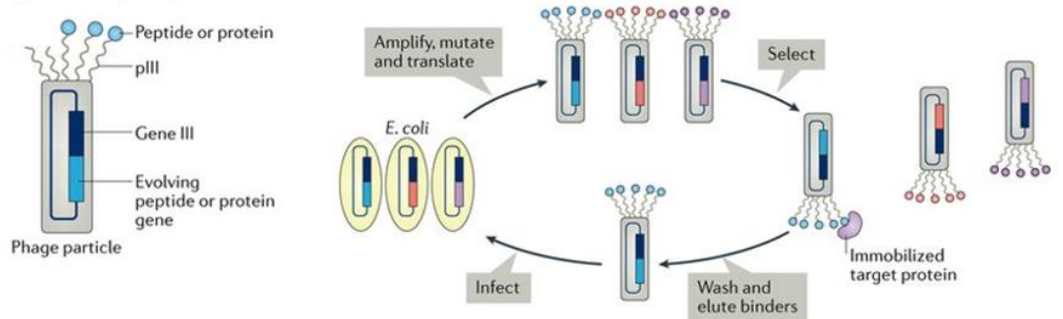


Figure 2.1. Display technologies for the production and high-throughput screening of cyclic peptides. (A)

In mRNA display, a library of transcribed mRNA is coupled to a pyromycin molecule via a DNA spacer. Upon *in vitro* translation of the mRNA library, pyromycin is coupled to the end of the peptide chain, thus linking each translated peptide with its encoding mRNA. The library is then screened for binding affinity to an immobilized target and selected hits are reversely transcribed to their respective cDNA and amplified via PCR. After multiple rounds of selection the isolated hits are identified via DNA sequencing. (B) Ribosome display is similar to mRNA display with the main difference being that the library of transcribed mRNAs is coupled to an RNA spacer without a stop codon. This spacer provides sufficient distance between the ribosome and the translated peptide sequence, thus enabling its interaction with protein targets, and at the same time prevents the release of the peptide from the tRNA/mRNA/ribosome complex, thus enabling phenotype-to-genotype conjugation. (C) In phage display, a bacteriophage library is initially produced by coupling a gene library, which encodes the peptides of interest, to the gene of a bacteriophage cell surface protein, such as pIII. *E. coli* cells are then infected with the constructed phage library, whose DNA is amplified and translated, leading to the display of the peptide library onto the phages' surface. Phages are then released from the bacteria and screened for binding activity to an immobilized protein target. The selected phages subsequently re-infect *E. coli* cells and after multiple rounds, the enriched population is identified via DNA sequencing. Adapted from [267].

Using display technologies, libraries of up to 10^{13} members have been developed and screened with great success [240,259,266]. Despite their many advantages and successful application against various targets, one important drawback of these approaches is that they can only be used for the identification of strong binders and no information on their biological function can be acquired. For example, Kiessling and co-workers utilized the phage display technology for the identification of A β aggregation modulators using either monomeric or aggregated A β [268]. However, none of the peptides that could bind to monomeric A β showed any effect on its aggregation, while those with binding affinity to the aggregated states in fact accelerated the aggregation reaction. This illustrates that there is no guarantee that the selected hits from display assays will have the desirable activity, which ultimately increases the cost, complexity and time of the screening process significantly [241].

An alternative genetic method for cyclic peptide production is the *in vivo* split-intein circular ligation of peptides and proteins (SICLOPPS) [269,270]. Inteins are protein splicing elements that are widely used in protein engineering and biotechnology in applications such as tag-less protein purification, segmental isotopic labelling of large proteins for nuclear magnetic resonance (NMR) studies, *in vitro* and *in vivo* protein semi-synthesis, development of molecular switches using conditional protein splicing and last but not least head-to-tail cyclization of peptides and proteins [271,272]. Similarly with RNA splicing, protein slicing is an autocatalytic process, in which an intervening polypeptide (intein) is self-excised while concomitantly ligating its two flanking sequences (exteins) with a native amide bond (Figure 2.2A).

While most inteins are encoded by one gene forming a contiguous domain, some inteins are naturally, or engineered to be, encoded by two separate genes forming a split intein, which comprises an N-terminal (I_N) and a C-terminal domain (I_C) [273]. These domains remain inactive until encountering their counterpart, whereupon they undergo protein trans-splicing (Figure 2.2B). The SICLOPPS technique utilises a rearranged split intein so that the I_C precedes the I_N flanking a peptide sequence in the form I_C -peptide- I_N [274]. Upon interaction of the two intein domains, the two ends of the intervening peptide are ligated and a head-to-tail cyclic peptide is released (Figure 2.2C). Notably, SICLOPPS has been recently combined with ribosome display, leading to the production of bicyclic peptides in the cytoplasm of *E. coli* [275], while inteins have also been used for the semi-synthetic production of macrocyclic organic-peptide hybrids (MOrPHs) [276,277] and bicyclic organo-peptide hybrids (BOrPHs) [278].

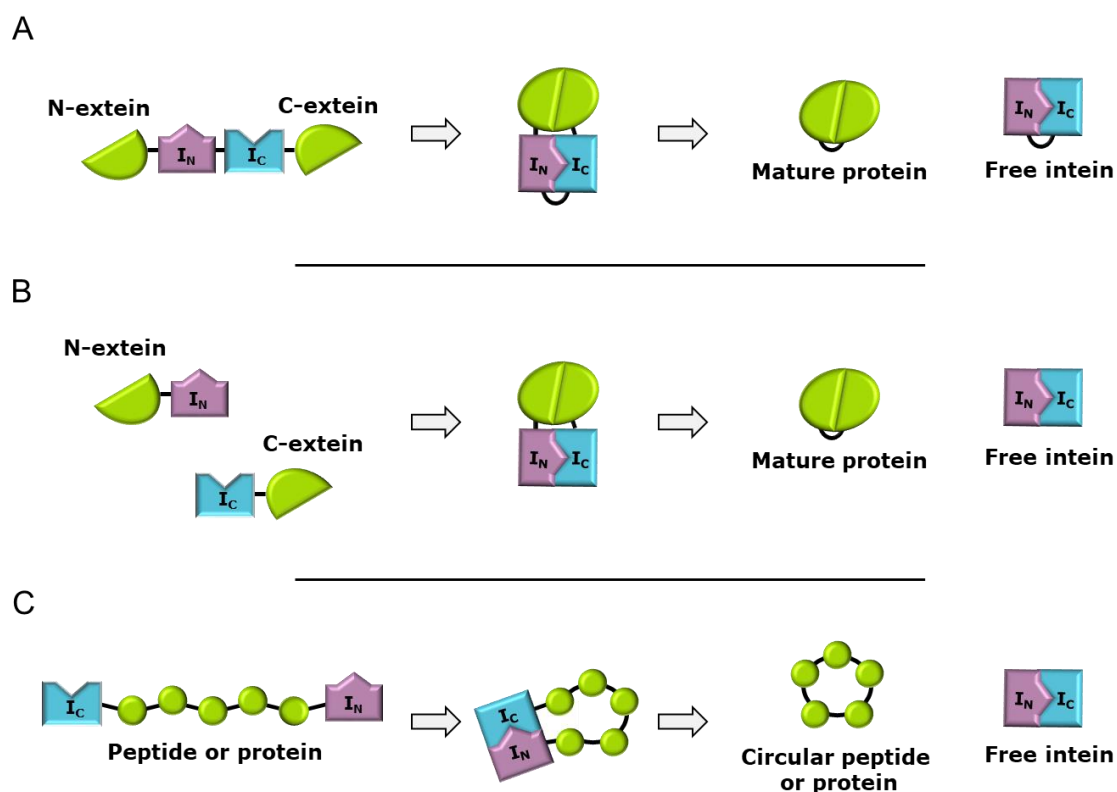


Figure 2.2. Schematic of protein splicing in (A) cis, (B) trans and (C) using the SICLOPPS technology.

The SICLOPPS technology has been successfully employed up to now in *E. coli*, yeasts, human cells and *C. elegans*, and libraries with up to 10^7 members for *E. coli*, 10^6 for yeasts and 10^5 for human cells have been constructed and subsequently screened against different targets [274]. However, compared to other aforementioned techniques, such as mRNA display, the theoretical diversity of the cyclic peptide libraries produced via SICLOPPS is significantly smaller, restricted by the number of possible combinations of amino acids allowed by the specific library design and the maximum transformation efficiency of the host organism ($\sim 10^9$ - 10^{10} transformants for *E. coli*, $\sim 10^6$ transformants for *Saccharomyces cerevisiae*, $\sim 10^4$ transformants for mammalian cells). Nevertheless, this method is highly valuable as the cyclic peptide libraries are produced inside cells, and therefore can be screened using cell-based functional assays [274,279]. This is an important advantage as cyclic peptides produced by this approach can be selected according to their bioactivity and not just their

binding affinity to a given target, eliminating the possibility of identifying cyclic peptides that are strong binders but are either completely inactive or with the opposite effect than the originally intended.

A noteworthy application of the SICLOPPS technology in early stage drug discovery is for the identification of cyclic peptides that inhibit α -syn toxicity, which is associated with PD [280]. Lindquist and co-workers applied the SICLOPPS technology in *S. cerevisiae* by constructing a library of 5 million cyclic octapeptides and screening them for suppression of α -syn toxicity. This methodology resulted in the identification of two potent suppressors that were also able to rescue neuronal degeneration in a *C. elegans* model of synucleinopathy. Furthermore, the peptidic nature of the selected compounds allowed the rapid determination of structure-activity relationships, resulting in the identification of a common bioactive motif [280]. Recently, one of the selected bioactive sequences was also successfully grafted onto an alternative, more rigid macrocyclic peptide scaffold called cyclotide, using intein-mediated protein trans-splicing, demonstrating that cyclotide-based libraries can also be used for phenotypic screening in yeast [281].

2.3. Construction of a combinatorial cyclic peptide library with expanded diversity using the SICLOPPS technology

The head-to-tail cyclic oligopeptide library presented herein was constructed using the SICLOPPS technology and the *Synechocystis* sp PCC6803 DnaE intein (Ssp DnaE) [282]. This intein was selected due to its low toxicity compared to other novel inteins and its relatively low sensitivity to the extein sequence, thus allowing the successful cyclization of a wide range of peptides [283]. The only requirement in order for the splicing mechanism to occur is the

presence of a nucleophile cysteine, serine or threonine as the first amino acid of the extein sequence (Figure 2.3).

We chose to study cyclic tetra-, penta-, hexa- and heptapeptide sequences with the general formula NuX₁X₂..X_N, where X is any one of the 20 natural amino acids, N=3-6, and Nu=Cys, Ser, or Thr. These molecular libraries have an average molecular weight of <770 Da, thus potentially exhibiting the advantages of both small molecules and larger peptides that are typically <900 Da [284]. Specifically, tetrapeptides have an average molecular weight of ~ 440 Da and generally comply with Lipinski's rule of 5 for druggability [285], while penta-, hexa- and heptapeptides occupy, on average, an area of chemical space beyond the rule of 5 (bRo5 space), where different rules for drug-likeness apply [286,287]. Furthermore, the maximum theoretical diversity of the combined library is more than 200 million different sequences (Table 2.1), therefore enabling the investigation of a greatly expanded area of chemical space [238].

Table 2.1. Theoretical diversity of the constructed combinatorial cyclo-NuX₁X₂X₃-X₆ oligopeptide library.

Peptide Type	General Formula	Theoretic Diversity
Tetrapeptides	cyclo-NuX ₁ X ₂ X ₃	3x20 ³ =24,000
Pentapeptides	cyclo-NuX ₁ X ₂ X ₃ X ₄	3x20 ⁴ =480,000
Hexapeptides	cyclo-NuX ₁ X ₂ X ₃ X ₄ X ₅	3x20 ⁵ =9,600,000
Heptapeptides	cyclo-NuX ₁ X ₂ X ₃ X ₄ X ₅ X ₆	3x20 ⁶ =192,000,000
Combined Library	cyclo-NuX ₁ X ₂ X ₃ -X ₆	202,104,000

Standard SICLOPPS pathway

Side reactions

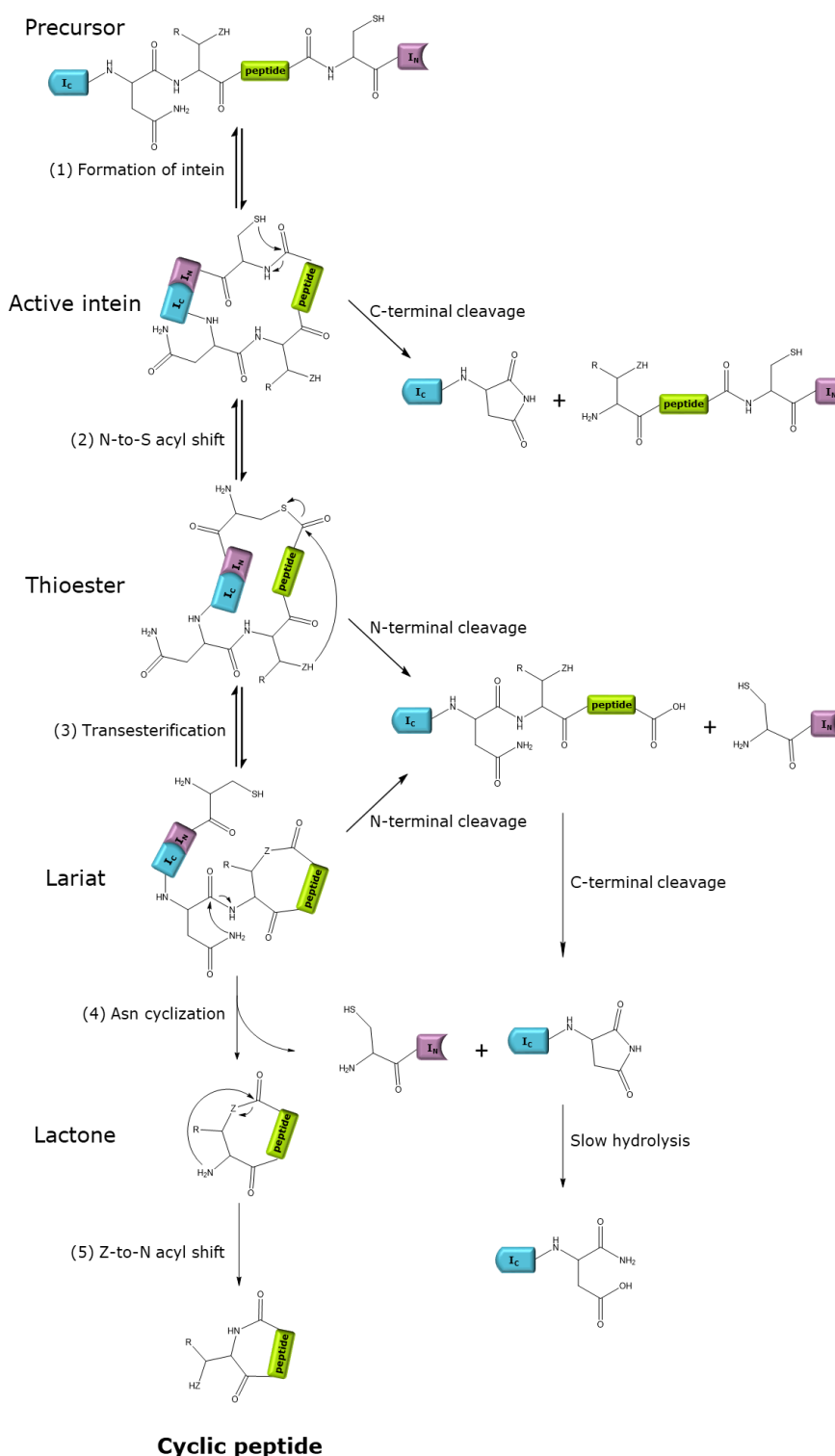


Figure 2.3. SICLOPPS mechanism standard pathway and side reactions. (1) Formation of an active intein; (2) N-to-S acyl shift yields a thioester intermediate; (3) Transesterification with a side-chain nucleophile (Cys, Ser or Thr, Z=O or S and R=H or CH₃ accordingly) to produce a lariat intermediate; (4) Asn side chain cyclization releases the lactone product; (5) Z-to-N acyl shift produces the thermodynamically favored cyclic peptide. Adapted from [288].

The libraries of genes encoding these cyclic oligopeptide libraries were constructed using degenerate PCR primers, in which the randomized amino acids (X) were encoded using random NNS codons, where N=A, T, G, or C and S=G or C. The degenerate codons encode all of the 20 natural amino acids without including the stop codons UAA and UGA. Cys, Ser, and Thr were encoded on these primers by utilizing the codons UGC, AGC, and ACC, respectively, which are the most frequently utilized codons for these amino acids in *E. coli*. The generated peptide-encoding gene libraries were cloned into the vector pSICLOPPS [282], as described in the Materials and Methods section, to form the pSICLOPPS-NuX₁X₂X₃X₄-X₆ vector library (Figure 2.4). These vectors express a combinatorial library of fusion proteins comprising four parts: (i) the N-terminal domain of the Ssp DnaE intein, (ii) a NuX₁X₂X₃-X₆ oligopeptide sequence, (iii) the C-terminal domain of the Ssp DnaE intein, and (iv) a chitin-binding domain (CBD) for immunodetection and/or purification, under the control of the P_{BAD} promoter and its inducer L(+)-arabinose (Figure 2.4). Cloning of the resulting gene libraries into the pSICLOPPS plasmid yielded a total of 1.3×10^9 independent transformants as judged by plating experiments after serial dilutions.

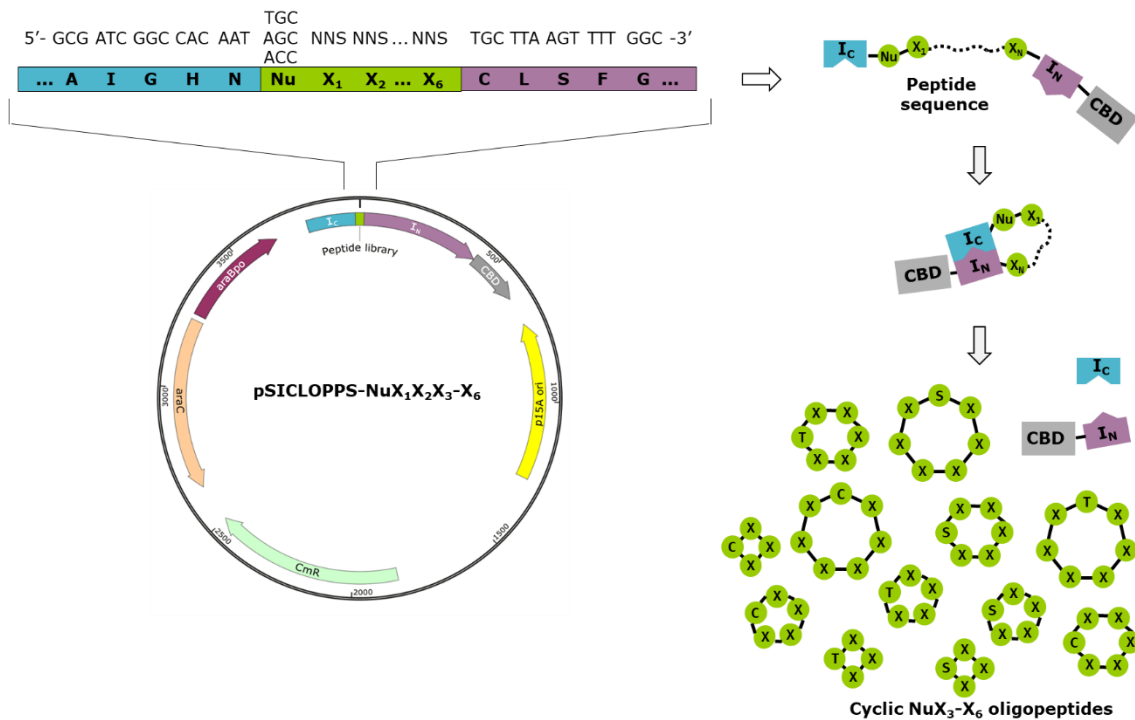


Figure 2.4. Schematic of the generation of the cyclic-NuX₁X₂X₃-X₆ peptide libraries via the SICLOPPS technology. (Left) Representation of the pSICLOPPS-NuX₁X₂X₃-X₆ vector library encoding the combinatorial heptapeptide library cyclo-NuX₁X₂X₃-X₆. Nu: Cys, Ser, or Thr; X: any of the 20 natural amino acids; NNS: randomized codons, where N=A, T, C or G and S=G or C; IC: C-terminal domain of the Ssp DnaE split-intein; IN: N-terminal domain of the Ssp DnaE split-intein; CBD: chitin-binding domain. (Right) Peptide cyclization using the SICLOPPS construct. Upon interaction between the two intein domains IC and IN, the encoded IC-NuX₁X₂X₃-X₆-IN-CBD fusions undergo intein splicing and peptide cyclization, leading to the production of the cyclo-NuX₁X₂X₃-X₆ library.

2.4. Quality assessment of the constructed combinatorial library of random cyclic oligopeptides.

In order to assess the quality of our constructed library, we randomly selected a total of 274 clones from the combined pSICLOPPSNuX₁X₂X₃-X₆ library and performed colony PCR and diagnostic digestion to verify insert size. Furthermore, we performed sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) and western blot analyses to verify production of the precursor fusion protein (molecular mass ~25 kDa), and processing of the

intein by monitoring the appearance of a lower molecular weight band (molecular mass ~20 kDa), which corresponds to the N-terminal intein domain fused to CBD (I_N -CBD) and indicates successful intein splicing and cyclic peptide formation (Figure 2.5). This analysis revealed that all constructed sub-libraries, apart from cyclo-ThrX₁X₂X₃X₄X₅, appear to produce a sufficient number of cyclic peptides in order to cover the theoretical diversity of each sub-library by at least one time (Table 2.2).

DNA sequencing of the peptide-encoding regions of the pSICLOPPS plasmid from 23 randomly selected clones revealed the presence of all three Nu amino acids Cys, Ser, and Thr at position 1 and a good representation of the twenty natural amino acids at all other positions within the tetra-, penta- and hexapeptide sequence, albeit with an over-representation of codons corresponding to Gly (Table 2.1 and Table 2.3).

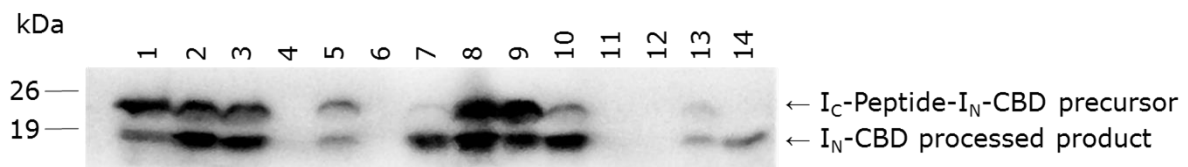


Figure 2.5. Indicative western blot analysis using an anti-CBD antibody of fourteen randomly selected individual clones from the constructed cyclo-NuX₁X₂X₃X₄X₅ hexapeptide sub-library, demonstrating that individual clones can exhibit variable levels of expression. Lanes 4, 6, 11 and 12 correspond to clones that contain stop codons or frameshifts and, thus, do not express a full-length IC-peptide-IN-CBD tetrapartite fusion or generate cyclic peptide product.

Table 2.2. Quality assessment of the constructed library via molecular biology techniques

Sub-library	General Formula	Transformants	PCR analysis and diagnostic digestion	SDS-PAGE analysis		Actual Diversity	Library Coverage
				tetra-partite fusion precursor	IN-CBD product		
Tetrapeptide	CysX ₁ X ₂ X ₃	770,000	8/13	8/18	6/18	256,667	32.1
	SerX ₁ X ₂ X ₃	710,000	11/13	10/18	7/18	276,111	34.5
	ThrX ₁ X ₂ X ₃	500,000	11/13	16/18	13/18	361,111	45.1
	combined NuX ₁ X ₂ X ₃	1,980,000	30/39	34/54	26/54	893,889	37.2
Pentapeptide	CysX ₁ X ₂ X ₃ X ₄	1,236,000	11/15	13/18	13/18	892,667	5.6
	SerX ₁ X ₂ X ₃ X ₄	604,000	17/25	7/18	6/18	201,333	1.3
	ThrX ₁ X ₂ X ₃ X ₄	836,000	9/10	15/18	15/18	696,667	4.4
	combined NuX ₁ X ₂ X ₃ X ₄	2,676,000	37/50	35/54	34/54	1,790,667	11.2
Hexapeptide	Cys-X ₁ X ₂ X ₃ X ₄ X ₅	12,870,000	9/15	15/18	14/18	10,010,000	3.1
	Ser-X ₁ X ₂ X ₃ X ₄ X ₅	8,784,000	12/15	10/18	8/18	3,904,000	1.22
	Thr-X ₁ X ₂ X ₃ X ₄ X ₅	5,316,000	8/15	14/18	5/18	1,476,667	0.5
	combined Nu-X ₁ X ₂ X ₃ X ₄ X ₅	26,970,000	29/45	39/54	27/54	15,390,667	4.8
Heptapeptide	Cys-X ₁ X ₂ X ₃ X ₄ X ₅ X ₆	410,450,000	42/90	37/90	37/90	168,740,556	2.6
	Ser-X ₁ X ₂ X ₃ X ₄ X ₅ X ₆	504,233,600	19/30	17/30	16/30	268,924,587	4.2
	Thr-X ₁ X ₂ X ₃ X ₄ X ₅ X ₆	319,000,000	21/30	20/30	14/30	148,866,667	2.3
	combined Nu-X ₁ X ₂ X ₃ X ₄ X ₅ X ₆	1,233,683,600	82/150	74/150	67/150	586,531,809	9.2

Table 2.3. Sequencing results of the peptide-encoding regions of 23 randomly selected clones from the constructed pSICLOPPS-NuX₁X₂X₃, pSICLOPPS-NuX₁X₂X₃X₄, and pSICLOPPS-NuX₁X₂X₃X₄X₅ vector sub-libraries.

Clone number	DNA sequence of peptide-encoding gene	Encoded peptide sequence
C4-1	TGC GGC AAG GTG	CGKV
C4-2	TGC CGC CAC CGG	CRHR
C4-3	AGC GCG TCC GGG	SASG
C4-4	AGC ACG CGC CGG	STRR
C4-5	ACC AAC TGG GTC	TNWV
C4-6	ACC AGG GCC TCC	TRAS
C4-7	AGC CGG GTG CTC	SRVL
C4-8	ACC AAC TGG CCG	TNWP
C5-1	TGC AAC TTG GTC TGG	CNLVW
C5-2	TGC TGC GCG GCG GGG	CCAAG
C5-3	TGC GCG TCG CGG GGG	CASRG
C5-4	AGC TTC GTG GAG GGG	SFVEG
C5-5	ACC TGC CCC GTG TAG	TCPV*
C5-6	ACC CCG GCG CGG TGC	TPARC
C5-7	ACC TCG GGC GCG TAG	TSGA*
C6-1	TGC GGG CGG GGG TGG ACG	CGRGWT
C6-2	TGC TGC AGC GGC TGC CGG	CCSGCR
C6-3	TGC AAG TCG GGG CAC GGC	CKSGHG
C6-4	AGC TTG GTG CCG TAC CTG	SLVPYL
C6-5	AGC GCC TAG GGC GGG CCC	SA*GGP
C6-6	AGC GAG GGG GGG GGG G	Frame shift
C6-7	ACC TCG CTC TAG TCC CAC	TSL*SH
C6-8	ACC AGG GGG GGC AGG GGG	TRGGRG
		*: stop codon

To evaluate the quality of the constructed libraries further, we characterized in more detail the amino acid diversity encoded within the generated oligopeptide libraries by performing deep sequencing analysis of the peptide-encoding region of the combined pSICLOPPS-NuX₁X₂X₃-X₆ vector library. Out of the ~3.7 million DNA sequences that were analysed, ~76% were unique and ~94% of those were found to encode unique cyclic peptide sequences (Table 2.4). Again, all amino acids were found to be encoded at every positions of the generated library, albeit with an over-representation of residues corresponding to Gly and Arg (Figure 2.6).

Table 2.4. Deep sequencing analysis of the peptide-encoding regions of ~3.7 million clones from the constructed pSICLOPPS-NuX₁X₂X₃-X₆ library

	Number of reads	Unique DNA sequences	Unique peptide sequences
cyclo-CysX ₁ X ₂ X ₃ sub-library	619	530 (86%)	438 (83%)
cyclo-SerX ₁ X ₂ X ₃ sub-library	1,348	1,018 (76%)	729 (72%)
cyclo-ThrX ₁ X ₂ X ₃ sub-library	1,197	1,100 (92%)	867 (79%)
Combined cyclo-NuX ₁ X ₂ X ₃ sub-library	3,164	2,648 (84%)	2,034 (77%)
cyclo-CysX ₁ X ₂ X ₃ X ₄ sub-library	1,456	1,313 (90%)	1,150 (88%)
cyclo-SerX ₁ X ₂ X ₃ X ₄ sub-library	2,765	2,304 (83%)	1,761 (76%)
cyclo-ThrX ₁ X ₂ X ₃ X ₄ sub-library	2,127	1,902 (89%)	1,626 (85%)
Combined cyclo-NuX ₁ X ₂ X ₃ X ₄ sub-library	6,348	5,519 (87%)	4,537 (82%)
cyclo-CysX ₁ X ₂ X ₃ X ₄ X ₅ sub-library	93,785	76,341 (81%)	60,493 (79%)
cyclo-SerX ₁ X ₂ X ₃ X ₄ X ₅ sub-library	73,570	63,493 (86%)	48,843 (77%)
cyclo-ThrX ₁ X ₂ X ₃ X ₄ X ₅ sub-library	83,401	73,929 (89%)	58,666 (79%)
Combined cyclo-NuX ₁ X ₂ X ₃ X ₄ X ₅ sub-library	250,756	213,763 (85%)	168,002 (79%)
cyclo-CysX ₁ X ₂ X ₃ X ₄ X ₅ X ₆ sub-library	1,305,675	1,023,580 (78%)	978,803 (96%)
cyclo-SerX ₁ X ₂ X ₃ X ₄ X ₅ X ₆ sub-library	1,318,365	885,393 (67%)	824,134 (93%)
cyclo-ThrX ₁ X ₂ X ₃ X ₄ X ₅ X ₆ sub-library	769,605	652,099 (85%)	636,647 (98%)
Combined cyclo-NuX ₁ X ₂ X ₃ X ₄ X ₅ X ₆ sub-library	3,393,645	2,561,072 (75%)	2,439,584 (95%)
Combined cyclo-NuX₁X₂X₃-X₆ library	3,653,913	2,783,002 (76%)	2,614,157 (94%)

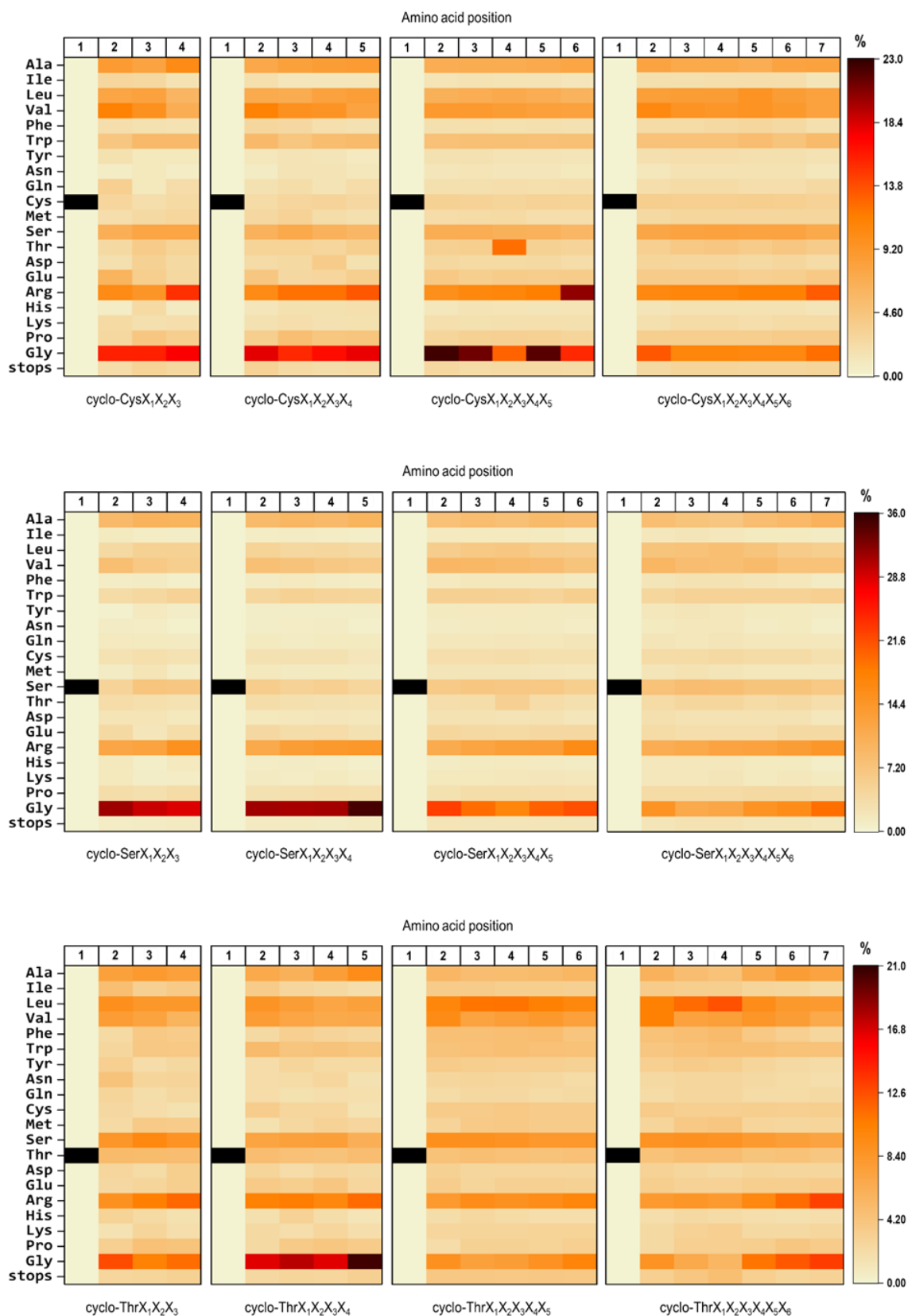


Figure 2.6. Heat map representation of the amino acid distribution at each position of the constructed cyclo-Cys₁X₂X₃-X₆ (top), cyclo-Ser₁X₂X₃-X₆ (middle) and cyclo-Thr₁X₂X₃-X₆ (bottom) sub-libraries, as demonstrated by the deep sequencing analysis results.

However the deep sequencing analysis revealed another important finding: the cyclo-CysX₁X₂X₃X₄X₅ and cyclo-SerX₁X₂X₃X₄X₅ sub-libraries are highly enriched with two DNA sequences that encode peptides CGGTGR and SGGTGR and which appear approximately 700 and 40 times more frequently than the succeeding sequence (Table 2.5). This defect could be attributed to an inherited preference in the degenerate PCR primers, or the ability of these almost identical sequences to promote *E. coli* fitness, leading to their enrichment during cell growth. Despite this drawback however, our results suggest that we have constructed a very high-diversity library, which should be encoding the vast majority of the theoretically possible tetra-, penta-, hexa- and heptapeptide cyclo-NuX₁X₂X₃-X₆ sequences.

Table 2.5. Top 10 most frequent sequences of the cyclo-CysX₁X₂X₃X₄X₅ and cyclo-SerX₁X₂X₃X₄X₅ sub-libraries as demonstrated by the deep sequencing analysis

cyclo-CysX ₁ X ₂ X ₃ X ₄ X ₅			cyclo-SerX ₁ X ₂ X ₃ X ₄ X ₅		
Reads	DNA sequence	Protein sequence	Reads	DNA sequence	Protein sequence
8153	TGCGGCGGCACCGGGCGC	CGGTGR	1459	AGCGGCGGCACCGGGCGC	SGGTGR
12	TGCGCCGGCACCGGGCGC	CAGTGR	38	AGCATGCCGTCCAGGGCG	SMPSRA
12	TGCGGCGGCACCGGGCGG	CGGTGR	32	AGCGGCGGGCACTTGTGG	SGGHLW
9	TGCCAGCAGATCATGCAG	CQQIMQ	21	AGCGGGGGGGGGGGGGGG	SGGGGG
8	TGCGAGGGCACGCCGCG	CEGTPA	19	AGCTGGCCGCGGGGCGCC	SWPRGA
6	TGCGGGAACCACGCGGGC	CGNHAG	18	AGCACGGGGTTGAGGATG	STGLRM
6	TGCGGCAACCAGAACTCG	CGNQNS	17	AGCGGGGCGATGCGGCTC	SGAMRL
6	TGCCCCTGCTGCGCCGC	CPLLRR	16	AGCTGGCGGAGCGGGGG	SWRSGG
6	TGCGACAGGGGGTTGTTC	CDRGLF	15	AGCTTCTGGGCGTGGAGG	SFWAWR
5	TGCATCGACGCGGAGTCG	CIDAES	14	AGCGGCTCCTAGGGGGTG	SGS*GV

2.5. Discussion

In conclusion, we have utilized a straightforward biotechnology technique called SICLOPPS, which permits the *in vivo* biosynthesis of libraries of cyclic peptides in a one-pot reaction requiring just the pSICLOPPS vector, a few degenerate PCR primers and a couple of simple molecular biology steps [274]. Using this technology we constructed a gene library that encodes more than 200 million cyclic oligopeptides with the general formula cyclo-NuX₁X₂X₃-X₆, where Nu is Cys, Ser, or Thr and X is any one of the 20 natural amino acids. This is, to our knowledge, the largest library of cyclic peptides ever produced using SICLOPPS.

The quality of this library was then assessed using: (i) molecular biology techniques, including colony PCR, diagnostic digestion and SDS-PAGE/western blot analysis and (ii) deep sequencing analysis. These techniques complemented each other very well: on the one hand colony PCR and diagnostic digestion offer the possibility of detecting the presence of the correct insert and estimating the cloning success rate and SDS-PAGE/western blot analysis permit the detection of the overexpressed precursor protein and its subsequent processing to yield the cyclic peptide. On the other hand, deep sequencing analysis offers the possibility of evaluating the diversity of the constructed library and identifying biases in the peptide sequences in great detail.

Overall, our results suggest that we have constructed a very high-diversity library, which should be encoding the vast majority of the theoretically possible >200 million cyclo-NuX₁X₂X₃-X₆ sequences.

Chapter 3 – Development of a generalized genetic screen for monitoring protein misfolding and aggregation

3.1. Screens for the identification of aggregation inhibitors as potential therapeutics against PMDs

Once a drug target has been identified, the next step in the drug discovery pipeline is the development and implementation of screening assays in order to identify hits and, subsequently, therapeutic leads. As PMDs have been associated with the misfolding and aggregation of one or more MisPs, targeting this aggregation process has been the most widely investigated therapeutic approach. Indeed, an extensive number of screening assays have been developed against PMDs which can be divided in three categories: *in silico*, *in vitro* and cell-based screening.

3.1.1. *In silico* screening assays for the identification of protein aggregation inhibitors

The most broadly used *in silico* method is structure-based design and high-throughput screening (HTS). It consists of using the protein's structural data from X-ray, NMR and cryogenic electron microscopy (cryo-EM) studies and performing molecular docking to detect possible ligand binding in a database of test compounds. Fersht and co-workers have successfully used this method to screen a collection of ~ 2 million commercially available drug-like small molecules and resulted in 80 potential binders against p53C(Y220C). One of these,

named PhiKan059, exhibited desirable *in vitro* results [185]. Subsequently, the same group screened a library of designed PhiKan059 analogues and identified PhiKan083, which exhibited improved binding affinity, validated by X-ray studies. Importantly, PhiKan083 was found able to stabilize p53C(Y220C) *in vitro*, increasing its half-life significantly [185]. Similarly, Amaro and co-workers have identified a cavity between loop L1 and sheet S3 of p53 and virtually screened 1,324 compounds identifying stictic acid as a chemical rescuer of the misfolding of various p53 mutants [289]. Indeed, *in vitro* and *in vivo* studies demonstrated that stictic acid was able to stabilize thermodynamically p53 mutants and to partially restore p53 activity in cancer cell lines.

On a related note, Vendruscolo, Knowles, Dobson and co-workers have utilized a quasi-structure-based design approach, where instead of using the structure of primary nuclei of A β , which are very difficult to characterize experimentally, they utilized the knowledge gained from a known A β aggregation inhibitor, bexarotene, and screened for small molecules that shared similar chemical properties [290]. This process resulted in the identification of 12 small molecules, whose aggregation inhibitory effects were validated by *in vitro* chemical kinetic experiments and *in vivo* *C. elegans* models.

Another *in silico* method widely used in drug discovery is fragment-based design, which refers to the production of compound fragments from reported active small-molecules and their subsequent screening using compound repositories and databases to identify potential therapeutics that are chemically related to the initial fragments. Using this approach Vendruscolo, Knowles, Dobson and co-workers have generated three distinct fragment-based libraries with potential aggregation inhibitors of A β , tau and α -syn, containing 16,850, 11,800 and 14,735 compounds, respectively [291]. *In vitro* studies of the A β library resulted in the identification of the aforementioned compound bexarotene as a potent A β aggregation inhibitor [212].

In conclusion, *in silico* assays have the great advantage that they can easily and cost-efficiently screen a huge number of compounds to identify novel therapeutics. Furthermore, their ability to screen available databases with approved medications allows for drug repositioning, which results in a significant reduction in drug development time and cost. However, these assays only identify putative hits and all results need to be subsequently validated by *in vitro* and *in vivo* assays. Moreover, the limited structural information for the majority of disease-associated MisPs renders structure-based design unsuitable for most PMDs.

3.1.2 *In vitro* screening assays for the identification of protein aggregation inhibitors

A wide range of techniques have been developed for monitoring protein misfolding *in vitro*, including ones that enable aggregate or protein characterization, such as Thioflavin T (ThT) and differential scanning fluorimetry (DSF), methods that provide secondary structure information, such as circular dichroism (CD) and Fourier transform infrared spectroscopy (FTIR), and techniques that enable the visual representation of the aggregated samples, such as transmission electron microscopy (TEM) and atomic force microscopy (AFM) [292].

ThT binding assays have been excessively used for the identification of aggregation inhibitors against PMDs. Vendruscolo, Knowles, Dobson and co-workers have recently described in detail the microscopic processes involved in amyloid formation of A β 42 as well as the rate constants and reaction rates governing each process, thus offering a significant advancement in the field [94]. This knowledge was later utilized to identify inhibitors of A β 42 with diverse inhibitory mechanisms [212,290,291,293]. Recently the same group has also developed a novel approach termed structure-kinetic activity relationship (SKAR), where a parent molecule is chemically modified creating a library of derivative compounds, which are

then screened using a ThT assay, ultimately resulting in the rational design of molecules with enhanced potency [294]. At the same time, Ventura and co-workers have developed an in-plate ThT assay for the rapid identification of α -syn aggregation inhibitors [295]. Using this assay they were able to screen 14,000 compounds resulting in the discovery of 47 novel strong modulators of α -syn aggregation, one of which was shown to protect dopaminergic neurons from α -syn-associated toxicity in a *C. elegans in vivo* PD model [295,296].

Another method that has been used for *in vitro* screening against MisP aggregation is DSF, a fluorescence-based assay that enables the identification of compounds that bind to and stabilize a test protein. Using DSF, Pérez and co-workers have screened a commercial library of 10,000 compounds and identified eight small molecules that stabilize mutants of phosphomannomutase 2 (PMM2) that have been associated with congenital disorder of glycosylation (CDG), and restore the enzyme's activity in a cellular model of the disease [297].

One of the most high-throughput *in vitro* methods used for the discovery of aggregation inhibitors involves screening of phage display libraries. As discussed earlier, this technique involves the construction of a rationally or randomly designed library, which is screened *in vitro* for the identification of binders to an immobilized protein target [259]. Using this technique, Rotter and co-workers constructed two peptide phage display libraries and identified several peptides that could bind to mutant p53 and induce proper folding and restore p53's activity [298]. Furthermore, Kamijo and co-workers constructed a randomized heptapeptide library and identified eight peptide that could bind to A β and inhibit its aggregation, while Linse and co-workers screened two commercially available libraries of recombinant antibodies and identified secondary nucleation inhibitors of A β aggregation.

In conclusion, although *in vitro* assays offer important information on the biophysical and biochemical properties of the test proteins and can result in the identification of well-

characterized inhibitors, they require the availability of purified test protein with reproducible behaviour, which in the case of MisPs can be very challenging. Moreover, *in vitro* assays are usually performed in multi-well formats which increases the time and cost of screening or requires a high degree of automation. Finally, as notably demonstrated by Kiessling and co-workers, the inconsistency of display technologies to identify binders with the desirable biological activity ultimately renders them as costly and inefficient as other conventional *in vitro* techniques [241,268].

3.1.3. Cell based screens for the identification of protein aggregation inhibitors

While a wide range of techniques exist for the investigation of protein misfolding *in vitro*, the complexity of the cellular environment makes the development of mammalian cell-based aggregation screening assays considerably more challenging. Indeed various presumed protein aggregation inhibitors identified through mammalian assays, were later proven to be false positives, due to non-specific biological actions. One notable example is the first p53-reactivating compound, CP-31398, which although initially reported to stabilize p53 and promote p53-mediated apoptosis [299], it was later shown to intercalate with DNA and promote cytotoxicity in a p53-independent manner [300].

Microbial hosts, such as bacteria and yeasts, offer the opportunity of studying protein aggregation in simplified, but still physiologically relevant conditions, as several protein folding and misfolding features can be reliably generated in both prokaryotic and eukaryotic microorganisms. Most importantly, microbial hosts enable the development of high-throughput and affordable screens, which result in biologically active compound hits. Indeed, a wide range of microbial screens have been developed for monitoring and rescuing protein misfolding, which have been thoroughly reviewed by Skretas and co-workers [96]. A notable example is the development of a yeast model of PD, where the increased expression of α -syn resulted in

the formation of α -syn inclusion bodies and severe cytotoxicity [301]. This system was used to screen various small molecule libraries [301-303], as well as an cyclic octapeptide library produced via the aforementioned SICLOPPS technology [280], resulting in the identification of several biologically active compounds against PD.

An engineering approach that has been repeatedly used for the identification of protein aggregation inhibitors is coupling a disease-associated MisP with a reporter protein (RP), in such a way that function of RP will depend on the correct folding of MisP. These fusions can be either end-to-end or insertional, meaning that the MisP is inserted into an internal position of RP. A widely used end-to-end MisP fusion is with a fluorescent protein, such as the green fluorescent protein (GFP). Hecht and co-workers first fused A β 42 to GFP and showed that bacterial overexpression of A β 42-GFP resulted in intracellular insoluble aggregates and lack of green fluorescence [304]. Subsequently, they performed a genetic screen resulting in the identification of A β 42 variants, which contrary to the wild-type protein, exhibited reduced aggregation propensity and increased green fluorescence [304] and also, utilized this methodology to screen two small-molecule libraries, resulting in the identification of various A β aggregation inhibitors, one of which was able to reduce toxicity of A β 42 in an *in vivo* *Drosophila* model of AD [305,306]. Using the same principles, Moffet and co-workers have screened two combinatorial peptide library and identified three peptide inhibitors of A β aggregation with diverse mechanisms [232].

Interestingly, the GFP-based screen has been utilized to monitor the aggregation of other disease-associated MisPs, besides A β 42. Specifically, Moffet and co-workers overexpressed an IAPP-GFP fusion in *E. coli* cells and performed a genetic screen, which resulted in the identification of IAPP variants that, compared to the wild-type protein, exhibited reduced protein aggregation and increased green fluorescence [307]. Furthermore, Fersht and co-workers demonstrated that the thermodynamic stability of various p53 mutants correlates

well with the aggregation propensity of their GFP fusions and their cellular fluorescence, when overexpressed in *E. coli* [308].

Another noteworthy example is the insertional fusion of MisPs to the β -lactamase (*bla*) enzyme, which provides bacterial resistance to the antibiotic ampicillin. Bardwell and co-workers first demonstrated that inserting an unstable protein between residues 196 and 197 of *bla* results in increased protein degradation and ampicillin susceptibility [309]. On the contrary, when the inserted protein is correctly folded the two *bla* domains re-associate and confer antibiotic resistance. Using this tripartite fusion system, Radford and co-workers identified one inhibitors of IAPP and demonstrated that this system can be applied for the identification of aggregation inhibitors against various MisPs associated with different PMDs [310]. Importantly, by coupling this system with electrospray ionization- ion mobility spectrometry-mass spectrometry (ESI-IMS/MS) they were able to validate compound-protein binding and determine the different aggregation species that were present in the sample, thus enabling the selection of hits with the desired mechanism of action [310].

An alternative approach for monitoring protein aggregation *in vivo* is by utilizing protein dyes with the ability to specifically bind to hydrophobic surfaces of amyloid or amyloid-like aggregates and exhibit increased fluorescence. In this way, by measuring cell fluorescence using fluorimetry or flow cytometry, one can monitor the aggregate formation inside living cells without the need of purified protein preparations [311]. Thioflavin S (ThS) and the recently developed fluorescent dye Proteostat, have both been successfully used for monitoring amyloid formation in microorganisms and screening for aggregation inhibitors [311-316].

3.2. Development of a generalized genetic screen for monitoring protein aggregation *in vivo*

As protein misfolding and aggregation are common defining features of all PMDs, we attempted to develop a genetic screen with potentially wide applicability, that would facilitate monitoring of the misfolding and aggregation of different MisPs associated with PMDs. As the aforementioned GFP assay has been successfully employed for monitoring the aggregation of A β 42, IAPP and certain p53 mutants, we reasoned that by generalizing this system we could provide an important tool for identifying protein aggregation inhibitors against multiple PMDs that remain incurable.

Specifically, due to the aggregation propensity of disease-associated MisPs, we hypothesized that overexpression of MisP-GFP fusions in *E. coli* cells would result in the accumulation of insoluble inclusion bodies exhibiting decreased fluorescence. Contrary, conditions that rescue MisP misfolding and/or inhibit its aggregation would result in the formation of soluble MisP-GFP fusions exhibiting enhanced fluorescence (Figure 3.1).

In order to test this hypothesis we employed the MisP-GFP assay for monitoring the misfolding and aggregation of p53, SOD1 and polyQ-HTT_{ex1} variants, which have been associated with cancer, ALS and HD respectively.

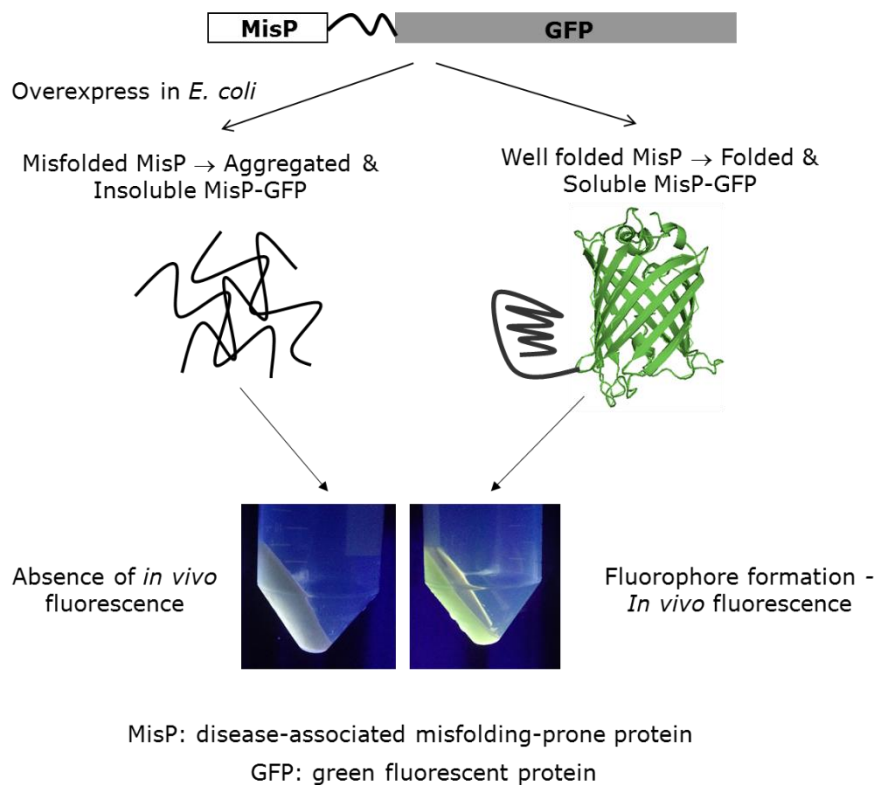


Figure 3.1. Schematic of the MisP-GFP genetic system for monitoring MisP folding and aggregation.

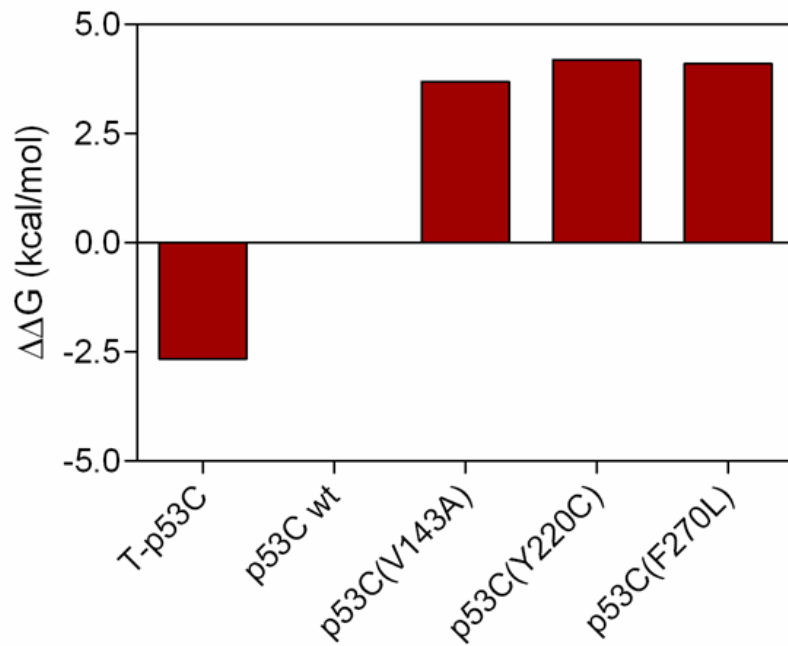
3.2.1. Monitoring the aggregation of p53 oncogenic variants

Structural mutations in the core domain of p53 - that destabilize and subsequently inactivate the protein - appear in ~15-20% of human cancer cases and therefore constitute important targets for cancer therapy [317]. For this reason, we tested whether the MisP-GFP assay could be used for monitoring the aggregation of three p53 core domain structural mutants, namely, the substitution of valine at position 143 by alanine (V143A), the substitution of tyrosine at position 220 by cysteine (Y220C) and the substitution of phenylalanine at position 270 by leucine (F270L).

We started by introducing these mutations into the core domain of wild-type p53 (p53C wt) and a highly stabilized variant of p53 (T-p53C), which has wild-type-like properties and, as demonstrated by X-ray studies, almost identical structure, apart from the four point

mutations (M133L, V203A, N239Y and N268D) that confer additional stability and make the protein easier to work with [318,319]. Then, we produced recombinant fusions of the constructed p53 variants with GFP and monitored their bacterial fluorescence and protein aggregation. Interestingly, we found that the fluorescence intensity of the constructed p53C-GFP fusions correlated very well with their thermodynamic stability and the accumulating amounts of soluble protein (Figure 3.3 and Figure 3.3). These results indicate that the MisP-GFP assay is able to reliably monitor p53C misfolding and aggregation and therefore could be utilized for the discovery of potential rescuers of destabilized forms of p53.

A



B

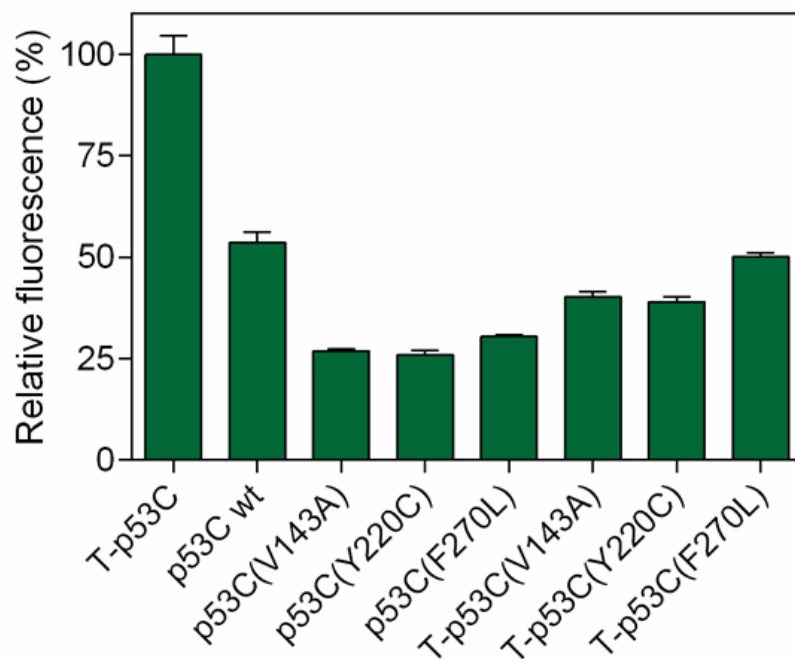


Figure 3.2. Monitoring the aggregation of p53 oncogenic variants. (A) Thermodynamic stability of the p53 core domain variants. $\Delta\Delta G$ represents the change in the free energy of urea-induced unfolding caused by mutations in wild-type p53 ($\Delta\Delta G = \Delta G_{wt} - \Delta G_{mut}$) as determined by Fersht and co-workers. [121,320]. (B) Relative fluorescence of *E. coli* BL21(DE3) cells overexpressing p53C-GFP fusions from pETp53C-GFP. Mean values \pm s.e.m. are reported.

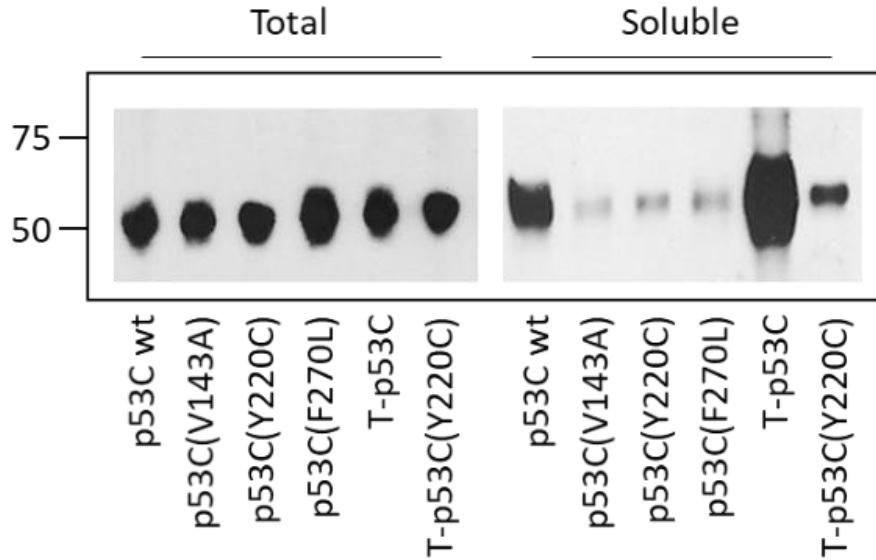


Figure 3.3. Solubility analysis of *E. coli* BL21(DE3) cells overexpressing p53C-GFP fusions. Total (left) and soluble (right) lysates of cells overexpressing different p53C-GFP fusions produced as in Figure 3.2B were analysed by SDS-PAGE and visualized by western blotting using the anti-GFP antibody.

3.2.2. Monitoring the aggregation of SOD1 variants

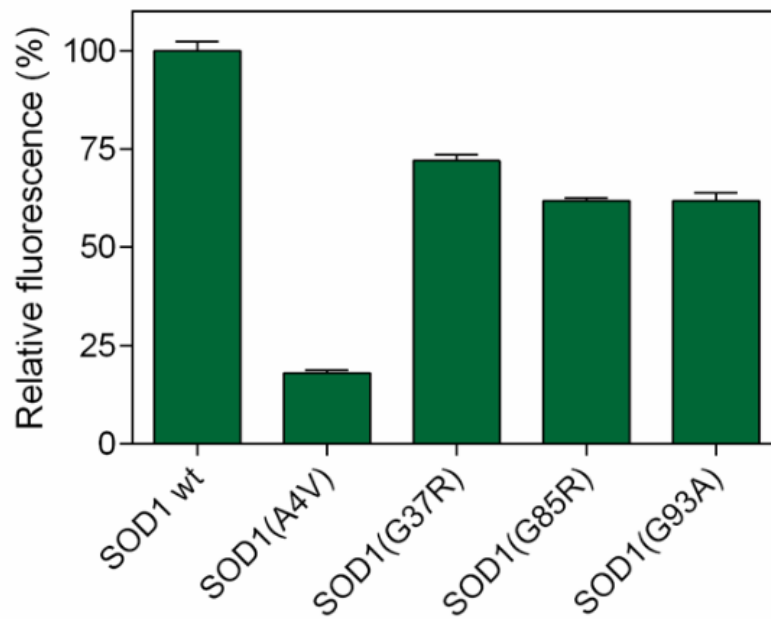
As mutations in the *SOD1* gene that cause thermal destabilization and/or aggregation of SOD1 have been widely associated with familial ALS (fALS) [187,321,322], we next tested whether we could detect mutation-induced folding changes to SOD1 using the MisP-GFP assay. For this reason, we produced end-to-end fusions of SOD1 variants with GFP in *E. coli* cells in order to and monitored their fluorescence and the aggregation status of the overexpressed fusions. We tested four SOD1 variants, namely the substitution of alanine at position 4 by valine (A4V), of glycine at position 37 by arginine (G37R), of glycine at position 85 by arginine (G85R) and of glycine at position 93 by alanine (G93A). These were selected from a set of over 150 fALS-associated SOD1 mutations as they exhibit different aggregation propensities and disease phenotypes (Table 3.1) [323,324].

Table 3.1. Statistics for fALS patients with the herein studied SOD1 mutations. Adapted from [323].

Mutation	Mean age at onset (years)	Mean life expectancy (years)
SOD1(A4V)	47.0 ± 13.7	1.5
SOD1(G37R)	29.3 ± 1.2	18
SOD1(G85R)	55.5 ± 12.6	6.5
SOD1(G93A)	43.1 ± 16.6	2.5

Indeed, using the MisP-GFP assay, we were able to follow SOD1 aggregation efficiently as all variants exhibited decreased levels of GFP fluorescence and solubility, compared to the generally non-pathogenic wild-type SOD1 (Figure 3.4). Interestingly, SOD1(A4V), which has been associated with the most aggressive type of fALS (Table 3.1) exhibited the most pronounced decrease in GFP fluorescence, in relation to the accumulating amount of soluble protein (Figure 3.4). Notably, when the SOD1 variants were produced in unfused, GFP-free form, their soluble levels were also well-correlated with their aggregation propensity as all pathogenic variants exhibited decreased levels of solubility compared to the wild-type protein (Figure 3.5).

A



B

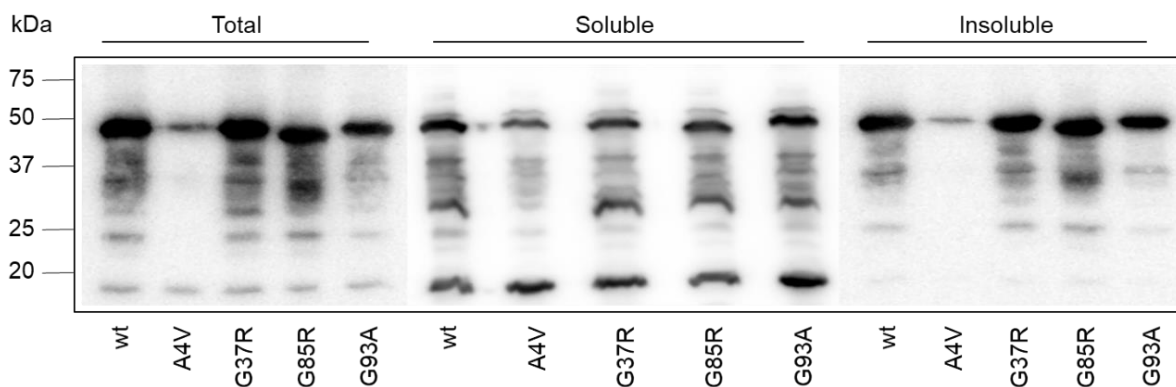


Figure 3.4. Monitoring the folding and misfolding of SOD1 variants using the MisP-GFP assay. (A) Relative fluorescence of *E. coli* BL21(DE3) cells overexpressing SOD1-GFP fusions from pETSOD1-GFP. Mean values \pm s.e.m. are reported from one experiment performed in triplicates. (B) Solubility analysis as in (A), using SDS-PAGE/western blotting and probing with the anti-GFP antibody.

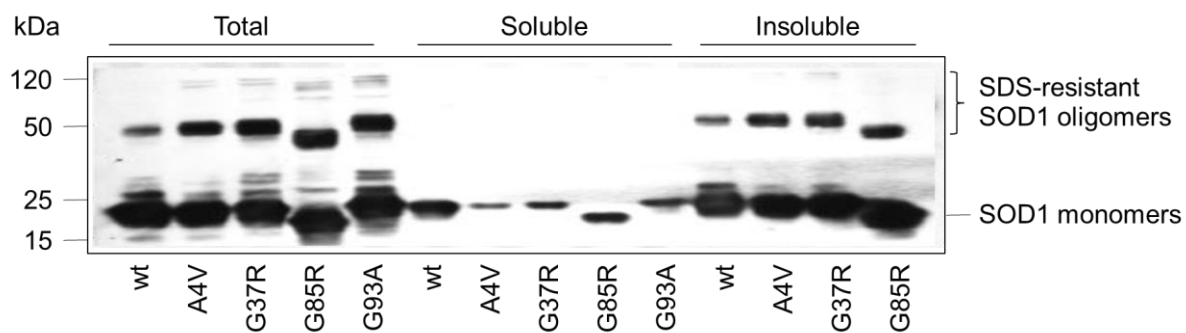


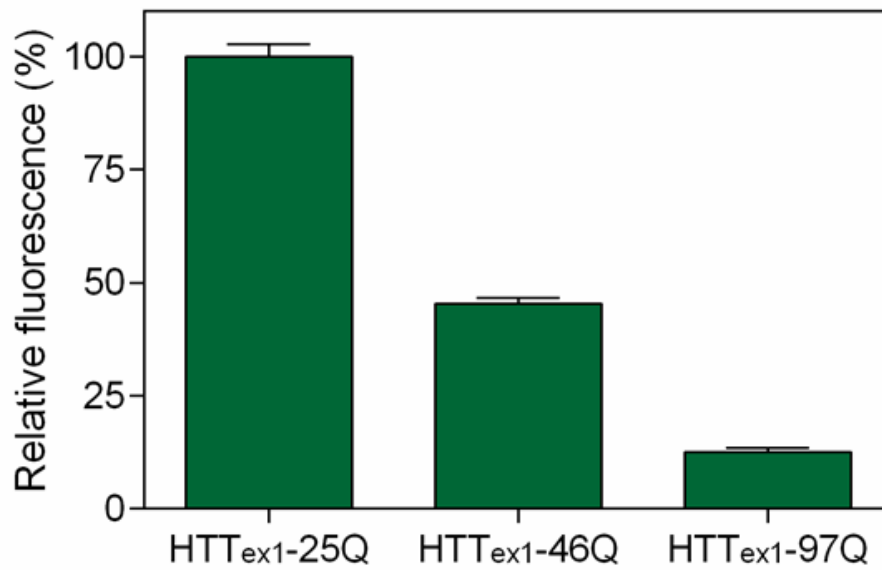
Figure 3.5. Solubility analysis of SOD1 variants overexpressed in *E. coli* Origami2(DE3). Production of the GFP-free SOD1 variants was performed from the corresponding pETSOD1 vectors. Total (left), soluble (middle) and insoluble (right) lysates were analysed by SDS-PAGE and visualized by western blotting using the anti-His antibody. Experiments were performed by Stefania Panoutsou at the NHRF.

3.2.3. Monitoring the aggregation of polyQ-HTT_{ex1} variants

HD is an inherited neurodegenerative disease caused by polyglutaminated (polyQ) expansions at the N-terminal region of huntingtin (HTT), a highly polymorphic protein normally containing less than 35 glutamine residues [325]. These polyQ N-terminal regions, which are encoded within the first exon of the corresponding gene (HTT_{ex1}), have been found to assemble into neurotoxic aggregates, leading to neurodegeneration [326]. Importantly, the severity of the HD phenotype is inversely correlated with the length of the polyQ expansion, as individuals with longer polyQ expansions develop the disease at younger ages [327].

In order to test whether we could monitor the aggregation of polyQ variants with different lengths using the MisP-GFP assay, we produced end-to-end fusions of polyQ-HTT_{ex1} with GFP in *E. coli* cells and monitored their fluorescence and the aggregation status of the overexpressed proteins. We assessed three polyQ-HTT_{ex1} variants, one containing a normal 25Q expansion and two expansions resulting in HD pathogenesis, namely 46Q and 97Q [328]. Interestingly, the levels of GFP fluorescence of *E. coli* cells and the amounts of soluble polyQ-GFP fusions produced were inversely proportional to the polyQ length (Figure 3.6), suggesting that this system is capable of monitoring the aggregation of polyQ-HTT_{ex1} variants and could be utilized for discovery of potential polyQ-HTT_{ex1}-aggregation inhibitors.

A



B

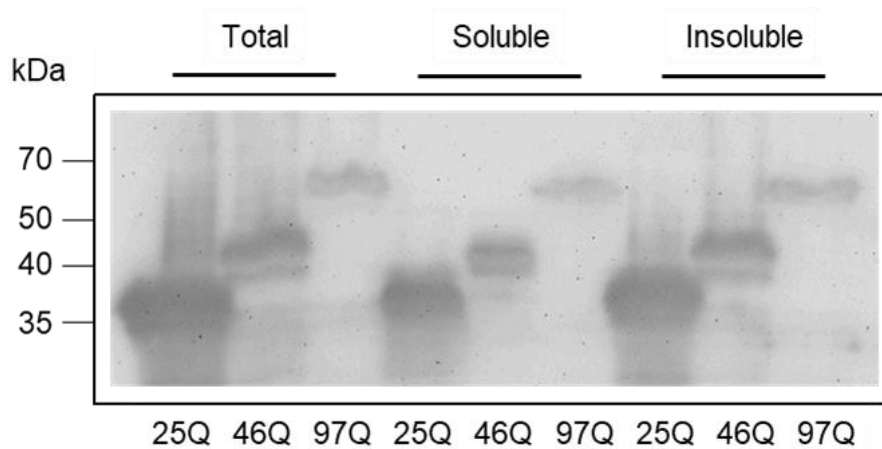


Figure 3.6. Monitoring the aggregation of polyQ-HTT_{ex1} variants using the MisP-GFP assay. (A) Relative fluorescence of *E. coli* BL21(DE3) cells overexpressing polyQ-HTT_{ex1}-GFP fusions from pETHHTT_{ex1}-GFP. Mean values \pm s.e.m. are reported from one experiment performed in triplicates. (B) Solubility analysis of polyQ-HTT_{ex1}-GFP variants overexpressed as in (A) using SDS-PAGE/western blotting and probing with the anti-GFP antibody.

3.2.4. Optimization of the MisP-GFP assay

As the expression of MisP-GFP has been associated with cell toxicity [329], we opted to determine the optimal conditions for the expression of different MisP-GFP constructs. We started by testing a range of different expression vectors, inducer concentrations, incubation temperatures and incubation periods for the production of fusions of GFP with p53C or SOD1 variants (Figure 3.7 and Figure 3.8).

In the case of p53C, these experiments revealed that conditions that enable rapid protein production, such as use of T7 promoter, incubation at 37 °C and increased inducer concentration, result in increased fluorescence intensity differences between the well-folded protein and the destabilized variants (Figure 3.7). Contrary, in the case of SOD1, conditions that detain protein expression, such as decreased inducer concentration and incubation at lower temperatures, result in enhanced differences in the fluorescence intensity of wild-type SOD1 and the disease-associated variants (Figure 3.8). These results indicate that, while the MisP-GFP assay can be utilized against different protein targets, the MisP-GFP expression conditions should be independently optimized whenever a new target is pursued.

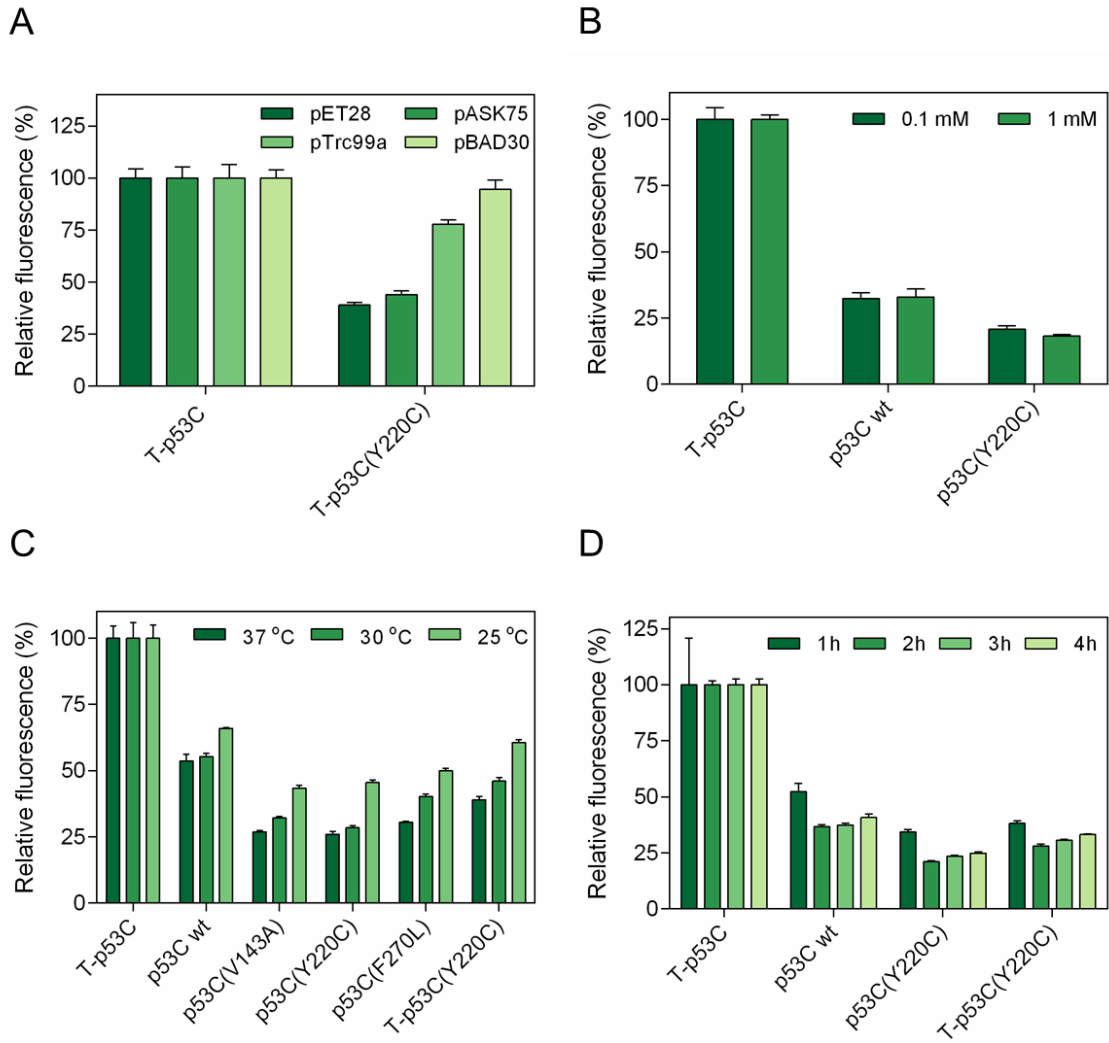


Figure 3.7. Effect of different optimization parameters on the bacterial fluorescence of *E. coli* BL21(DE3) cells producing p53C-GFP fusions. (A) Investigation of different expression vectors, (B) IPTG concentrations, (C) incubation temperatures and (D) incubation periods. In all panels, over-expression was performed using the pET28 vector, unless otherwise stated and the fluorescence of the bacterial population producing T-p53C was arbitrarily set to 100. Mean values \pm s.e.m. are presented. Each experiment was performed in triplicates.

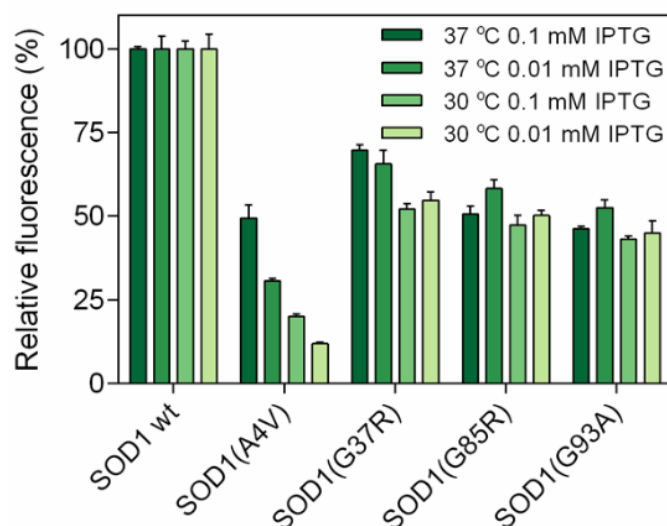


Figure 3.8. Effect of varying incubation temperatures and IPTG concentrations on the bacterial fluorescence of *E. coli* Origami2(DE3) cells overexpressing SOD1-GFP fusions from the pET28 vector. The fluorescence of the bacterial population producing SOD1 wt was arbitrarily set to 100. Mean values \pm s.e.m. are presented. Each experiment was performed in triplicates.

In parallel, we tested whether the GFP reporter is the most ideal fluorescence partner for monitoring protein misfolding and aggregation. For this, we produced recombinant fusions of p53C variants with the blue fluorescent protein (BFP) or the red fluorescent protein (RFP), and measured their bacterial fluorescence and protein aggregation (Figure 3.9). Interestingly, GFP and BFP fusions with T-p53C and T-p53C(Y220C) exhibited similar fluorescence intensity differences, indicating that both fluorescent proteins can be appropriately employed for monitoring protein misfolding and aggregation (Figure 3.9). Contrary, the differences in T-p53C-RFP and T-p53C(Y220C)-RFP fluorescence intensities were decreased compared to the GFP and BFP fusions, indicating that this fluorescent partner is less efficient for this purpose (Figure 3.9).

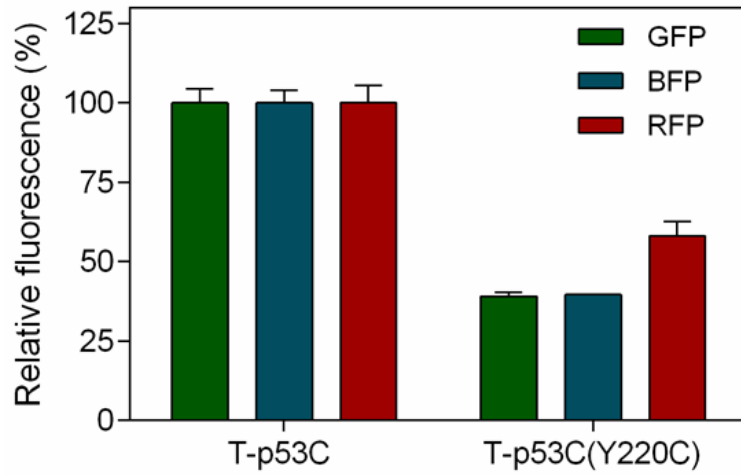


Figure 3.9. Effect of different fluorescent protein partners for monitoring protein aggregation in *E. coli* Tuner(DE3) cells. Protein production was performed using the pET28 vector and the bacterial population producing T-p53C was arbitrarily set to 100. Mean values \pm s.e.m. are presented. Each experiment was performed in triplicates.

3.3. Discussion

In conclusion, in this chapter we describe the development of a generalized genetic assay for monitoring the misfolding and aggregation of MisPs associated with different PMDs. As a large number of PMDs still remain incurable, the development of such assays constitute powerful tools for addressing this unmet medical need and facilitate the identification of rescuers of pathogenic protein misfolding and/or aggregation.

The herein described genetic assay is based on the recombinant production in *E. coli* cells of end-to-end fusions of disease-associated MisPs with the fluorescent protein GFP. In this manner, misfolding and/or aggregation of the MisP directly affects the folding and fluorescence of GFP, and therefore, by measuring the bacterial fluorescence one can easily monitor the folding status of the investigated MisP.

We illustrate the generality of this assay by targeting three unrelated proteins, p53, SOD1 and HTT, whose variants have been associated with cancer, ALS and HD respectively. In all cases, we found that the bacterial fluorescence of *E. coli* cells producing different MisP-GFP fusions as well as their accumulating soluble protein levels, strongly correlate with the aggregation propensity of the investigated MisPs and their associated pathogenicity.

Overall, we demonstrate that the herein described genetic assay can be utilized for distinguishing between pathogenic and non-pathogenic MisP variants associated with numerous PMDs, and, importantly, for the discovery of potential rescuers of pathogenic protein misfolding and/or aggregation.

Chapter 4 - Development of an integrated bacterial system for the discovery of potential PMD therapeutics

In order to accelerate the early drug discovery process and address the imperative medical need for potential therapeutics against PMDs, ultrahigh-throughput methodologies that allow deeper investigation of chemical space are urgently required. Towards this, we developed a novel ultrahigh-throughput biotechnology system, where *E. coli* cells are genetically engineered in order to perform two simultaneous tasks: (i) produce expanded molecular libraries, such as the one presented in Chapter 2 comprising more than 200 million cyclic peptides and (ii) enable their direct functional screening using the MisP-GFP genetic assay described in Chapter 3.

This approach enables the identification of cyclic peptides with the ability to bind to the MisP of interest and rescue its misfolding and/or aggregation, by monitoring their effect on the fluorescence of the recombinant MisP-GFP fusions. Specifically, since the bacterial fluorescence of *E. coli* cells producing MisP-GFP fusions is strongly correlated to the aggregation propensity of the MisPs, cells producing such cyclic peptides will exhibit significantly increased levels of GFP fluorescence. Importantly, as both cyclic peptide production and their functional screening are performed inside living cells, bacterial clones producing such cyclic peptides can be isolated in an ultrahigh-throughput manner using fluorescence-activated cell sorting (FACS), therefore enabling the rapid and facile identification of bioactive hits (Figure 4.1).

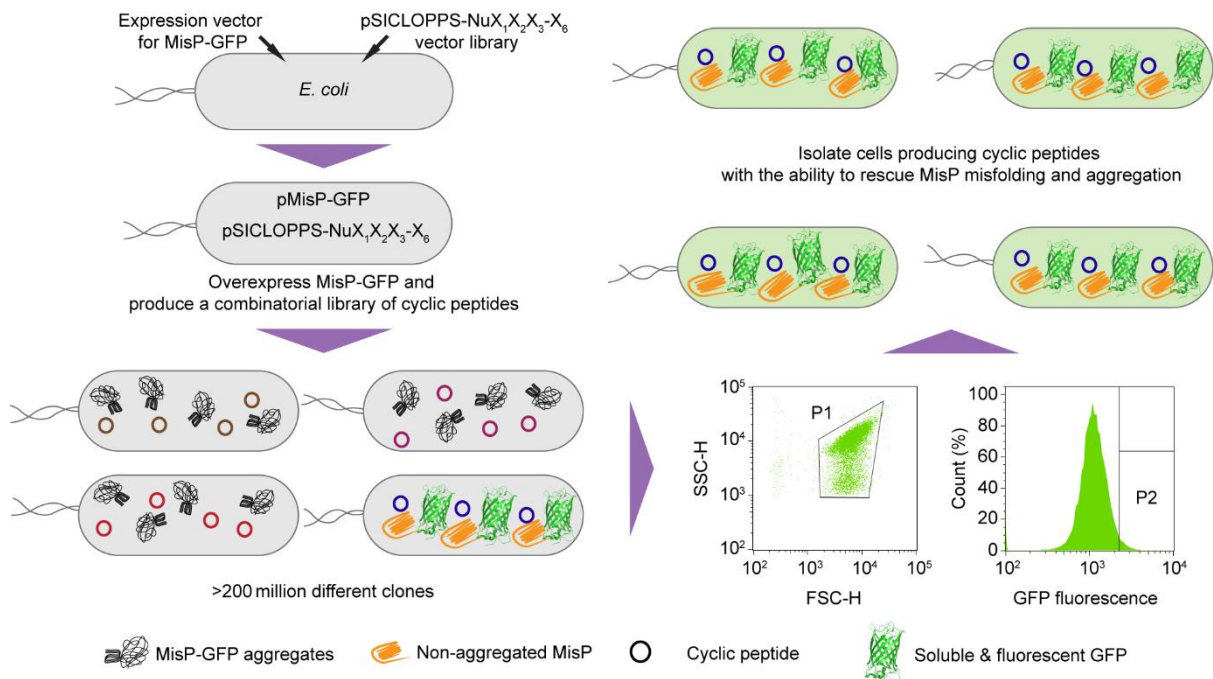


Figure 4.1. Schematic of the utilized bacterial platform for discovering inhibitors of protein aggregation.

pMisP-GFP: plasmid encoding a misfolded protein-GFP fusion; pSICLOPPS-NuX₁X₂X₃-X₆: vector library encoding the combinatorial tetra-, penta-, hexa- or heptapeptide library cyclo-NuX₁X₂X₃-X₆; Nu: Cys, Ser, or Thr; X: any of the 20 natural amino acids; FSC-H: forward scatter; SSC-H: side scatter; P: sorting gate.

In this chapter we present the employment of this ultrahigh-throughput system for identifying folding rescuers and/or aggregation inhibitors against four targets: p53C(Y220C), A β 42, SOD1(A4V) and HTT_{ex1-97Q}, which have been associated with cancer, AD, ALS and HD respectively.

4.1. Targeting cancer - Identification of p53C(Y220C) folding rescuers using the generated ultrahigh-throughput biotechnology platform

For the identification of cyclic peptides that would stabilize the p53C(Y220C) variant we performed three attempts. In the first attempt we utilized a two-plasmid system comprising a pET28 expression vector producing the p53C(Y220C)-GFP fusion, i.e. pETp53C(Y220C)-GFP, and the pSICLOPPS vector library described in Chapter 2, that produced the cyclic tetra-, penta- and hexapeptides, i.e. pSICLOPPS-NuX₁X₂X₃-X₆ (Figure 4.2). Protein expression was performed at 37 °C using 0.1 mM isopropyl-β-D-thiogalactoside (IPTG) and 0.002 % arabinose as described in the Materials and Methods section. Cells were gated on a side-scatter (SSC-H) versus forward-scatter (FSC-H) plot in order to eliminate non-cellular events, and were subjected to FACS sorting for the isolation of the bacterial population exhibiting the top ~2% fluorescence. Under these conditions and after two rounds of sorting, the selected bacterial population exhibited a very small fluorescence increase compared to the initial population (~10%), while no further fluorescence increase was observed after additional rounds (Figure 4.3).

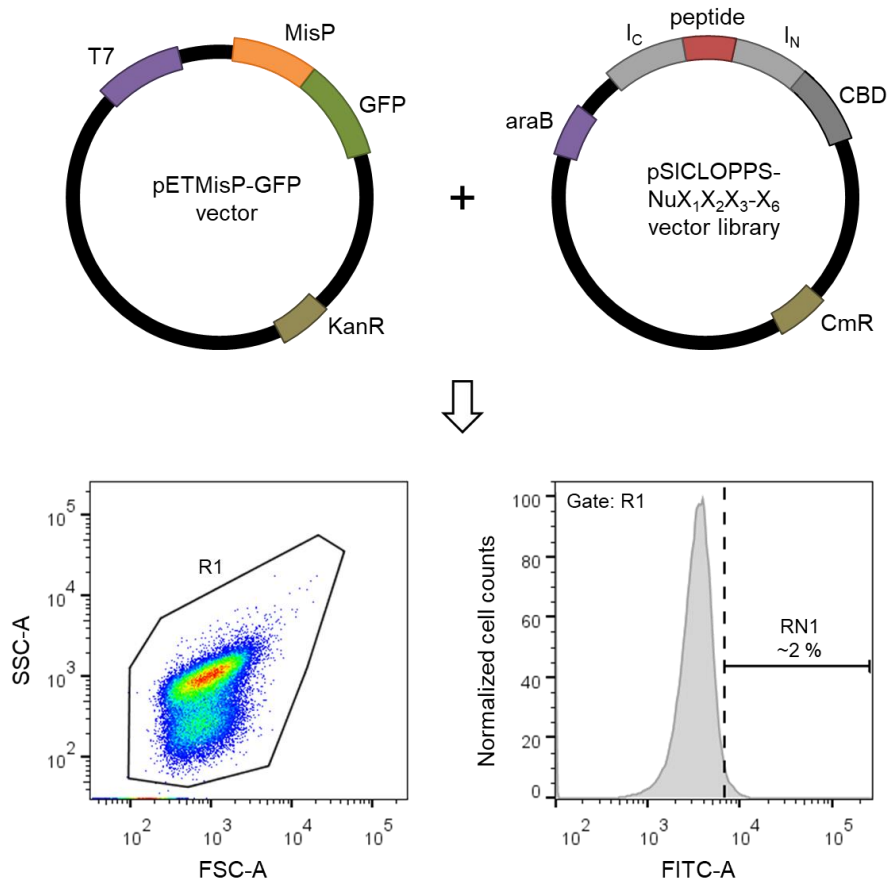


Figure 4.2. Schematic of the utilized two-plasmid system and the gating strategy for FACS sorting for the identification of rescuers of MisP-GFP misfolding and aggregation. pETMisP-GFP: plasmids encoding a MisP-GFP fusion; pSICLOPPS-NuX₁X₂X₃-X₆: vector library encoding the combinatorial tetra-, penta-, hexa- or heptapeptide library cyclo-NuX₁X₂X₃-X₆; Nu: Cys, Ser, or Thr; X: any of the 20 natural amino acids; FSC-A: forward scatter; SSC-A: side scatter; R1 and RN1: sorting gates.

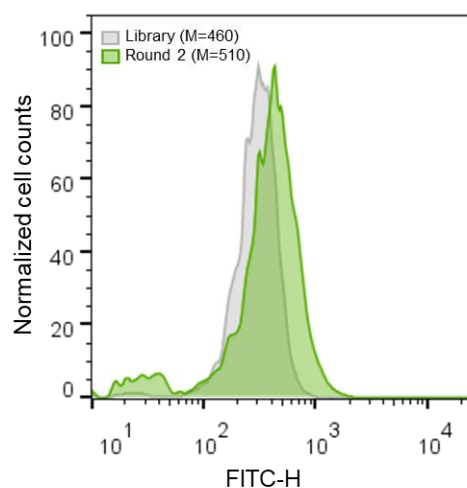


Figure 4.3. Histograms of BL21(DE3) cells co-expressing p53C(Y220C) and the initial cyclo-NuX₁X₂X₃-X₅ oligopeptide library (grey) or the enriched library after the second round of sorting (green). M=mean GFP fluorescence in arbitrary units. FACS sorting was performed at the University of Texas at Austin.

We next isolated 140 individual cyclic peptide sequences from the sorted population and evaluated their ability to enhance the fluorescence levels of T-p53C(Y220C)-GFP. We found that 21 of them exhibited reproducibly a significant fluorescence increase when T-p53C(Y220C)-GFP was produced from the pETT-p53C(Y220C)-GFP vector, in which case protein overexpression is controlled by the T7 promoter (Figure 4.4A). However, these effects were found to be promoter-specific, as the selected peptide clones were unable to enhance the fluorescence of T-p53C(Y220C)-GFP fusions when produced from the pASKT-p53C(Y220C)-GFP vector, in which case protein production is controlled by the tetracycline (Tet) promoter (Figure 4.4B). Furthermore, the observed phenotypes were found to be p53-non-specific as the selected peptide clones were also able to increase the fluorescence levels of A β 42-GFP fusions when produced from the pETA β 42-GFP vector (Figure 4.4A). In total, these results indicate that the selected cyclic peptide clones are probably false-positive hits and function in a p53-non-specific manner by probably modulating the protein expression by the T7 RNA polymerase/promoter system.

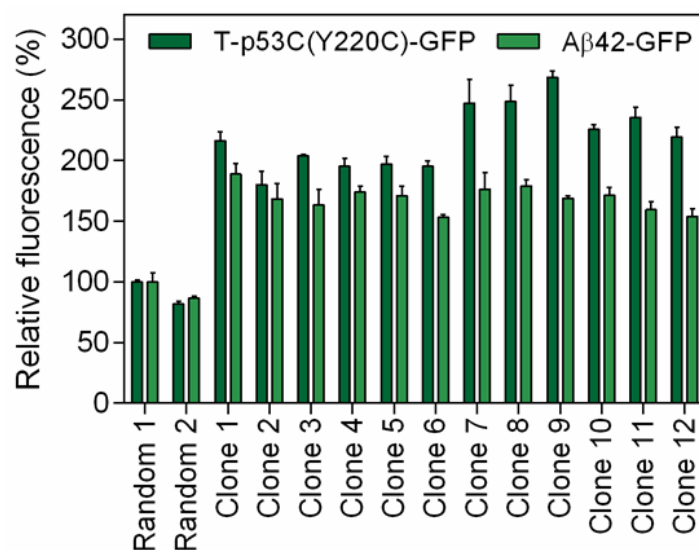
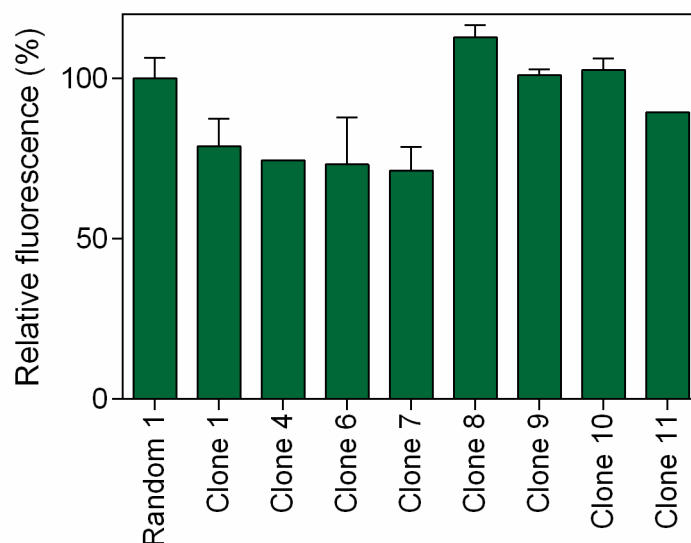
A**B**

Figure 4.4. First attempt of FACS sorting for the identification of p53C(Y220C) folding rescuers. (A) Fluorescence of *E. coli* BL21(DE3) cells co-expressing T-p53C(Y220C)-GFP or Aβ42-GFP in the presence of 12 individually selected cyclic peptide sequences from the second round of FACS sorting shown in (A). MisP-GFP production was performed from the pET28 vector. (B) As in (A) but using the pASK75 vector for T-p53C(Y220C)-GFP production. In both panels, the fluorescence of bacterial cells producing T-p53C(Y220C)-GFP or Aβ42-GFP together with a random cyclic peptide from the initial unsorted cyclo-NuX₁X₂X₃-X₅ library (Random 1) was arbitrarily set to 100. Mean ± sem are presented of one experiment performed in triplicates.

In our second attempt, we modified our system in order to eliminate the possibility of isolating false-positive cyclic peptides with a p53-non-specific effect. For this reason, we started by overexpressing the T-p53C(Y220C)-GFP fusion from the pETT-p53C(Y220C)-GFP vector and isolated the bacterial population exhibiting enhanced fluorescence via FACS (Figure 4.5, top). Then, after two rounds of sorting we isolated the so-far enriched population and re-transformed the selected pSICLOPPS vectors into *E. coli* cells that carry the pASKT-p53C(Y220C)-GFP vector and thus produce the same GFP fusion but from the pASK75 vector (Figure 4.5, bottom). This procedure could enable the isolation of a bacterial population that exhibits increased levels of fluorescence irrespectively of the promoter utilized for overexpression.

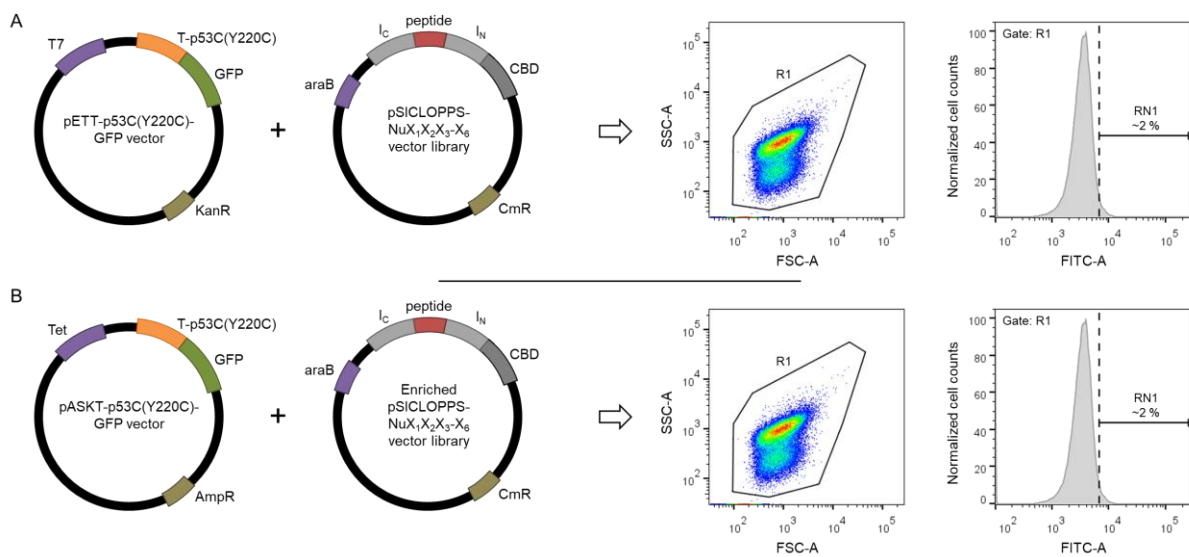


Figure 4.5. Schematic of the utilized two-plasmid system and the gating strategy utilized in our second attempt of FACS sorting for the identification of T-p53C(Y220C) folding rescuers. Initially, overexpression of T-p53C(Y220C)-GFP was performed under the control of the T7 promoter (A) and then under the control of the Tet promoter (B). pETT-p53C(Y220C)-GFP and pASKT-p53C(Y220C)-GFP: plasmids encoding the T-p53C(Y220C)-GFP fusion under the control of the T7 or Tet promoter respectively; pSICLOPPS-NuX₁X₂X₃-X₆: vector library encoding the combinatorial tetra-, penta-, hexa- or heptapeptide library cyclo-NuX₁X₂X₃-X₆; Nu: Cys, Ser, or Thr; X: any of the 20 natural amino acids; FSC-A: forward scatter; SSC-A: side scatter; R1 and RN1: sorting gates.

After a total of five rounds of sorting we observed an almost three-fold increase of the bacterial fluorescence, which was unaltered after additional rounds (Figure 4.6). We isolated 530 individual bacterial clones from the enriched bacterial population and 84 of them exhibited a significant increase in T-p53C(Y220C)-GFP fluorescence. Importantly, most of the selected bacterial clones exhibited a p53-specific effect as indicated by assessing their effect on the levels of A β 42-GFP fluorescence (Figure 4.7A).

Next, we opted to determine whether the observed phenotype was dependent on the ability of the Ssp DnaE intein to splice and form a cyclic peptide. For this reason, we introduced the double amino acid substitution H24L/F26A into the C-terminal domain of the Ssp DnaE intein, which is known to abolish asparagine cyclization at the I_C/extein junction and prevent extein splicing and peptide cyclization [280]. We found one clone (clone 2), which was able to increase the fluorescence of T-p53C(Y220C)-GFP in a splicing-dependent manner, indicating that intein processing and the possible production of the cyclic peptide is necessary for the observed phenotype (Figure 4.7B). However, the above mentioned results were not always reproducible. Importantly, sequencing of the selected clone revealed that it encoded the CGGTGR peptide sequence, which was the most abundant sequence of the unsorted pSICLOPPS vector library, appearing ~700 times more frequently than its succeeding sequence (Table 2.5). These observations suggested that the selected clone could, in fact, be a false positive and that FACS sorting was not successful in providing an enriched sub-library, with enhanced bacterial fluorescence.

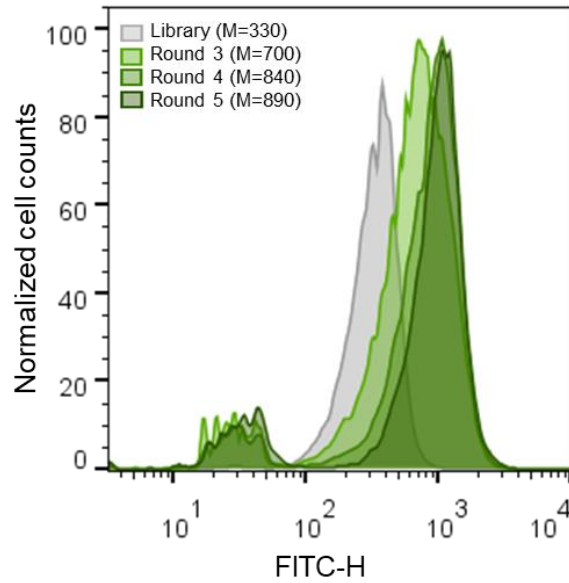


Figure 4.6. Histograms of BL21(DE3) cells co-expressing T-p53C(Y220C) and the initial cyclo-NuX₁X₂X₃-X₅ oligopeptide library or the enriched libraries after three, four or five rounds of sorting. M=mean GFP fluorescence in arbitrary units. FACS sorting was performed at the University of Texas at Austin.

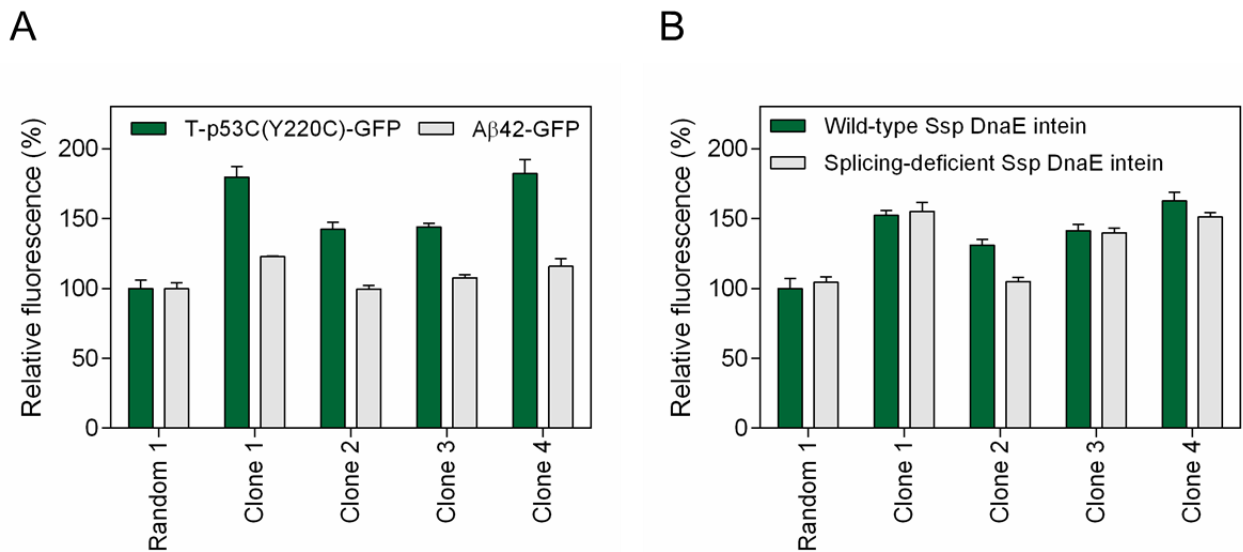


Figure 4.7. Second attempt of FACS sorting for the identification of p53C(Y220C) folding rescuers (A) Relative fluorescence of *E. coli* BL21(DE3) overexpressing T-p53C(Y220C)-GFP (green bars) or Aβ42-GFP (white bars) and four individually selected cyclic peptide clones from the fifth round of FACS sorting shown in (Figure 4.6). (B) Relative fluorescence of *E. coli* BL21(DE3) cells overexpressing T-p53C(Y220C)-GFP and the four selected cyclic peptide clones as in (A) and utilizing either a wild-type split Ssp DnaE intein (green bars) or the splicing-deficient variant H24L/F26A32 (white bars). The fluorescence of bacterial cells producing T-p53C(Y220C)-GFP or Aβ42-GFP together with a random cyclic peptide (Random 1) was arbitrarily set to 100. Mean ± sem are presented of one experiment performed in triplicates.

As in the first two attempts the isolated bacterial population mostly exhibited a p53-non-specific fluorescence increase, we decided to further modify our system in order to select the bacterial population that exhibits a T-p53C(Y220C)-GFP fluorescence increase but at the same time has no effect on the fluorescence of A β 42-GFP. For this reason, we chose to insert the T-p53C(Y220C)-BFP construct into the pCDF-1b vector and generate a three-plasmid system comprising of pCDFT-p53C(Y220C)-BFP, pETA β 42-GFP and the pSICLOPPS-NuX₁X₂X₃-X₆ vector libraries described in Chapter 2, that produced the cyclic tetra-, penta-, hexa- and heptapeptide libraries (Figure 4.9).

We decided to utilize the pCDF-1b vector as: i) the difference between the levels of fluorescence of T-p53C-BFP and T-p53C(Y220C)-BFP were comparable to the ones observed when the pET28 vector was used (Figure 3.7 and Figure 4.8), ii) pCDF is compatible with the pET28 and pSICLOPPS vectors used so far and iii) it contains the same promoter as pETA β 42-GFP, and therefore by selecting a bacterial population with increased BFP fluorescence but unaffected GFP fluorescence, the possibility of promoter-specific false-positives is eliminated.

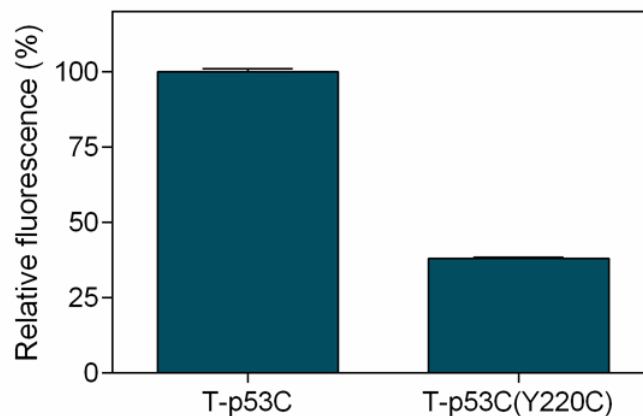


Figure 4.8. Relative fluorescence of *E. coli* Tuner(DE3) cells overexpressing T-p53C-BFP and T-p53(Y220C)-BFP from the pCDFT-p53C-BFP and pCDFT-p53(Y220C)-BFP vectors. The fluorescence of bacterial cells producing T-p53C-BFP was arbitrarily set to 100. Mean \pm sem are presented of one experiment performed in triplicates.

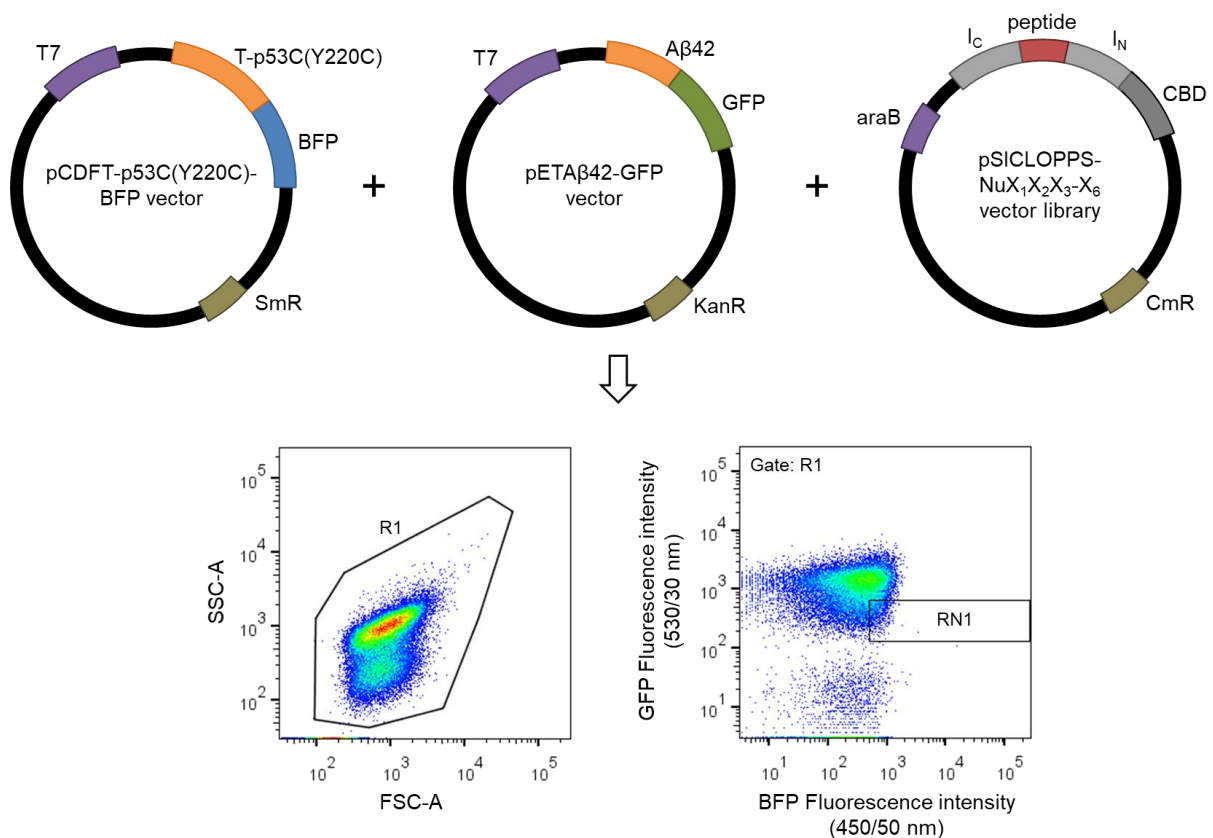


Figure 4.9. Schematic of the utilized three-plasmid system and the gating strategy for FACS sorting for the identification of p53C(Y220C) folding rescuers. pCDFT-p53C(Y220C)-BFP: plasmid encoding the T-p53C(Y220C)-BFP fusion; pETA β 42-GFP: plasmid encoding the A β 42-GFP fusion; pSICLOPPS-NuX₁X₂X₃-X₆: vector library encoding the combinatorial tetra-, penta-, hexa- or heptapeptide library cyclo-NuX₁X₂X₃-X₆; Nu: Cys, Ser, or Thr; X: any of the 20 natural amino acids; FSC-A: forward scatter; SSC-A: side scatter; R1 and RN1: sorting gates.

Using this three-plasmid system, we simultaneously overexpressed the T-p53C(Y220C)-BFP and A β 42-GFP fusions using 0.1 mM IPTG, as well as the cyclo-NuX₁X₂X₃-X₆ oligopeptide libraries using 0.005 % arabinose as described in the Materials and Methods section. For FACS sorting, we first gated cells based on their SSC-H and FSC-H properties, and then we isolated the bacterial population exhibiting the top ~2% of BFP fluorescence, as well as an unaltered GFP fluorescence (Figure 4.9). After five rounds of sorting for the tetra-, penta- and heptapeptide libraries as well as four rounds of sorting for the hexapeptide library we observed a significant increase of the bacterial BFP fluorescence, which

was unaltered after additional rounds (Figure 4.10). Importantly, this bacterial population did not exhibit an increase of the GFP fluorescence, which corresponds to the A β 42-GFP fusion.

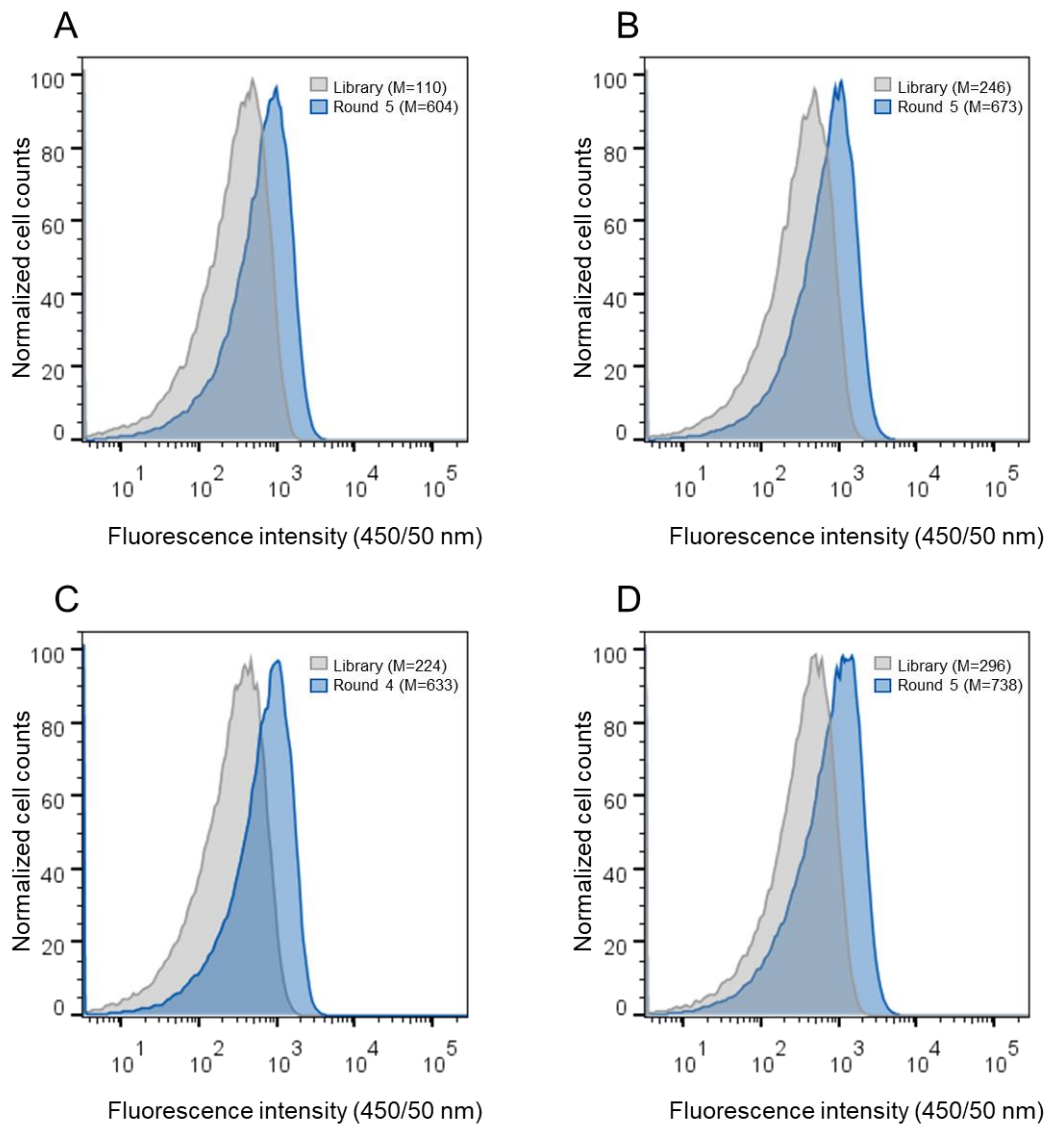


Figure 4.10. FACS sorting of *E. coli* Tuner(DE3) cells overexpressing T-p53(Y220C)-BFP, A β 42-GFP and the cyclo-NuX₁X₂X₃-X₆ libraries. Histograms of cells co-expressing T-p53(Y220C)-BFP, A β 42-GFP and (A) the cyclo-NuX₁X₂X₃ tetrapeptide library, (B) the cyclo-NuX₁X₂X₃X₄ pentapeptide library, (C) the cyclo-NuX₁X₂X₃X₄X₅ hexapeptide library and (D) the cyclo-NuX₁X₂X₃X₄X₅X₆ heptapeptide library. In all panels, the initial library is shown in grey and the enriched sorted library in blue. M=mean BFP fluorescence in arbitrary units. FACS sorting was performed at the University of Texas at Austin.

We isolated 200 individual bacterial clones from each enriched bacterial population and identified six bacterial clones that were able to increase the fluorescence and solubility of T-p53C(Y220C)-GFP in a p53-specific manner, that was also dependent on peptide cyclization as illustrated using the splicing deficient Ssp DnaE intein [280] (Figure 4.11 and Figure 4.12). DNA sequencing of the peptide-encoding region of the selected clones revealed the presence of four tetrapeptide sequences, termed p53C4-4, p53C4-16, p53C4-19 and p53C4-21 (p53-targeting cyclic 4-peptide number 4, 16, 19 and 21 respectively), one pentapeptide sequence termed p53C5-18 (p53-targeting cyclic 5-peptide number 18) and one heptapeptide sequence termed p53C7-10 (p53-targeting cyclic 7-peptide number 10) (Table 4.1). The selected cyclic peptides were further evaluated to determine their ability to bind to p53C(Y220C), increase its thermodynamic stability and restore its apoptotic function (Chapter 5).

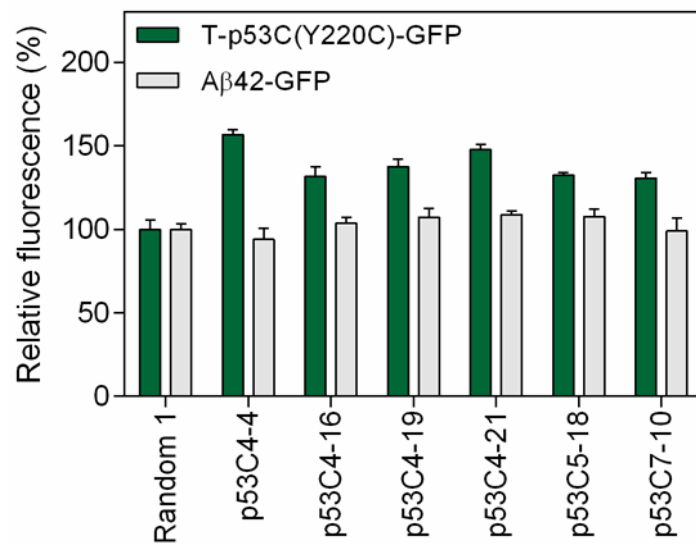
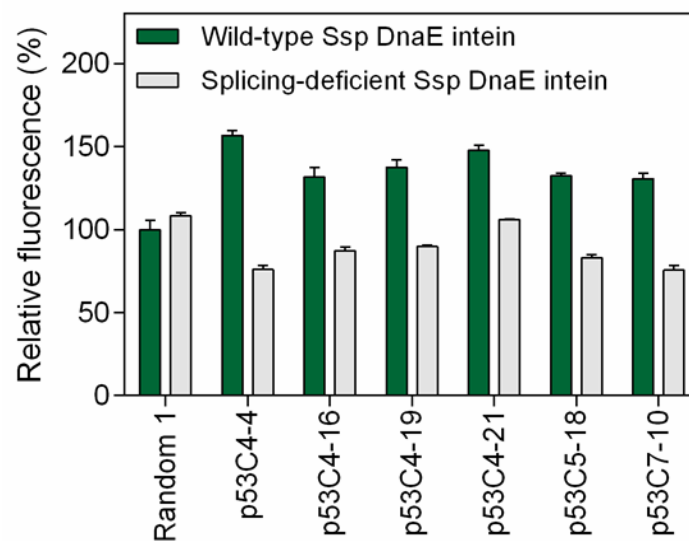
A**B**

Figure 4.11. Identification of p53C(Y220C) folding rescuers via FACS sorting. (A) Relative fluorescence of *E. coli* Tuner(DE3) cells overexpressing T-p53(Y220C)-GFP (green bars) or Aβ42-GFP (white bars) and six selected cyclic peptide clones isolated from the sorted populations shown in Figure 4.10. (B) Relative fluorescence of *E. coli* Tuner(DE3) cells overexpressing T-p53(Y220C)-GFP and the six selected cyclic peptide clones from (A) and utilizing either a wild-type split Ssp DnaE intein (green bars) or the splicing-deficient variant H24L/F26A32 (white bars). The fluorescence of bacterial cells producing T-p53(Y220C)-GFP or Aβ42-GFP together with a random cyclic peptide (Random 1) was arbitrarily set to 100. Mean ± sem are presented of one experiment performed in triplicates.

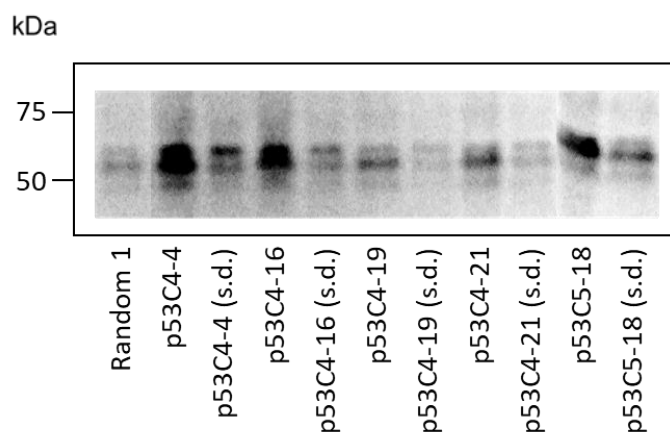


Figure 4.12. Solubility analysis of *E. coli* Tuner(DE3) cells overexpressing T-p53C(Y220C)-GFP and the five individual cyclic peptide sequences. Western blot analysis of soluble lysates of cells overexpressing T-p53C(Y220C)-GFP and five of the selected cyclic peptide sequences tested in (Figure 4.11), utilizing either a wild-type split Ssp DnaE intein or the splicing-deficient variant H24L/F26A32 (denoted as s.d.). The predicted molecular mass of the T-p53C(Y220C)-GFP fusion is ~53 kDa. Blotting was performed using the anti-GFP antibody.

Table 4.1. Cyclic oligopeptide sequences encoded by the selected bacterial clones exhibiting enhanced T-p53C(Y220C)-GFP fluorescence

Name of isolated clone	DNA sequence of peptide-encoding region	Encoded peptide sequence
p53C4-4	TGC TTC TCC TCC	CFSS
p53C4-16	TGC CTC GAG CAG	CLEQ
p53C4-19	TGC CTC CGG CGC	CLRR
p53C4-21	TGC GCG GCC CAG	CAAQ
p53C5-18	ACC CGG GGG GGC TGC	TRGGC
p53C7-10	AGC AAG CGG AGC GGG ATG CAG	SKRSGMQ

4.2. Targeting AD – Identification of A β 42 aggregation inhibitors using the generated ultrahigh-throughput biotechnology platform

In order to identify putative A β 42 aggregation inhibitors, in this thesis we screened the constructed cyclo-NuX₁X₂X₃X₄X₅X₆ heptapeptide sub-library, described in Chapter 2, as screening of the cyclic tetra-, penta- and hexapeptides for the same target has been the subject of the PhD thesis of Ilias Matis from our lab [330].

For this reason, we utilized a two-plasmid system comprising a pET28 expression vector producing the A β 42-GFP fusion and the pSICLOPPS vector library described in Chapter 2, which produced the cyclo-NuX₁X₂X₃X₄X₅X₆ heptapeptide sub-library (Figure 4.2). Protein expression was carried out at 37 °C using 0.1 mM IPTG and 0.005 % arabinose as described in the Materials and Methods section. Cells were gated on a side-scatter (SSC-H) versus forward-scatter (FSC-H) plot in order to eliminate non-cellular events, and were subjected to FACS sorting for the isolation of the bacterial population exhibiting the top ~2% fluorescence. After seven rounds of sorting we observed an almost six-fold increase of the bacterial GFP fluorescence compared to the initial library (Figure 4.13). No further substantial increase in fluorescence was observed after additional rounds of sorting.

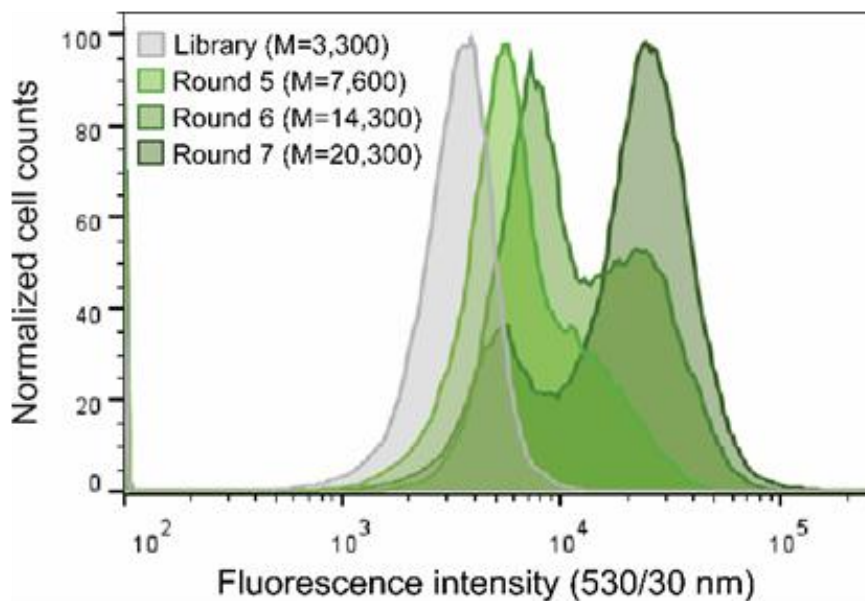
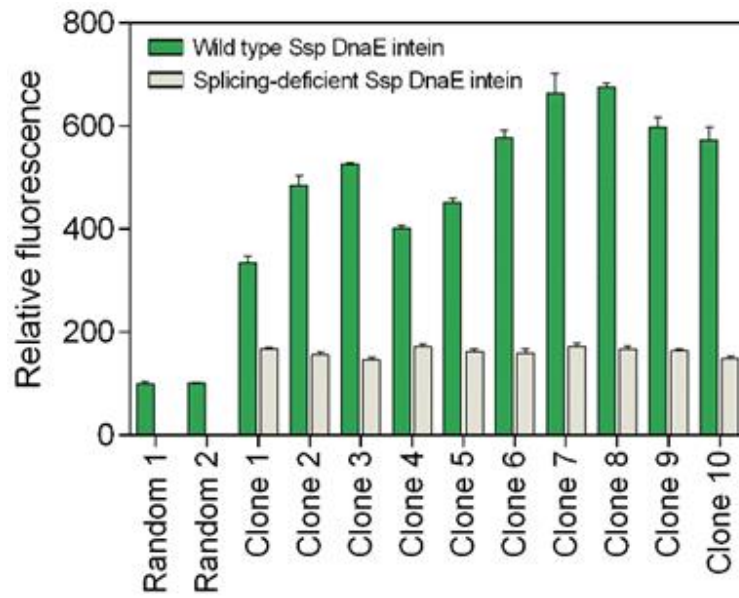


Figure 4.13. FACS sorting of *E. coli* Tuner(DE3) cells overexpressing A β 42-GFP and the combined cyclic heptapeptide library. M: mean GFP fluorescence in arbitrary units. Experiments were performed at the University of Texas at Austin.

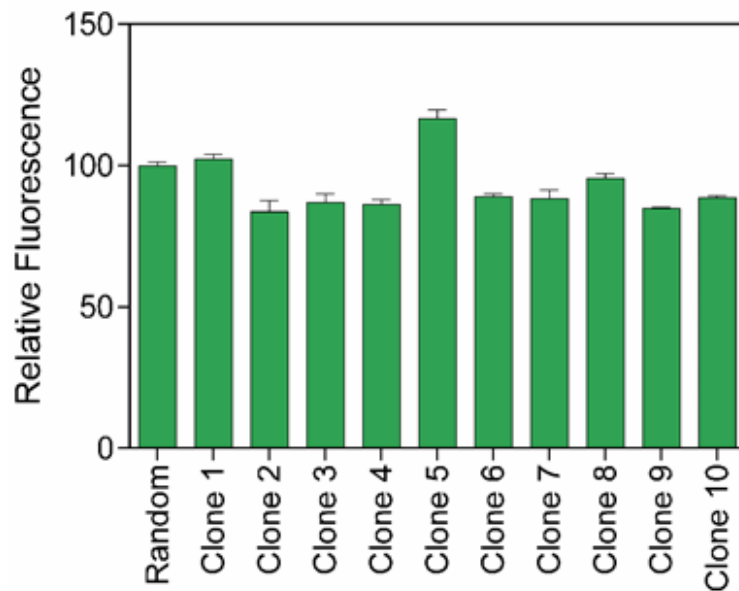
We randomly selected ten individual clones from the sorted population, isolated their peptide-encoding vectors, re-transformed them into fresh *E. coli* Tuner(DE3) cells carrying pETA β 42-GFP and induced protein production from both vectors. Interestingly, we found that in the presence of all selected cyclic-heptapeptides, the fluorescence of A β 42-GFP was dramatically increased compared to cells expressing the same A β 42-GFP fusion in the presence of two cyclic peptide sequences that were randomly selected from the initial unsorted cyclo-NuX₁X₂X₃-X₅ library (Figure 4.14A). Furthermore, all isolated clones expressed a full-length intein-peptide fusion (~25 kDa), which could undergo processing to yield a lower molecular weight band corresponding to excised I_N-CBD (~20 kDa), thus suggesting successful intein processing and possible formation of a cyclic peptide product (Figure 4.14C). Importantly, the observed phenotypic effects were found to be A β 42-specific, as the selected peptide cloned did not enhance the levels of T-p53C(Y220C)-GFP, and also they were dependent on peptide

cyclization as illustrated using the splicing deficient Ssp DnaE intein [280] (Figure 4.14 A and B).

A



B



C

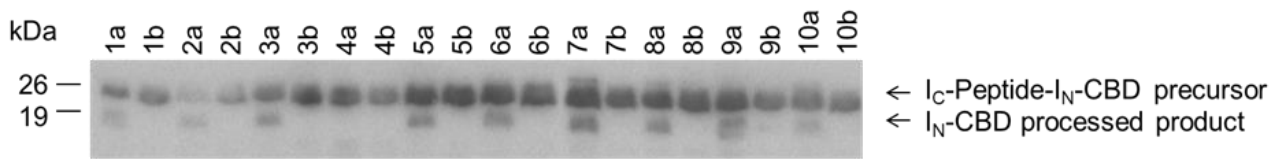


Figure 4.14. Identification of A β 42 aggregation inhibitors via FACS sorting. (A) Relative fluorescence of *E. coli* Tuner(DE3) cells overexpressing A β 42-GFP and ten randomly selected cyclic heptapeptide clones isolated after the seventh round of FACS sorting shown in (Figure 4.13) and utilizing either the wild-type split Ssp DnaE intein (green bars) or the splicing-deficient variant H24L/F26A32 (white bars). Two randomly selected cyclic peptide sequences (random 1 and 2) from the initial unsorted cyclo-NuX₁X₂X₃-X₅ library were used as a negative control. The fluorescence of the bacterial population producing cyclic peptide random 1 was arbitrarily set to 100. Mean values \pm s.e.m. are presented (n= 3 independent experiments, each one performed in three replicates). (B) Relative fluorescence of *E. coli* Tuner(DE3) cells overexpressing p53C(Y220C)-GFP and the ten selected cyclic heptapeptide clones tested in (A). The fluorescence of the bacterial population producing the random cyclic peptide was arbitrarily set to 100. Experiments were carried out in triplicates and the reported values correspond to the mean value \pm s.e.m. (C) Western blot analysis of the ten selected clones (A) using an anti-CBD antibody. The upper band of \sim 25 kDa corresponds to the I_C-peptide sequence-I_N-CBD precursor, while the lower band of \sim 20 kDa corresponds to the processed I_N-CBD product, whose appearance is an indication of successful cyclic peptide formation. For each clone, (a) represents a wild-type intein where cyclic peptide formation is allowed, while (b) represents the H24L/F26A splicing deficient variant where no cyclic peptide is produced.

Analysis of the expressed A β 42-GFP fusions by SDS-PAGE and western blotting revealed that the bacterial clones expressing the selected cyclic heptapeptides produce significantly increased levels of soluble A β 42-GFP compared to the random cyclic peptides, despite the fact that accumulation of total A β 42-GFP protein remained at similar levels (Figure 4.15, top). Furthermore, when the same cell lysates were analyzed by native PAGE and western blotting, we observed that co-expression of the selected cyclic peptides reduced the accumulation of higher-order A β 42-GFP aggregates, which could not enter the gel, and increased the amounts of species with higher electrophoretic mobility (Figure 4.15, bottom left). These higher electrophoretic mobility species correspond to the fraction of the total A β 42-GFP that exhibits fluorescence (Figure 4.15, bottom right). Since the solubility and

fluorescence of bacterially expressed A β 42-GFP has been found to be inversely proportional to the aggregation propensity of A β 42 [304,305], the results described above suggest that A β 42 aggregation is significantly decreased in the presence of the selected cyclic heptapeptides.

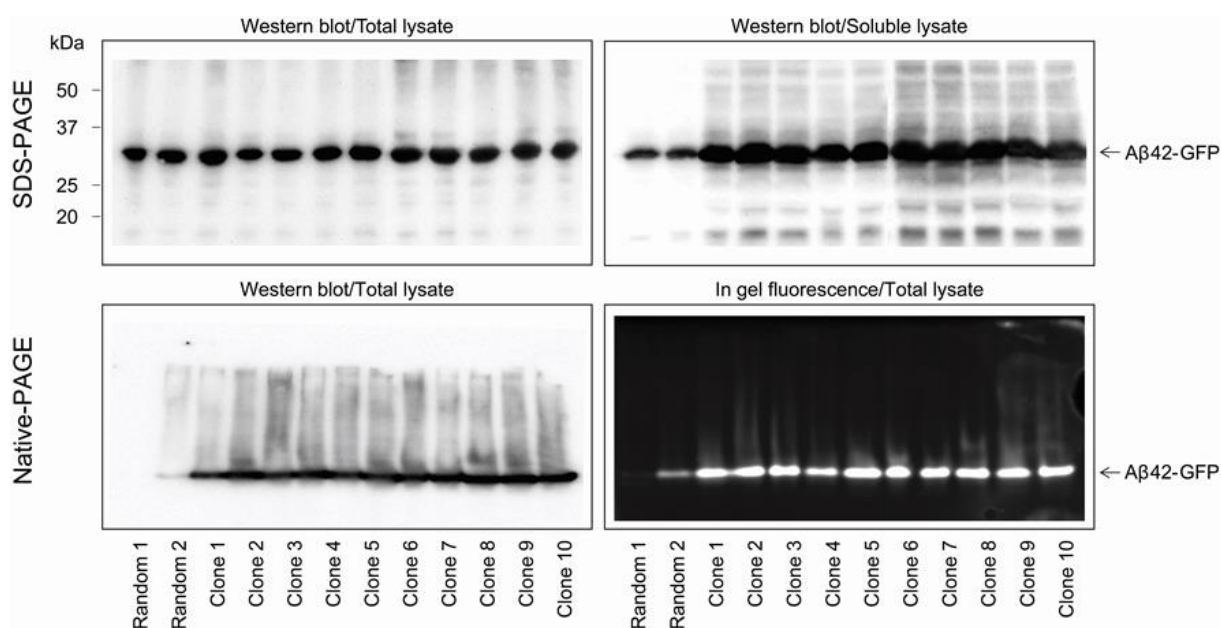


Figure 4.15. Solubility analysis of *E. coli* Tuner(DE3) cells overexpressing A β 42-GFP and the ten individual cyclic peptide sequences. (top) Western blot analysis of total (left) and soluble (right) lysates of cells overexpressing A β 42-GFP and the ten individual cyclic peptide sequences tested in Figure 4.14. The predicted molecular mass of the A β 42-GFP fusion is ~32 kDa. (bottom) Western blotting using the anti-A β antibody 6E10 (left) and in-gel fluorescence (right) analyses of total lysates following native-PAGE of cells co-expressing A β 42-GFP and the ten individual cyclic peptide sequences tested in Figure 4.14.

Importantly, similar results were acquired when A β 42 was produced in unfused, GFP-free form. When we tested the effects of the selected cyclic heptapeptides on A β 42 aggregation with an *in vivo* assay using whole-cell staining of intracellular formation of A β 42 aggregates with thioflavin S (ThS) [313], we observed that co-production of the selected peptides resulted in decreased levels of ThS fluorescence, further indicating reduced aggregate formation (Figure 4.16).

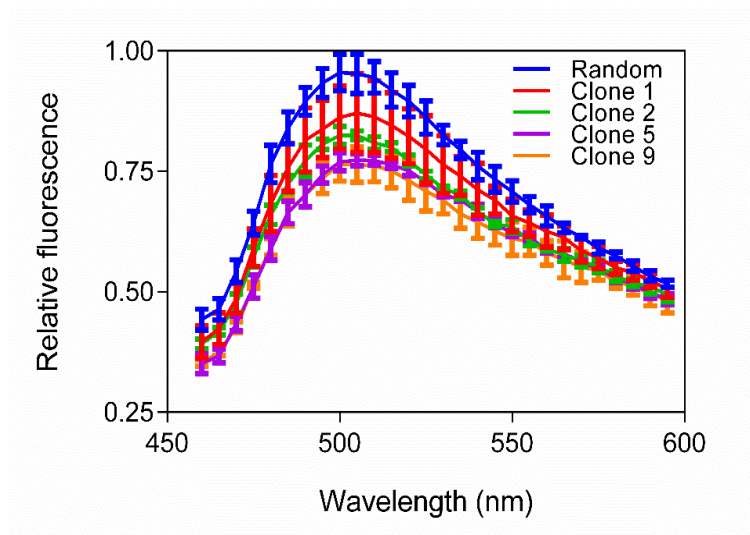


Figure 4.16. Emission spectra of *E. coli* Tuner(DE3) cells overexpressing A β 42 along with four of the selected cyclic heptapeptide sequences tested in (B) and stained with ThS. The maximum fluorescence of the bacterial population producing random 1 was arbitrarily set to 100. Mean values \pm s.e.m. are presented (n= 1 experiment performed in three replicates).

DNA sequencing of the ten selected clones revealed five distinct cyclic heptapeptide sequences: cyclo-CKVWQLL (present six times among the sequenced clones), cyclo-CRVWTEL, cyclo-CKVWMPL, cyclo-CRVWQTV and cyclo-CRIVPSL (**Table 4.2**). The selected cyclic peptides were further evaluated *in vitro* and *in vivo* to determine their ability to inhibit the aggregation of A β 42 and the associated neurotoxicity (Chapter 6).

Table 4.2. Cyclic heptapeptide sequences encoded by the selected bacterial clones exhibiting enhanced A β 42-GFP fluorescence

Isolated Clone #	DNA sequence of peptide-encoding region	Encoded peptide sequence
1	TGC AGG GTG TGG ACG GAG TTG	CRVWTEL
2	TGC AAG GTG TGG CAG TTG TTG	CKVWQLL
3	TGC AAG GTG TGG CAG TTG TTG	CKVWQLL
4	TGC ATC GTC GTC CCG TCG ATC	CIVVPSI
5	TGC AAG GTC TGG ATG CCG CTC	CKVWMPL
6	TGC AAG GTG TGG CAG TTG TTG	CKVWQLL
7	TGC AAG GTG TGG CAG TTG TTG	CKVWQLL
8	TGC AAG GTG TGG CAG TTG TTG	CKVWQLL
9	TGC CGC ATC GTC CCC AGC TTG	CRIVPSL
10	TGC AAG GTG TGG CAG TTG TTG	CKVWQLL

4.3. Targeting ALS – Identification of SOD1(A4V) folding rescuers using the generated ultrahigh-throughput biotechnology platform

For the identification of cyclic peptides that would rescue the misfolding of SOD1(A4V) we utilized the two-plasmid system previously described, comprising of a pET28 expression vector producing the SOD1(A4V)-GFP fusion and the pSICLOPPS vector library producing the cyclic tetra-, penta-, hexa- and heptapeptides (Figure 4.2). Protein expression was performed at 37 °C using 0.01 mM IPTG and 0.005 % arabinose as described in the Materials and Methods section, while FACS gating was performed as mentioned previously (Figure 4.2). After four rounds of sorting we observed a staggering, more than 70-fold increase of the bacterial GFP fluorescence compared to the initial library (Figure 4.17), while no further substantial increase in fluorescence was observed after additional rounds of sorting. The evaluation of the tetra-, penta- and hexapeptide sorted libraries has been the subject of Stefania Panoutsou's PhD thesis [330], while the heptapeptide sorted library are currently evaluated as part of Maria Giannakou's PhD thesis at the National Hellenic Research Foundation (NHRF).

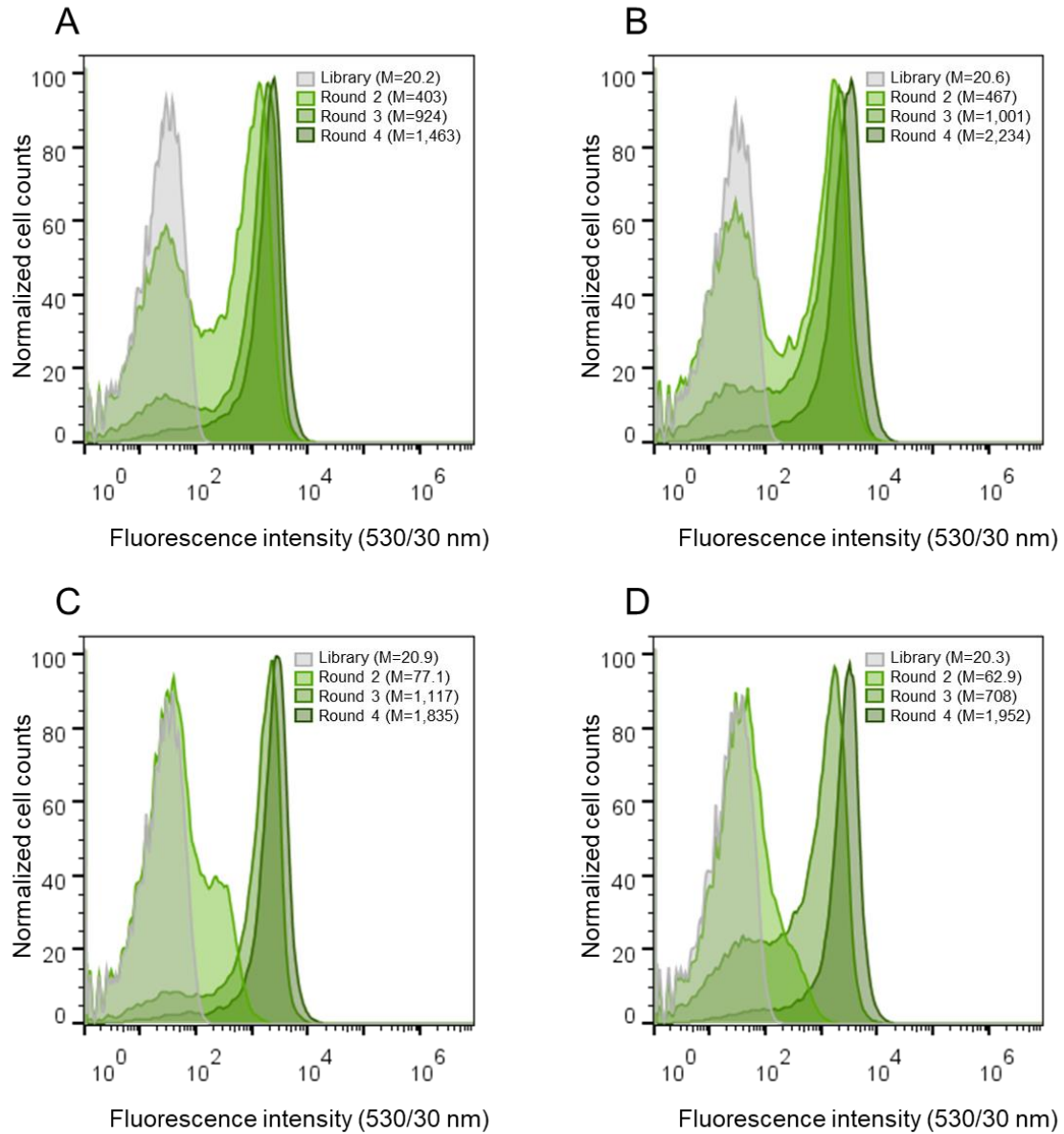


Figure 4.17. FACS sorting for the identification of SOD1(A4V) folding rescuers. Histograms of Origami2(DE3) cells co-expressing SOD1(A4V) and (A) the cyclo-NuX₁X₂X₃ tetrapeptide library, (B) the cyclo-NuX₁X₂X₃X₄ pentapeptide library, (C) the cyclo-NuX₁X₂X₃X₄X₅ hexapeptide library and (D) the cyclo-NuX₁X₂X₃X₄X₅X₆ heptapeptide library. In all panels, the initial library is shown in grey and the enriched sorted libraries in green. M=mean GFP fluorescence in arbitrary units. FACS sorting was performed at the University of Texas at Austin.

4.4. Targeting HD – Identification of HTT_{ex1}-97Q aggregation inhibitors using the generated ultrahigh-throughput biotechnology platform

For the identification of cyclic peptides that would inhibit the aggregation of HTT_{ex1}-97Q we utilized the same plasmid system and gating strategy as in the case of SOD1, with the only difference being the production of HTT_{ex1}-97Q-GFP by the pET28 vector using 0.1 mM IPTG (Figure 4.2). After six rounds of sorting we observed an almost five-fold increase of the bacterial GFP fluorescence compared to the initial library (Figure 4.18), while no further substantial increase in fluorescence was observed after additional rounds of sorting. The sorted populations are currently being evaluated for the discovery of putative aggregations inhibitors of HTT_{ex1}-97Q.

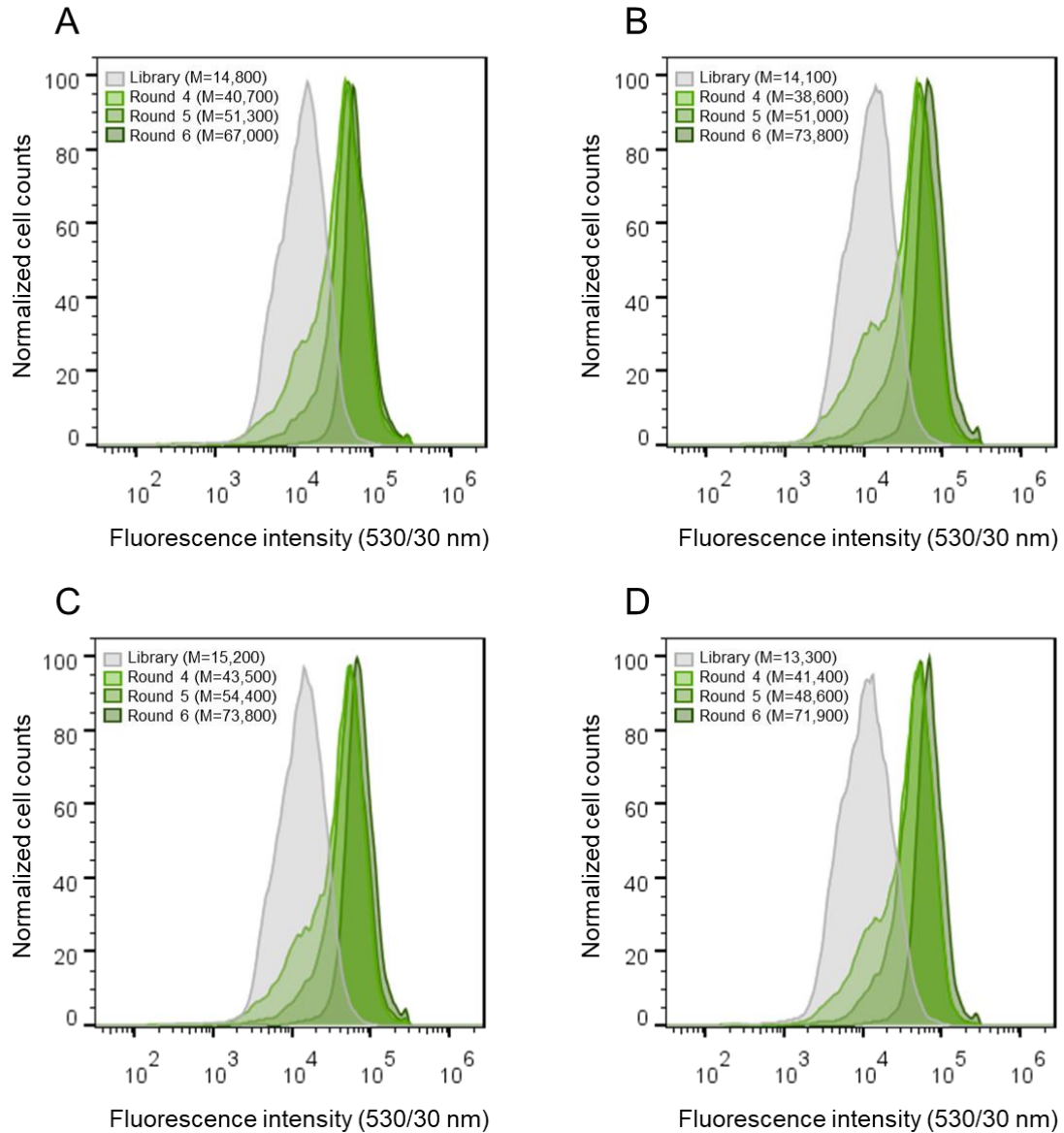


Figure 4.18. FACS sorting for the identification of HTT_{ex1-97Q} aggregation inhibitors. Histograms of Tuner(DE3) cells co-expressing HTT_{ex1-97Q} and (A) the cyclo-NuX₁X₂X₃ tetrapeptide library, (B) the cyclo-NuX₁X₂X₃X₄ pentapeptide library, (C) the cyclo-NuX₁X₂X₃X₄X₅ hexapeptide library and (D) the cyclo-NuX₁X₂X₃X₄X₅X₆ heptapeptide library. In all panels, the initial library is shown in grey and the enriched sorted libraries in green. M=mean GFP fluorescence in arbitrary units. FACS sorting was performed at the University of Texas at Austin.

4.5. Discussion

In this chapter we present the development of an ultrahigh-throughput *in vivo* screening system that enables: (i) the construction of cyclic peptide libraries with expanded diversities and (ii) their direct functional screening using a FACS-based assay that enables the rapid and facile identification of macrocycles with the ability to rescue pathogenic protein misfolding and aggregation.

We employed this system against four MisP targets: p53C(Y220C), A β 42, SOD1(A4V) or HTT_{ex1}-97Q, and isolated four distinct bacterial populations that produce cyclic peptides with the ability to rescue protein misfolding and aggregation of the respective MisP of interest. Notably, while for the majority of the targets this process was very straightforward, in the case of p53C(Y220C), it was rather laborious. However, the high adaptability of the reported bacterial system that enables the isolation of bacterial populations with specific properties, finally allowed us to isolate a bacterial population with the desired effects on p53C(Y220C) misfolding and aggregation.

The cyclic peptides identified against p53C(Y220C) and A β 42 are further evaluated in Chapters 5 and 6 respectively, while evaluation of the selected rescuers against SOD1(A4V) will be presented in the PhD theses of S. Panoutsou and M. Giannakou. Finally, the identified cyclic peptides against HTT_{ex1}-97Q aggregation are currently being evaluated in our laboratory.

Chapter 5 – Evaluation of the selected cyclic peptides on their ability to rescue the misfolding of p53(Y220C) *in vitro* and restore its pro-apoptotic function in cancer cells lines

5.1. Biosynthesis and isolation of selected cyclic peptides from bacterial cells

First, we sought to chemically synthesize the selected cyclic peptides, in order to further evaluate their ability to bind to T-p53C(Y220C), increase its thermodynamic stability and restore its apoptotic function. However, the head-to-tail cyclization of the p53C4-16, p53C5-18 and p53C7-10 proved very laborious and time consuming, while chemical synthesis of cyclic peptides p53C4-14, p53C4-19 and p53C4-21 has not been possible so far. Since in the literature efforts to synthesize head-to-tail cyclic peptides with small ring sizes using traditional synthetic approaches have been repeatedly highly challenging [331,332], we decided to produce the selected cyclic peptides recombinantly in *E. coli* cells and then isolate them using affinity chromatography. As we did not wish to affect the ring size of the selected cyclic peptides by introducing a tag, such as the HPQ motif, that would enable direct purification of the desirable cyclic peptides [280,333], we opted to isolate the biosynthesized cyclic peptides in an indirect manner, by taking advantage of the intein's potential for tag-less protein purification [334]. For this reason, we overexpressed the tetra-partite fusion protein Ic-peptide-

I_N-CBD containing the selected peptide sequences in *E. coli* and attempted to immobilize the precursor protein onto a chitin resin, thus allowing the tag-less purification of the cyclic peptide upon intein splicing (Figure 5.1A and B). However, the overexpressed fusion exhibited low affinity for the chitin beads, as illustrated by western blotting of the samples collected during the purification process (Figure 5.1C), rendering purification of the cyclic peptide by this method very difficult.

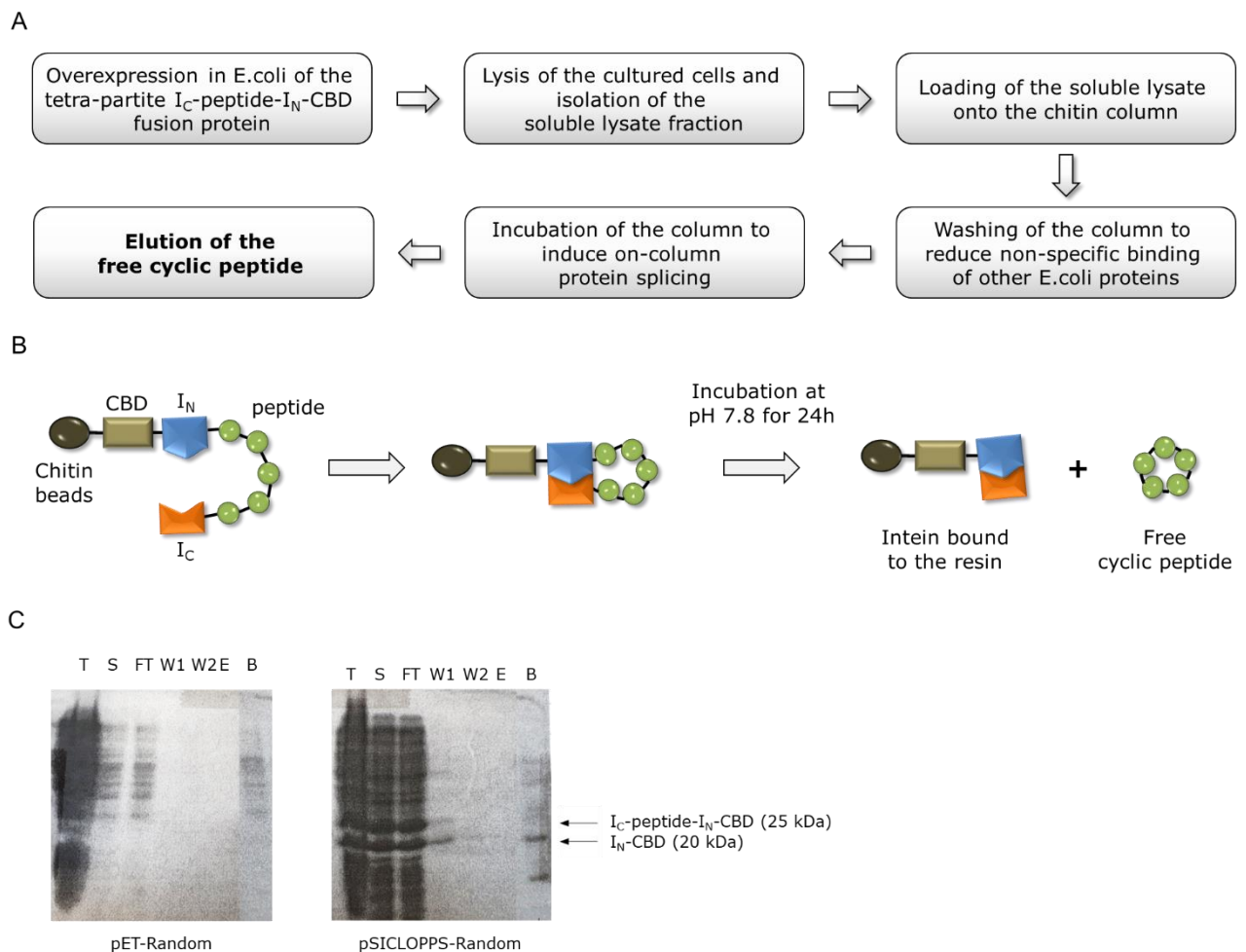


Figure 5.1. Cyclic peptide production and purification via on-column protein splicing using chitin affinity chromatography. (A) Flow chart of the on-column purification process. (B) Schematic of the on-column intein-mediated circular ligation reaction. (C) Western blotting using the anti-CBD antibody of samples collected during the cyclic peptide purification procedure described in (A). Protein overexpression was performed using the pET28 (left) or pSICLOPPS (right) vector. T: total lysate, S: soluble lysate, FT: flow-through, W1: first wash, W2: second wash, E: elution and B: chitin beads after cyclic peptide elution.

For this reason, we introduced a polyhistidine tag upstream of the tetra-partite fusion and attempted protein purification in a similar manner but using instead immobilized metal affinity chromatography (IMAC) and a nickel resin (Ni-NTA) (Figure 5.2A). Indeed, by using this method we were able to selectively capture the precursor protein onto the nickel resin and upon incubation at room temperature and at a pH permissive for intein splicing for 72 h we observed significant intein processing, as indicated by western blotting of the samples collected during the purification process (Figure 5.2B). Importantly, mass spectrometry of the elution fractions after the on-column splicing reaction verified the presence of the free cyclic peptides (Figure 5.3).

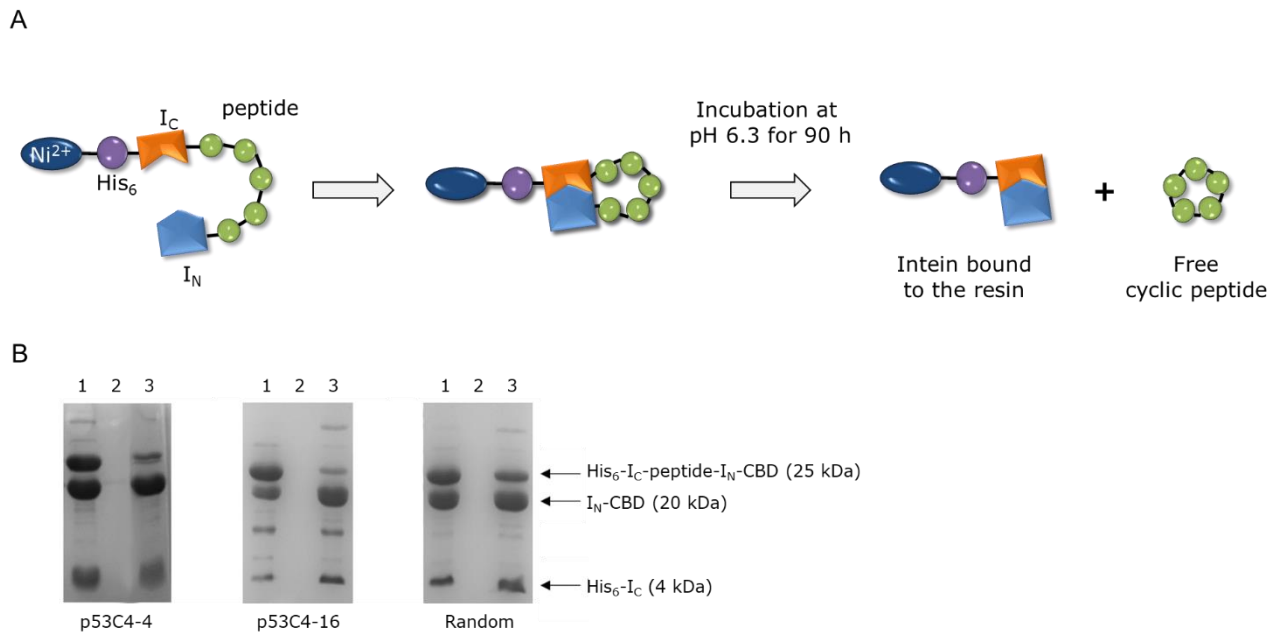


Figure 5.2. Cyclic peptide production and purification via on-column protein splicing using IMAC. (A) Schematic of the on-column intein-mediated circular ligation reaction. (C) SDS-PAGE of samples collected during the cyclic peptide purification procedure of p53C4-4, p53C4-16 and a random peptide from the unsorted pSICLOPPS library, as described in (Figure 5.1A) and with the only difference being the utilization of Ni-NTA instead of a chitin resin. In all cases, the 1st and 3rd lane depict samples containing nickel resin loaded with the soluble lysates, before and after incubation at RT for 90 h in order to induce protein splicing. The 2nd lane represents the elution sample containing the free cyclic peptide.

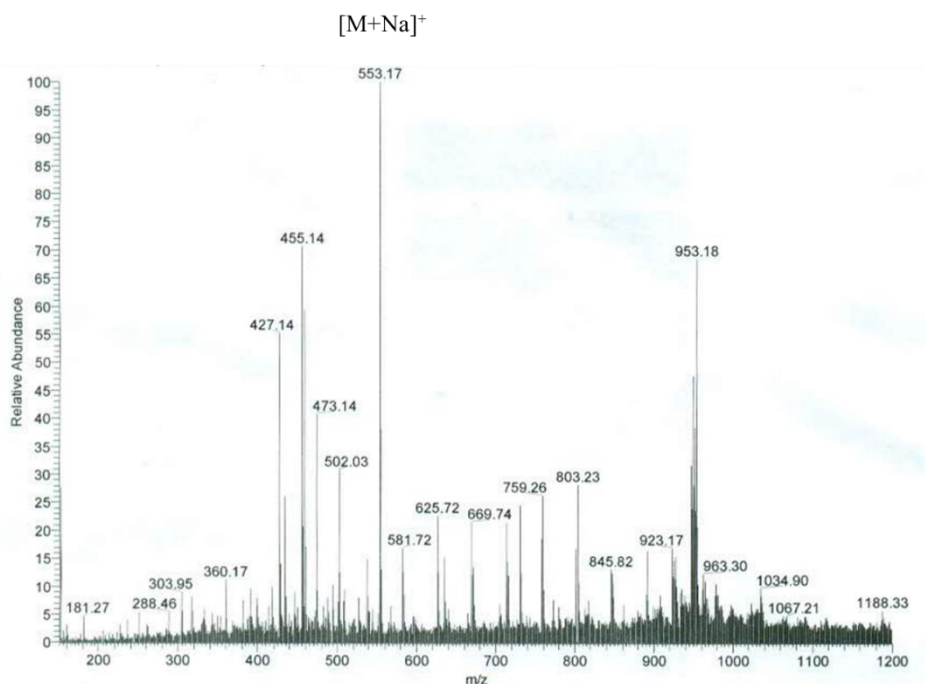


Figure 5.3. Full scan spectrum (ESI-MS) of the sample containing the free cyclic peptide after on-column intein-mediated circular ligation. The main [M+Na]⁺ peak represents the single-charged sodium adduct of the cyclic peptide. Mass spectrometry experiments were performed by Ms. Eleni Siapi at the NHRF.

5.2. Evaluation of the selected cyclic peptides' ability to promote cell death of a cancer cell line expressing the p53(Y220C) mutation

In order to evaluate the cyclic peptides' ability to promote cancer cell death, we investigated the effect of the selected peptides on the cell viability of the human melanoma cell line WM164 carrying the p53(Y220C) mutation [335]. Cells were grown in 96-well plates for 24 h and then incubated for 48 h in the presence of 1% or 10% v/v of elution fractions after on-column purification of the selected cyclic peptides (Figure 5.4). Interestingly, all cyclic peptides except the heptapeptide p53C7-10, were found to reduce the viability of WM164 cells when administered at the highest concentration. Importantly, p53C4-16 exhibited a significant decrease of cell viability of ~55 % compared to the cells treated with random peptide sample (Figure 5.4). However, as we were unable to precisely measure the concentration of the cyclic

peptides after on-column intein splicing nor determine their ability to permeate cell membranes, the reduced effect of the rest of the compounds may be due to insufficient intracellular cyclic peptide levels.

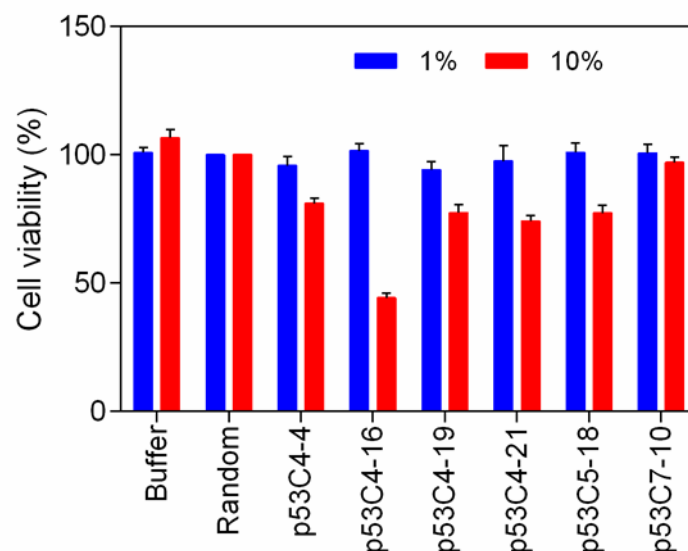


Figure 5.4. Viability of melanoma WM164 (p53(Y220C)) cells after treatment for 48 h with 1% or 10% v/v of the selected cyclic peptides, as determined by the MTT assay. Cell viability of samples treated with a random peptide from the unsorted pSICLOPPS library was arbitrarily set to 100. Mean values \pm sd are presented (n=6 replicate wells per condition). Experiments were performed by Dr. Zacharoula Linardaki under the supervision of Dr. Vassiliki Pletsa at the NHRF.

Encouraged by the results presented above, we synthesized p53C4-16 by solid-phase chemical synthesis at mg scale (Figure 5.5). To further evaluate the efficacy of p53C4-16, we investigated increasing concentrations of synthetic p53C4-16 for their ability to promote cell death of the WM164 cancer cell line (Figure 5.6). Notably, the selected cyclic peptide was able to affect cell viability in a dose-dependent manner and at concentrations \sim 100 μ M (Figure 5.6A). Furthermore, when co-administered with the widely utilized chemotherapeutic drug cisplatin, p53C4-16 was able to further decrease cell viability, resulting in almost complete cell death (Figure 5.6B).

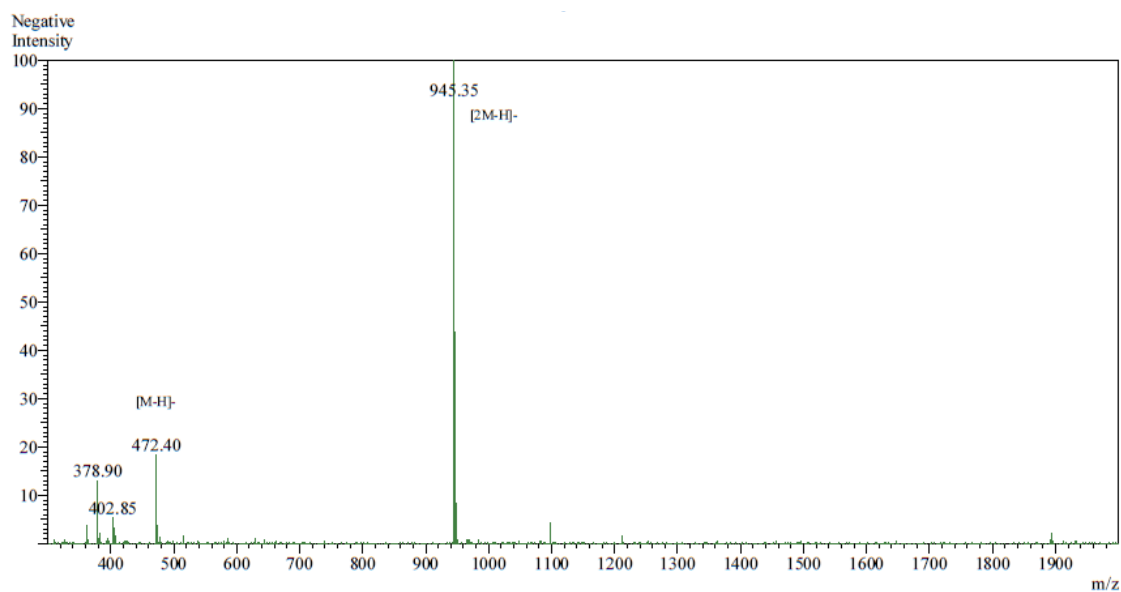


Figure 5.5. Full scan spectrum (ESI-MS) of the synthesized p53C4-16 by solid-phase chemical synthesis.

The [M-H]⁻ peak represents the negatively charged cyclic peptide, after one proton is removed. The [2M-H]⁻ peak represents the dimer formed by the association of [M-H]⁻ with its neutral counterpart, commonly created due to high sample concentration. Solid-phase synthesis and MS analysis were performed by Genscript (USA).

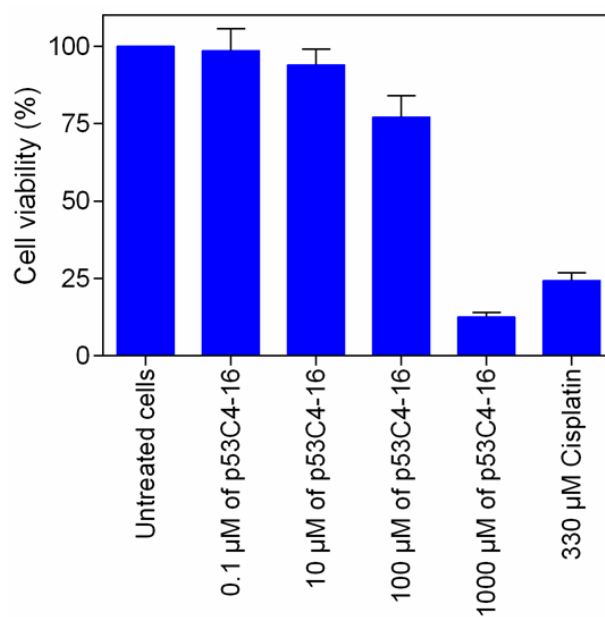
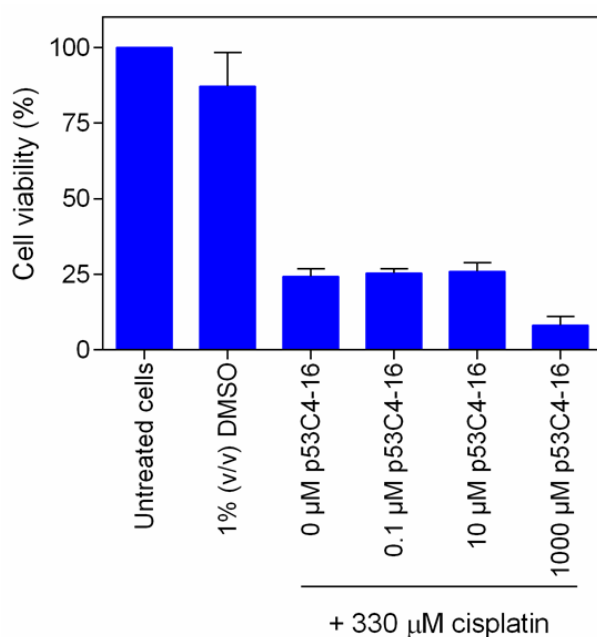
A**B**

Figure 5.6. Viability of melanoma WM164 (p53(Y220C)) cells after treatment for 48 h with different concentration of p53C4-16, as determined by the MTT assay in (A) the absence or (B) the presence of 330 μM (100 μg/ml) cisplatin. p53C4-16 was able to reduce cell viability in a dose-dependent manner and also enhance the apoptotic effect of the known chemotherapeutic agent cisplatin. Viability of untreated cells was arbitrarily set to 100. Mean values \pm sd are presented (n=6 replicate wells per condition). Experiments were performed by Dr. Zacharoula Linardaki under the supervision of Dr. Vassiliki Pletsa at the NHRF.

In order to support that p53C4-16 has selective toxicity towards cancer cell lines, we determined its effect on the viability of the highly sensitive human mesenchymal stem cells derived from Wharton's Jelly (WJ-MSC) [336]. Indeed, while at the highest concentration p53C4-16 exhibits a small cell viability decrease (Figure 5.7), this effect is not considered significant enough to support a general cytotoxic effect of the selected cyclic peptide, supporting the notion that p53C4-16 is able to selectively promote cell death of cancer cells.

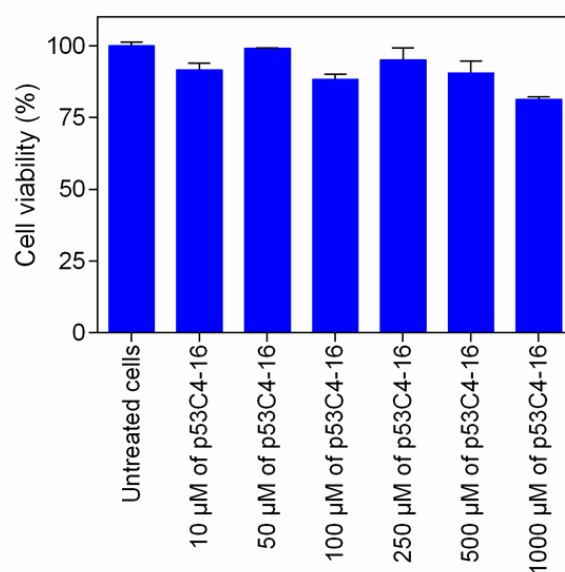


Figure 5.7. Viability of human mesenchymal cells derived from Wharton's Jelly after treatment for 3 h with different concentration of p53C4-16, as determined by the MTT assay. Viability of untreated cells was arbitrarily set to 100. Mean values \pm sd are presented (n=3 replicate wells per condition). Experiments were performed by Elena Taki under the supervision of Dr. Vassilis Zoumpourlis at the NHRF.

5.3. Biosynthesis and purification of T-p53C and T-p53C(Y220C) for use in *in vitro* experiments

For the execution of the following *in vitro* experiments, we firstly needed to isolate the proteins of interest in a pure form. For this reason, we produced fusions of T-p53C or T-p53(Y220C) along with a His₆ affinity tag in *E. coli*. Upon overexpression, the proteins of interest were purified using IMAC (Figure 5.8), followed by size-exclusion chromatography using the HiLoad 16/600 Superdex 200 pg column (Figure 5.9) as described in the Materials and Methods section. This procedure resulted in the preparation of T-p53C and T-p53C(Y220C) at high purity.

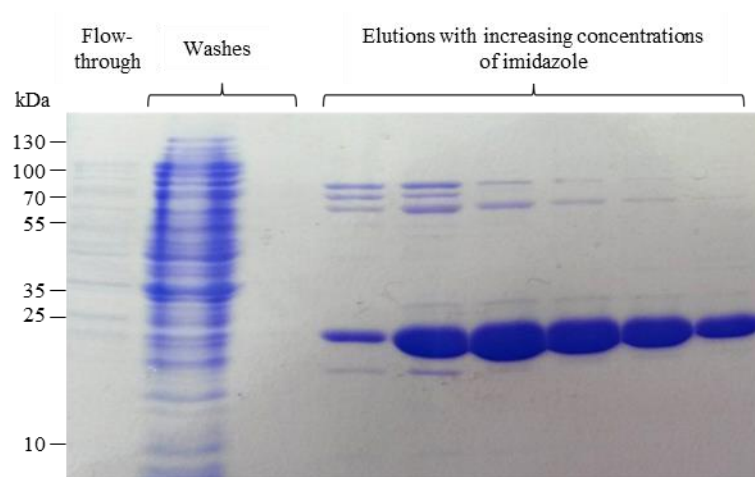


Figure 5.8. SDS-PAGE of samples collected during protein purification of T-p53C(Y220C)-His₆ by Ni-IMAC after overexpression in *E. coli* Tuner(DE3).

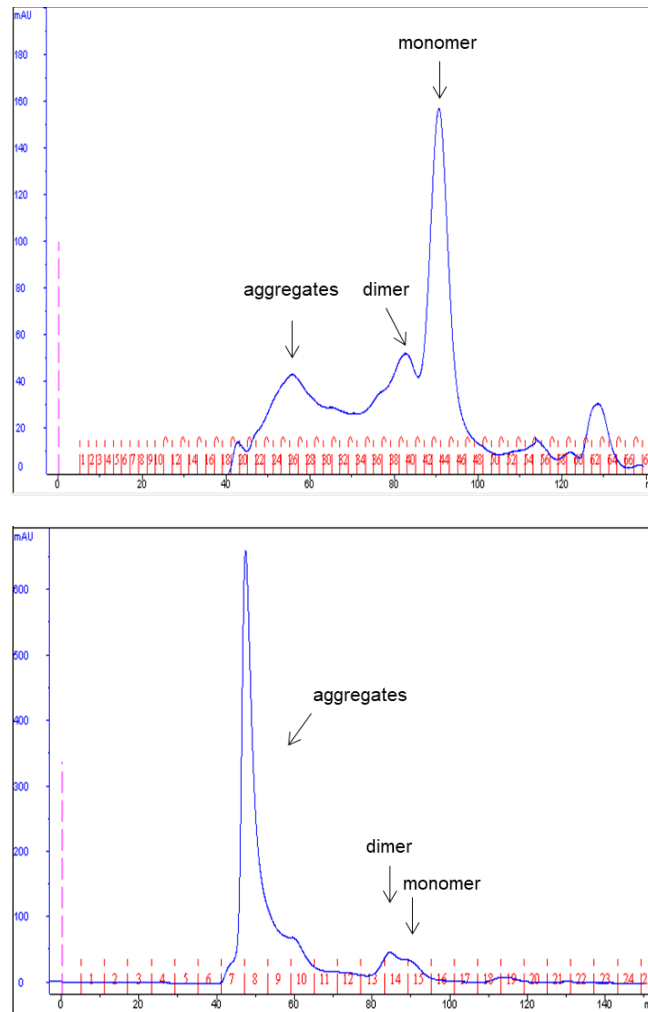


Figure 5.9. Size-exclusion chromatography (SEC) of T-p53C-His₆ (top) and T-p53C(Y220C)-His₆ (bottom) after the IMAC purification shown in Figure 5.8. As expected, in the samples before SEC, T-p53C-His₆ was mostly present in a monomeric form, while the aggregation prone T-p53C(Y220C)-His₆ was present in higher order aggregates.

5.4. Evaluation of the selected cyclic peptides' ability to increase the thermodynamic stability of T-p53C(Y220C) *in vitro*

In order to investigate the potential of the selected cyclic peptides to increase the thermodynamic stability of T-p53C(Y220C), we used differential scanning fluorimetry (DSF) and the dye SYPRO orange, which exhibits a significant fluorescence increase upon binding to hydrophobic protein regions (Figure 5.10). This method enables the estimation of a protein's

melting temperature (T_m), which is proportional to its thermodynamic stability, by simply monitoring the sample's fluorescence levels during thermal unfolding [337].

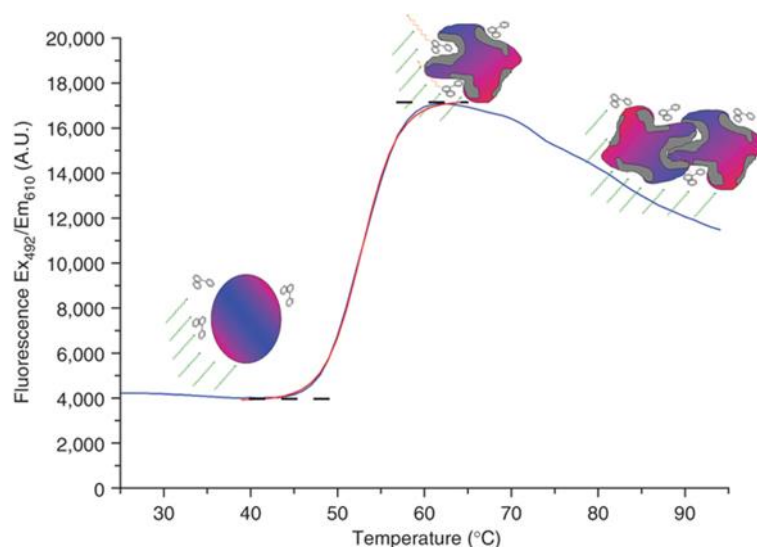


Figure 5.10. Characteristic diagram of fluorescence intensity vs. temperature of a representative globular protein in the presence of SYPRO orange. Initially, SYPRO orange is unable to bind to the folded protein thus emitting low levels of fluorescence (represented by green arrows). Upon thermal denaturation, the dye binds to the protein's exposed hydrophobic regions (depicted in grey) and emits increased levels of fluorescence at 610 nm (represented by red arrows). Further temperature increase results in protein aggregation and precipitation and gradual fluorescence decrease. Adapted from [337].

We first opted to determine the T_m of T-p53C and T-p53C(Y220C) in the absence of any test compound using DSF. For this reason we prepared samples of T-p53C and T-p53C(Y220C) in the presence of SYPRO orange and measured their fluorescence while gradually increasing the samples' temperature using a real-time PCR instrument. This process generated a sigmoidal curve whose inflection point (T_m) was calculated by determining the maximum of the first derivative $[d(\text{RFU})/dT]$ (Figure 5.11) and for more accurate results by fitting the data to the Boltzmann equation using the OriginPro software (Table 5.1 and Appendix A).

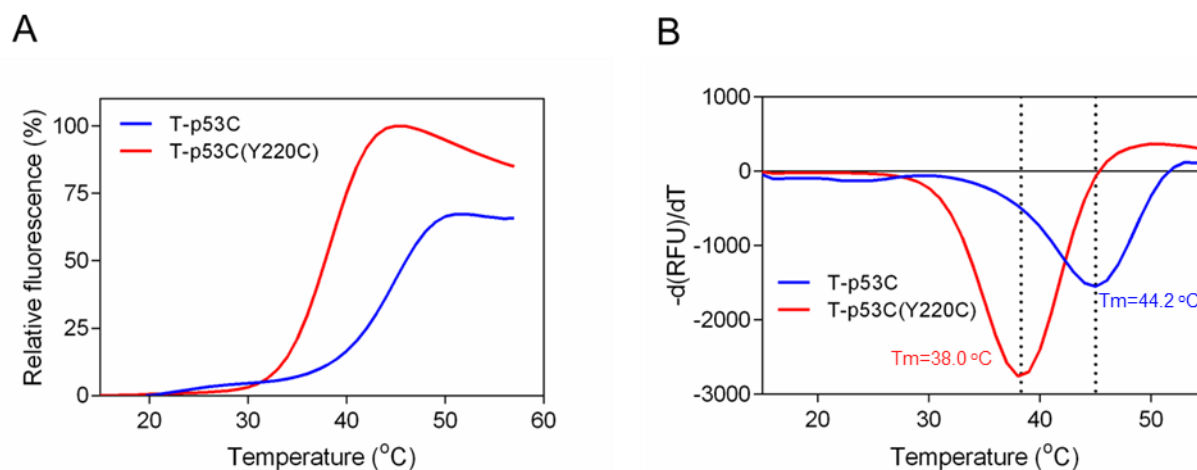


Figure 5.11. Thermal denaturation analysis of T-p53C and T-p53C(Y220C) using DSF. (A) Relative fluorescence of T-p53C and T-p53C(Y220C) in the presence of SYPRO orange and during gradual temperature increase. (B) Plot of the first derivative of (A) enabling the estimation of T_m values. In both panels mean values of one experiment performed in triplicates are presented.

Table 5.1. Calculated T_m values of T-p53C and T-p53C(Y220C) after fitting to the Boltzmann equation

Sample	Calculated T_m (°C)			Mean T_m (°C)
T-p53C	44.01 ± 0.12	44.22 ± 0.11	44.23 ± 0.09	44.15 ± 0.11
T-p53C(Y220C)	37.95 ± 0.03	37.98 ± 0.03	38.01 ± 0.04	37.98 ± 0.03

Next, we studied the effect of p53C4-16 on the T_m of T-p53C(Y220C). Interestingly, by increasing the concentration of p53C4-16, we noticed a gradual shift of the fluorescence curve of T-p53C(Y220C) in the presence of SYPRO orange, which reached its maximum at 15 molar equivalents of p53C4-16 (Figure 5.12). At this concentration, p53C4-16 was able to increase the T_m of T-p53C(Y220C) by ~ 0.8 °C, which is comparable to other published small molecule re-activators of T-p53C(Y220C) [186]. In addition, in the presence of increasing concentrations of p53C4-16 we noticed a gradual decrease in the levels of fluorescence,

indicating that the exposed hydrophobic surfaces of T-p53C(Y220C) were reduced and thus, that the protein is more stable in the presence of p53C4-16 (Figure 5.12).

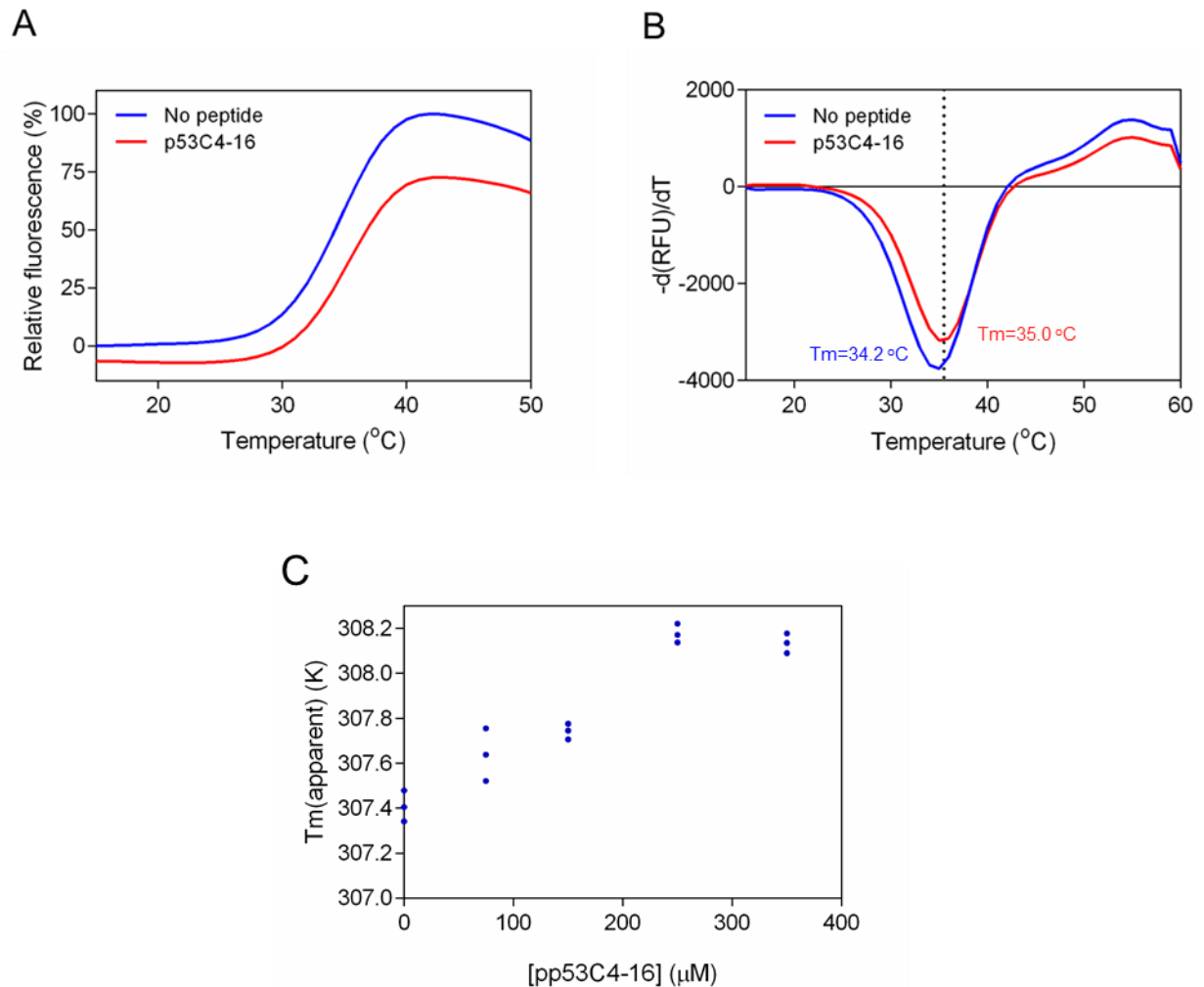


Figure 5.12. Effect of p53C4-16 to the thermal stability of T-p53C(Y220C) using DSF. (A) Relative fluorescence of 15 μM T-p53C(Y220C) in the absence or presence of 250 μM p53C4-16 as measured using DSF and the SYPRO orange dye. (B) Plot of the first derivative of (A) enabling the estimation of T_m values. In both panels mean values of one experiment performed in triplicates are presented. (C) Concentration dependent thermal stabilization of T-p53C(Y220C) by increasing concentrations of p53C4-16 as measured by DSF.

Table 5.2. Calculated T_m values of T-p53C(Y220C) in the absence or presence of increasing concentration of p53C4-16, after fitting to the Boltzmann equation.

Sample	Calculated T _m (°C)			Mean T _m (°C)
No peptide	34.19 ± 0.05	34.26 ± 0.06	34.33 ± 0.08	34.26 ± 0.06
75 μM p53C4-16	34.60 ± 0.06	34.37 ± 0.05	34.49 ± 0.06	34.49 ± 0.06
150 μM p53C4-16	34.56 ± 0.06	34.59 ± 0.05	34.63 ± 0.05	34.59 ± 0.06
250 μM p53C4-16	35.02 ± 0.05	35.07 ± 0.06	34.99 ± 0.06	35.03 ± 0.06
350 μM p53C4-16	34.94 ± 0.04	34.99 ± 0.04	35.03 ± 0.05	34.98 ± 0.04

Similarly, we monitored the thermal denaturation of T-p53C(Y220C) in the presence of p53C5-18 and p53C7-10. Neither of the cyclic peptides were found able to increase the protein's T_m even when administered at a 20-fold molar excess in the DSF assay (Figure 5.13, Figure 5.14 and Table 5.3). However, the fluorescence levels of samples containing T-p53C(Y220C), SYPRO orange and increasing concentrations of p53C5-18 were gradually decreased compared to the no peptide samples, indicating that the exposed hydrophobic surfaces of T-p53C(Y220C) were reduced and thus, that the protein is more stable in the presence of p53C5-18 (Figure 5.13).

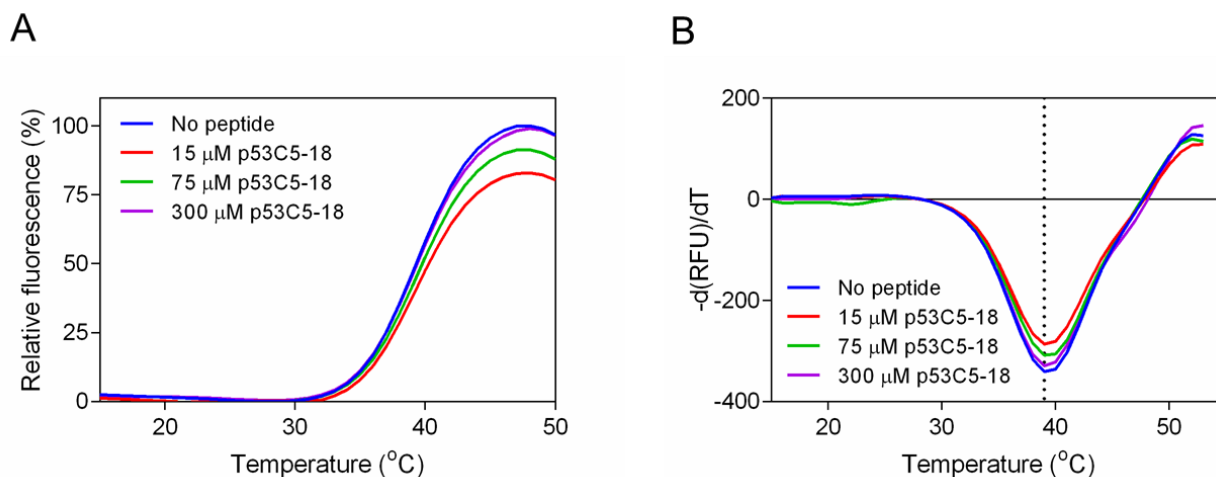


Figure 5.13. Effect of p53C5-18 to the thermal stability of T-p53C(Y220C) using DSF. (A) Relative fluorescence of 15 μM T-p53C(Y220C) in the absence or presence of increasing concentrations of p53C5-18 as measured by DSF and the SYPRO orange dye. (B) Plot of the first derivative of (A) enabling the estimation of T_m values. In both panels mean values of one experiment performed in triplicates are presented.

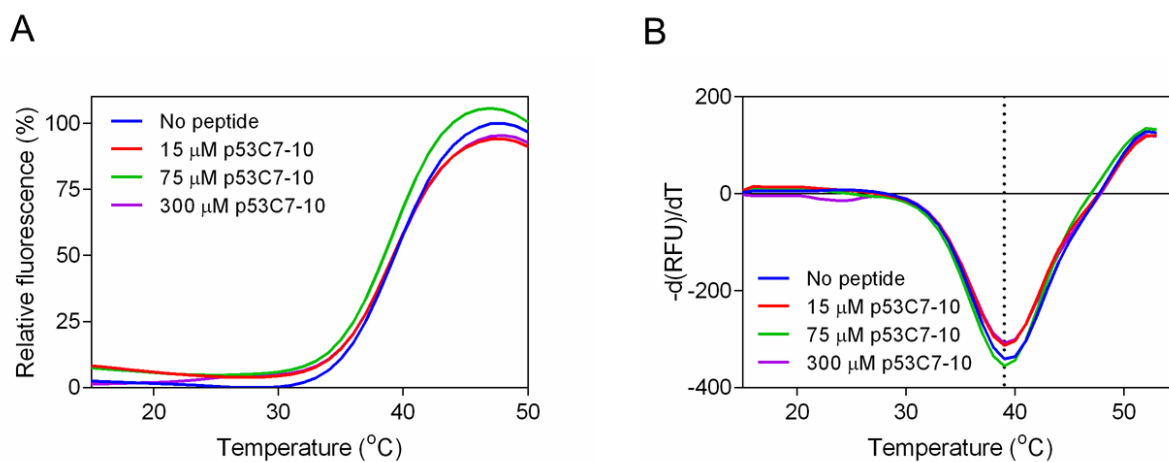


Figure 5.14. Effect of p53C7-10 to the thermal stability of T-p53C(Y220C) using DSF. (A) Relative fluorescence of 15 μM T-p53C(Y220C) in the absence or presence of increasing concentrations of p53C7-10 as measured using DSF and the SYPRO orange dye. (B) Plot of the first derivative of (A) enabling the estimation of T_m values. In both panels mean values of one experiment performed in triplicates are presented.

Table 5.3. Calculated T_m values of T-p53C(Y220C) in the absence or presence of increasing concentration of p53C5-18 or p53C7-10, after fitting to the Boltzmann equation.

Sample	Calculated T _m (°C)			Mean T _m (°C)
No peptide	39.80 ± 0.06	39.24 ± 0.04	39.27 ± 0.04	39.43 ± 0.04
15 μM p53C5-18	39.41 ± 0.03	39.75 ± 0.05	39.13 ± 0.04	39.43 ± 0.04
75 μM p53C5-18	39.21 ± 0.04	39.96 ± 0.04	39.22 ± 0.03	39.46 ± 0.04
300 μM p53C5-18	39.18 ± 0.03	39.36 ± 0.03	39.67 ± 0.04	39.40 ± 0.04
15 μM p53C7-10	39.23 ± 0.02	39.06 ± 0.03	39.17 ± 0.03	39.15 ± 0.03
75 μM p53C7-10	39.07 ± 0.04	38.76 ± 0.07	39.00 ± 0.03	38.94 ± 0.05
300 μM p53C7-10	39.31 ± 0.04	39.31 ± 0.04	39.25 ± 0.03	39.29 ± 0.04

5.5. Evaluation of the selected cyclic peptides' ability to inhibit T-p53C(Y220C) aggregation *in vitro*

It has been previously shown that destabilized p53 mutants aggregate *in vitro* and *in vivo* creating amyloid-like fibrils that can bind thioflavin T (ThT) [338]. Indeed, rescuers of p53 aggregation have been previously found to decrease tumour cell viability, supporting the notion that the identification of p53 aggregation inhibitors could be a potentially therapeutic approach for various carcinomas [227]. With this in mind, we opted to evaluate the ability of the selected cyclic peptides to inhibit the aggregation of T-p53C(Y220C) *in vitro* by monitoring the fluorescence levels of protein samples in the presence of ThT.

First, we monitored the aggregation process of protein samples containing 20 μM of ThT and 15 μM of monomeric T-p53C or T-p53C(Y220C) at 37 °C, by measuring ThT fluorescence every 5 min and for a period of 10 h. As expected, we found that T-p53C(Y220C)

self-assembled into amyloid-like aggregates over time, while T-p53C exhibited no significant aggregation and ThT fluorescence increase (Figure 5.15A). Next, we studied the ability of the selected cyclic peptides to inhibit the aggregation of T-p53C(Y220C) by incubating monomeric T-p53C(Y220C) with increasing concentrations of p53C4-16, p53C5-18 and p53C7-10. Interestingly, we found that, while both p53C5-18 and p53C7-10 showed no significant increase of the proteins T_m , they were both able to inhibit its aggregation significantly at 5 molar equivalents of each cyclic peptide (Figure 5.15B and C), while not affecting ThT fluorescence when incubated in the absence of T-p53C(Y220C) (Figure 5.15E). In contrast, p53C4-16, which was able to increase the T_m of T-p53C(Y220C) (Figure 5.15D), was found unable to significantly affect its aggregation propensity. The latter observation is somewhat surprising, as thermal stabilization of T-p53C(Y220C) should theoretically also result in a reduction in its aggregation propensity. However, taking into account that the aggregation assays were performed at 37 °C the reported thermodynamic stabilization of T-p53C(Y220C) may not be sufficient to inhibit its aggregation at this rather elevated temperature.

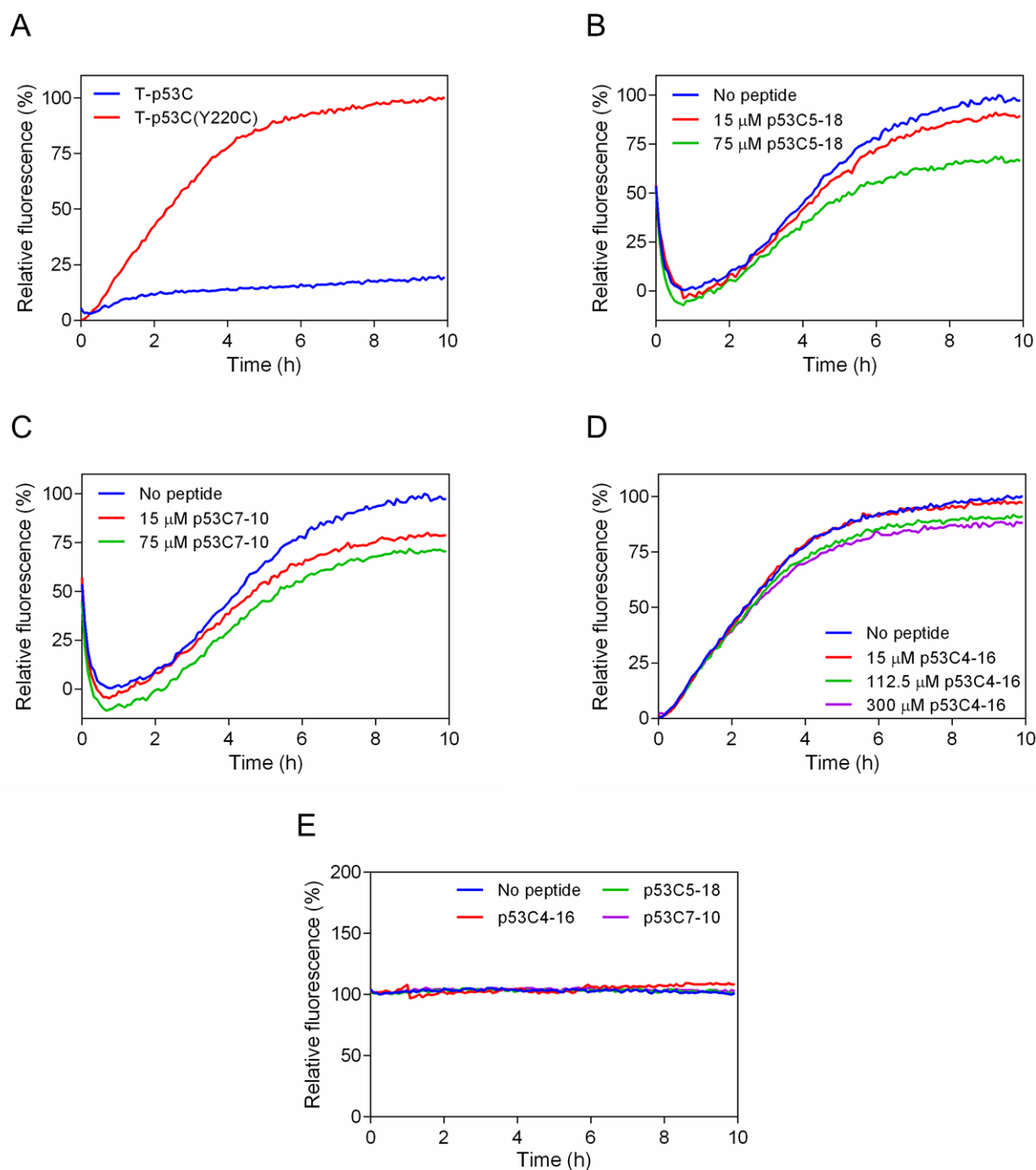


Figure 5.15. Kinetic studies of T-p53C and T-p53C(Y220C) in the absence and presence of the selected peptides. Relative fluorescence of protein samples containing 20 μ M ThT and (A) 15 μ M of T-p53C or T-p53C(Y220C) in the absence of any cyclic peptide, (B) 15 μ M T-p53C(Y220C) in the absence and presence of 1 and 5 molar equivalents of p53C5-18, (C) 15 μ M T-p53C(Y220C) in the absence and presence of 1 and 5 molar equivalents of p53C7-10, (D) 15 μ M T-p53C(Y220C) in the absence and presence of 1, 7.5 and 20 molar equivalents of p53C4-16 and (E) 300 μ M of each selected cyclic peptide in the absence of T-p53C(Y220C). In each panel mean values of one experiment performed in triplicates is presented.

5.6. Discussion

In this chapter we have presented the evaluation of selected hits from the engineered ultrahigh-throughput bacterial screening system for their ability to increase the thermodynamic stability of T-p53C(Y220C) and reduce its aggregation propensity *in vitro*, as well as reduce the cell proliferation of cancer cell lines carrying the p53(Y220C) mutation.

Initially, the selected cyclic peptides were recombinantly produced in *E. coli* and isolated using a one-step intein-mediated purification process and subsequently evaluated for their ability to promote cell death of a cancer cell line carrying the p53(Y220C) mutation. All selected cyclic peptides except of the larger heptapeptide p53C7-10, were found to decrease cancer cell proliferation, with p53C4-16 exhibiting the greatest effect. Synthetic p53C4-16 was subsequently administered to cancer and normal cells in increasing concentrations and was found to selectively decrease tumour cell viability in a dose-dependent manner.

Besides p53C4-16, the rest of the cyclic peptides were not further tested in cancer cell lines, mainly due to persisting difficulties associated with chemical synthesis of these cyclic peptides. While such experiments are planned for the near future, we hypothesize that their preliminary reduced effect, or lack thereof, could be a result of insufficient biosynthetic production, or in the case of p53C7-10, poor cell membrane permeability due to its larger ring size and molecular mass (MW= 775 Da). Therefore, we believe that all selected cyclic peptides have the potential to affect cell viability and should be further tested in cancer cell lines carrying the p53(Y220C) mutation.

In parallel, we report the assessment of the three synthetically produced cyclic peptides p53C4-16, p53C5-18 and p53C7-10, as stabilizers of T-p53C(Y220C) and/or aggregation inhibitors. Interestingly, p53C4-16 was able to increase the protein's T_m by approximately 1 °C, and had a small effect on its aggregation propensity. Contrary, both p53C5-18 and p53C7-

10 were able to inhibit T-p53C(Y220C) aggregation, but not affect its T_m . These observations are very interesting as they indicate that the selected cyclic peptides from the bacterial screen could in fact rescue the protein's misfolding by different mechanisms, acting as either protein stabilizers or aggregation inhibitors.

Chapter 6 – Evaluation of the selected cyclic peptides on their ability to affect A β 42 aggregation and its associated neurotoxicity *in vitro* and *in vivo*

6.1. High-throughput analysis of the isolated hits

As multiple distinct cyclic heptapeptide sequences were identified among the ten selected clones initially tested (**Table 4.2**), we hypothesized that numerous A β 42-targeting macrocyclic sequences may exist among the selected peptide pool. To determine the entire ensemble of potentially bioactive cyclic heptapeptides present in our library, we performed deep sequencing analysis of the heptapeptide-encoding regions in >0.4 million pSICLOPPS-NuX₁X₂X₃X₄X₅X₆ vectors contained in the selected bacterial population after the seventh round of sorting (Figure 4.13). This analysis revealed 416 distinct cyclic heptapeptide sequences appearing at least twenty times within the sorted population, thus indicating that their presence in the selected pool is not coincidental (Appendix B). Indeed, cloning of three randomly chosen cyclic heptapeptide sequences appearing in the sorted pool only with very low frequencies, revealed that they are also efficient in increasing the fluorescence of bacterially expressed A β 42-GFP (Figure 6.1).

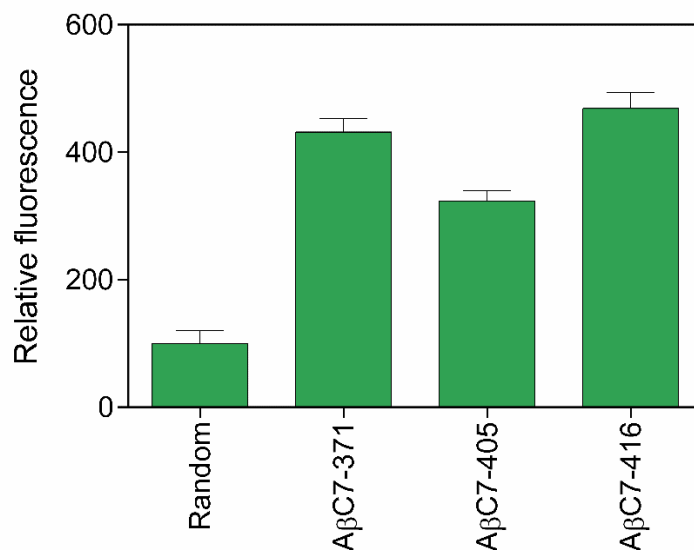
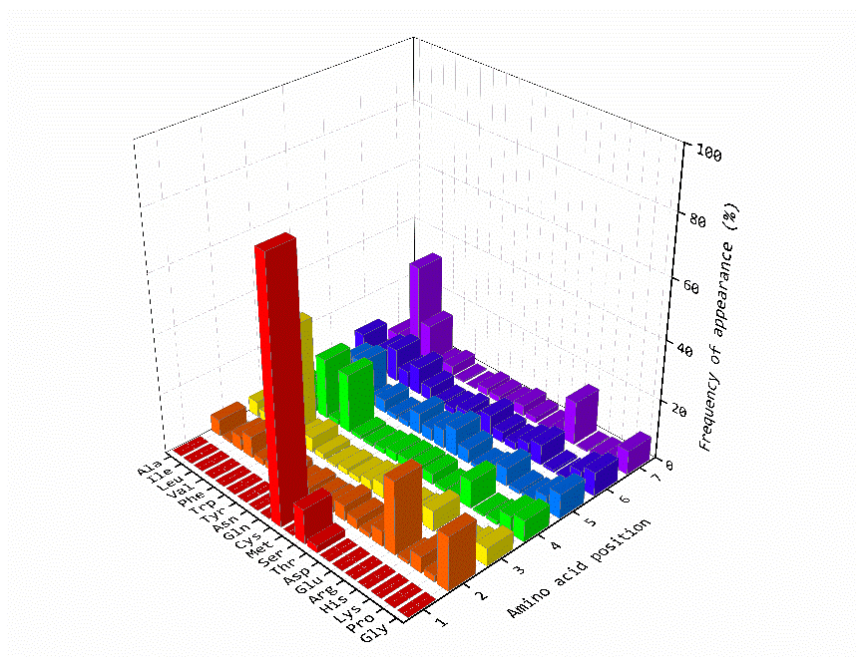


Figure 6.1. Low frequency cyclic heptapeptides from the sorted pool affect the aggregation of Aβ42-GFP. Relative fluorescence of *E. coli* Tuner(DE3) cells co-expressing Aβ42-GFP and three cyclic heptapeptide sequences appearing in the sorted pool only at low frequencies as shown in Appendix B. The fluorescence of the bacterial population producing the random cyclic peptide was arbitrarily set to 100. Experiments were carried out in triplicates and the reported values correspond to the mean value ± s.e.m.

We next performed sequence analysis of the selected cyclic heptapeptides. We found that Cys was the nucleophilic amino acid that was present at position 1 in the vast majority of the selected cyclic heptapeptides (99.6% of all selected sequences) (Figure 6.2A). Furthermore, we observed that the frequency of appearance of only a very small number of specific amino acids was enriched at each position among the selected sequences: Arg and Lys at position 2; Val at position 3; Trp and Thr at position 4; Ile, Gln, Cys, Met, Ser, Thr and Pro at position 5; Ala, Leu, Val, Glu, Lys and Pro at position 6; and Ile, Leu and Pro at position 7 (Figure 6.2B and Table 6.1). On the contrary, the majority of amino acids, including the ones that were present in higher abundance in the initial library, were strongly de-enriched (Figure 6.2B and Table 6.1), thus indicating a highly efficient selection process.

A



B

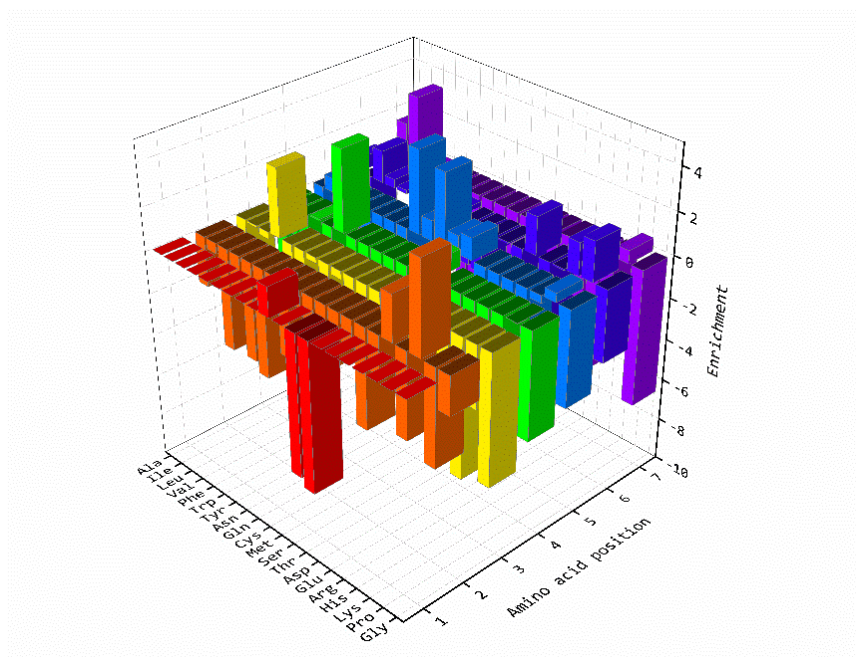
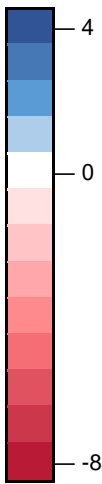


Figure 6.2. Sequence analysis of the selected heptapeptides. (A) Frequency of appearance of the 20 natural amino acids at each position of the heptapeptide sequences selected after the seventh round of sorting (Figure 4.13). (B) Enrichment of the 20 natural amino acids at each position of the heptapeptide sequences selected after the seventh round of sorting (Figure 4.13). Values represent the log₂-fold change of the amino acid frequency of appearance of the peptides from the sorted pool compared to the initial library.

Table 6.1. Enrichment (blue) and depletion (red) of the 20 amino acids in each position of the heptapeptide sequences. Values represent the log2 fold change of the amino acid distribution of peptides from the selected pool compared to the initial library.

		Position						
		1	2	3	4	5	6	7
Amino Acid	A		-2.4	-5.1	-5	-2.9	0.6	-0.6
	I		-0.5	-0.1	-1.3	0.7	-3.7	1.8
	L		-4.8	-0.3	-3.2	-1.9	1.5	3.3
	V		-1.8	3.2	0.1	-1.6	0.4	0.1
	F		-4.7	-3.9	-4.6	-5	0.2	-0.3
	W		-5.2	-2.9	3.8	-2.1	-1.9	-3.2
	Y		-1.2	-0.3	-4.8	-2.2	0.1	-7.3
	N		-3.5	-2	-3.9	-1.2	-2.1	-7.4
	Q		-0.8	-0.4	-4.9	3.9	0.2	-1.5
	C	1.4	-1.3	-3.4	-5.6	0.9	-2	-4.8
	M		-3	-2.5	-2.5	3.4	-2.1	-2.6
	S	-6.7	-0.9	-1.6	-4.8	0.8	-0.7	-6.7
	T	-7.1	-1.3	-4.2	0.8	1	-0.1	-1.2
	D		-4.7	-4.4	-5.9	-6.6	-2.3	-8.1
	E		-1.8	-3.6	-6.2	-4.7	1.8	-6.9
	R		2.1	-5	-2.5	-6.2	-4.1	-2.7
	H		-4	-5.6	-5.7	-6.2	-0.1	-5.6
	K		4.2	-6.6	-5.6	-0.4	1.5	-3.3
	P		-4.5	-2.9	0	0.4	1.9	0.6
	G		-1.8	-6.2	-5.1	-4.6	-3.4	-6.5



In order to identify potential relationships among the selected cyclic heptapeptides, we carried out sequence similarity analysis and hierarchical clustering. As the similarity analysis is performed using linear sequences, all possible circular permutations of each selected cyclic heptapeptide were taken into consideration (Figure 6.3). From the 416 cyclic heptapeptides selected, 323 of them formed 1,467 unique pairs with more than 70% sequence identity and formed twenty distinct clusters with similar sequence characteristics (Figure 6.4 and Table 6.2). Clusters I and II were the most dominant, comprising 75.0% and 4.6% of the selected bacterial clones respectively, as well as 25.7% and 6% of the unique cyclic heptapeptide sequences selected (Figure 6.5, Table 6.2 and Appendices C, D). The majority of peptides from Clusters I and II appeared to belong to a cyclo-CxVWxxx and a cyclo-CxxVPSx motif, respectively, in agreement with our previous observations (Table 4.2).

Cyclic heptapeptide



Linear representations

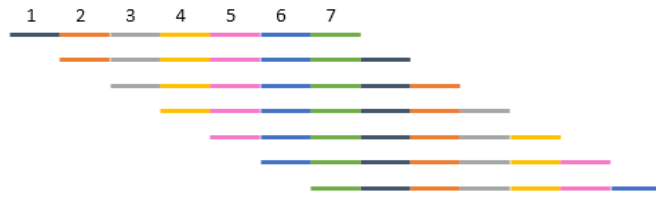


Figure 6.3. Schematic of the linear representations (circular permutants) of a cyclic heptapeptide.

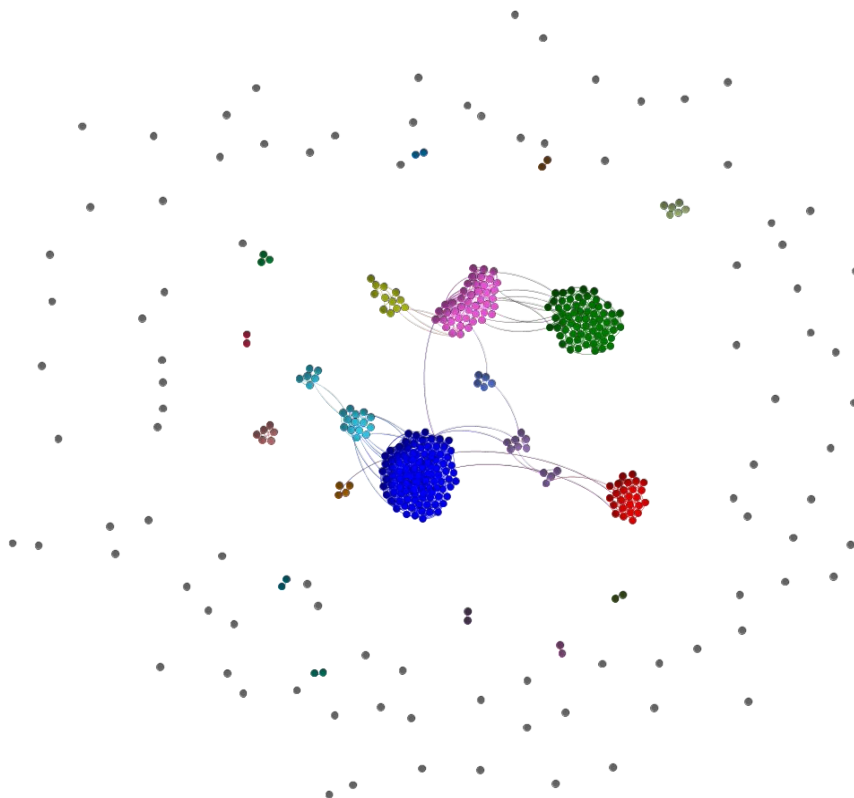


Figure 6.4. Network visualization of all clusters identified using the Girvan-Newman algorithm of the Gephi software. Nodes represent different cyclic peptide sequences and solid lines connect pairs of peptides that share at least 70% sequence identity. Grey nodes represent cyclic heptapeptides that do not share at least 70% homology with any other cyclic peptide from the selected pool.

Table 6.2. Distribution of the heptapeptide sequences in the different clusters identified.

Number	Cluster name	Number of reads	Number of unique peptides	Cluster Reads/Total reads (%)	Peptides/Total peptides (%)
1	I	303,245	107	75.01	25.72
2	II	19,690	25	4.87	6.01
3	III	16,667	64	4.12	15.38
4	IV	6,047	44	1.5	10.58
5	V	5,362	20	1.33	4.81
6	VI	2,837	12	0.7	2.88
7	VII	2,192	6	0.54	1.44
8	VIII	1,987	5	0.49	1.2
9	IX	1,700	2	0.42	0.48
10	X	1,545	6	0.38	1.44
11	XI	1,290	4	0.32	0.96
12	XII	611	2	0.15	0.48
13	XIII	606	11	0.15	2.64
14	XIV	482	2	0.12	0.48
15	XV	258	3	0.06	0.72
16	XVI	235	2	0.06	0.48
17	XVII	95	2	0.02	0.48
18	XVIII	86	2	0.02	0.48
19	XIX	54	2	0.01	0.48
20	XX	42	2	0.01	0.48
21	Singletons	39,249	93	9.71	22.36
Sum		404,280	416	100	100

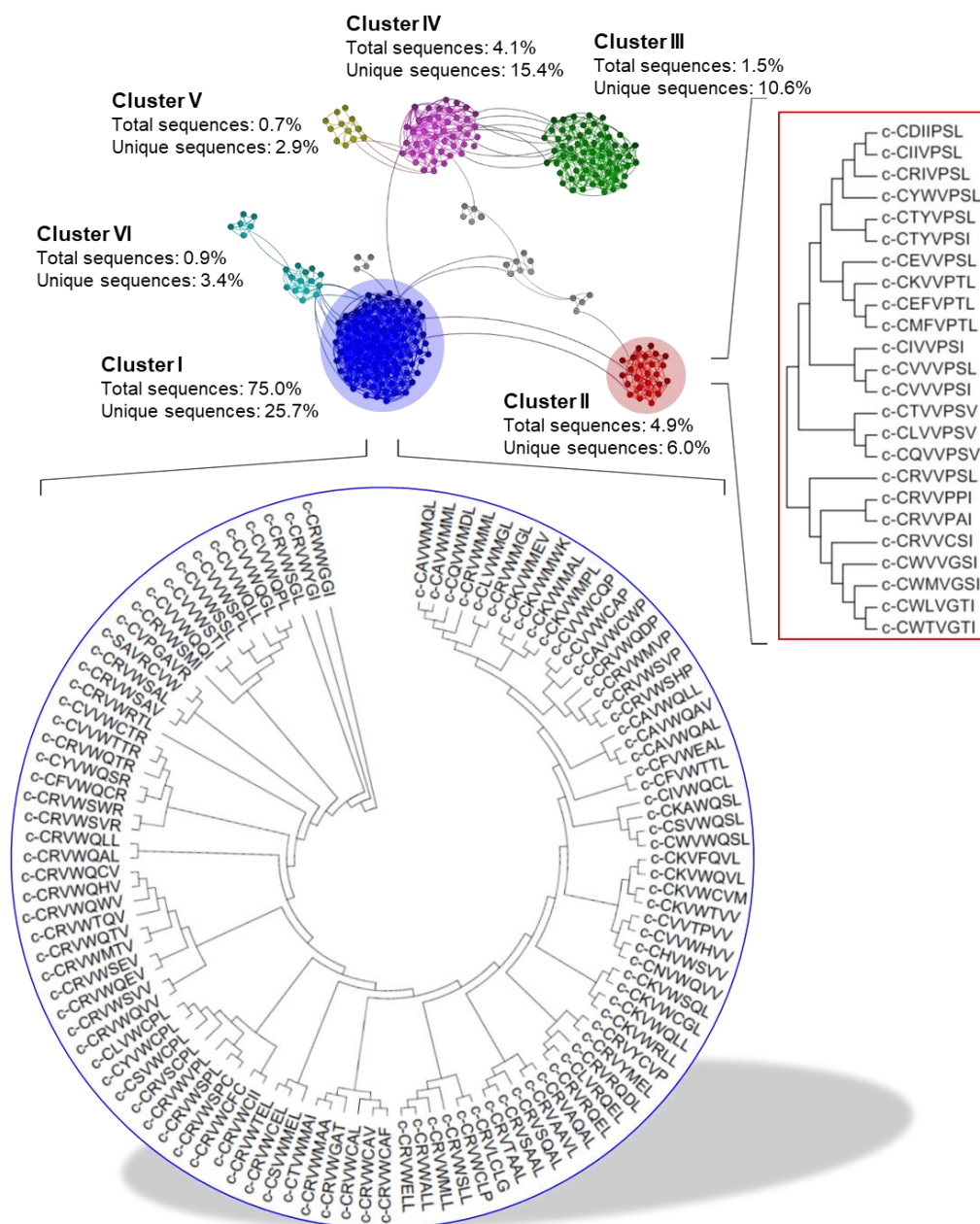


Figure 6.5. Visualization of the main clusters formed by the selected cyclic heptapeptides according to their sequence similarities, as in Figure 6.4 . The sequences of the members of the two most dominant clusters (Clusters I and II) are shown in the corresponding dendrograms.

Two of the selected heptapeptides, cyclo-CKVWQLL and cyclo-CRIVPSL, termed A β C7-1 and A β C7-14 (A β -targeting cyclic 7-peptide number 1 and 14), respectively (Figure 6.6), were chosen for subsequent analysis and were synthesized chemically in mg quantities. These cyclic peptides were selected because they were both encountered in the post-selection

pool investigated initially (Table 4.2) and, more importantly, they were the most frequently encountered members among the two most dominant clusters (Clusters I and II) (Appendices C, D).

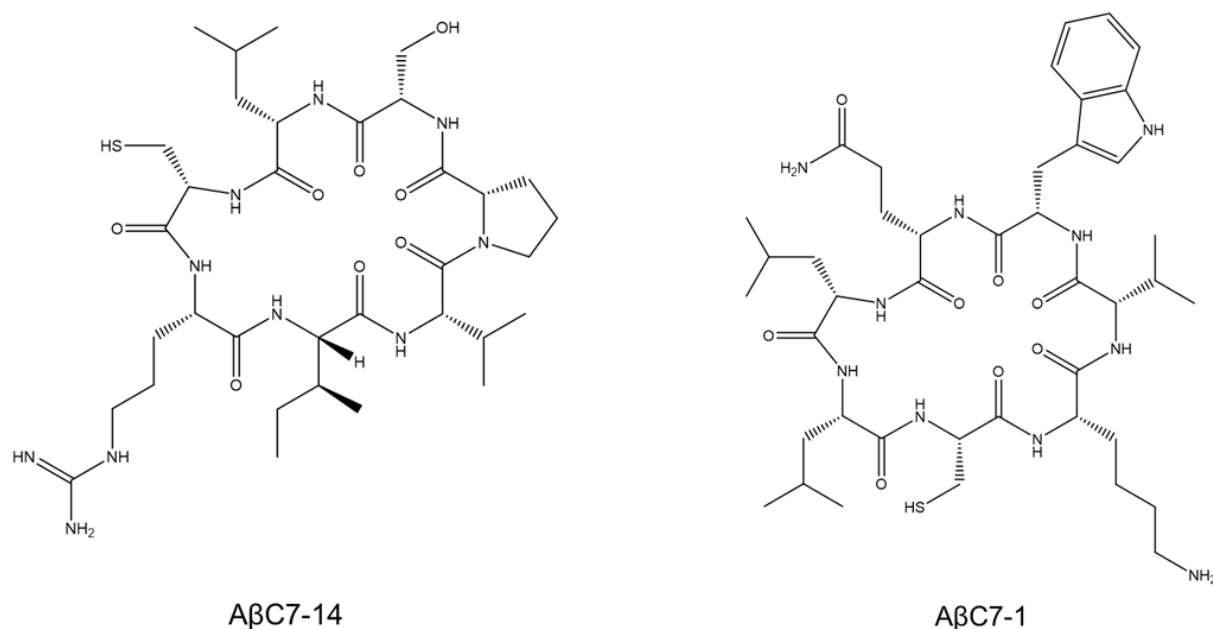
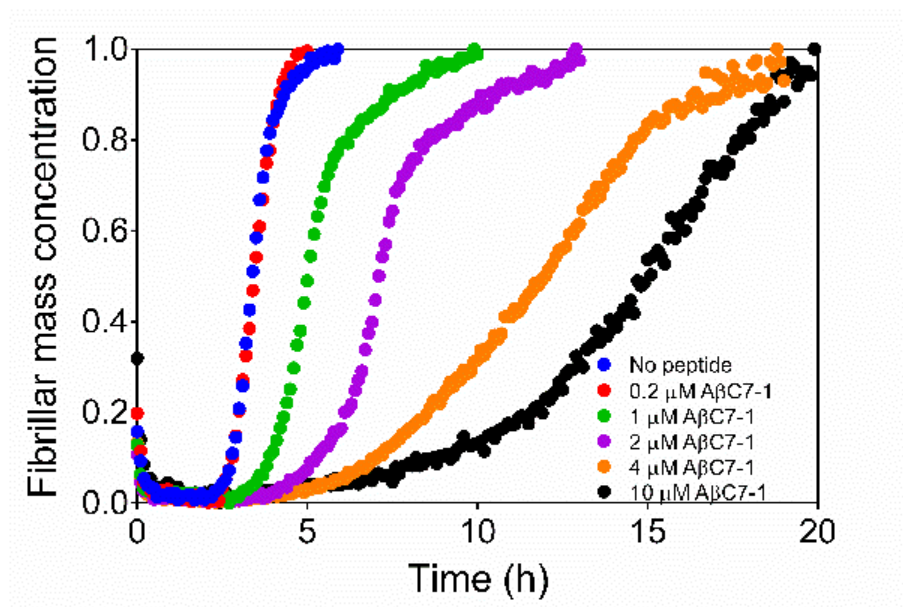


Figure 6.6. Chemical structures of the selected cyclic heptapeptides AβC7-1 and AβC7-14.

6.2. *In vitro* evaluation of AβC7-1 and AβC7-14

The ability of the selected cyclic heptapeptides to inhibit the Aβ42 aggregation was determined using a highly reproducible approach previously described based on monitoring the kinetics of Aβ42 aggregation by thioflavin T (ThT) staining [339,340]. Monomeric Aβ42 was purified after recombinant production in *E. coli* and kinetic reactions were initiated using 2 μM Aβ42 in the absence and presence of AβC7-1 and AβC7-14. Both cyclic heptapeptides inhibited the aggregation of Aβ42 very effectively at sub-stoichiometric ratios as low as 0.5 molar equivalents for AβC7-1 and 0.1 for AβC7-14 (Figure 6.7).

A



B

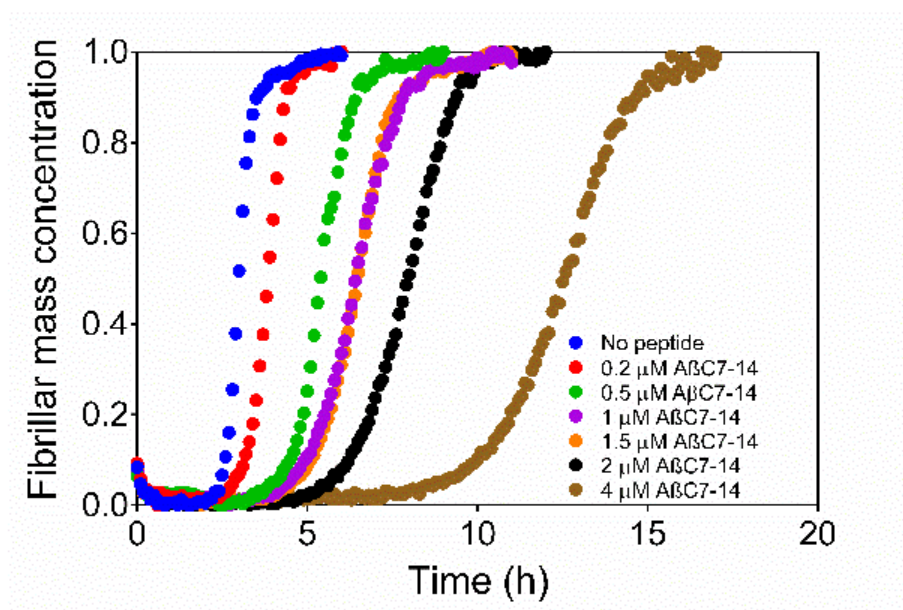
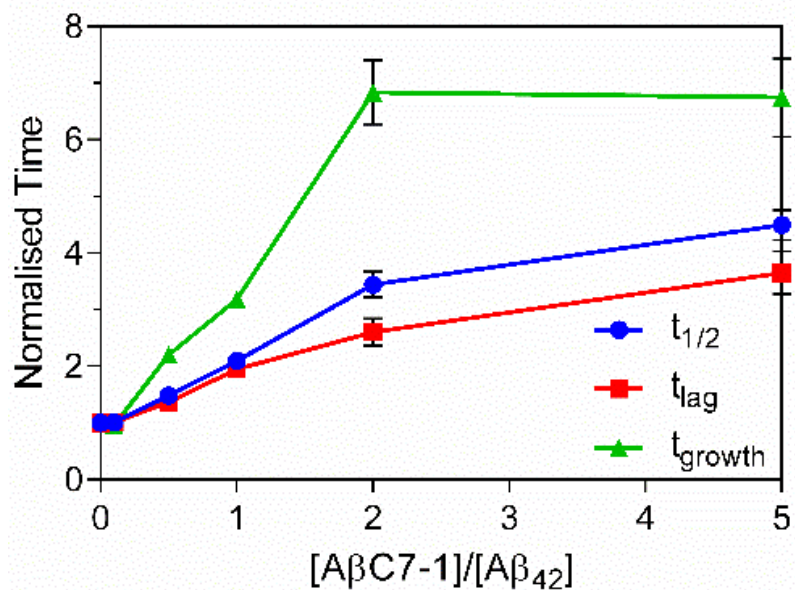


Figure 6.7. Kinetic profiles of the aggregation of 2 μM Aβ42 in the absence and presence of (A) AβC7-1 or (B) AβC7-14, at different molar ratios. Mean values of one experiment performed in triplicates are presented. Experiments were performed in collaboration with Sean Chia and under the supervision of Dr. Johnny Habchi and Prof. Michele Vendruscolo at the Centre for Misfolding Diseases of the University of Cambridge.

Specifically, we found that both the t_{lag} (time required for the ThT fluorescence to reach 10% of the total amplitude) and t_{growth} (transition time from 10% to 90% of the total ThT fluorescence amplitude) of the A β 42 aggregation reaction were increased in the presence of the two selected macrocycles, suggesting that they are able to inhibit both the primary and secondary nucleation steps of A β 42 aggregation, albeit to a different extent (Figure 6.8).

A



B

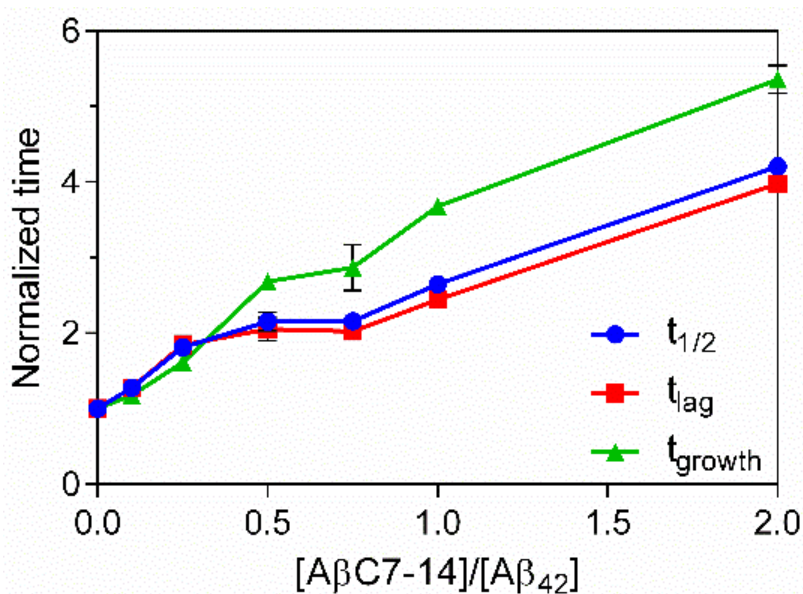


Figure 6.8. Normalised $t_{1/2}$, t_{lag} and t_{growth} values of the kinetic aggregation reactions from (Figure 6.7) in the presence of (A) $A\beta C7-1$ or (B) $A\beta C7-14$. Mean values \pm sem of one experiment performed in triplicates are presented. Experiments were performed in collaboration with Sean Chia and under the supervision of Dr. Johnny Habchi and Prof. Michele Vendruscolo at the Centre for Misfolding Diseases of the University of Cambridge.

Furthermore, we found that A β 42 fibrils formed after the completion of the aggregation reaction in the absence and presence of both A β C7-1 and A β C7-14 were similar in both size and morphology (Figure 6.9). Thus, it is likely that these selected macrocycles are not binding irreversibly to A β 42 species, and redirecting the aggregation process towards off-pathway aggregates. Importantly, the observed deceleration of A β 42 aggregation by the selected macrocycles could also be observed in the absence of ThT, when the progress of the aggregation was monitored by extracting aliquots at different time points and probing fibril formation by dot blotting using the fibril-specific OC antibody (Figure 6.10), therefore excluding the possibility of interference of the two selected compounds with ThT binding to A β 42 fibrils.

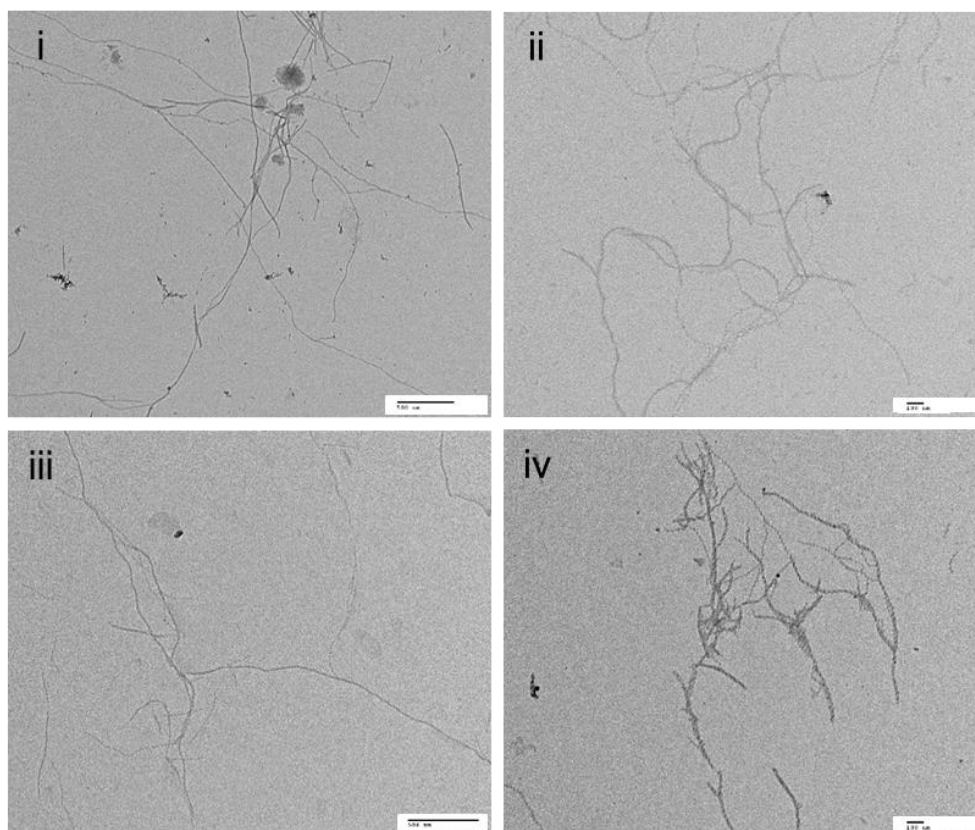


Figure 6.9. Representative TEM images of 2 μM A β 42 fibrils in the absence (left) and presence (right) of either 10 μM A β C7-1 (top) or 4 μM A β C7-14 (bottom). Samples were taken at the reaction end point as measured by thioflavin T (Figure 6.7). Experiments were performed in collaboration with Sean Chia and Rodrigo Lessa Cataldi, and under the supervision of Dr. Johnny Habchi and Prof. Michele Vendruscolo at the Centre for Misfolding Diseases of the University of Cambridge.

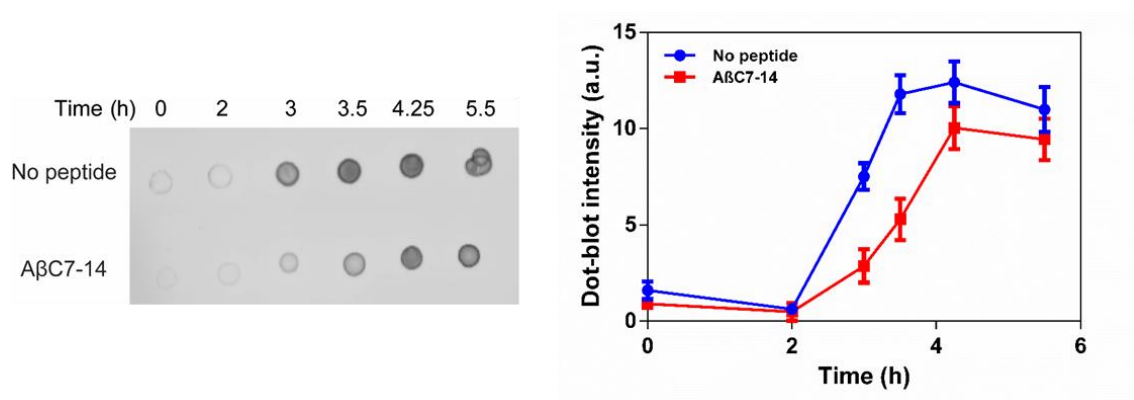


Figure 6.10. Time course of the aggregation of 2 μM A β 42 in the presence and absence of 0.5 μM A β C7-14 using a dot blot assay and the A β 42 fibril-specific antibody OC. Mean values \pm sem of one experiment performed in triplicates are presented. Experiments were performed in collaboration with Sean Chia and under the supervision of Dr. Johnny Habchi and Prof. Michele Vendruscolo at the Centre for Misfolding Diseases of the University of Cambridge.

6.3. *In vivo* evaluation of A β C7-1 and A β C7-14

To evaluate the effects of A β C7-1 and A β C7-14 *in vivo*, we tested their impact on A β 42 aggregation and its associated toxicity in an established *C. elegans* model of Alzheimer's disease. We used GMC101, a transgenic strain expressing human A β 42 in body-wall muscle cells under the control of a heat-inducible promoter [341]. Upon temperature up-shift, these nematodes (hereafter referred to as A β worms) exhibit muscle-localized A β 42 aggregation and eventually the emergence of a paralysis phenotype [341]. Since the *in vitro* results suggested that the two compounds affect the early stages of A β 42 aggregation, A β C7-1 and A β C7-14 were administered to the A β worms before aggregation was initiated. The fitness of the A β worms - defined as the frequency and speed of body bends - was monitored in the absence and presence of A β C7-1 and A β C7-14 and compared to wild-type nematodes, which do not express A β 42. Both peptides increased the motility and speed of the A β worms throughout their lifetime, with higher activity at day 7 of adulthood, probably due to peptide degradation and/or depletion at later days (Figure 6.12).

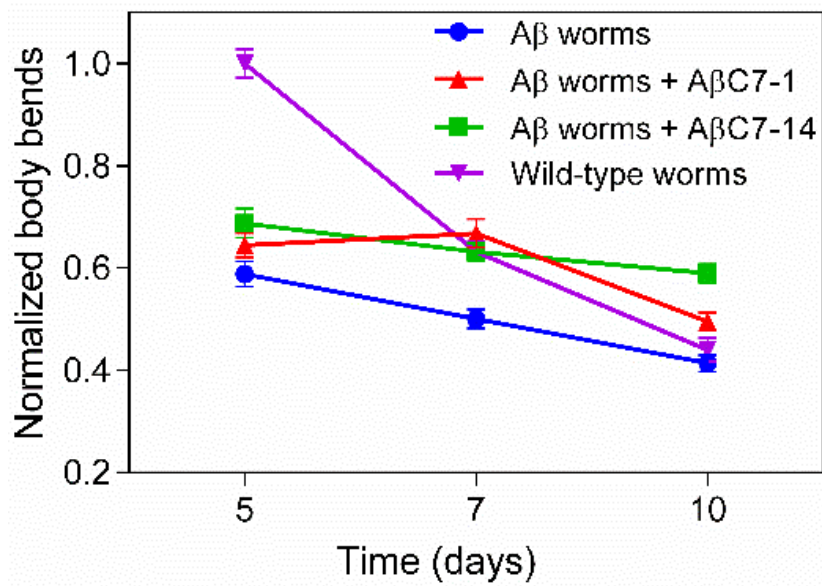
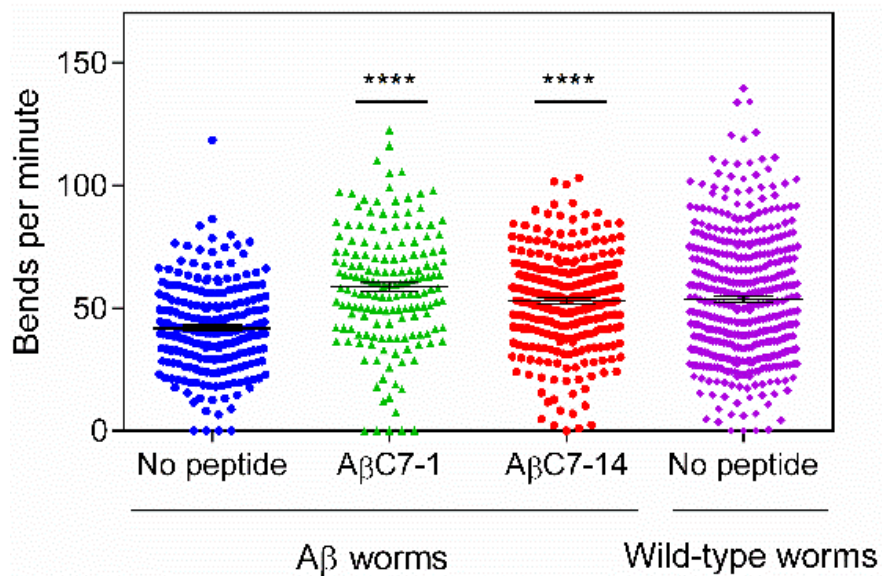
A**B**

Figure 6.11. A β C7-1 and A β C7-14 increase the motility of A β worms. (A) Normalised motility of A β and wild-type worms in the absence and presence of 40 μ M A β C7-1 and 5 μ M A β C7-14 during 5-10 d of adulthood. (B) Motility of individual A β and wild-type worms in the absence and presence of A β C7-1 and A β C7-14 at 7 d of adulthood. In all panels \sim 200 worms were analysed on average and mean values \pm s.e.m. are presented (n=number of worms tested in one experiment). Experiments were performed under the supervision of Michele Perni and Prof. Michele Vendruscolo at the Centre for Misfolding Diseases of the University of Cambridge.

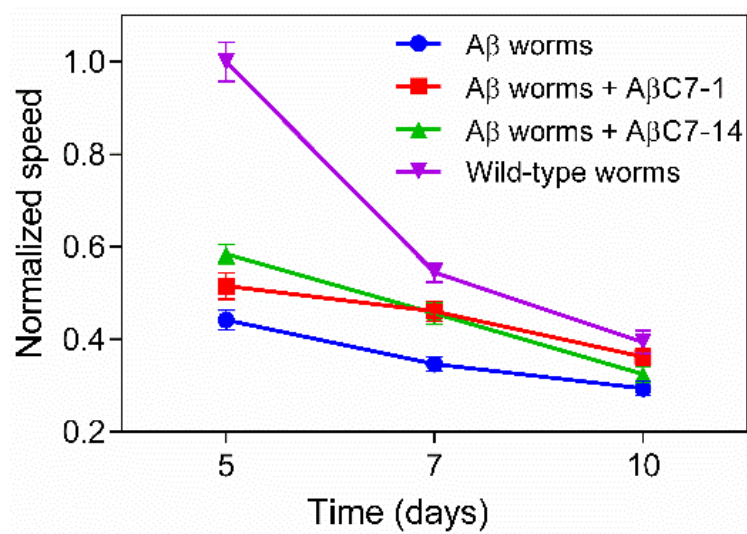
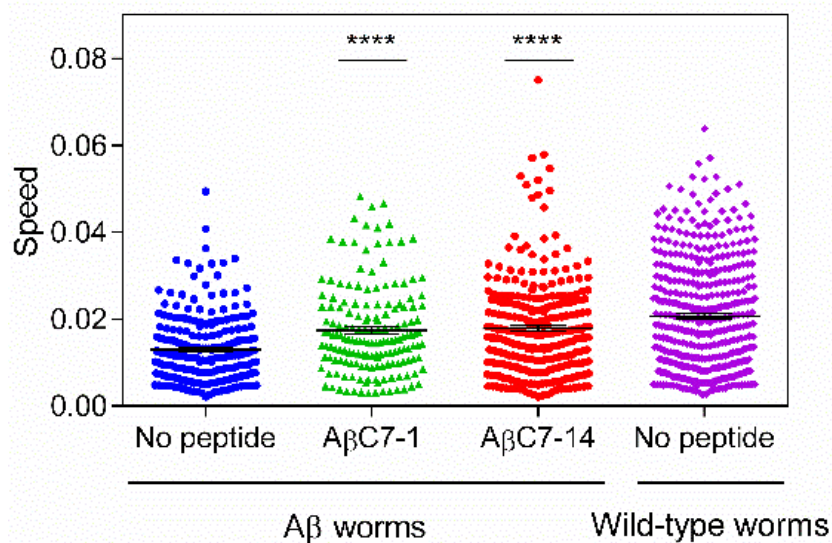
A**B**

Figure 6.12. A β C7-1 and A β C7-14 increase the speed of movement of A β worms. (A) Normalised speed of movement of A β and wild-type worms as in Figure 6.11. (B) Speed of individual A β and wild-type worms as in Figure 6.11. In all panels \sim 200 worms were analysed on average and mean values \pm s.e.m. are presented (n =number of worms tested in one experiment). Statistical significance is denoted by: **** P \leq 0.0001, for differences to the “No peptide A β worms” sample. Experiments were performed under the supervision of Michele Perni and Prof. Michele Vendruscolo at the Centre for Misfolding Diseases of the University of Cambridge.

Notably, both peptides were able to restore the total fitness of the A β worms to approximately the levels of the wild-type animals (Figure 6.13).

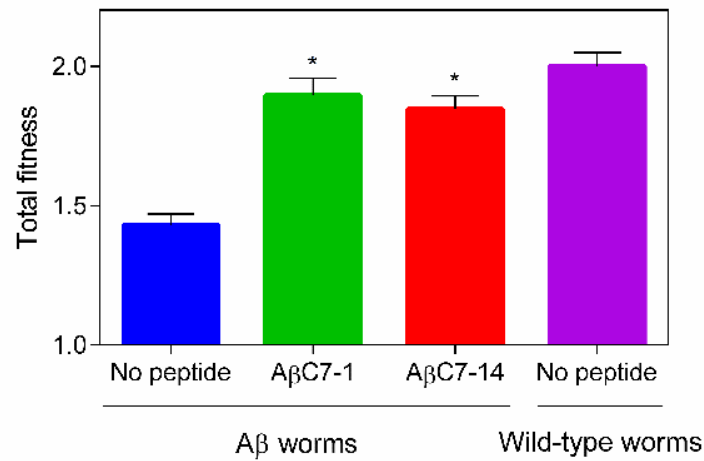


Figure 6.13. Total fitness of individual A β and wild-type worms in the absence and presence of A β C7-1 and A β C7-14 at 7 d of adulthood. Mean values \pm s.e.m. are presented. Statistical significance is denoted by: * $P \leq 0.05$ for differences to the “No peptide A β worms” sample. Experiments were performed under the supervision of Michele Perni and Prof. Michele Vendruscolo at the Centre for Misfolding Diseases of the University of Cambridge.

Furthermore, A β worms treated with either one of the selected cyclic peptides produced 50-60% fewer A β 42 aggregates, as determined by imaging of the worms using the amyloid-specific dye NIAD-4 (Figure 6.14).

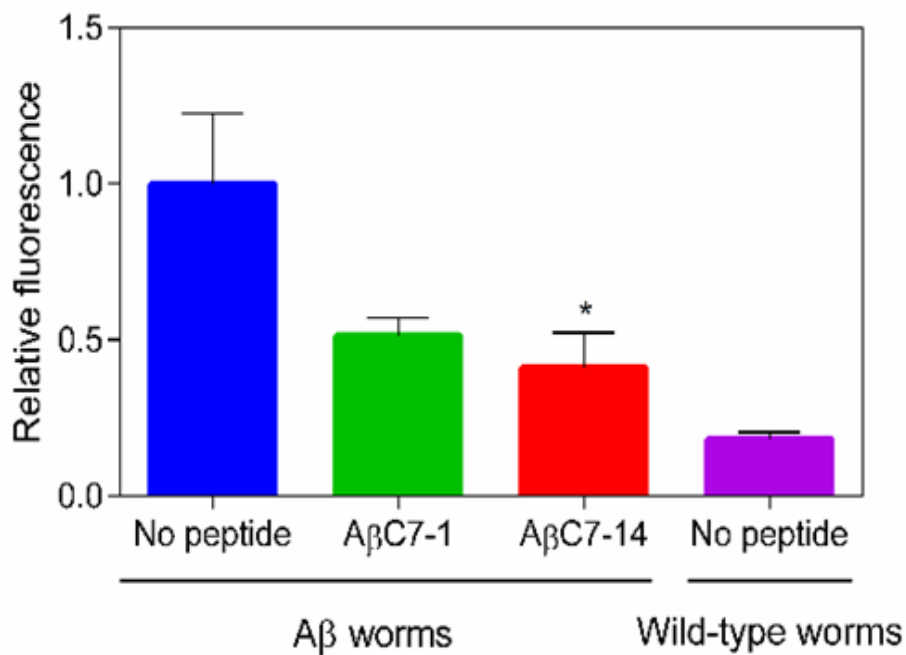
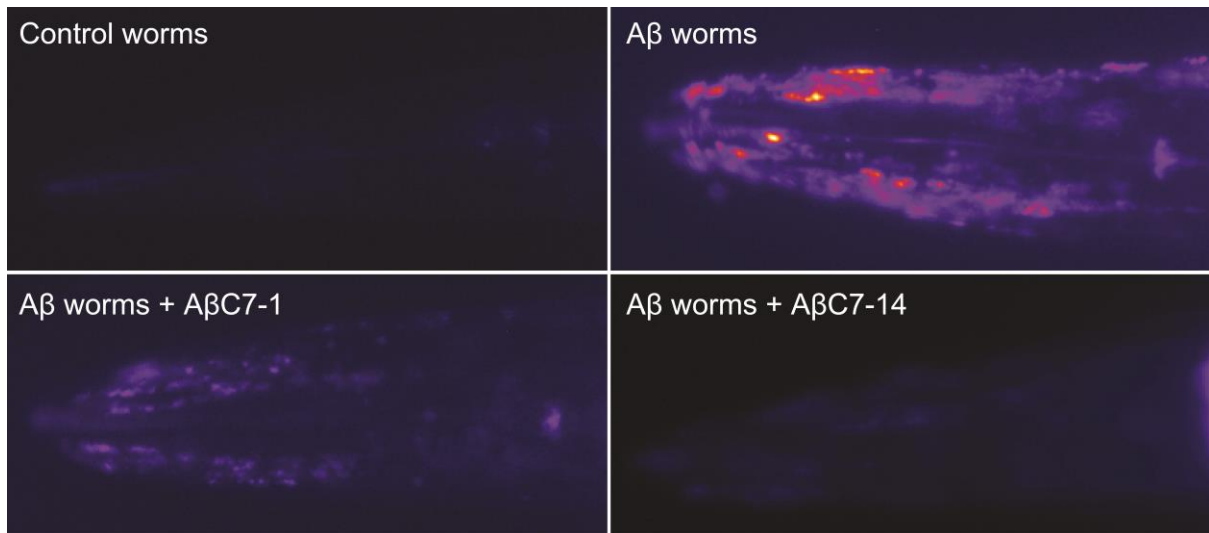


Figure 6.14. Imaging of A β and wild-type worms in the absence and presence of A β C7-1 and A β C7-14 at 7 d of adulthood. (top) Representative images from A β and wild-type worms at 7 d of adulthood in the presence and absence of 40 μ M A β C7-1 and 5 μ M A β C7-14. (bottom) Relative fluorescence of A β 42 and wild-type worms at 7 d of adulthood showing a 50-60% decrease in A β aggregate formation in the presence of A β C7-1 and A β C7-14. 25 worms were analysed in total. Mean values \pm s.e.m. are presented (n=number of worms tested in one experiment). Statistical significance is denoted by: * $P \leq 0.05$ for differences to the “No peptide A β worms” sample. Experiments were performed under the supervision of Michele Perni and Prof. Michele Vendruscolo at the Centre for Misfolding Diseases of the University of Cambridge.

In order to exclude the possibility of promoter- or strain-specific effects, we treated the transgenic *C. elegans* strain CL4176 with A β C7-1 and A β C7-14, which expresses human A β 42 in its body wall muscle cells under a different promoter [342]. Consistently with our previous observations, the administration of either one of the cyclic peptides resulted in a significant delay in the emergence of its characteristic paralysis phenotype (Figure 6.15). These results demonstrate the protective effect of the two cyclic peptides in the context of an animal, as shown by decrease of A β 42 deposits, increased locomotion, delay of paralysis and recovery of total fitness.

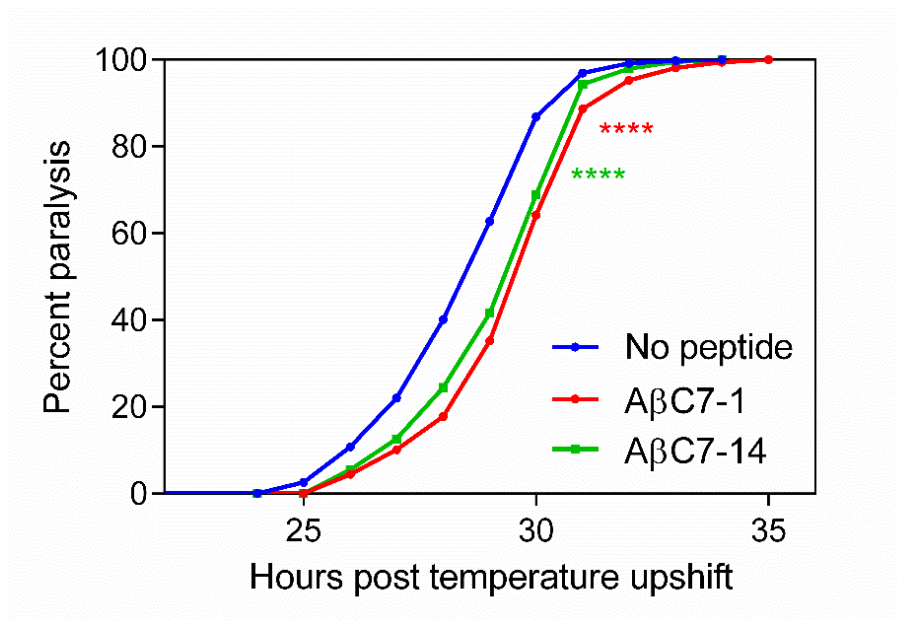


Figure 6.15. Paralysis curves of *C. elegans* CL4176 expressing human A β 42 and treated with 10 μ M of A β C7-1 and A β C7-14. No peptide: mean=28.4 \pm 0.2, n=863/867; A β C7-1: mean=29.5 \pm 0.1, n=867/873, P<0.0001; and A β C7-14: mean=29.2 \pm 0.1, n=887/893, P<0.0001. Experiments were performed by Nikoletta Papaevgeniou under the supervision of Dr. Niki Chondrogianni at the NHRF.

6.4. Structure-activity relationships of A β C7-1 and A β C7-14

In order to identify the functionally important residues within the selected peptides, we performed nucleophile substitutions at position 1 and Ala scanning mutagenesis at positions 2-7 for both A β C7-1 and A β C7-14. Then, we compared the effect of these amino acid substitutions on the levels of bacterially expressed A β 42-GFP fluorescence and aggregation with those of the selected sequences (positive control) and of random cyclic peptide sequences (negative control). For both A β C7-1 and A β C7-14, the substitution of Cys at position 1 with Ser resulted in ~50% reduction in fluorescence, while the substitution with Thr resulted in levels of A β 42-GFP fluorescence and aggregation similar to those corresponding to the selected sequence (Figure 6.16 and Figure 6.18). The latter observation is somewhat surprising, considering the dominant appearance of Cys1 sequences among the selected cyclic heptapeptide pool (Figure 6.2A). However, it may be related to previous results from our lab, where FACS sorting of the cyclic tetra-, penta- and hexapeptides resulted in the identification of cyclic pentapeptides against A β 42 aggregation corresponding to the cyclo-TXXXXR motif (X is anyone of the 20 natural amino acids) [330]. Furthermore, for both peptides, Ala scanning mutagenesis at the majority of the positions 2-7 resulted in significant A β 42-GFP fluorescence decrease and concomitant increase in aggregation (Figure 6.16 and Figure 6.18). Specifically, for A β C7-1, substitutions at positions 2, 3, 4 and 7 resulted in a ~30-70% decrease in A β 42-GFP fluorescence, while for A β C7-14, substitutions at all positions except Ser6, resulted in a ~45-80% decrease (Figure 6.16, Figure 6.17 and Figure 6.18).

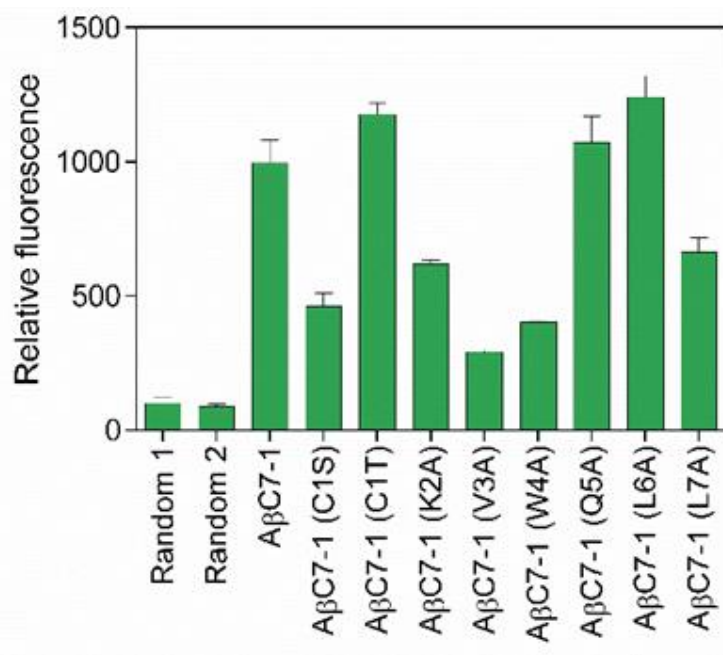


Figure 6.16. Relative fluorescence of *E. coli* Tuner(DE3) cells overexpressing Aβ42-GFP and AβC7-1 or the indicated variants thereof as measured by flow cytometry. The fluorescence of the bacterial population co-producing the random cyclic peptide was arbitrarily set to 100. Experiments were carried out in triplicates (n=1 independent experiments) and the reported values correspond to the mean value ± s.e.m.

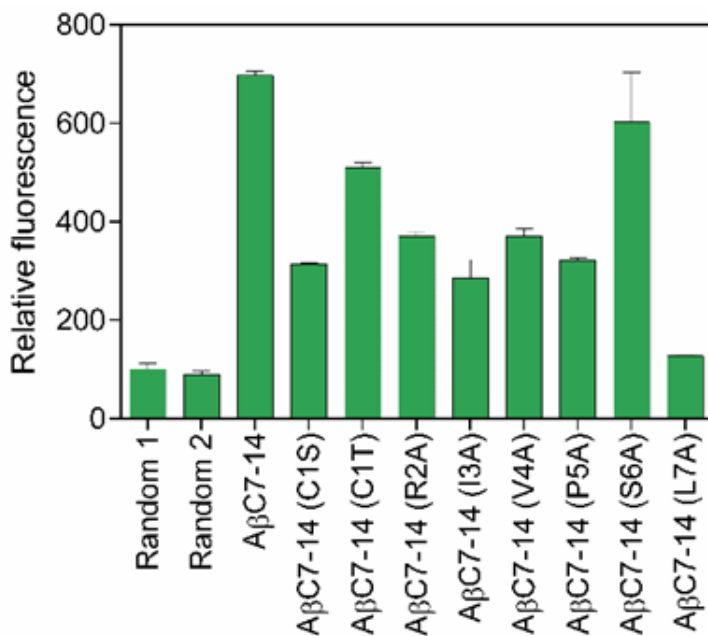


Figure 6.17. Relative fluorescence of *E. coli* Tuner(DE3) cells overexpressing Aβ42-GFP and AβC7-14 or the indicated variants thereof as measured by flow cytometry. The fluorescence of the bacterial population co-producing the random cyclic peptide was arbitrarily set to 100. Experiments were carried out in triplicates (n=1 independent experiments) and the reported values correspond to the mean value ± s.e.m.

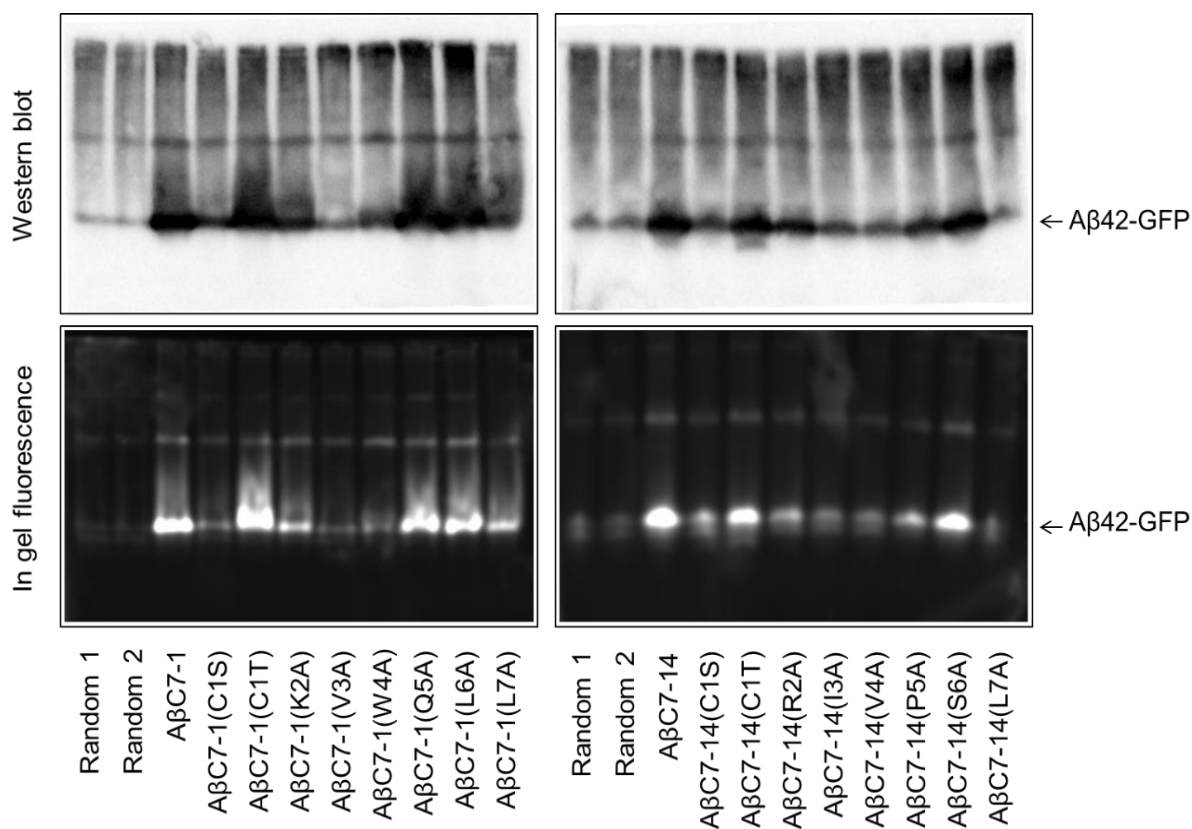


Figure 6.18. Solubility analysis of *E. coli* Tuner(DE3) cells overexpressing A β 42-GFP and the two selected cyclic peptide sequences along with the indicated variants. Total lysates of cells co-expressing A β 42-GFP and the indicated variants of A β C7-1 (left) or A β C7-14 (right) as in Figure 6.16 and Figure 6.17, were separated by native-PAGE and visualized by western blotting using the anti-A β antibody 6E10 (top) and in-gel fluorescence (bottom).

These observations indicate that a number of residues in both selected cyclic heptapeptides are important for optimal aggregation inhibition activity. Indeed, when we performed sequence analysis of all the selected sequences belonging to either Cluster I or Cluster II, we found that the peptides appearing most frequently in each cluster have strong preferences for specific amino acids at each position. More specifically, for Cluster I, Arg and Lys at position 2 appeared in >90% of the selected peptides, while Val at position 3, Trp at position 4, Gln, Cys, Ser, Met and Thr at position 5 and Ile, Val and Leu at position 7 appeared in >99% of the selected clones (Figure 6.19A and Appendix C). Similarly, for Cluster II, the

frequency of appearance of Arg, Ile, Val and Gln at position 2 was ~93%, whereas for Ile and Val at position 3, Val at position 4, Pro at position 5, Ser and Ala at position 6 and Ile, Leu and Val at position 7 was >97% (Figure 6.19B and Appendix D). Taken together, our results indicate that the most bioactive motifs against A β misfolding and aggregation in the investigated macrocycle library are cyclo-(C,T)(R,K)VW(Π ,A,M)X(Ψ ,P) and cyclo-(C,T) Δ (I,V)VP(S,A) Ψ , for Clusters I and II, respectively, where X is anyone of the 20 natural amino acids; Π : any one of the polar amino acids Q, C, S and T; Δ : R, I, V or Q; and Ψ : any one of the aliphatic amino acids L, V and I.

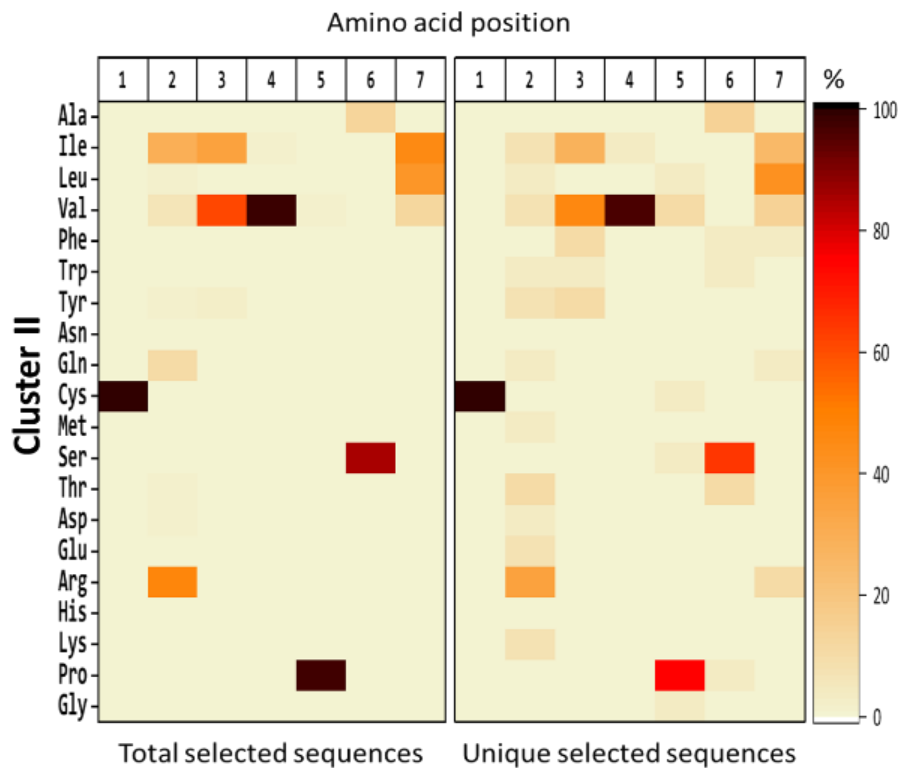
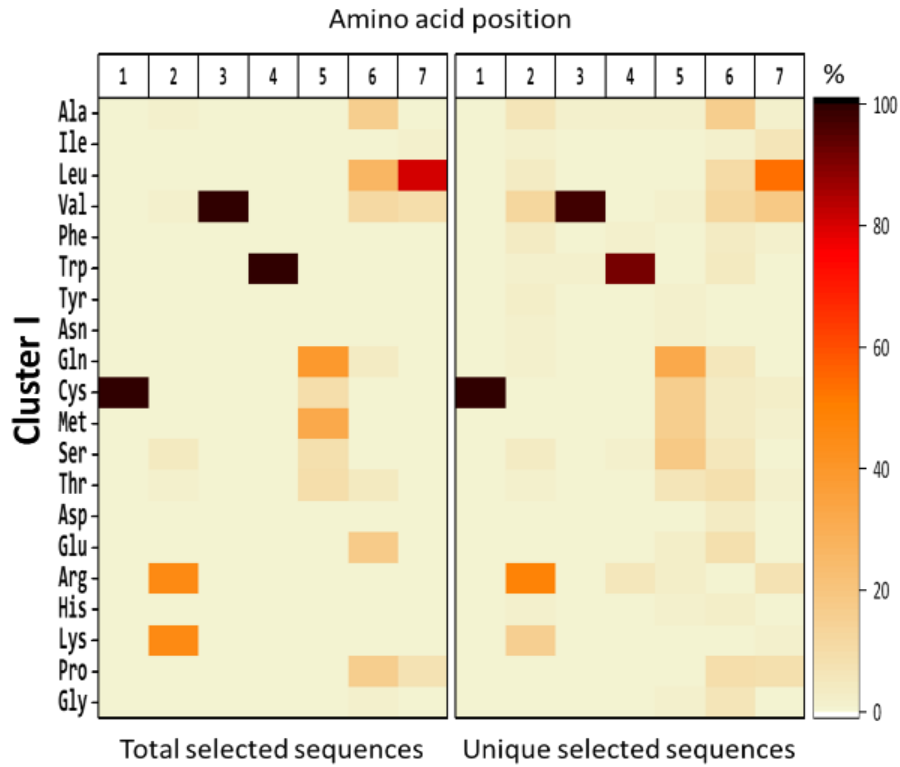


Figure 6.19. Heat map representation of the amino acid distribution at each position of the peptide sequences corresponding to Cluster I (top) and Cluster II (bottom), as demonstrated by the deep sequencing analysis results. In each panel, the total (left) or the unique (right) heptapeptide sequences were included in the analysis.

6.5. Discussion

In this chapter we present the evaluation of the selected population of cyclic peptides after FACS screening against A β 42 aggregation. By coupling the screening methodology described in Chapter 4 with high-throughput sequencing analysis of the isolated hits, we were able to identify over 400 cyclic heptapeptides with the potential to inhibit A β 42 aggregation. Using sequence similarity analysis these heptapeptides were divided into twenty clusters, with distinct sequence characteristics.

In order to verify the bioactivity of the selected heptapeptides, we chose the most frequently encountered member from each of the two most dominant clusters (Clusters I and II) for further *in vitro* and *in vivo* analysis. Indeed, by utilizing a highly reproducible approach, developed by our collaborators at the University of Cambridge, for monitoring the kinetics of A β 42 aggregation *in vitro*, we were able to verify that both cyclic heptapeptides inhibit the aggregation of A β 42 very effectively at sub-stoichiometric ratios. Furthermore, the two selected cyclic peptides appear to inhibit both the primary and secondary nucleation steps of A β 42 aggregation, albeit to a different extent, suggesting that members from different clusters may function by different inhibitory mechanisms.

Subsequently, both selected heptapeptides were evaluated in two established *C. elegans* models of AD, and were found capable of protecting the animals from the A β 42 aggregation associated cytotoxicity, as they were able to decrease the A β 42 deposits found in the worms' body wall muscle cells, increase their locomotion and delay their paralysis.

After verifying that representative members of the two most dominant clusters were bioactive *in vitro* and *in vivo*, we used a combination of site-directed mutagenesis and deep sequencing analysis to rapidly define the sequence motifs providing optimal bioactivity among each selected cluster. For Clusters I and II these were found to be cyclo-

(C,T)(R,K)VW(Π ,A,M)X(Ψ ,P) and cyclo-(C,T) Δ (I,V)VP(S,A) Ψ , respectively, where X is anyone of the 20 natural amino acids; Π : any one of the polar amino acids Q, C, S and T; Δ : R, I, V or Q; and Ψ : any one of the aliphatic amino acids L, V and I.

Finally, while members from the remaining eighteen clusters were not evaluated *in vitro* or *in vivo*, we believe that these clusters also include cyclic peptides with increased bioactivity and therefore should be tested in future experiments. Interestingly, none of the identified sequence motifs bare any similarity to neither the sequence of A β 42 nor to other macrocyclic A β 42 aggregation modulators that we have previously identified using the same method [330], thus highlighting that they comprise novel potentially therapeutic entities - and not extensions of previously identified bioactive sequences -, whose identification would have been impossible by rational design.

Chapter 7 – Conclusion

The work presented in this thesis describes the development of a generalized ultrahigh-throughput system for monitoring the folding of disease-associated MisPs and identifying cyclic peptides with the ability to rescue the associated protein misfolding and aggregation. Specifically, combinatorial libraries of head-to-tail cyclic tetra-, penta-, hexa- and heptapeptides are biosynthesized in the cytoplasm of *E. coli* cells and simultaneously screened for their ability to correct the problematic folding of disease-associated MisPs, using an ultrahigh-throughput screening assay that links the aggregation of MisP with a selectable phenotype. This approach offers a number of important advantages:

First, it allows the facile screening of macrocyclic libraries with expanded diversities, whose sizes are only limited by the theoretical diversity of the specific library design and the transformation efficiency in *E. coli*. In this thesis, we present the straightforward construction of a high-quality cyclic peptide library with more than 200 million members. By following the same approach, larger libraries with up to 10^{10} members can be easily constructed. Importantly, as *E. coli* cells can be engineered to accommodate a wide range of artificial amino acids, the already constructed gene libraries can be encoded in such environments and produce macrocycles with even higher diversities, as well as better pharmacokinetic properties. In addition, while in this thesis we present the construction of a library of head-to-tail cyclic peptides, the described system can be easily adapted to construct libraries of side-chain-to-backbone cyclic peptides [343], bicyclic peptides [275], lasso peptides[344], θ -defensins[345] and cyclotides[346] among others [347,348], further increasing the diversity of the available test molecules. Therefore, by screening libraries with greatly expanded diversities, the chances

of identifying molecules with the desired properties, especially against difficult targets such as protein-protein interactions and protein aggregation, are significantly enhanced.

Additionally, the herein described system enables the direct functional screening of the constructed library, thus allowing the identification of hits with aggregation inhibitory activity. Compared to other high-throughput approaches frequently utilized for the same purpose, such as phage display, this bacterial system decreases the possibility of identifying strong binders that are either completely inactive or with the opposite effect than the originally intended, and therefore the overall cost and time of the selection process is limited significantly. To corroborate this, even in some cases where the bacterial screening may result in false-positive hits, these can be easily and cost-effectively identified without the need of purified target proteins or synthetic compounds. Furthermore, the incorporation of FACS-based selection into our system has enabled the high-throughput identification of hits, in a greatly accelerated manner compared to other functional multi-well assays, which become impractical for libraries with more than 10^6 - 10^7 members. Specifically, in the case of A β 42, we were able to screen our complete library of >200 million compounds and identify hundreds of bioactive hits in just a few days.

Moreover, our approach enables the simultaneous biosynthesis of the molecular libraries and their direct functional screening inside living cells, therefore eliminating the need for organic synthesis of the test compounds and purification of the target MisP. This advantage is highly important as the synthesis of cyclic peptides using traditional methods is very costly and laborious. Specifically, the average time required for the synthesis of the cyclic peptides that were further evaluated in this thesis was approximately four months, whereas in the case of the three selected cyclic tetrapeptides p53C4-4, p53C4-19 and p53C4-21, all organic synthesis efforts have failed so far. Contrary, using our system we were able to evaluate the aforementioned peptides' ability to rescue protein misfolding and aggregation inside bacterial

cells, as well as isolate biosynthesized cyclic peptides for further *in vitro* evaluation. Furthermore, our system also eliminates the need for reproducible purification of the target protein, which in the case of MisPs can be very laborious and frequently result in inaccurate findings [340].

Another advantage of the described approach is that it enables the identification of bioactive molecules in a fully unbiased manner, without requiring knowledge of the 3D structure of the targeted MisP. This point is very important as this information is lacking for many disease-associated MisPs. However, even in the few cases where protein structural information is available, rational design of potential therapeutic compounds frequently result in dead-ends. Notably, in the case of p53(Y220C), even though many compounds have been designed to target the aberrantly formed crevice in the surface of the oncogenic variant, none of them have made it to the market still. While there are no indications that the herein identified cyclic peptides against p53(Y220C) will be more successful, by screening molecules in an unbiased manner and selecting hits exclusively based on their bioactivity, the selection process is not restricted and a wider range of potentially therapeutic compounds can be identified. Therefore, the chances of discovering compounds with the desired properties are ultimately increased. To further highlight the importance of unbiased screening, in the case of A β 42, the herein presented aggregation inhibitors bear no resemblance to neither the sequence of the target protein nor to other previously described aggregation modulators and thus their identification would have been impossible by rational design.

Furthermore, the presented system is highly versatile and can be adapted as required in order to provide hits against different MisPs. Indeed, while for the majority of the investigated MisPs the identification of bioactive hits was very straightforward, in the case of p53(Y220C) the initial employment of this system resulted in false positive hits. However, after carefully selected modifications this system was able to provide a number of hits with the desired

bioactivity. Therefore, we anticipate that even for MisPs that are more demanding, this system will be capable of providing potentially therapeutic compounds.

Last but not least, this bacterial system is widely applicable against various PMDs as it takes advantage of their common molecular origin, i.e. protein aggregation. Herein, we have utilized this system to target four different disease-associated proteins, namely p53(Y220C), A β 42, SOD1(A4V) and HTT_{ex1}-97Q. In the case of p53(Y220C), screening of the cyclic tetra-, penta-, hexa- and heptapeptide libraries resulted in the identification of six putative rescuers of p53(Y220C) with drug-like characteristics, i.e. four cyclic tetrapeptides and one cyclic pentapeptide that comply by Lipinski's rule of five [285], as well as one cyclic heptapeptide that falls into the bRo5 space, where different rules for drug-likeness apply [286,287]. Further evaluation of three of the selected hits indicated that these function by different rescuing mechanisms; two of the selected cyclic peptides were shown to inhibit the aggregation of p53(Y220C) *in vitro* without affecting its thermodynamic stability whereas the third was found unable to affect the protein's aggregation process, but capable of increasing its thermodynamic stability and promoting cell death in cancer cell lines carrying this mutation.

Furthermore, in the case of A β 42, as part of this thesis we present the screening of the cyclic heptapeptide library and the identification of over 400 putative inhibitors of A β 42 aggregation, which were subsequently divided into twenty clusters with distinct sequence characteristics. Importantly, the bioactivity of two members from the two most prominent clusters were verified *in vitro* and in two *C. elegans* models of AD, and suggested that members from different clusters may function by preferentially inhibiting different stages of A β 42 aggregation. After structure-activity-relationship analysis, we were able to determine the sequence motifs providing optimal bioactivity among each selected cluster, which in the case of the two most prominent clusters were found to be cyclo-(C,T)(R,K)VW(II,A,M)X(Ψ ,P) and cyclo-(C,T) Δ (I,V)VP(S,A) Ψ , where X is anyone of the 20 natural amino acids; II: any one of

the polar amino acids Q, C, S and T; Δ : R, I, V or Q; and Ψ : any one of the aliphatic amino acids L, V and I. Furthermore, screening and evaluation of the cyclic tetra-, penta- and hexapeptide libraries, which has been the focus of the PhD thesis of I. Matis, has similarly resulted in the identification of hundreds of putative modulators of A β 42 aggregation, two of which were evaluated *in vivo* and *in vitro*. Similarly, the identified hits have been divided into distinct clusters and are currently investigated for their ability to modulate A β 42 aggregation in diverse mechanisms.

Finally, in the case of SOD1(A4V) and HTT_{ex1-97Q}, we were able to select cyclic peptide candidate folding rescuers and/or aggregation inhibitors, which are currently under investigation in our laboratory. Importantly, in the case of SOD1(A4V), the screening and evaluation of the cyclic tetra-, penta- and hexapeptide libraries has been the focus of the PhD thesis of Ms. Stefania Panoutsou and has so far resulted in the identification of a number of cyclic peptides with the ability to rescue the misfolding of SOD1(A4V) *in vitro*.

Overall, the work presented in this thesis describes the development of a generalized and highly adaptable ultrahigh-throughput system that enables the discovery and characterization of potent aggregation inhibitors in a fully unbiased manner. Importantly, this system allows the facile investigation of more than 200 million different molecules and thus comprises the largest, to our knowledge, direct functional screen against protein aggregation that has been reported to date.

Chapter 8 – Future perspectives

As mentioned in the previous chapter, the herein presented bacterial system enables the ultrahigh-throughput screening of more than 200 million head-to-tail cyclic peptides for the identification of rescuers of protein misfolding and aggregation. While this system has enabled the largest – to our knowledge – direct functional screening of lower-molecular-weight entities against protein aggregation to date, it can be further expanded in order to allow the biosynthesis and simultaneous screening of libraries with even greater diversities. Specifically, we are currently in the process of constructing a library of more than 5.6 billion different cyclic peptides that also include non-canonical amino acids. Screening of this library will result in the investigation of a greatly expanded chemical space and the identification of high quality leads with desired pharmacokinetic properties.

Furthermore, while we present herein the utilization of the described system against four target MisPs, a wider range of MisPs should be investigated in order to verify the universality of this approach. These should include MisPs associated with diverse mechanisms of pathogenesis (i.e. dominant-negative effect, aberrant localization, enhanced degradation, etc.) and different molecular forms (i.e. soluble or membrane-bound). Indeed, in an effort to achieve this, future targets of our laboratory include pathogenic variants of human rhodopsin, a membrane-bound protein associated with retinitis pigmentosa.

With respect to the herein presented rescuers of p53(Y220C) misfolding, it should be noted that further studies are required in order to confirm reactivation of the target protein. First, direct binding of the selected cyclic peptides to the mutant protein should be demonstrated using X-ray crystallography or NMR and the ability of the selected peptides to restore the mutated protein back to its folded state should be visualized using conformational

antibodies that exclusively identify correctly folded p53. Next, all of the selected peptides should be further investigated in cancer cell lines, in order to more thoroughly assess their ability to promote cell death and also activate p53 target genes, such as p21. Importantly, in case of negative results in cell-based assays, diverse delivery methods should be investigated in order to verify that the test compounds are able to permeate the cell membranes and reach their protein target.

Additionally, it would be highly interesting to assess whether the effects of these cyclic peptides are p53-mutant-specific or whether they have a broader applicability towards other p53 oncogenic variants. Furthermore, in the case of the three cyclic tetrapeptides that have proven very difficult to produce via organic synthesis, their biosynthesis and purification via on-column protein splicing should be optimized, in order to overcome this bottleneck and allow their further *in vitro* evaluation. This optimization will also prove very useful for future peptide hits, allowing an accelerated and more affordable cyclic peptide production process. Finally, if the aforementioned studies do not confirm reactivation of p53(Y220C) by either one of the six selected cyclic peptides, the herein described bacterial system should be further modified in order to allow high-throughput screening based on other p53 physiological functions, such as binding to target DNA [349].

With respect to the A β 42 aggregation inhibitors, members from the rest of the identified clusters should be evaluated *in vitro* and *in vivo* in order to determine their efficacy and ascertain whether they have different mechanisms of action (i.e. inhibit, modulate or reverse aggregation) or preferentially affect different steps of the aggregation process (i.e. primary nucleation, secondary nucleation or elongation), as initially suggested by the herein presented results. Furthermore, while the two selected cyclic heptapeptides A β C7-1 and A β C7-14 show high efficacy against A β 42 aggregation *in vitro* and *in vivo*, this is only the start of a long road towards drug development. Indeed, both leads should be further evaluated in respect of

pharmacokinetics and toxicology, and optimized accordingly with the ultimate goal being their preclinical development and subsequent clinical evaluation. These studies have indeed commenced for previously selected A β 42 aggregation modulators, with promising results thus far.

Regarding the identified SOD1(A4V) and HTT_{ex1-97Q} folding rescuers and/or aggregation inhibitors, further *in vitro* and *in vivo* evaluation is required in order to verify their bioactivity and identify consensus motifs of the selected sequences. This investigation is currently in progress in our laboratory and preliminary results appear highly promising.

Finally, although this system has already proven very efficient for targeting four different PMDs, an effort should be made to examine its potential against even more demanding MisPs, such as membrane proteins that are notoriously difficult to study. By successfully identifying hits against such proteins, we would further demonstrate the ability of this system to target PMDs in a generalized and highly affordable manner, characterized by unprecedented simplicity and speed.

Chapter 9 - Materials and Methods

Reagents and chemicals

All enzymes for recombinant DNA used in this study were purchased from New England Biolabs with the exception of alkaline phosphatase FastAP which was obtained from ThermoFisher Scientific. Recombinant plasmids were purified using Macherey-Nagel NucleoSpin Plasmid and Qiagen Plasmid Mini and Midi kits. PCR products and DNA extracted from agarose gels were purified using Macherey-Nagel Nucleospin Gel and PCR Clean-up. Synthetic p53C4-16 was purchased from Genscript (USA), p53C5-18 and p53C7-10 from Caslo ApS (Denmark) and A β C7-1 and A β C7-14 from Proteogenix (France). Before each experiment, p53C4-16, p53C5-18 and p53C7-10 were freshly dissolved in H₂O to form 10 mM solutions, A β C7-1 was dissolved in 50% acetonitrile/0.1% Tween-20 and A β C7-14 was dissolved in 30% acetonitrile/0.025% Tween-20 to form 10 mM solutions. Under these conditions, all cyclic peptides were stably in a monomeric form, as verified by dynamic light scattering analyses. The pSICLOPPS, pETA β 42-GFP and pET-Sac-A β (M1-42) vectors were kind gifts from Prof. S. Benkovic (University of Pennsylvania), Prof. M. H. Hecht (Princeton University) and Prof. S. Linse (Lund University) respectively.

Construction of the combinatorial cyclic peptide library

Initially, we constructed twelve distinct combinatorial cyclic peptide sub-libraries: the cyclo-CysX₁X₂X₃, cyclo-SerX₁X₂X₃, and cyclo-ThrX₁X₂X₃ tetrapeptide sub-libraries (pSICLOPPS-CysX₁X₂X₃, pSICLOPPS-SerX₁X₂X₃, and pSICLOPPS-ThrX₁X₂X₃ vector sub-libraries), the cyclo-CysX₁X₂X₃X₄, cyclo-SerX₁X₂X₃X₄, and cyclo-ThrX₁X₂X₃X₄ cyclic

pentapeptide sub-libraries (pSICLOPPS-CysX₁X₂X₃X₄, pSICLOPPS-SerX₁X₂X₃X₄, and pSICLOPPS-ThrX₁X₂X₃X₄ vector sub-libraries) the cyclo-CysX₁X₂X₃X₄X₅, cyclo-SerX₁X₂X₃X₄X₅, and cyclo-ThrX₁X₂X₃X₄X₅ cyclic hexapeptide sub-libraries (pSICLOPPS-CysX₁X₂X₃X₄X₅, pSICLOPPS-SerX₁X₂X₃X₄X₅, and pSICLOPPS-ThrX₁X₂X₃X₄X₅ vector sub-libraries) and the cyclo-CysX₁X₂X₃X₄X₅X₆, cyclo-SerX₁X₂X₃X₄X₅X₆, and cyclo-ThrX₁X₂X₃X₄X₅X₆ cyclic heptapeptide sub-libraries (pSICLOPPS-CysX₁X₂X₃X₄X₅X₆, pSICLOPPS-SerX₁X₂X₃X₄X₅X₆, and pSICLOPPS-ThrX₁X₂X₃X₄X₅X₆ vector sub-libraries) (Appendix E). These vectors express libraries of fusion proteins comprising four parts: (i) the C-terminal domain of the split Ssp DnaE intein (I_C), (ii) a tetra-, penta-, hexa- or heptapeptide sequence, (iii) the N-terminal domain of the split Ssp DnaE intein (I_N), and (iv) a chitin-binding domain (CBD) under the control of the PBAD promoter and its inducer L(+)-arabinose (Figure 2.4). The libraries of genes encoding these combinatorial libraries of random cyclic oligopeptides were constructed using the degenerate forward primers GS032, GS033, GS034, GS072, GS073, GS074, GS075, GS076, GS077, GS078, GS079, GS080 individually in pair with the reverse primer GS035, and pSICLOPPS as a template (Appendix F). Cys, Ser, and Thr were encoded in these primers by the codons TGC, AGC, and ACC, respectively, which are the most frequently utilized ones for these amino acids in *E. coli*, while the randomized amino acids (X) were encoded using random NNS codons, where N=A, T, G, or C and S=G or C, as described previously [282]. A second PCR reaction was performed in each case to eliminate mismatches using the aforementioned amplified DNA fragments as templates and the forward primers GS069, GS070 and GS071 for each one of the peptide sub-libraries starting with Cys, Ser or Thr, respectively, together with the reverse primer GS035. The resulting PCR products were then digested with BglII and HindIII for 5 h and inserted into a similarly digested and dephosphorylated pSICLOPPSKanR vector (see below). The ligation reactions were optimised at a 12:1 insert:vector molar ratio and performed for 4 h at 16 °C. Approximately

0.35, 0.7, 3.5 and 10 µg of the pSICLOPPSKanR vector were used for each one of the tetra-, penta-, hexa- and heptapeptide libraries, respectively. The ligated DNA was then purified using spin columns (Macherey-Nagel, Germany), transformed into electro-competent MC1061 cells prepared in-house, plated onto Luria-Bertani (LB) agar plates containing 25 µg/mL chloramphenicol and incubated at 37 °C for 14-16 h. This procedure resulted in the construction of the combined pSICLOPPS-NuX1X2X3-X5 library with a total diversity of about 1,234,000,000 independent transformants, as judged by plating experiments after serial dilutions.

Plasmid constructions

Plasmids involved in cyclic peptide production

For the construction of the pSICLOPPS vectors encoding for variants of the selected cyclic peptides, the auxiliary pSICLOPPSKanR vector was generated initially by PCR amplification of the gene encoding aminoglycoside 3'-phosphotransferase (KanR - the enzyme conferring resistance to the antibiotic kanamycin) from pET28a(+) using primers GS043-DG002, digestion with BglII and HindIII and insertion into similarly digested pSICLOPPS vector.

For the construction of the auxiliary pSICLOPPS(H24L/F26A)KanR vector, the gene encoding the C-terminal domain of the split Ssp DnaE intein (I_C) was first mutated using primers GS037 and DD015 and the pSICLOPPS vector as a template. The I_C(H24L/F26A) PCR product was then digested with NcoI and BglII, the KanR gene was digested from pSICLOPPSKanR with BglII and HindIII and the pSICLOPPS vector was digested with NcoI-HindIII. Three-way ligation of the aforementioned digested products (I_C(H24L/F26A), KanR

and pSICLOPPS) resulted in the production of the desired pSICLOPPS(H24L/F26A)KanR vector.

For the construction of pSICLOPPS(H24L/F26A)-p53C4-4, pSICLOPPS(H24L/F26A)-p53C4-16, pSICLOPPS(H24L/F26A)-p53C4-19, pSICLOPPS(H24L/F26A)-p53C4-21, pSICLOPPS(H24L/F26A)-p53C5-18, pSICLOPPS(H24L/F26A)-p53C7-10, pSICLOPPS(H24L/F26A)-A β C7-1, pSICLOPPS(H24L/F26A)-A β C7-2, pSICLOPPS(H24L/F26A)-A β C7-3, pSICLOPPS(H24L/F26A)-A β C7-7, pSICLOPPS(H24L/F26A)-A β C7-14, pSICLOPPS(H24L/F26A)-Random1 and pSICLOPPS(H24L/F26A)-Random2, plasmids pSICLOPPS-p53C4-4, pSICLOPPS-p53C4-16, pSICLOPPS-p53C4-19, pSICLOPPS-p53C4-21, pSICLOPPS-p53C5-18, pSICLOPPS-p53C7-10, pSICLOPPS-A β C7-1, pSICLOPPS-A β C7-2, pSICLOPPS-A β C7-3, pSICLOPPS-A β C7-7, pSICLOPPS-A β C7-14, pSICLOPPS-Random1 and pSICLOPPS-Random2 were digested with BglII and HindIII and the resulting inserts were ligated into a similarly digested pSICLOPPS(H24L/F26A)KanR vector.

For the construction of pSICLOPPS-A β C7-371, pSICLOPPS-A β C7-405 and pSICLOPPS-A β C7-416, forward primers DD175, DD176 and DD178 were used individually, in pair with the reverse primer GS035 and using pSICLOPPS as a template and the resulting PCR products were digested with BglII and HindIII and ligated into a similarly digested pSICLOPPSKanR.

For the construction of pSICLOPPS-A β C7-1(C1S), pSICLOPPS-A β C7-1(C1T), pSICLOPPS-A β C7-1(K2A), pSICLOPPS-A β C7-1(V3A), pSICLOPPS-A β C7-1(W4A), pSICLOPPS-A β C7-1(Q5A), pSICLOPPS-A β C7-1(L6A), pSICLOPPS-A β C7-1(L7A), pSICLOPPS-A β C7-14(C1S), pSICLOPPS-A β C7-14(C1T), pSICLOPPS-A β C7-14(R2A), pSICLOPPS-A β C7-14(I3A), pSICLOPPS-A β C7-14(V4A), pSICLOPPS-A β C7-14(P5A),

pSICLOPPS-A β C7-14(S6A) and pSICLOPPS-A β C7-14(L7A), forward primers DD129, DD130, DD131, DD132, DD133, DD134, DD135, DD136, DD137, DD138, DD139, DD140, DD141, DD142, DD143 and DD144 were used individually, in pair with the reverse primer GS035 and using pSICLOPPS-A β C7-1 or pSICLOPPS-A β C7-14 as a template accordingly. The resulting PCR products were digested with BglII and HindIII and ligated into a similarly digested pSICLOPPSKanR.

For the construction of pET-Random1, the gene encoding the I_C-peptide-I_N-CBD fusion was digested using NcoI-HindIII and ligated into similarly digested pET28a(+).

For the construction of pSICLOPPS-His₆-Random1, pSICLOPPS-His₆-p53C4-4, pSICLOPPS-His₆-p53C4-16, pSICLOPPS-His₆-p53C4-19, pSICLOPPS-His₆-p53C4-21, pSICLOPPS-His₆-p53C5-18, pSICLOPPS-His₆-p53C7-10, the genes encoding the His₆-I_C-peptide-I_N-CBD fusions were amplified by PCR using primers DD052 and GS035 and the respective pSICLOPPS vectors as templates. The PCR products were then digested with NcoI and HindIII and ligated into similarly digested pSICLOPPSKanR.

Plasmids involved in MisP production

For the construction of pETp53Cwt-GFP, a truncated human *TP53* gene encoding the DNA-binding (core) domain of p53 (p53C, amino acids 94-312) was assembled by PCR using the primers GS118, GS119, GS120, GS121, GS122, GS123, GS124, GS125, GS126, GS127, GS128, GS129, GS130, GS131, GS132, and GS133. The PCR product was then digested with NdeI and BamHI, and was inserted into similarly digested pETA β 42-GFP vector. For constructing pETT-p53C-GFP, the M133L, V203A, N239Y and N268D mutations were introduced into the p53C-encoding gene by overlap PCR using the pETp53Cwt-GFP as a template and primers GS003, GS004, GS011, GS012, GS013, GS014, GS015, GS016, GS042b

and GS042c. The PCR product was digested with NdeI and BamHI and inserted into similarly digested pETA β 42-GFP.

For the construction of pETp53C(V143A)-GFP, pETp53C(Y220C)-GFP and pETp53C(F270L)-GFP, the p53C-encoding gene was mutated by overlap PCR starting from pETp53Cwt-GFP as a template and using primers GS003, GS004, GS007, GS008, GS041, GS042a, DD011 and DD012 accordingly. The PCR product was digested with NdeI and BamHI and inserted into similarly digested pETA β 42-GFP. For the construction of pETT-p53C(V143A)-GFP and pETT-p53C(Y220C)-GFP, the same methodology was followed with the only difference being the utilization of pETT-p53C-GFP instead of pETp53Cwt-GFP as a template, whereas for the construction of pETT-p53C(F270L)-GFP a different set of primers were also required, namely primers DD009 and DD010.

For the construction of pASKT-p53C-GFP and pASKT-p53C(Y220C)-GFP, the T-p53C-GFP and T-p53C(Y220C)-GFP genes were amplified by PCR from the respective pET vectors using primers GS002 and GS003 and ligated into pASK75 using the restriction sites XbaI-HindIII.

For the construction of pTrcT-p53C-GFP, pTrcT-p53C(Y220C)-GFP, pBADT-p53C-GFP and pBADT-p53C(Y220C)-GFP, the T-p53C-GFP and T-p53C(Y220C)-GFP genes were digested from the respective pASK vectors using XbaI and HindIII and ligated into similarly digested pTrc99A or pBAD30.

For the construction of pETT-p53C-BFP and pETT-p53C(Y220C)-BFP, the BFP gene was amplified by PCR from the pBADcstA-BFP using the primers DD006 and DD007, digested with BamHI and XhoI and inserted into similarly digested pETT-p53C-GFP and pETT-p53C(Y220C)-GFP. Similarly, for the construction of pETT-p53C-RFP and pETT-p53C(Y220C)-RFP, the RFP gene was amplified from the T_YES_Isaacs_1_pSTC1 (ref)

using the primers DD079 and DD083, digested with BamHI and XhoI and inserted into similarly digested pETT-p53C-GFP and pETT-p53C(Y220C)-GFP.

For the construction of pCDFT-p53C-BFP and pCDFT-p53C(Y220C)-BFP, the *TP53* genes were amplified by PCR using the primers DD008 and GS004, as well as the mutagenic primers DD070 and DD071. This resulted in the construction of *TP53* genes that lack the intermediate NcoI site and also contain an extra Val residue after the first Met, so that the rest of the amino acids remain in frame. These PCR products were then digested with NcoI and BamHI, the BFP gene was digested from the pETT-p53C-BFP with BamHI and HindIII and the pCDF-1b vector was digested with NcoI and HindIII. Three-way ligation of the aforementioned digested products (T-p53C or T-p53C(Y220C), BFP and pCDF) resulted in the production of the desired vectors.

For the construction of pETT-p53C-His₆ and pETT-p53C(Y220C)-His₆, the *TP53* genes were amplified by PCR using the primers GS003 and DD002 and the respective pETp53C-GFP vectors as templates. The PCR products were then digested with NdeI and XhoI and ligated into a similarly digested pETAβ42-GFP.

For the construction of pETSOD1wt-GFP, the human *SOD1* cDNA was generated by PCR-mediated gene assembly using the primers GS100, GS101, GS102, GS103, GS104, GS105, GS106, GS107, GS108, GS109, GS110, and GS111. The assembled gene was then digested with NdeI and BamHI, and inserted into similarly digested pETAβ42-GFP vector, in place of Aβ42. For pETSOD1(A4V)-GFP, *SOD1* was amplified by PCR from the pETSOD1wt-GFP vector using the mutagenic forward primer GS059 and the reverse primer GS060. The resulting PCR product was then digested with NdeI and BamHI, and inserted into similarly digested pETAβ42-GFP. For pETSOD1(G37R)-GFP, pETSOD1(G85R)-GFP and pETSOD1(G93A)-GFP construction, *SOD1* was mutated by overlap PCR starting from

pETSOD1wt-GFP as a template and using the primers: GS058, GS059, GS112, GS113, GS114, GS115, GS116 and GS117, accordingly. All *SOD1* PCR products were then digested with NdeI and BamHI, and inserted into similarly digested pETA β 42-GFP vector.

For the construction of pETA β 42, the A β 42 gene was amplified using primers DD004 and IM022 as well as pETA β 42-GFP [304] as a template. The resulting PCR product was digested with NcoI and XhoI and inserted into a similarly digested pET28a(+).

Cyclic peptide library screening for the identification of p53C(Y220C) folding rescuers

FACS sorting utilizing a two-plasmid system comprising of pETp53C(Y220C)-GFP and the combined pSICLOPPS-NuX₁X₂X₃-X₅ vector library

Electrocompetent *E. coli* BL21(DE3) cells carrying the expression vector pETp53C(Y220C)-GFP, which produces p53C(Y220C)-GFP under the control of the strong bacteriophage T7 promoter, were co-transformed with the combined pSICLOPPS-NuX₁X₂X₃-X₅ vector library, producing the cyclic tetra-, penta- and hexapeptides. Approximately 10⁸ transformants carrying both vectors were harvested, pooled together and grown in LB liquid medium containing 0.002% L(+)-arabinose -the inducer of cyclic peptide production- at 37 °C with shaking. When the optical density at 600 nm (OD₆₀₀) of the bacterial culture was about 0.5, 0.1 mM IPTG was added to the medium to induce overexpression of the p53C(Y220C)-GFP reporter. After about 2 h of induction at 37 °C, the fluorescence (FITC-H; 530/30 nm) of 10,000 cells was recorded using a Becton–Dickinson FACS Aria system (BD Biosciences) and analysed using FlowJo software (FlowJo). Then, around 10⁸ cells were gated on a side-scatter (SSC-H) versus forward-scatter (FSC-H) plot to eliminate events that included non-cellular

particles and subjected to FACS sorting for the isolation of the bacterial population exhibiting the top 1–3% fluorescence using the BD FACSAria sorter and the FACSDiva software (BD Biosciences). The isolated cells were re-grown and screened for one additional round in an identical manner, at which point DNA was isolated from the enriched pool using a Qiagen Plasmid Mini Kit.

FACS sorting sequential integration of two two-plasmid systems comprising of pETT-p53C(Y220C)-GFP or pASKT-p53C(Y220C)-GFP and the combined pSICLOPPS-NuX₁X₂X₃–X₅ vector library

Sorting experiments was performed as described in the above section with the following modifications: (i) the pETT-p53C(Y220C)-GFP, which produces the T-p53C(Y220C)-GFP instead of the p53C(Y220C)-GFP was initially used, (ii) after two rounds of sorting, DNA of the enriched population was isolated and re-transformed into BL21(DE3) cells carrying the pASKT-p53C(Y220C)-GFP vector, which produces T-p53C(Y220C)-GFP under the control of the tetracycline promoter and (iii) after three additional rounds of sorting DNA was isolated from the enriched pool using a Qiagen Plasmid Mini Kit.

FACS sorting utilizing a three-plasmid system comprising of pCDFT-p53C(Y220C)-BFP, pETA β 42-GFP and the pSICLOPPS-NuX₁X₂X₃–X₆ vector libraries

Electrocompetent *E. coli* Tuner(DE3) cells carrying the expression vectors pCDFT-p53C(Y220C)-BFP and pETA β 42-GFP, were co-transformed with the four pSICLOPPS vector libraries, combined according to their size, i.e. tetra-, penta-, hexa- and heptapeptides. Approximately 10⁶, 10⁷, 10⁸ or 10⁹ transformants carrying pCDFT-p53C(Y220C)-BFP and pETA β 42-GFP together with pSICLOPPS-NuX₁X₂X₃, pSICLOPPS-NuX₁X₂X₃X₄,

pSICLOPPS-NuX₁X₂X₃X₄X₅ or pSICLOPPS-NuX₁X₂X₃X₄X₅X₆ respectively, were harvested, pooled together and diluted to an OD₆₀₀ of 0.1 in LB liquid medium containing 0.005% L(+)-arabinose to induce cyclic peptide production. Cultures were incubated at 37 °C with shaking until an OD₆₀₀ of 0.4-0.5, at which point 0.1 mM IPTG was added to the medium to induce simultaneous overexpression of both T-p53(Y220C)-BFP and Aβ₄₂-GFP.

After two hours of induction at 37 °C the fluorescence of 50,000 cells was recorded using a BD FACSAria II system (BD Biosciences, USA). For BFP fluorescence a 405 nm solid state laser was used for excitation and a 450/40 nm band pass filter for detection, while for GFP fluorescence a 488 nm solid state laser was used for excitation and a 530/30 nm band pass filter for detection. Then, ~3x 10⁶, 10⁷, 10⁸ or 10⁹ cells were gated on a SSC-H vs. FSC-H plot in order to eliminate non-cellular events, and were subjected to FACS sorting by selecting the bacterial population exhibiting the top ~2 % of BFP fluorescence but unaltered GFP fluorescence. The isolated cells were re-grown and screened in an identical manner for three additional rounds in the case of the hexapeptide library and four additional rounds in the case of the tetra-, penta- and heptapeptide libraries. Finally, DNA from the enriched populations was isolated using a Qiagen Plasmid Mini Kit.

Cyclic peptide library screening for the identification of Aβ₄₂ aggregation inhibitors

Electrocompetent *E. coli* Tuner (DE3) cells carrying the expression vector pETAβ₄₂-GFP [304], were co-transformed with the combined pSICLOPPS-NuX₁X₂X₃X₄X₅X₆ vector library. Approximately 10⁹ transformants carrying both vectors were harvested, pooled together and diluted to OD₆₀₀ of 0.1 in LB liquid medium containing 0.005% L(+)-arabinose to induce cyclic peptide production. Cultures were incubated at 37 °C with shaking until an

OD₆₀₀ of 0.4-0.5, at which point 0.1 mM IPTG was added to the medium to induce overexpression of the Aβ₄₂-GFP reporter. Fluorescence of 50,000 cells was recorded after two hours of induction at 37 °C using a BD FACSAria II system (BD Biosciences, USA) with a 488 nm solid state laser for the excitation of GFP and a 530/30 band pass filter for detection. Then, ~3x10⁹ cells were gated on a SSC-H vs. FSC-H plot in order to eliminate non-cellular events, and were subjected to FACS sorting for the isolation of the bacterial population exhibiting the top ~2% fluorescence. The isolated cells were re-grown and screened for six additional rounds in an identical manner, at which point DNA was isolated from the enriched pool using a Qiagen Plasmid Mini Kit.

Cyclic peptide library screening for the identification of SOD1(A4V) folding rescuers

Electrocompetent *E. coli* Origami2(DE3) cells carrying the expression vector pETSOD1(A4V)-GFP, were co-transformed with the four pSICLOPPS vector libraries, combined according to their size, i.e. tetra-, penta-, hexa- and heptapeptides. Approximately 10⁶, 10⁷, 10⁸ or 10⁹ transformants carrying pETSOD1(A4V)-GFP and pSICLOPPS-NuX₁X₂X₃, pSICLOPPS-NuX₁X₂X₃X₄, pSICLOPPS-NuX₁X₂X₃X₄X₅ or pSICLOPPS-NuX₁X₂X₃X₄X₅X₆ respectively, were harvested, pooled together and diluted to an OD₆₀₀ of 0.1 in LB liquid medium containing 0.005% L(+)-arabinose to induce cyclic peptide production. Cultures were incubated at 37 °C with shaking until an OD₆₀₀ of 0.4-0.5, at which point 0.01 mM IPTG was added to the medium to induce overexpression of SOD1(A4V)-GFP. Fluorescence was recorded as described for Aβ₄₂-GFP. Then, ~3x 10⁶, 10⁷, 10⁸ or 10⁹ cells were gated on a SSC-H vs. FSC-H plot in order to eliminate non-cellular events, and were subjected to FACS sorting for the isolation of the bacterial population exhibiting the top ~2% fluorescence. The isolated

cells were re-grown and screened for three additional rounds in an identical manner, at which point DNA was isolated from the enriched pool using a Qiagen Plasmid Mini Kit.

Cyclic peptide library screening for the identification of HTT_{ex1}-97Q aggregation inhibitors

For HTT sorting the same procedure as for SOD1(A4V) was followed with the only differences being: (i) the use of *E. coli* Tuner(DE3) cells, (ii) the induction of HTT_{ex1}-97Q-GFP expression from the pETHHTT_{ex1}-97Q-GFP vector using 0.1 mM IPTG and (iii) the performance of six rounds of sorting in total.

Protein/cyclic peptide production in liquid cultures.

E. coli cells freshly transformed with the appropriate expression vector(s) were used for all protein production experiments. Single bacterial colonies were used to inoculate overnight liquid LB cultures containing the appropriate antibiotics for plasmid maintenance (100 µg/mL ampicillin, 40 µg/mL chloramphenicol, 50 µg/mL kanamycin or 100 µg/mL streptomycin (Sigma)) at 37 °C. These cultures were used with a 1:100 dilution to inoculate fresh LB cultures in all cases.

For p53C-GFP, p53C-BFP, p53C-RFP, Aβ42-GFP and polyQ-HTT_{ex1}-GFP production, BL21(DE3) or Tuner(DE3) cells transformed with the corresponding vector were grown in 5 ml liquid LB cultures containing the relevant antibiotic at 37 °C to an OD₆₀₀ of ~0.4 with shaking, at which point MisP-GFP production was initiated by the addition of the appropriate inducer (0.1 mM IPTG for pET-, pTrc- and pCDF-based vectors, 0.02 % L(+)-arabinose for pBAD-based vectors and 0.2 µg/ml anhydrotetracycline for pASK-based

vectors). Recombinant protein production was performed at 37 °C for 2 h, 30 °C for 5 h or 25 °C for 16 h, unless otherwise stated.

For SOD1-GFP production, BL21(DE3) or Origami 2(DE3) cells transformed with the corresponding SOD1-GFP-encoding vector were grown in 5 ml liquid LB cultures containing 50 µg/mL kanamycin , 200 µM CuCl₂ and 200 µM ZnCl₂ at 37 °C to an OD₆₀₀ of ~0.4 with shaking, at which point SOD1-GFP production was induced by the addition of 0.01 mM IPTG at 37 °C for 2 h, unless otherwise stated.

For recombinant protein production of individual members from the pSICLOPPS libraries, MC1061 cells transformed with the corresponding pSICLOPPS vector were grown in 5 ml liquid LB cultures containing 40 µg/mL chloramphenicol at 37 °C to an OD₆₀₀ of ~0.4 with shaking, at which point cyclic peptide production was induced by the addition of 0.0002% L(+)-arabinose at 37 °C for 3 h, unless otherwise stated.

For simultaneous production of MisP-GFP fusions and selected cyclic peptides, BL21(DE3) or Tuner(DE3) cells transformed with the corresponding vectors were grown in 5 ml liquid LB cultures containing the relevant antibiotics and 0.02% L(+)-arabinose at 37 °C to an OD₆₀₀ of ~0.4 with shaking, at which point MisP-GFP production was initiated by the addition of 0.1 mM IPTG for 2 h.

Bacterial cell fluorescence

After protein overexpression, bacterial cells corresponding to 1 mL culture with OD₆₀₀=1 were harvested by centrifugation, re-suspended in 100 µL phosphate-buffered saline (PBS), transferred to a 96-well FLUOTRAC 200 plate (Greiner Bio One International, Austria) and after excitation at 488 nm, their fluorescence was measured at 510 nm using a TECAN

Safire II-Basic plate reader (Tecan, Austria). For ThS fluorescence measurements, a similar procedure was followed with the exception that cells were re-suspended in 250 μ M ThS in PBS, equilibrated for 15 min and then emission spectra were recorded at a range of 460-600 nm. For flow cytometric measurements, after protein overexpression, fluorescence of 50,000 cells resuspended in PBS was recorded at 530/30 nm after GFP excitation at 488 nm, using a BD FACSAria II system.

In-gel fluorescence and western blot analyses

After protein overexpression, cells corresponding to 1 mL culture with $OD_{600}=1$ were harvested by centrifugation and re-suspended in 100 μ L PBS. Samples were lysed by brief sonication cycles on ice and the resulting lysates (total fraction) were clarified by centrifugation at 13,000 rpm for 25 min (soluble fraction). Samples were analyzed by native or SDS-PAGE, on 10 or 15% gels and without prior boiling (for in-gel fluorescence) or after boiling of the samples for 10 min (for western blotting). In-gel fluorescence was analyzed on a ChemiDoc-It² Imaging System equipped with a CCD camera and a GFP filter (UVP, UK), after exposure for about 3 sec. For western blotting, proteins were transferred to polyvinylidene fluoride (PVDF) membranes (Merck, Germany) for 1 h at 12 V on a semi-dry blotter (Thermo Fisher, USA). Membranes were blocked with 5% non-fat dry milk in Tris-buffered saline containing 0.1% Tween-20 (TBST) for 1 h at room temperature. After washing with TBST three times, membranes were incubated with the appropriate primary antibody, re-washed with TBST and then, if necessary, incubated with a secondary antibody. Both incubations were performed at room temperature for 1 h. The proteins were then visualized using the ChemiDoc-It² Imaging System (UVP, UK).

The utilized antibodies were a mouse anti-A β (6E10) (Covance, USA) at 1:2,000 dilution, a mouse anti-CBD antibody (New England Biolabs, USA) at 1:100,000 dilution, a horseradish peroxidase(HRP)-conjugated goat anti-mouse secondary antibody (BioRad) at 1:4000 dilution, a mouse anti-GFP at 1:20,000 dilution (Clontech, USA) and a mouse monoclonal (HRP)-conjugated anti-polyhistidine antibody (Sigma, USA) at 1:2,500 dilution. All antibodies were diluted in 0.5% non-fat dry milk in TBST.

High-throughput sequencing analysis

For the high-throughput sequencing analysis three DNA samples were produced: the first contained a combined pSICLOPPS-NuX₃-X₅ vector library with approximately equal amounts of each one of the tetra-, penta- and hexapeptide sub-libraries, the second contained the pSICLOPPS-NuX₁X₂X₃X₄X₅X₆ vector library and the third contained the enriched peptide library against A β 42 aggregation after the 7th round of sorting. All DNA samples were digested with NcoI and BsrGI and the resulting ~250 bp products that contained the variable peptide-encoding region were isolated. The analysis was performed at the Genomics core facility of the Biomedical Sciences Research Center “Alexander Fleming” (Athens, Greece) using an Ion Torrent high-throughput sequencing platform. From the obtained data, all the sequences with mismatches outside of the variable peptide-encoding region were removed, and only the 12-, 15-, 18- or 21-bp-long peptide-encoding sequences with NNS codons were subjected to further analysis. For the enriched peptide library, all sequences including stop codons were also discarded from subsequent analysis.

Peptide sequence similarity analysis and clustering

Sequence similarity analysis was performed using the Immune Epitope Database (IEDB) clustering tool (<http://tools.iedb.org/cluster2/>) and the fully interconnected clusters (cliques) method [350]. This approach allows all peptides in a clique to share a minimal level of homology, while at the same time one peptide can be part of multiple cliques [350]. As sequence similarity analysis is performed using linear sequences, the circular permutants of each cyclic heptapeptide appearing at least 20 times within the sorted population were identified and taken into consideration, tallying up to 2912 linear representations for the 416 cyclic heptapeptides. From this analysis, 5087 cliques sharing at least 70% sequence homology were identified and after re-integration of the different circular permutants to their original cyclic peptide sequence, 617 unique cliques remained. From the 416 distinct cyclic heptapeptides, 323 were covered in the cliques forming a total of 1467 unique pairs with more than 70% sequence homology. The remaining 93 cyclic peptides did not share a minimal level of 70% homology with any other of the peptides. The results were then presented in an undirected network graph using the Gephi graph visualization software [351] and cluster identification was performed using the Girvan-Newman algorithm [352].

Cyclic peptide on-column purification

Chitin affinity chromatography

Single bacterial colonies from *E. coli* BL21(DE3) cells freshly transformed with pET-Random1 or pSICLOPPS-Random1 were used to inoculate overnight liquid LB cultures containing 50 µg/mL kanamycin or 40 µg/ml chloramphenicol at 37 °C. These cultures were used with a 1:100 dilution to inoculate 500 ml LB cultures with the appropriate antibiotic and

when cell density reached an OD₆₀₀ of 0.8-1.2, protein overexpression was induced by the addition of 1 mM IPTG or 0.2 % L(+)-arabinose respectively, at 37, 20 or 15 °C for 2, 16 or 40 h, respectively. Cell pellets were collected after overexpression of the tetrapartite fusion I_C-Random1-I_N-CBD, re-suspended in 40 mL chitin buffer (25 mM Tris-HCl, 500 mM NaCl, pH 7.0 or 20 mM Tris-HCl, 500 mM NaCl, 1 mM TCEP pH 7.8) and lysed by brief sonication steps on ice. The soluble cell lysates were collected after centrifugation at 13,000 × g for 30 min at 4 °C, and mixed with 2 mL chitin resin (New England Biolabs, USA) at 4 °C for 16 h on a roller mixer before loading onto a 5 mL polypropylene chromatography column (Pierce, USA). The flow-through buffer was re-loaded onto the column three times and column-bound protein was washed with 100 mL chitin buffer. Then, 1 ml of chitin buffer were added onto the column and left overnight at room temperature for the intein splicing and peptide cyclization process to occur. Following on-column intein processing and cyclization, the elution fraction containing the cyclic peptide were collected, while the I_N-CBD product remained bound to the resin.

Immobilized metal affinity chromatography

Single bacterial colonies from *E. coli* BL21(DE3) cells freshly transformed the appropriate pSICLOPPS-His₆ vector were used to inoculate overnight liquid LB cultures containing 40 µg/ml chloramphenicol at 37 °C. These cultures were used with a 1:100 dilution to inoculate 1.5 L LB cultures supplemented with 40 µg/ml chloramphenicol and when cell density reached an OD₆₀₀ of 0.8-1.0, protein overexpression was induced by the addition of 0.2 % L(+)-arabinose respectively, at 37 °C for 3 h, or 20 °C for 16 h, respectively. Cell pellets were collected after overexpression of the tetrapartite fusion I_C-peptide-I_N-CBD, re-suspended in 20 mL lysis buffer (50 mM NaH₂PO₄, 300 mM NaCl, 10 mM imidazole, pH 8.0) and lysed by brief sonication steps on ice. The soluble cell lysates were collected after centrifugation at

13,000 × g for 30 min at 4 °C, and mixed with 2 mL Ni-NTA agarose resin (Qiagen, Germany) at 4 °C for 1 h on a roller mixer before loading onto a 5 mL polypropylene chromatography column (Pierce, USA). Column-bound protein was washed with 20 mL wash buffer (50 mM NaH₂PO₄, 300 mM NaCl, 20 mM imidazole, pH 8.0) and left for 90 h in 2 ml of splicing buffer (50 mM NaH₂PO₄, 300 mM NaCl, 5 mM DTT, pH 6.3 or 20 mM Pipes, 200 mM NaCl, 1 mM DTT, pH 6.0) at room temperature for the intein splicing and peptide cyclization process to occur. Following on-column intein processing and cyclization, the elution fraction containing the cyclic peptide were collected, while the His₆-I_C and the I_N-CBD products remained bound to the resin, due to non-covalent interaction between the I_C and I_N fragments.

T-p53C and T-p53(Y220C) overexpression and purification by IMAC and SEC

Single bacterial colonies from *E. coli* Tuner(DE3) cells freshly transformed with pETT-p53C-His₆ or pETT-p53C(Y220C)-His₆ were used to inoculate overnight liquid LB cultures containing 50 µg/mL kanamycin at 37 °C. These cultures were used with a 1:100 dilution to inoculate 2 L cultures with Terrific Broth medium (TB), supplemented with 50 µg/mL kanamycin and 200 µM ZnCl₂, and when cell density reached an OD₆₀₀ of 0.8-1.2, protein overexpression was induced by the addition of 1 mM IPTG at 18 °C for about 16 h. Cell pellets were collected after T-p53C or T-p53C(Y220C) overexpression by centrifugation, re-suspended in 20 mL lysis buffer (50 mM NaH₂PO₄, 300 mM NaCl, 10 mM imidazole, pH 8.0) and lysed by brief sonication steps on ice. The soluble cell lysates were collected after centrifugation at 13,000 × g for 25 min at 4 °C, and mixed with 1 mL Ni-NTA agarose resin (Qiagen, Germany) for 1 h at 4 °C on a roller mixer before loading onto a 5 mL polypropylene chromatography column (Pierce, USA). Column-bound protein was washed twice with 10 mL

wash buffer (50 mM NaH₂PO₄, 300 mM NaCl, 20 mM imidazole, pH 8.0) and eluted in three fractions, each consisting of 1 mL elution buffer (50 mM NaH₂PO₄, 300 mM NaCl, 250 mM imidazole, pH 8.0). The eluted proteins were then purified further by size-exclusion chromatography (SEC) using a HiLoad 16/600 Superdex 200 pg column (GE Healthcare, USA), to isolate the monomeric protein fractions in phosphate buffer (25 mM NaH₂PO₄, 150 mM NaCl, 20 mM imidazole, pH 7.2). The purified proteins were quantified from the absorbance of the integrated peak area using $\epsilon_{280} = 19,535 \text{ M}^{-1} \text{ cm}^{-1}$ for T-p53C and $\epsilon_{280} = 18,045 \text{ M}^{-1} \text{ cm}^{-1}$ for T-p53C(Y220C).

Preparation of A β 42 samples

The recombinant A β (M1-42) peptide (MDAEFRHDSGYEVHHQKLVFFAEDVGS NKGAIIGLMVGGVVIA), herein termed A β 42, was expressed in *E. coli* using the pET-Sac-Abeta(M1-42) vector and purified as described previously [212]. The lyophilized peptide was then dissolved in 6 M GuHCl, further purified using a Superdex 75 10/300 GL column (GE Healthcare) and eluted in 20 mM sodium phosphate buffer, pH 8, supplemented with 200 μ M EDTA and 0.02% NaN₃. The centre of the peak was collected and A β 42 concentration was determined from the absorbance of the integrated peak area using $\epsilon_{280} = 1,490 \text{ M}^{-1} \text{ cm}^{-1}$ [353].

Differential Scanning Fluorimetry (DSF)

Thermal denaturation analysis for T-p53C and T-p53C(Y220C) was performed after addition of appropriate amounts of synthetic cyclic peptides to 15 μ M of monomeric protein in phosphate buffer (25 mM NaH₂PO₄, 150 mM NaCl, 20 mM imidazole, pH 7.2) supplemented with 10x SYPRO orange (Thermo Scientific). The samples were prepared in low-binding

Eppendorf tubes on ice in triplicates, transferred to 96-well plates at 25 μL per well and incubated in a CFX96TM Real-Time PCR Detection System at a temperature range of 15-70 $^{\circ}\text{C}$ gradually increased at a rate of 1 $^{\circ}\text{C}/\text{min}$. Fluorescence was recorded after each temperature increase and the negative first derivative of each measurement $[-d(\text{RFU})/dT]$ was calculated by the RT-PCR instrument. The derived data were analyzed using Prism (GraphPad Software Inc, USA) and the melting temperature (T_m) was estimated as the minimum of the $-d(\text{RFU})/dT$ plot. For more accurate results the OriginPro software was also used to determine each T_m by fitting the fluorescence data to the Boltzmann equation.

T-p53C and T-p53C(Y220C) aggregation kinetics experiments

Kinetic experiments for T-p53C and T-p53C(Y220C) aggregation was performed after addition of appropriate amounts of synthetic cyclic peptides to 15 μM of monomeric protein in phosphate buffer (25 mM NaH_2PO_4 , 150 mM NaCl , 20 mM imidazole, pH 7.2) supplemented with 20 μM ThT (Sigma). Samples were prepared in triplicates in low-binding Eppendorf tubes on ice and transferred to a 96-well half-area, low-binding, clear bottom, polyethylene glycol coating plate (Corning 3881) at 80 μL per well. The 96-well plate was then placed at 37 $^{\circ}\text{C}$ under quiescent conditions on a TECAN Safire II-Basic plate reader and after excitation at 440 nm, ThT fluorescence was measured at 480 nm, through the bottom of the plate.

A β 42 aggregation kinetics experiments

Kinetic experiments were performed as described previously [212]. Briefly, appropriate amounts of the synthetic cyclic peptides were added to 2 μM of monomeric A β 42 to obtain the desired cyclic peptide:A β 42 molar ratios and samples were supplemented with 20 μM ThT,

1% v/v acetonitrile and 0.025% or 0.1% v/v Tween-20 for A β C7-14 and A β C7-1 respectively. All samples were prepared in low-binding Eppendorf tubes on ice using careful pipetting to avoid introduction of air bubbles and each sample was pipetted into three wells of a 96-well half-area, low-binding, clear bottom, polyethylene glycol coating plate (Corning 3881), at 80 μ L per well. The 96-well plate was then placed at 37 °C under quiescent conditions on a plate reader (Fluostar Omega, Fluostar Optima or Fluostar Galaxy; BMG Labtech) and after excitation at 440 nm, ThT fluorescence was measured at 480 nm, through the bottom of the plate.

Transmission Electron Microscopy

TEM experiments were performed as described previously [290]. Briefly, 5 μ L aliquots from each sample were removed after the aggregation reaction had reached the final plateau and placed on a carbon support film on 400-mesh copper grid (EM Resolutions Ltd.). After adsorption, grids were negatively stained with 2% w/v uranyl acetate and images were recorded using a FEI Tecnai G₂ transmission electron microscope (Cambridge Advanced Imaging Centre) and analyzed using the SIS Megaview II Image Capture system (Olympus).

Dot-blot assay

The dot-blot assay was performed as described previously [212] while samples were prepared as for the kinetic experiments albeit in the absence of ThT. 2 μ L aliquots of each sample were removed at different time points and spotted onto a nitrocellulose membrane (0.2 μ m, Whatman). The membranes were then dried, blocked with 5% non-fat dry milk in TBST and incubated with an A β 42 fibril-specific primary antibody (OC, Millipore) and Alexa-Fluor

488-conjugated secondary antibody (Life Technologies) according to the manufacturer's instructions. Fluorescence detection was performed using the Typhoon Trio Imager (GE Healthcare) and fluorescence quantification was performed using the ImageJ software (National Institutes of Health).

Cell viability experiments

The WM164 human melanoma cell line was grown in Dulbecco's Modified Eagle Medium (DMEM) (Gibco, UK), supplemented with 10 % fetal bovine serum (FBS) and 1 % penicillin/streptomycin (Pen-Strep). Cells were cultured in 96-well plates at a concentration of 5000 cells/well and grown at 37 °C. After 24 h, synthetic or recombinantly produced cyclic peptides, diluted in fresh medium, were added to the WM164 cells and incubated at 37 °C for 48 h in the absence or presence of 100 µg/ml cisplatin. After treatment, MTT solution (5 mg/mL in PBS) was added to the medium at a final concentration of 0.5 mg/mL and the plates were incubated for an additional 3 h at 37 °C. Then, 100 µl of solubilization solution (DMSO) was added to each well to dissolve the dark blue formazan crystals formed and the absorbance was measured at 570 nm using a TECAN Safire II-Basic plate reader. For cell viability measurements of human mesenchymal stem cells derived from Wharton's Jelly (WJ-MSC), a similar procedure was followed with slight modifications.

***C. elegans* motility assay**

Strains. The following strains were used for these experiments: (1) GMC101; genotype *dvIs100 [unc-54p::A-beta-1-42::unc-54 3'- UTR + mtl-2p::GFP]*; *mtl-2p::GFP* constitutively expresses the green fluorescent protein (GFP) in intestinal cells; *unc-54p::A-beta-1-42*

expresses the human A β 42 peptide in body wall muscle cells, resulting in A β 42 aggregation and worm paralysis after temperature up-shift from 20° to 25°C [341]. (2) N2, wild-type *C. elegans* var Bristol, herein referred to as control worms[354]].

Propagation procedures. *C. elegans* worms were propagated using standard conditions and as described previously [212,354]. Briefly, the worms were treated with hypochlorite bleach, eggs were hatched overnight in M9 buffer (3 g/L KH₂PO₄, 6 g/L Na₂HPO₄, 5 g/L NaCl and 1 mM MgSO₄), and then distributed on nematode growth medium (NGM) [1 mM CaCl₂, 1 mM MgSO₄, 5 mg/ml cholesterol, 250 mM KH₂PO₄ (pH 6), 17 g/L agar, 3 g/L NaCl and 7.5 g/L casein] plates seeded with the *E. coli* OP50 cells and incubated at 20°C. Upon reaching the L4 stage, ~700 worms were placed on NGM plates containing the desired concentration of the cyclic peptides in 1% acetonitrile as well as 75 μ M 5-fluoro-2'-deoxyuridine (FUDR) to inhibit growth of offspring. The plates were then transferred to 24 °C in order to promote A β 42 expression and aggregation.

Motility assay. On days 3 through 10 of adulthood, worms were collected using M9 buffer and distributed on unseeded 9-cm NGM plates. The worms' movements were recorded at 30 frames per second for 1 min using a homemade microscopic setup and the body bends were quantified using a custom-tracking algorithm as described previously [212]. In total, ~2300 worms were analysed per drug with an average of ~200 worms per experiment. Total fitness refers to the sum of the mobility and speed of the worms.

Aggregate quantification. Staining and microscopy were performed as described previously [212]. Briefly, live animals were stained by incubating with 1 mM NIAD-4 (0.1% DMSO in M9 buffer) for 6 hours at room temperature and then transferred on NGM plates to allow destaining for about 16 hours. Stained worms were then anaesthetized by adding 40 mM NaN₃ and mounted on 2% agarose pads on glass microscope slides. Images were captured

using a Zeiss Axio Observer D1 fluorescence microscope (Carl Zeiss Microscopy GmbH) with a 20× objective and a 49004 ET-CY3/TRITC filter (Chroma Technology Corp.) and fluorescence intensity was calculated using the ImageJ software (National Institutes of Health). Only the head region of the worms was examined due to the high background signal in the intestines.

***C. elegans* paralysis assay**

The paralysis assay was performed using standard procedures and as described previously [330]. Briefly, synchronized L4 larvae CL4176 {smg-1(cc546) I; dvIs27 [myo-3::A β (1-42)-let 3'UTR(pAF29); pRF4 (rol-6(su1006))} [342] (~150-300 per condition) were transferred to NGM plates, that were seeded with *E. coli* OP50 and contained 10 μ M of the cyclic peptides in 1% acetonitrile, and were incubated at 16 °C for 48 h before transgene induction via temperature up-shift to 25 °C. 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. Scoring of paralyzed animals was initiated 24 h after temperature up-shift. Worms were considered as paralyzed upon failure to move their half end-body upon prodding, while animals that died were excluded from the assay. The log-rank (Mantel–Cox) test was used to evaluate differences between paralysis curves and to determine *P* values for all independent data. *n* in paralysis figures is the number of animals that paralyzed over the total number of animals used (the number of paralyzed animals plus the number of dead and censored animals). Median paralysis values are expressed as mean \pm s.e.m.

Statistical analyses

Statistical analyses were performed using Prism (GraphPad Software Inc, USA) and mean values were compared using unpaired two-tailed t-tests. For animal experiments, group sizes were chosen based on prior experience and literature precedence, so that sufficient numbers remained at the endpoints of the experiment. No samples, worms or data points were excluded from the reported analyses.

References

1. Elliott, W.H. & Elliott, D.C. *Biochemistry and molecular biology*, (Oxford University Press, New York, 2009).
2. Anfinsen, C.B., Haber, E., Sela, M. & White, F.H. The kinetics of formation of native ribonuclease during oxidation of the reduced polypeptide chain. *Proc Natl Acad Sci U S A* **47**, 1309-1314 (1961).
3. Levinthal, C. How to fold gracefully. in *Mossbauer Spectroscopy in Biological Systems*, Vol. 67 22-24 (University of Illinois Press, Urbana., Allerton House, Monticello, Illinois., 1969).
4. Levinthal, C. Are there pathways for protein folding? *J Chim Physique* **65**, 44-45 (1968).
5. Wetlaufer, D.B. Nucleation, rapid folding and globular intrachain regions in proteins. *Proc Natl Acad Sci U S A* **70**(1973).
6. Go, N. Theoretical studies of protein folding. *Annu Rev Biophys Bioeng* **12**, 183–210 (1983).
7. Tsong, T.Y., Baldwin, R.L. & McPhie, P. A sequential model of nucleation-dependent protein-folding: Kinetic studies of ribonuclease A. *J Mol Biol* **63**, 453-475 (1972).
8. Karplus, M. & Weaver, D.L. Protein-folding dynamics. *Nature* **260**, 404-406 (1976).
9. Karplus, M. & Weaver, D.L. Folding dynamics: The diffusion collision model and experimental data. *Protein Sci* **3**, 650-668 (1994).
10. Kim, P.S. & Baldwin, R.L. Specific intermediates in the folding reactions of small proteins and the mechanism of folding. *Annu Rev Biochem* **51**, 459-489 (1982).

11. Harrison, S.C. & Durbin, R. Is there a single pathway for the folding of a polypeptide chain? *Proc Natl Acad Sci U S A* **82**, 4028–4030 (1985).
12. Baldwin, R.L. How does protein folding get started? *Trends Biochem Sci* **14**, 291-294 (1989).
13. Kauzmann, W. Some factors in the interpretation of protein denaturation. *Adv Protein Chem* **14**, 1-63 (1959).
14. Tanford, C., Buckley III, C.E., De, P.K. & Lively, E.P. Effect of ethylene glycol on the conformation of gamma-globulin and beta-lactoglobulin. *J Biol Chem* **237**, 1168-1171 (1962).
15. Fersht, A.R. Optimization of rates of protein folding: The nucleation-condensation mechanism and its implications. *Proc Natl Acad Sci U S A* **92**, 10869-10873 (1995).
16. Fersht, A.R. Nucleation mechanisms in protein folding. *Curr Opin Struct Biol* **7**, 3-9 (1997).
17. Sali, A., Shakhnovich, E. & Karplus, M. How does a protein fold? *Nature* **369**, 248-251 (1994).
18. Bryngelson, J.D., Onuchic, J.N., Socci, N.D. & Wolynes, P.G. Funnels, pathways, and the energy landscape of protein folding: A synthesis. *Proteins* **21**, 167-195 (1995).
19. Wolynes, P., Onuchic, J. & Thirumalai, D. Navigating the folding routes. *Science* **267**, 1619-1620 (1995).
20. 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 U S A* **92**, 3626-3630 (1995).
21. Socci, N.D., Onuchic, J.N. & Wolynes, P.G. Diffusive dynamics of the reaction coordinate for protein folding funnels. *J Chem Phys* **104**, 5860-5868 (1996).
22. Dobson, C.M. Protein folding and misfolding. *Nature* **426**, 884-890 (2003).

23. Dill, K.A. & Chan, H.S. From Levinthal to pathways to funnels. *Nat Struct Biol* **4**, 10-19 (1997).
24. Gazit, E. The "correctly folded" state of proteins: Is it a metastable state? *Angew Chem Int Ed* **41**, 257-259 (2002).
25. Baldwin, A.J., et al. Metastability of native proteins and the phenomenon of amyloid formation. *J Am Chem Soc* **133**, 14160-14163 (2011).
26. Clark, P. Protein folding in the cell: Reshaping the folding funnel. *Trends Biochem Sci* **29**, 527-534 (2004).
27. Jahn, T.R. & Radford, S.E. Folding versus aggregation: Polypeptide conformations on competing pathways. *Arch Biochem Biophys* **469**, 100-117 (2008).
28. Ellis, R.J. & Minton, A.P. Join the crowd. *Nature* **425**, 27-28 (2003).
29. Chow, M.K., Ellisdon, A.M., Cabrita, L.D. & Bottomley, S.P. Polyglutamine expansion in ataxin-3 does not affect protein stability. *J Biol Chem* **279**, 47643-47651 (2004).
30. Chiti, F. & Dobson, C.M. Protein misfolding, amyloid formation, and human disease: A summary of progress over the last decade. *Annu Rev Biochem* **86**, 27-68 (2017).
31. Uversky, V.N., Oldfield, C.J. & Dunker, A.K. Intrinsically disordered proteins in human diseases: Introducing the D2 concept. *Annu Rev Biophys* **37**, 215-246 (2008).
32. De Baets, G., et al. An evolutionary trade-off between protein turnover rate and protein aggregation favors a higher aggregation propensity in fast degrading proteins. *PLoS Comput Biol* **7**, e1002090 (2011).
33. Hipp, M.S., Park, S.-H. & Hartl, F.U. Proteostasis impairment in protein-misfolding and -aggregation diseases. *Trends Cell Biol* **24**, 506-514 (2014).
34. Ritossa, F.M. A new puffing pattern induced by a temperature shock and DNP in *Drosophila*. *Experientia* **18**, 571-573 (1962).
35. Lindquist, S. The heat-shock response. *Annu Rev Biochem* **55**, 1151-1191 (1986).

36. Brehme, M., et al. A chaperome subnetwork safeguards proteostasis in aging and neurodegenerative disease. *cell rep* **9**, 1135-1150 (2014).
37. Balchin, D., Hayer-Hartl, M. & Hartl, F.U. In vivo aspects of protein folding and quality control. *Science* **353**, aac4354 (2016).
38. Dupré, D.J., Hammad, M.M., Holland, P. & Wertman, J. Role of chaperones in G protein coupled receptor signaling complex assembly. in *GPCR Signalling Complexes – Synthesis, Assembly, Trafficking and Specificity* 131-154 (ed. Dupré, D.J., Hébert, T.T., Jockers, R., Springer Science+Business Media, Dordrecht, 2012).
39. Treusch, S., Cyr, D.M. & Lindquist, S. Amyloid deposits: Protection against toxic protein species? *Cell Cycle* **8**, 1668-1674 (2009).
40. Wilson, M.R. & Easterbrook-Smith, S.B. Clusterin is a secreted mammalian chaperone. *Trends Biochem Sci* **25**, 95-98 (2000).
41. Göthel, S.F. & Marahiel, M.A. Peptidyl-prolyl cis-trans isomerases, a superfamily of ubiquitous folding catalysts. *Cell Mol Life Sci* **55**, 423-436 (1999).
42. Freedman, R.B., Hirst, T.R. & Tuite, M.F. Protein disulphide isomerase: building bridges in protein folding. *Trends Biochem Sci* **19**, 331-336 (1994).
43. Mizuno, Y., Hattori, N., Mori, H., Suzuki, T. & Tanaka, K. Parkin and Parkinson's disease. *Curr Opin Neurol* **14**, 477-482 (2001).
44. Platt, F.M., d' Azzo, A., Davidson, B.L., Neufeld, E.F. & Tiffit, C.J. Lysosomal storage diseases. *Nat Rev Dis Primers* **4**, 27 (2018).
45. Soti, C. & Csermely, P. Aging and molecular chaperones. *Exp Gerontol* **38**, 1037–1040 (2003).
46. Cuervo, A.M. & Dice, J.F. Age-related decline in chaperone-mediated autophagy. *J Biol Chem* **275**, 31505-31513 (2000).

47. Saez, I. & Vilchez, D. The mechanistic links between proteasome activity, aging and age-related diseases. *Curr Genomics* **15**, 38-51 (2014).
48. Carrell, R.W. & Lomas, D.A. Conformational disease. *Lancet* **350**, 134-138 (1997).
49. Marinko, J.T., et al. Folding and misfolding of human membrane proteins in health and disease: From single molecules to cellular proteostasis. *Chem Rev* (2019).
50. Hou, Z.S., Ulloa-Aguirre, A. & Tao, Y.X. Pharmacoperone drugs: Targeting misfolded proteins causing lysosomal storage-, ion channels-, and G protein-coupled receptors-associated conformational disorders. *Expert Rev Clin Pharmacol* **11**, 611-662 (2018).
51. Petrov, D., Mansfield, C., Moussy, A. & Hermine, O. ALS clinical trials review: 20 years of failure. Are we any closer to registering a new treatment? *Front Aging Neurosci* **68**(2017).
52. Briggs, R., Kennelly, S.P. & O'Neill, D. Drug treatments in Alzheimer's disease. *Clin Med (Lond)* **16**, 247-253 (2016).
53. Oertel, W.H. Recent advances in treating Parkinson's disease. *FI000Res* **6**, 260 (2017).
54. Coelho, T., et al. Mechanism of action and clinical application of tafamidis in hereditary transthyretin amyloidosis. *Neurol Ther* **5**, 1-25 (2016).
55. Warnock, D.G., et al. Oral migalastat HCl leads to greater systemic exposure and tissue levels of active α -galactosidase A in Fabry patients when co-administered with infused agalsidase. *PLoS One* **10**, e0134341 (2015).
56. Condren, M.E. & Bradshaw, M.D. Ivacaftor: A novel gene-based therapeutic approach for cystic fibrosis. *J Pediatr Pharmacol Ther* **18**, 8-13 (2013).
57. Taylor-Cousar, J.L., et al. Tezacaftor-ivacaftor in patients with cystic fibrosis homozygous for Phe508del. *N Engl J Med* **377**, 2013-2023 (2017).
58. Vertex Pharmaceuticals Incorporated. (2018) *FDA approves SYMDEKOTM (tezacaftor/ivacaftor and ivacaftor) to treat the underlying cause of cystic fibrosis in*

people ages 12 and older with certain mutations in the CFTR gene. Available at <https://investors.vrtx.com/static-files/aa8947ae-56ea-41a3-998a-33a3d60bc437>.

Accessed 03/03/2019.

59. Bulloch, M.N., Hanna, C. & Giovane, R. Lumacaftor/ivacaftor, a novel agent for the treatment of cystic fibrosis patients who are homozygous for the F580del CFTR mutation. *Expert Rev Clin Pharmacol* **10**, 1055-1072 (2017).
60. Carlile, G.W., et al. A novel triple combination of pharmacological chaperones improves F508del-CFTR correction. *Sci Rep* **8**, 11404 (2018).
61. Valastyan, J.S. & Lindquist, S. Mechanisms of protein-folding diseases at a glance. *Dis Model Mech* **7**, 9-14 (2014).
62. Watson, M.S., et al. Cystic fibrosis population carrier screening: 2004 revision of American College of Medical Genetics mutation panel. *Genet Med* **6**, 387-391 (2004).
63. Wang, X., et al. Hsp90 cochaperone Aha1 downregulation rescues misfolding of CFTR in cystic fibrosis. *Cell* **127**, 803-815 (2006).
64. Meacham, G.C., Patterson, C., Zhang, W., Younger, J.M. & Cyr, D.M. The Hsc70 co-chaperone CHIP targets immature CFTR for proteasomal degradation. *Nat Cell Biol* **3**, 100-105 (2001).
65. Balch, W.E., Roth, D.M. & Hutt, D.M. Emergent properties of proteostasis in managing cystic fibrosis. *Cold Spring Harb Perspect Biol* **3**, a004499 (2011).
66. Michelin, K., et al. Biochemical study on β -glucosidase in individuals with Gaucher's disease and normal subjects. *Clin Chim Acta* **343**, 145-153 (2004).
67. Ron, I. & Horowitz, M. ER retention and degradation as the molecular basis underlying Gaucher disease heterogeneity. *Hum Mol Genet* **14**, 2387-2398 (2005).
68. Joerger, A.C. & Fersht, A.R. Structural biology of the tumor suppressor p53 and cancer-associated mutants. *Adv Cancer Res* **97**(2007).

69. Milner, J. & Medcalf, E.A. Cotranslation of activated mutant p53 with wild type drives the wild-type p53 protein into the mutant conformation. *Cell* **65**, 765-774 (1991).
70. Vijayakumaran, R., Tan, K.H., Miranda, P.J., Haupt, S. & Haupt, Y. Regulation of mutant p53 protein expression. *Front Oncol* **5**, 1-8 (2015).
71. Xu, J., et al. Gain of function of mutant p53 by coaggregation with multiple tumor suppressors. *Nat Chem Biol* **7**, 285-295 (2011).
72. Khani, P., et al. Keratins and epidermolysis bullosa simplex. *J Cell Physiol* **234**, 289-297 (2018).
73. Werner, N.S., et al. Epidermolysis bullosa simplex-type mutations alter the dynamics of the keratin cytoskeleton and reveal a contribution of actin to the transport of keratin subunits. *Mol Biol Cell* **15**, 990-1002 (2004).
74. Coulombe, P.A. & Lee, C.H. Defining keratin protein function in skin epithelia: Epidermolysis bullosa simplex and its aftermath. *J Invest Dermatol* **132**, 763-775 (2012).
75. Perlmutter, D.H. Alpha-1-antitrypsin deficiency: Importance of proteasomal and autophagic degradative pathways in disposal of liver disease-associated protein aggregates. *Annu Rev Med* **62**(2011).
76. Salsi, V., Vigano, M.A., Cocchiarella, F., Mantovani, R. & Zappavigna, V. Hoxd13 binds in vivo and regulates the expression of genes acting in key pathways for early limb and skeletal patterning. *Dev Biol* **317**, 497-507 (2008).
77. Albrecht, A.N., et al. A molecular pathogenesis for transcription factor associated poly-alanine tract expansions. *Hum Mol Genet* **13**, 2351-2359 (2004).
78. Mendes, H.F., van der Spuy, J., Chapple, J.P. & Cheetham, M.E. Mechanisms of cell death in rhodopsin retinitis pigmentosa: implications for therapy. *Trends Mol Med* **11**, 177-185 (2005).

79. Gragg, M. & Park, P.S.-H. Misfolded rhodopsin mutants display variable aggregation properties. *Biochim Biophys Acta Mol Basis Dis* **1864**, 2938-2948 (2018).
80. Li, J., Parker, B., Martyn, C., Natarajan, C. & Guo, J. The PMP22 gene and its related diseases. *Mol Neurobiol* **47**, 673-698 (2012).
81. Dickson, K.M., et al. Association of calnexin with mutant peripheral myelin protein-22 ex vivo: A basis for “gain-of-function” ER diseases. *Proc Natl Acad Sci U S A* **99**, 9852–9857 (2002).
82. Strupp, M., Zwergal, A. & Brandt, T. Episodic ataxia type 2. *Neurotherapeutics* **4**, 267-273 (2007).
83. Mezghrani, A., et al. A destructive interaction mechanism accounts for dominant-negative effects of misfolded mutants of voltage gated calcium channels. *J Neurosci* **28**, 4501-4511 (2008).
84. Ingram, V.M. Gene mutations in human haemoglobin: The chemical difference between normal and sickle cell haemoglobin. *Nature* **180**, 326-328 (1957).
85. Dean, J. & Schechter, A.N. Sickle-cell anemia: Molecular and cellular bases of therapeutic approaches (first of three parts). *N Engl J Med* **299**, 752-763 (1978).
86. Dong, L.M. & Weisgraber, K.H. Human apolipoprotein E4 domain interaction. Arginine 61 and glutamic acid 255 interact to direct the preference for very low density lipoproteins. *J Biol Chem* **271**, 19053-19057 (1996).
87. Chen, H.K., et al. Apolipoprotein E4 domain interaction mediates detrimental effects on mitochondria and is a potential therapeutic target for Alzheimer disease. *J Biol Chem* **286**, 5215-5221 (2011).
88. Nathan, B.P., et al. Differential effects of apolipoproteins E3 and E4 on neuronal growth in vitro. *Science* **264**, 850-852 (1994).

89. Ma, J.Y., Yee, A., Brewer, H.B.J., Das, S., & Potter, H. Amyloid-associated proteins alpha 1-antichymotrypsin and apolipoprotein E promote assembly of Alzheimer beta-protein into filaments. *Nature* **372**, 92-94 (1994).
90. Belsches-Jablonski, A.P., Demory, M.L., Parsons, J.T. & Parsons, S.J. The Src pathway as a therapeutic strategy. *Drug Discov Today Ther Strateg* **2**, 313-321 (2005).
91. Knowles, T.P.J., Vendruscolo, M. & Dobson, C.M. The amyloid state and its association with protein misfolding diseases. *Nat Rev Mol Cell Biol* **15**, 384-396 (2014).
92. Eisenberg, D. & Jucker, M. The amyloid state of proteins in human diseases. *Cell* **148**, 1188-1203 (2012).
93. Dobson, C.M. The amyloid phenomenon and its links with human disease. *Cold Spring Harb Perspect Biol* **9**(2017).
94. Cohen, S.I., Vendruscolo, M., Dobson, C.M. & Knowles, T.P. From macroscopic measurements to microscopic mechanisms of protein aggregation. *J Mol Biol* **421**, 160-171 (2012).
95. Karamanos, T.K., Kalverda, A.P., Thompson, G.S. & Radford, S.E. Mechanisms of amyloid formation revealed by solution NMR. *Prog Nucl Magn Reson Spectrosc* **88-89**, 86-104 (2015).
96. 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* **24**, 2055-2075 (2018).
97. Scannevin, R.H. Therapeutic strategies for targeting neurodegenerative protein misfolding disorders. *Curr Opin Chem Biol* **44**, 66-74 (2018).
98. Blancas-Mejia, L.M., et al. Immunoglobulin light chain amyloid aggregation. *Chem Commun (Camb)* **54**, 10664-10674 (2018).

99. Saraiva, M.J.M. Transthyretin mutations in health and disease. *Hum Mutat* **5**, 191-196 (1995).
100. Meehan, S., et al. Amyloid fibril formation by lens crystallin proteins and its implications for cataract formation. *J Biol Chem* **279**, 3413-3419 (2004).
101. Marin-Argany, M., et al. Cell damage in light chain amyloidosis: Fibril internalization, toxicity and cell-mediated seeding. *J Biol Chem* **291**, 19813-19825 (2016).
102. Reixach, N., Deechongkit, S., Jiang, X., Kelly, J.W. & Buxbaum, J.N. Tissue damage in the amyloidoses: Transthyretin monomers and nonnative oligomers are the major cytotoxic species in tissue culture. *Proc Natl Acad Sci U S A* **101**, 2817-2822 (2004).
103. Soto, C. Transmissible proteins: Expanding the prion heresy. *Cell* **149**, 968-977 (2012).
104. Meyer-Luehmann, M., et al. Exogenous induction of cerebral beta-amyloidogenesis is governed by agent and host. *Science* **313**, 1781-1784 (2006).
105. Fuster-Matanzo, A., Hernández, F. & Ávila, J. Tau spreading mechanisms; implications for dysfunctional tauopathies. *Int J Mol Sci* **19**, 645-659 (2018).
106. Iba, M., et al. Synthetic tau fibrils mediate transmission of neurofibrillary tangles in a transgenic mouse model of Alzheimer's-like tauopathy. *J Neurosci* **33**, 1024-1037 (2013).
107. Blanco, L.P., Evans, M.L., Smith, D.R., Badtke, M.P. & Chapman, M.R. Diversity, biogenesis and function of microbial amyloids. *Trends Microbiol* **20**, 66-73 (2011).
108. Platt, F.M. Emptying the stores: Lysosomal diseases and therapeutic strategies. *Nat Rev Drug Discov* **17**, 133-150 (2017).
109. Dobson, C.M. Principles of protein folding, misfolding and aggregation. *Semin Cell Dev Biol* **15**, 3-16 (2004).
110. Denny, R.A., Gavrin, L.K. & Saiah, E. Recent developments in targeting protein misfolding diseases. *Bioorg Med Chem Lett* **23**, 1935-1944 (2013).

111. Bieging, K.T. & Attardi, L.D. Deconstructing p53 transcriptional networks in tumor suppression. *Trends Cell Biol* **22**, 97-106 (2012).
112. Joerger, A.C. & Fersht, A.R. Structure-function-rescue: The diverse nature of common p53 cancer mutants. *Oncogene* **26**, 2226-2242 (2007).
113. Reinhardt, H.C. & Schumacher, B. The p53 network: cellular and systemic DNA damage responses in aging and cancer. *Trends Genet* **28**, 128-136 (2012).
114. Uversky, V.N. p53 proteoforms and intrinsic disorder: An illustration of the protein structure–function continuum concept. *Int J Mol Sci* **17**, 1874 (2016).
115. Khoo, K.H., Mayer, S. & Fersht, A.R. Effects of stability on the biological function of p53. *J Biol Chem* **284**, 30974-30980 (2009).
116. Shadfan, M., Lopez-Pajares, V. & Yuan, Z.M. MDM2 and MDMX: Alone and together in regulation of p53. *Transl Cancer Res* **1**, 88-89 (2012).
117. Joerger, A.C. & Fersht, A.R. The p53 pathway: Origins, inactivation in cancer, and emerging therapeutic approaches. *Annu Rev Biochem* **85**, 375-404 (2016).
118. Xu, J., et al. Heterogeneity of Li-Fraumeni syndrome links to unequal gain-of-function effects of p53 mutations. *Sci Rep* **4**, 4223 (2014).
119. Bullock, A.N. & Fersht, A.R. Rescuing the function of mutant p53. *Nat Rev Cancer* **1**, 68-76 (2001).
120. Brachmann, R.K., Yu, K., Eby, Y., Pavletich, N.P. & Boeke, J.D. Genetic selection of intragenic suppressor mutations that reverse the effect of common p53 cancer mutations. *EMBO J* **17**, 1847-1859 (1998).
121. Joerger, A.C., Ang, H.C. & Fersht, A.R. Structural basis for understanding oncogenic p53 mutations and designing rescue drugs. *Proc Natl Acad Sci U S A* **103**, 15056-15061 (2006).

122. Gaiddon, C., Lokshin, M., Ahn, J., Zhang, T. & Prives, C. A subset of tumor-derived mutant forms of p53 down-regulate p63 and p73 through a direct interaction with the p53 core domain. *Mol Cell Biol* **21**, 1874-1887 (2001).
123. Zheng, T., et al. Spliced MDM2 isoforms promote mutant p53 accumulation and gain-of-function in tumorigenesis. *Nat Commun* **4**(2013).
124. Alzheimer, A., Stelzmann, R.A., Schnitzlein, H.N. & Murtagh, F.R. An English translation of Alzheimer's 1907 paper, "Uber eine eigenartige Erkankung der Hirnrinde". *Clin Anat* **8**, 429-431 (1995).
125. Selkoe, D.J. Alzheimer's disease: Genes, proteins, and therapy. *Physiol Rev* **81**, 741-766 (2001).
126. Wimo, A., et al. The worldwide costs of dementia 2015 and comparisons with 2010. *Alzheimers Dement* **13**, 1-7 (2017).
127. Association, A.s. 2016 Alzheimer's disease facts and figures. *Alzheimers Dement* **12**, 459-509 (2016).
128. Zhu, X.C., et al. Rate of early onset Alzheimer's disease: A systematic review and meta-analysis. *Ann Transl Med* **3**, 38 (2015).
129. Guerreiro, R.J., Gustafson, D.R. & Hardy, J. The genetic architecture of Alzheimer's disease: Beyond APP, PSENs and APOE. *Neurobiol Aging* **33**, 437-456 (2010).
130. De Strooper, B., Iwatsubo, T. & Wolfe, M.S. Presenilins and γ -secretase: Structure, function, and role in Alzheimer Disease. *Cold Spring Harb Perspect Med* **2**, a006304 (2012).
131. Barão, S., Moechars, D., Lichtenthaler, S.F. & De Strooper, B. BACE1 physiological functions may limit its use as therapeutic target for Alzheimer's disease. *Trends Neurosci* **39**, 158-169 (2016).

132. Graham, W.V., Bonito-Oliva, A. & Sakmar, T.P. Update on Alzheimer's disease therapy and prevention strategies. *Annu Rev Med* **68**, 413-430 (2017).
133. BrightFocus Foundation. (2000) *Amyloid plaques and neurofibrillary tangles*. Available at <https://www.brightfocus.org/alzheimers-disease/infographic/amyloid-plaques-and-neurofibrillary-tangles>. Accessed 03/03/2019.
134. 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).
135. Benilova, I., Karran, E. & De Strooper, B. The toxic A β oligomer and Alzheimer's disease: An emperor in need of clothes. *Nat Neurosci* **15**, 349-357 (2012).
136. Lee, S.J., Nam, E., Lee, H.J., Savelieff, M.G. & Lim, M.H. Towards an understanding of amyloid- β oligomers: Characterization, toxicity mechanisms, and inhibitors. *Chem Soc Rev* **46**, 310-323 (2017).
137. Sengupta, U., Nilson, A.N. & Kaye, R. The role of amyloid- β oligomers in toxicity, propagation, and immunotherapy. *EBioMedicine* **6**, 42-49 (2016).
138. Kaye, R. & Lasagna-Reeves, C.A. Molecular mechanisms of amyloid oligomers toxicity. *J Alzheimers Dis* **33**, S67-78 (2013).
139. Tseng, B.P., Green, K.N., Chan, J.L., Blurton-Jones, M. & LaFerla, F.M. A β inhibits the proteasome and enhances amyloid and tau accumulation. *Neurobiol Aging* **29**, 1607-1618 (2007).
140. Benson, M.D. Liver transplantation and transthyretin amyloidosis. *Muscle Nerve* **47**, 157-162 (2013).
141. Stangou, A.J., et al. Hereditary fibrinogen A alpha-chain amyloidosis: Phenotypic characterization of a systemic disease and the role of liver transplantation. *Blood* **115**, 2998-3007 (2010).

142. Ashwin, P.T., Shah, S. & Wolfssohn, J.S. Advances in cataract surgery. *Clin Exp Optom* **92**, 333-342 (2009).
143. Platt, F.M. & Jeyakumar, M. Substrate reduction therapy. *Acta Paediatr* **97**, 88-93 (2008).
144. Brady, R.O. Enzyme replacement for lysosomal diseases. *Annu Rev Med* **57**, 283-296 (2006).
145. Zhang, Y.Q. & Sarge, K.D. Celastrol inhibits polyglutamine aggregation and toxicity though induction of the heat shock response. *J Mol Med (Berl)* **85**, 1421-1428 (2007).
146. Warrick, J.M., et al. Suppression of polyglutamine-mediated neurodegeneration in *Drosophila* by the molecular chaperone HSP70. *Nat Genet* **23**, 425-428 (1999).
147. Auluck, P.K., Chan, H.Y., Trojanowski, J.Q., Lee, V.M. & Bonini, N.M. Chaperone suppression of alpha-synuclein toxicity in a *Drosophila* model for Parkinson's disease. *Science* **295**, 865-868 (2002).
148. Baldo B., et al. A screen for enhancers of clearance identifies huntingtin as a heat shock protein 90 (Hsp90) client protein. *J Biol Chem* **287**, 1406-1414 (2012).
149. Ko, H.S., et al. CHIP regulates leucine-rich repeat kinase-2 ubiquitination, degradation, and toxicity. *Proc Natl Acad Sci U S A* **106**, 2897-2902 (2009).
150. Li, D., et al. Functional inactivation of endogenous MDM2 and CHIP by HSP90 causes aberrant stabilization of mutant p53 in human cancer cells. *Mol Cancer Res* **9**, 577-588 (2011).
151. Lee, H., et al. Ganetespib targets multiple levels of the receptor tyrosine kinase signaling cascade and preferentially inhibits ErbB2-overexpressing breast cancer cells. *Sci Rep* **8**, 6829 (2018).
152. Berger, Z., et al. Rapamycin alleviates toxicity of different aggregate-prone proteins. *Hum Mol Genet* **15**, 433-442 (2006).

153. Webb, J.L., Ravikumar, B., Atkins, J., Skepper, J.N. & Rubinsztein, D.C. Alpha-synuclein is degraded by both autophagy and the proteasome. *J Biol Chem* **278**, 25009–25013 (2003).
154. Cooley, C.B., et al. Unfolded protein response activation reduces secretion and extracellular aggregation of amyloidogenic immunoglobulin light chain. *Proc Natl Acad Sci U S A* **111**, 13046-13051 (2014).
155. Shoulders, M.D., et al. Stress-independent activation of XBP1s and/or ATF6 reveals three functionally diverse ER proteostasis environments. *Cell Rep* **3**, 1279–1292 (2013).
156. Ostrowitzki, S., Lasser, R.A., et al. A phase III randomized trial of gantenerumab in prodromal Alzheimer's disease. *Alzheimers Res Ther* **9**, 95 (2017).
157. Sevigny, J., et al. The antibody aducanumab reduces A β plaques in Alzheimer's disease. *Nature* **537**, 50-56 (2016).
158. Ultsch, M., et al. Structure of crenezumab complex with A β shows loss of β -hairpin. *Sci Rep* **6**, 39374 (2016).
159. Logovinsky, V., et al. Safety and tolerability of BAN2401 - a clinical study in Alzheimer's disease with a protofibril selective A β antibody. *Alzheimers Res Ther* **8**, 14 (2016).
160. Weihofen, A., et al. Development of an aggregate-selective, human-derived α -synuclein antibody BIIB054 that ameliorates disease phenotypes in Parkinson's disease models. *Neurobiol Dis* **124**, 276-288 (2018).
161. Schenk, D.B., et al. First-in-human assessment of PRX002, an anti-alpha-synuclein monoclonal antibody, in healthy volunteers. *Mov Disord* **32**, 211-218 (2017).

162. Maier, M., et al. A human-derived antibody targets misfolded SOD1 and ameliorates motor symptoms in mousemodels of amyotrophic lateral sclerosis. *Sci Transl Med* **10**, 470 (2018).
163. 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).
164. Coric, V., et al. Targeting prodromal Alzheimer disease with avagacestat: A randomized clinical trial. *JAMA Neurol* **72**, 1324-1333 (2015).
165. Egan, M.F., et al. Randomized trial of verubecestat for mild-to-moderate Alzheimer's disease. *N Engl J Med* **378**, 1691-1703 (2018).
166. Gafni, J., et al. Inhibition of calpain cleavage of huntingtin reduces toxicity: Accumulation of calpain/caspase fragments in the nucleus. *J Biol Chem* **279**, 20211-20220 (2004).
167. Graham, R.K., et al. Cleavage at the caspase-6 site is required for neuronal dysfunction and degeneration due to mutant huntingtin. *Cell* **125**, 1179-1191 (2006).
168. Berke, S.J., Schmied, F.A., Brunt, E.R., Ellerby, L.M. & Paulson, H.L. Caspase-mediated proteolysis of the polyglutamine disease protein ataxin-3. *J Neurochem* **89**, 908-918 (2004).
169. Haacke, A., et al. Proteolytic cleavage of polyglutamine-expanded ataxin-3 is critical for aggregation and sequestration of non-expanded ataxin-3. *Hum Mol Genet* **15**, 555-568 (2006).
170. Li, W., et al. Aggregation promoting C-terminal truncation of alpha-synuclein is a normal cellular process and is enhanced by the familial Parkinson's disease-linked mutations. *Proc Natl Acad Sci U S A* **102**, 2162-2167 (2005).

171. Liu, C.W., et al. A precipitating role for truncated alpha-synuclein and the proteasome in alpha-synuclein aggregation: Implications for pathogenesis of Parkinson disease. *J Biol Chem* **208**, 22670-22678 (2005).
172. Gamblin, T.C., et al. Caspase cleavage of tau: Linking amyloid and neurofibrillary tangles in Alzheimer's disease. *Proc Natl Acad Sci U S A* **100**, 10032-10037 (2003).
173. Yin, H. & Kuret, J. C-terminal truncation modulates both nucleation and extension phases of tau fibrillization. *FEBS Lett* **580**, 211-215 (2006).
174. Page, L.J., et al. Metalloendoprotease cleavage triggers gelsolin amyloidogenesis. *EMBO J* **24**, 4124-4132 (2005).
175. Avila, J. Tau phosphorylation and aggregation in Alzheimer's disease pathology. *FEBS Lett* **580**, 2922-2927 (2006).
176. Le Corre, S., et al. An inhibitor of tau hyperphosphorylation prevents severe motor impairments in tau transgenic mice. *Proc Natl Acad Sci U S A* **103**, 9673-9678 (2006).
177. Noble, W., et al. Inhibition of glycogen synthase kinase-3 by lithium correlates with reduced tauopathy and degeneration in vivo. *Proc Natl Acad Sci U S A* **102**, 6990-6995 (2005).
178. Bhat, R., et al. Structural insights and biological effects of glycogen synthase kinase 3-specific inhibitor AR-A014418. *J Biol Chem* **278**, 45937-45945 (2003).
179. Garber, K. Alnylam launches era of RNAi drugs. *Nat Biotechnol* **36**, 777-778 (2018).
180. Monteys, A.M., Ebanks, S.A., Keiser, M.S. & Davidson, B.L. CRISPR/Cas9 editing of the mutant huntingtin allele in vitro and in vivo. *Mol Ther* **25**, 12-23 (2017).
181. Bjursell, M., et al. Therapeutic genome editing with CRISPR/Cas9 in a humanized mouse model ameliorates α 1-antitrypsin deficiency phenotype. *EBioMedicine* **29**, 104-111 (2018).
182. Eisenstein, M. CRISPR takes on Huntington's disease. *Nature* **557**, S42-S43 (2018).

183. Meng, X., et al. Two small molecules restore stability to a subpopulation of the cystic fibrosis transmembrane conductance regulator with the predominant disease-causing mutation. *J Biol Chem* **292**, 3706-3719 (2017).
184. Donaldson, S.H., et al. Tezacaftor/ivacaftor in subjects with cystic fibrosis and F508del/F508del-CFTR or F508del/G551D-CFTR. *Am J Respir Crit Care Med* **197**, 214–224 (2018).
185. Boeckler, F.M., et al. Targeted rescue of a destabilized mutant of p53 by an in silico screened drug. *Proc Natl Acad Sci U S A* **105**, 10360-10365 (2008).
186. Liu, X., et al. Small molecule induced reactivation of mutant p53 in cancer cells. *Nucleic Acids Res* **41**, 6034-6044 (2013).
187. Rosen, D.R., et al. Mutations in Cu/Zn superoxide dismutase gene are associated with familial amyotrophic lateral sclerosis. *Nature* **362**, 59-62 (1993).
188. Ray, S.S., Nowak, R.J., Brown, R.H.J. & Lansbury, P.T.J. Small-molecule-mediated stabilization of familial amyotrophic lateral sclerosis-linked superoxide dismutase mutants against unfolding and aggregation. *Proc Natl Acad Sci U S A* **102**, 3639-3644 (2005).
189. 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).
190. Capper, M.J., et al. The cysteine-reactive small molecule ebselen facilitates effective SOD1 maturation. *Nat Commun* **9**, 1693 (2018).
191. Ahmad, B., Borana, M.S. & Chaudhary, A.P. Understanding curcumin-induced modulation of protein aggregation. *Int J Biol Macromol* **100**, 89-96 (2017).
192. Cornejo, A., et al. Rosmarinic acid prevents fibrillization and diminishes vibrational modes associated to β sheet in tau protein linked to Alzheimer's disease. *J Enzyme Inhib Med Chem* **32**, 945-953 (2017).

193. Chemerovski-Glikman, M., et al. Rosmarinic Acid Restores Complete Transparency of Sonicated Human Cataract Ex Vivo and Delays Cataract Formation In Vivo. *Sci Rep* **8**, 9341 (2018).
194. Taguchi, Y., et al. Specific biarsenical labeling of cell surface proteins allows fluorescent- and biotin-tagging of amyloid precursor protein and prion proteins. *Mol Biol Cell* **20**, 233-244 (2009).
195. Heiser, V., et al. Inhibition of huntingtin fibrillogenesis by specific antibodies and small molecules: Implications for Huntington's disease therapy. *Proc Natl Acad Sci U S A* **97**, 6739-6744 (2000).
196. Lendel, C., et al. On the mechanism of nonspecific inhibitors of protein aggregation: Dissecting the interactions of alpha-synuclein with Congo red and lacmoid. *Biochemistry* **48**, 8322-8334 (2009).
197. Seki, T., et al. Congo red, an amyloid-inhibiting compound, alleviates various types of cellular dysfunction triggered by mutant protein kinase C γ that causes spinocerebellar ataxia type 14 (SCA14) by inhibiting oligomerization and aggregation. *J Pharmacol Sci* **114**, 206–216 (2010).
198. Zhou, W., et al. At low concentrations, 3,4-dihydroxyphenylacetic acid (DOPAC) binds non-covalently to alpha-synuclein and prevents its fibrillation. *J Mol Biol* **388**, 597-610 (2009).
199. Swift, J., et al. Identification of aggregation inhibitors of the human antibody light chain repertoire by phage display. *Protein Eng Des Sel* **27**, 405-409 (2014).
200. Sato, M., et al. Site-specific inhibitory mechanism for amyloid β 42 aggregation by catechol-type flavonoids targeting the Lys residues. *J Biol Chem* **288**, 23212-23224 (2013).

201. Ehrnhoefer, D.E., et al. EGCG redirects amyloidogenic polypeptides into unstructured, off-pathway oligomers. *Nat Struct Mol Biol* **15**, 558-566 (2008).
202. Ferraz da Costa, D.C., et al. Resveratrol prevents p53 aggregation in vitro and in breast cancer cells. *Oncotarget* **9**, 29112-29122 (2018).
203. Ladiwala, A.R.A., et al. Resveratrol selectively remodels soluble oligomers and fibrils of amyloid A β into off-pathway conformers. *J Biol Chem* **285**, 24228-24237 (2010).
204. Friedemann, P. (2018) *Sunphenon EGCg (Epigallocatechin-Gallate) in the early stage of Alzheimer's disease*. Available at <https://ClinicalTrials.gov/show/NCT00951834>. Accessed 21/01/2019.
205. Dandona, P. (2012) *Resveratrol in type 2 diabetes and obesity*. Available at <https://ClinicalTrials.gov/show/NCT01158417>. Accessed 03/03/2019.
206. Delatycki, M. (2014) *A study of resveratrol as treatment for Friedreich ataxia*. Available at <https://ClinicalTrials.gov/show/NCT01339884>. Accessed 03/03/2019.
207. Hôpitaux de Paris. (2018) *Resveratrol and Huntington disease*. Available at <https://ClinicalTrials.gov/show/NCT02336633>. Accessed 03/03/2019.
208. Johns Hopkins University. (2018) *BDPP treatment for mild cognitive impairment (MCI) and prediabetes or type 2 diabetes mellitus (T2DM)*. Available at <https://ClinicalTrials.gov/show/NCT02502253>. Accessed 03/03/2019.
209. Taniguchi, S., et al. Inhibition of heparin-induced tau filament formation by phenothiazines, polyphenols, and porphyrins. *J Biol Chem* **280**, 7614-7623 (2005).
210. Ono, K. & Yamada, M. Antioxidant compounds have potent antifibrillogenic and fibril-destabilizing effects for alpha-synuclein fibrils in vitro. *J Neurochem* **97**, 105-115 (2006).

211. Hu, Q., et al. Baicalein inhibits α -synuclein oligomer formation and prevents progression of α -synuclein accumulation in a rotenone mouse model of Parkinson's disease. *Biochim Biophys Acta* **1862**, 1883-1890 (2016).
212. 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).
213. Song, Y., et al. Protective effect of centipede grass against Abeta oligomerization and Abeta-mediated cell death in PC12 cells. *Pharm Biol* **53**, 1260-1266 (2015).
214. Neuropore Therapies Inc. (2016) *Phase 1 study of NPT200-11 in healthy subjects*. Available at <https://ClinicalTrials.gov/show/NCT02606682>. Accessed 03/03/2019.
215. Wilcock, G.K., et al. Potential of low dose leuco-methylthioninium bis(hydromethanesulphonate) (LMTM) monotherapy for treatment of mild Alzheimer's disease: Cohort analysis as modified primary analysis in a phase 3 clinical trial. *J Alzheimers Dis* **61**, 435-457 (2018).
216. Soto, C. & Martin, Z. Therapeutic strategies against protein misfolding in neurodegenerative diseases. *Expert Opin Drug Discov* **4**, 71-84 (2009).
217. Pallitto, M.M., Ghanta, J., Heinzelman, P., Kiessling, L.L. & Murphy, R.M. Recognition sequence design for peptidyl modulators of beta-amyloid aggregation and toxicity. *Biochemistry* **38**, 3570-3578 (1999).
218. Findeis, M.A., et al. Characterization of cholyl-Leu-Val-Phe-Phe-Ala-OH as an inhibitor of amyloid beta-peptide polymerization. *Amyloid* **8**, 231-241 (2001).
219. Gordon, D.J., Sciarretta, K.L. & Meredith, S.C. Inhibition of β -amyloid(40) fibrillogenesis and disassembly of β -amyloid(40) fibrils by short β -amyloid congeners containing n-methyl amino acids at alternate residues. *Biochemistry* **40**, 8237-8245 (2001).

220. Soto, C., et al. Beta-sheet breaker peptides inhibit fibrillogenesis in a rat brain model of amyloidosis: Implications for Alzheimer's therapy. *Nat Med* **4**, 822-826 (1998).
221. Soto, C., Kindy, M.S., Baumann, M. & Frangione, B. Inhibition of Alzheimer's amyloidosis by peptides that prevent β -sheet conformation. *Biochem Biophys Res Commun* **226**, 672-680 (1996).
222. Findeis, M.A., et al. Modified-peptide inhibitors of amyloid beta-peptide polymerization. *Biochemistry* **38**, 6791-6800 (1999).
223. Esteras-Chopo, A., Pastor, M.T., Serrano, L. & López de la Paz, M. New strategy for the generation of specific D-peptide amyloid inhibitors. *J Mol Biol* **377**, 1372-1381 (2008).
224. Eskici, G. & Gur, M. Computational design of new peptide inhibitors for amyloid beta (A β) aggregation in Alzheimer's disease: Application of a novel methodology. *PLoS One* **8**, e66178 (2013).
225. Hughes, E., Burke, R.M. & Doig, A.J. Inhibition of toxicity in the b-amyloid peptide fragment β -(25–35) using N-methylated derivatives. *J Biol Chem* **275**, 25109-25115 (2000).
226. Li, H., Monien, B.H., Fradinger, E.A., Urbanc, B. & Bitan, G. Biophysical characterization of A β 42 C-terminal fragments-inhibitors of A β 42 neurotoxicity. *Biochemistry* **49**, 1259-1267 (2010).
227. Soragni, A., et al. A designed inhibitor of p53 aggregation rescues p53 tumor suppression in ovarian carcinomas. *Cancer Cell* **29**, 90-103 (2016).
228. Banerjee, V., et al. Superoxide dismutase 1 (SOD1)-derived peptide inhibits amyloid aggregation of familial amyotrophic lateral sclerosis SOD1 mutants. *ACS Chem Neurosci* **7**, 1595-1606 (2016).

229. Dul, J.L., Davis, D.P., Williamson, E.K., Stevens, F.J. & Argon, Y. Hsp70 and antifibrillogenic peptides promote degradation and inhibit intracellular aggregation of amyloidogenic light chains. *J Cell Biol* **152**, 705-716 (2001).
230. Scrocchi, L.A., et al. Design of peptide-based inhibitors of human islet amyloid polypeptide fibrillogenesis. *J Mol Biol* **318**, 697-706 (2002).
231. Selivanova, G., et al. Restoration of the growth suppression function of mutant p53 by a synthetic peptide derived from the p53 C-terminal domain. *Nat Med* **3**, 632-638 (1997).
232. Baine, M., et al. Inhibition of Abeta42 aggregation using peptides selected from combinatorial libraries. *J Pept Sci* **15**, 499-503 (2009).
233. Kawasaki, T., Onodera, K. & Kamijo, S. Selection of peptide inhibitors of soluble Abeta(1-42) oligomer formation by phage display. *Biosci Biotechnol Biochem* **74**, 2214-2219 (2010).
234. Zompra, A.A., Galanis, A.S., Werbitzky, O. & Albericio, F. Manufacturing peptides as active pharmaceutical ingredients. *Future Med Chem* **1**, 361-377 (2009).
235. Chiti, F. & Dobson, C.M. Protein misfolding, functional amyloid, and human disease. *Annu Rev Biochem* **75**(2006).
236. Hegde, R.N., Subramanian, A., Pothukuchi, P., Parashuraman, S. & Luini, A. Rare ER protein misfolding-mistrafficking disorders: Therapeutic developments. *Tissue Cell* **49**, 175-185 (2017).
237. Dobson, C.M. Alzheimer's disease: Addressing a twenty-first century plague. *Rend Fis Acc Lincei* **26**, 251-262 (2015).
238. Kogej, T., et al. Big pharma screening collections: More of the same or unique libraries? The AstraZeneca-Bayer Pharma AG case. *Drug Discov Today* **18**, 1014-1024 (2013).

239. 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).
240. Passioura, T., Katoh, T., Goto, Y. & Suga, H. Selection-based discovery of druglike macrocyclic peptides. *Annu Rev Biochem* **83**(2014).
241. Halford, B. How DNA-encoded libraries are revolutionizing drug discovery. *Chem Eng News* **95**, 28-33 (2017).
242. Nixon, A.E., Sexton, D.J. & Ladner, R.C. Drugs derived from phage display: from candidate identification to clinical practice. *MAbs* **6**, 73-85 (2013).
243. Lau, J.L. & Dunn, M.K. Therapeutic peptides: Historical perspectives, current development trends, and future directions. *Bioorg Med Chem* **26**, 2700-2707 (2018).
244. Tapeinou, A., Matsoukas, M.T., Simal, C. & Tselios, T. Cyclic peptides on a merry-go-round; towards drug design. *Biopolymers* **104**, 453-461 (2015).
245. Kesavan, K. Development of peptide-based diagnostic and therapeutic agents in oncology. in *Peptide-based Drug Discovery: Challenges and New Therapeutics* 326-363 (ed. Srivastava, V., The Royal Society of Chemistry, Croydon, UK, 2017).
246. Jad, Y.E., El-Faham, A., de la Torre, B.G. & Albericio, F. Solid-phase peptide synthesis, the state of the art: Challenges and opportunities. in *Peptide-based Drug Discovery: Challenges and New Therapeutics* 518-550 (ed. Srivastava, V., The Royal Society of Chemistry, Croydon, UK, 2017).
247. Joo, S.H. Cyclic peptides as therapeutic agents and biochemical tools. *Biomol Ther* **20**, 19-26 (2012).
248. Rezai, T., Yu, B., Millhauser, G.L., Jacobson, M.P. & Lokey, R.S. Testing the conformational hypothesis of passive membrane permeability using synthetic cyclic peptide diastereomers. *J Am Chem Soc* **128**, 2510-2511 (2006).

249. Kwon, Y.U. & Kodadek, T. Quantitative comparison of the relative cell permeability of cyclic and linear peptides. *Chem Biol* **14**, 671-677 (2007).
250. Wang, C.K. & Craik, D.J. Cyclic peptide oral bioavailability: Lessons from the past. *Biopolymers* **106**, 901-909 (2016).
251. Nielsen, D.S., et al. Orally absorbed cyclic peptides. *Chem Rev* **117**, 8094-8128 (2017).
252. Demmer, O., Frank, A.O. & Kessler, H. Design of cyclic peptides. in *Peptide and Protein Design for Biopharmaceutical Applications* (ed. Jensen, K.J., John Wiley & Sons Ltd, West Sussex, UK, 2009).
253. Foster, A.D., et al. Methods for the creation of cyclic peptide libraries for use in lead discovery. *J Biomol Screen* **20**, 563-576 (2015).
254. Goodnow, R.A.J., Dumelin, C.E. & Keefe, A.D. DNA-encoded chemistry: Enabling the deeper sampling of chemical space. *Nat Rev Drug Discov* **16**(2017).
255. Harris, P.A., et al. DNA-encoded library screening identifies benzo[b][1,4]oxazepin-4-ones as highly potent and monoselective receptor interacting protein 1 kinase inhibitors. *J Med Chem* **59**, 2163–2178 (2016).
256. Halford, B. Breakthroughs with bar codes DNA-encoded libraries help pharma find drug leads. *Chem Eng News* **95**, 28-33 (2017).
257. Nuevolution. (2017) *Nuevolution technology progress: Nuevolution scales its compound collection to 40 trillion using its Chemetics™ drug discovery platform*. Available at <https://nuevolution.com/nuevolution-technology-progress-nuevolution-scales-its-compound-collection-to-40-trillion-using-its-chemetics-drug-discovery-platform/>. Accessed 03/03/2019.
258. Boder, E.T. & Wittrup, K.D. Yeast surface display for screening combinatorial polypeptide libraries. *Nat Biotechnol* **15**, 553-557 (1997).

259. Sidhu, S.S., Lowman, H.B., Cunningham, B.C. & Wells, J.A. Phage display for selection of novel binding peptides. *Methods Enzymol* **328**, 333-363 (2000).
260. Getz, J.A., Schoep, T.D. & Daugherty, P.S. Peptide discovery using bacterial display and flow cytometry. in *Protein Engineering for Therapeutics*, Vol. 503 75–97 (eds. Wittrup, K.D. & Verdine, G.L., Academic Press, 2012).
261. Plückthun, A. Ribosome display: A perspective. in *Ribosome Display and Related Technologies*, Vol. 805 (eds. Douthwaite J. & R., J., Springer, New York, NY, 2012).
262. Josephson, K., Ricardo, A. & Szostak, J.W. mRNA display: From basic principles to macrocycle drug discovery. *Drug Discov Today* **19**, 388-399 (2014).
263. Goto, Y., Katoh, T. & Suga, H. Flexizymes for genetic code reprogramming. *Nat Protoc* **6**, 779-790 (2011).
264. Passioura, T. & Suga, H. A RaPID way to discover nonstandard macrocyclic peptide modulators of drug targets. *Chem Commun (Camb)* **53**, 1931-1940 (2017).
265. Bertoldo, D., et al. Phage selection of peptide macrocycles against β -catenin to interfere with Wnt signaling. *ChemMedChem* **11**, 834-839 (2016).
266. Hacker, D.E., Hoinka, J., Iqbal, E.S., Przytycka, T.M. & Hartman, M.C. Highly constrained bicyclic scaffolds for the discovery of protease-stable peptides via mRNA display. *ACS Chem Biol* **12**, 795-804 (2017).
267. Davis, A.M., Plowright, A.T. & Valeur, E. Directing evolution: The next revolution in drug discovery? *Nat Rev Drug Discov* **16**, 681-698 (2017).
268. Orner, B.P., Liu, L., Murphy, R.M. & Kiessling, L.L. Phage display affords peptides that modulate beta-amyloid aggregation. *J Am Chem Soc* **128**, 11882-11889 (2006).
269. Scott, C.P., Abel-Santos, E., Wall, M., Wahnon, D.C. & Benkovic, S.J. Production of cyclic peptides and proteins in vivo. *Proc Natl Acad Sci U S A* **96**, 13638-13643 (1999).

270. Scott, C.P., Abel-Santos, E., Jones, A.D. & Benkovic, S.J. Structural requirements for the biosynthesis of backbone cyclic peptide libraries. *Chemistry & Biology* **8**, 801-815 (2001).
271. Wood, D.W. & Camarero, J.A. Intein applications: From protein purification and labeling to metabolic control methods. *J Biol Chem* **289**, 14512-14519 (2014).
272. Perler, F.B. Protein splicing mechanisms and applications. *IUBMB Life* **57**, 469-476 (2005).
273. Volkmann, G. & Iwai, H. Protein trans-splicing and its use in structural biology: Opportunities and limitations. *Mol Biosyst* **6**, 2110-2121 (2010).
274. Tavassoli, A. SICLOPPS cyclic peptide libraries in drug discovery. *Curr Opin Chem Biol* **38**, 30-35 (2017).
275. Bionda, N. & Fasan, R. Ribosomal synthesis of natural-product-like bicyclic peptides in *Escherichia coli*. *Chembiochem* **16**, 2011-2016 (2015).
276. Frost, J.R., Smith, J.M. & Fasan, R. Design, synthesis, and diversification of ribosomally derived peptide macrocycles. *Curr Opin Struct Biol* **23**, 571-580 (2013).
277. Satyanarayana, M., Vitali, F., Frost, J.R. & Fasan, R. Diverse organo-peptide macrocycles via a fast and catalyst-free oxime/intein-mediated dual ligation. *Chem Commun* **48**, 1461-1463 (2012).
278. Smith, J.M., Hill, N.C., Krasniak, P.J. & Fasan, R. Synthesis of bicyclic organo-peptide hybrids via oxime/intein-mediated macrocyclization followed by disulfide bond formation. *Org Biomol Chem* **12**, 1135-1142 (2014).
279. Gietz, R.D. & Schiestl, R.H. High-efficiency yeast transformation using the LiAc/SS carrier DNA/PEG method. *Nat Protoc* **2**, 31-34 (2007).
280. 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).

281. Jagadish, K., et al. Recombinant expression and phenotypic screening of a bioactive cyclotide against alpha-synuclein-induced cytotoxicity in baker's yeast. *Angew Chem Int Ed Engl* **54**, 8390-8394 (2015).
282. Tavassoli, A. & Benkovic, S.J. Split-intein mediated circular ligation used in the synthesis of cyclic peptide libraries in *E coli*. *Nat Protoc* **2**, 1126-1133 (2007).
283. Townend, J.E. & Tavassoli, A. Traceless production of cyclic peptide libraries in *E. coli*. *ACS Chem Biol* **11**, 1624-1630 (2016).
284. AstraZeneca. (2019) *Small molecules*. Available at <https://www.astrazeneca.com/what-science-can-do/drug-modalities/small-molecule.html>. Accessed 03/03/2019.
285. Lipinski, C.A. Drug-like properties and the causes of poor solubility and poor permeability. *J Pharmacol Toxicol Methods* **44**, 235-249 (2000).
286. Villar, E.A., et al. How proteins bind macrocycles. *Nat Chem Biol* **10**, 723-731 (2014).
287. Poongavanam, V., Doak, B.C. & Kihlberg, J. Opportunities and guidelines for discovery of orally absorbed drugs in beyond rule of 5 space. *Curr Opin Chem Biol* **44**, 23-29 (2018).
288. Saleh, L. & Perler, F.B. Protein splicing in cis and in trans. *Chem Rec* **6**, 183-193 (2006).
289. Wassman, C.D., et al. Computational identification of a transiently open L1/S3 pocket for reactivation of mutant p53. *Nat Commun* **4**, 1407 (2013).
290. Habchi, J., et al. Systematic development of small molecules to inhibit specific microscopic steps of A β 42 aggregation in Alzheimer's disease. *Proc Natl Acad Sci U S A* **114**, E200-E208 (2017).
291. Joshi, P., et al. A fragment-based method of creating small-molecule libraries to target the aggregation of intrinsically disordered proteins. *ACS Comb Sci* **18**, 144-153 (2016).

292. Gregoire, S., Irwin, J. & Kwon, I. Techniques for monitoring protein misfolding and aggregation in vitro and in living cells. *Korean J Chem Eng* **29**, 693-702 (2012).
293. Limbocker, R., et al. Trodusquemine enhances A β 42 aggregation but suppresses its toxicity by displacing oligomers from cell membranes. *Nat Commun* **10**, 225 (2019).
294. Chia, S., et al. SAR by kinetics for drug discovery in protein misfolding diseases. *Proc Natl Acad Sci U S A* **115**, 10245-10250 (2018).
295. Pujols, J., et al. High-throughput screening methodology to identify alpha-synuclein aggregation inhibitors. *Int J Mol Sci* **18**, 478 (2017).
296. Pujols, J., et al. Small molecule inhibits α -synuclein aggregation, disrupts amyloid fibrils, and prevents degeneration of dopaminergic neurons. *Proc Natl Acad Sci U S A* **115**, 10481-10486 (2018).
297. Yuste-Checa, P., et al. Pharmacological chaperoning: A potential treatment for PMM2-CDG. *Hum Mutat* **38**, 160-168 (2017).
298. Tal, P., et al. Cancer therapeutic approach based on conformational stabilization of mutant p53 protein by small peptides. *Oncotarget* **7**, 11817–11837 (2016).
299. Foster, B.A., Coffey, H.A., Morin, M.J. & Rastinejad, F. Pharmacological rescue of mutant p53 conformation and function. *Science* **286**, 2507-2510 (1999).
300. Rippin, T.M., et al. Characterization of the p53-rescue drug CP-31398 in vitro and in living cells. *Oncogene* **21**, 2119-2129 (2002).
301. Su, L.J., et al. Compounds from an unbiased chemical screen reverse both ER-to-Golgi trafficking defects and mitochondrial dysfunction in Parkinson's disease models. *Dis Model Mech* **3**, 194-208 (2010).
302. Griffioen, G., et al. A yeast-based model of alpha-synucleinopathy identifies compounds with therapeutic potential. *Biochim Biophys Acta* **1762**, 312-318 (2006).

303. Tardiff, D.F., et al. Yeast reveal a "druggable" Rsp5/Nedd4 network that ameliorates alpha-synuclein toxicity in neurons. *Science* **342**, 979-983 (2013).
304. 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).
305. 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).
306. McKoy, A.F., Chen, J., Schupbach, T. & Hecht, M.H. A novel inhibitor of amyloid beta (A β) peptide aggregation: From high throughput screening to efficacy in an animal model of Alzheimer disease. *J Biol Chem* **287**, 38992-39000 (2012).
307. Fox, A., et al. Selection for nonamyloidogenic mutants of islet amyloid polypeptide (IAPP) identifies an extended region for amyloidogenicity. *Biochemistry* **49**, 7783-7789 (2010).
308. Mayer, S., Rüdiger, S., Ang, H.C., Joerger, A.C. & Fersht, A.R. Correlation of levels of folded recombinant p53 in Escherichia coli with thermodynamic stability in vitro. *J Mol Biol* **372**, 268-276 (2007).
309. Foit, L., et al. Optimizing protein stability in vivo. *Mol Cell* **36**, 861-871 (2009).
310. Saunders, J.C., et al. An in vivo platform for identifying inhibitors of protein aggregation. *Nat Chem Biol* **12**, 94-101 (2016).
311. Ventura, S. & Navarro, S. Screening Protein Aggregation in Cells Using Fluorescent Labels Coupled to Flow Cytometry. in *Protein Misfolding Diseases* 195-212 (ed. Gomes, C., Humana Press, New York, NY, 2019).
312. Espargaró, A., Medina, A., Di Pietro, O., Muñoz-Torrero, D. & Sabate, R. Ultra rapid in vivo screening for anti-Alzheimer anti-amyloid drugs. *Sci Rep* **6**, 23349 (2016).

313. Espargaró, A., Sabate, R. & Ventura, S. Thioflavin-S staining coupled to flow cytometry. A screening tool to detect in vivo protein aggregation. *Mol Biosyst* **8**, 2839-2844 (2012).
314. Navarro, S., Carija, A., Munoz-Torrero, D. & Ventura, S. A fast and specific method to screen for intracellular amyloid inhibitors using bacterial model systems. *Eur J Med Chem* **121**, 785-792 (2016).
315. Espargaró, 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).
316. 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).
317. Joerger, A.C. & Fersht, A.R. The tumor suppressor p53: From structures to drug discovery. *Cold Spring Harb Perspect Biol* **2**, a000919 (2010).
318. Joerger, A.C., Allen, M.D. & Fersht, A.R. Crystal structure of a superstable mutant of human p53 core domain. Insights into the mechanism of rescuing oncogenic mutations. *J Biol Chem* **279**, 1291-1296 (2004).
319. Nikolova, P.V., Henckel, J., Lane, D.P. & Fersht, A.R. Semirational design of active tumor suppressor p53 DNA binding domain with enhanced stability. *Proc Natl Acad Sci U S A* **95**, 14675-14680 (1998).
320. Bullock, A.N., Henckel, J. & Fersht, A.R. Quantitative analysis of residual folding and DNA binding in mutant p53 core domain: Definition of mutant states for rescue in cancer therapy. *Oncogene* **19**, 1245-1256 (2000).

321. Wang, Q., Johnson, J.L., Agar, N.Y. & Agar, J.N. Protein aggregation and protein instability govern familial amyotrophic lateral sclerosis patient survival. *PLoS Biol* **6**, e170 (2008).
322. Valentine, J.S., Doucette, P.A. & Zittin Potter, S. Copper-zinc superoxide dismutase and amyotrophic lateral sclerosis. *Annu Rev Biochem* **74**, 563-593 (2005).
323. Juneja, T., Pericak-Vance, M.A., Laing, N.G., Dave, S. & Siddique, T. Prognosis in familial amyotrophic lateral sclerosis: Progression and survival in patients with Glu100Gly and Ala4Val mutations in Cu,Zn superoxide dismutase. *Neurology* **48**, 55-57 (1997).
324. Prudencio, M., Hart, P.J., Borchelt, D.R. & Andersen, P.M. Variation in aggregation propensities among ALS-associated variants of SOD1: Correlation to human disease. *Hum Mol Genet* **18**, 3217-3226 (2009).
325. Albin, R.L. & Tagle, D.A. Genetics and molecular biology of Huntington's disease. *Trends Neurosci* **18**, 11-14 (1995).
326. Chen, M. & Wolynes, P.G. Aggregation landscapes of Huntingtin exon 1 protein fragments and the critical repeat length for the onset of Huntington's disease. *Proc Natl Acad Sci U S A* **114**, 4406-4411 (2017).
327. Walker, F.O. Huntington's disease. *Lancet* **369**, 218-228 (2007).
328. Meriin, A.B., et al. Huntington toxicity in yeast model depends on polyglutamine aggregation mediated by a prion-like protein Rnq1. *J Cell Biol* **157**, 997-1004 (2002).
329. Navarro, S., Villar-Piqué, A. & S., V. Selection against toxic aggregation-prone protein sequences in bacteria. *Biochim Biophys Acta* **1843**, 866-874 (2014).
330. 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).

331. White, C.J. & Yudin, A.K. Contemporary strategies for peptide macrocyclization. *Nat Chem* **3**, 509-524 (2011).
332. Schmidt, U. & Langner, J. Cyclotetrapeptides and cyclopentapeptides: Occurrence and synthesis. *J Pept Res* **49**, 67-73 (1997).
333. Naumann, T.A., Savinov, S.N. & Benkovic, S.J. Engineering an affinity tag for genetically encoded cyclic peptides. *Biotechnol Bioeng* **92**, 820-830 (2005).
334. Lahiry, A., Fan, Y., Stimple, S.D., Raith, M. & Wood, D.W. Inteins as tools for tagless and traceless protein purification. *J Chem Technol Biotechnol* **93**, 1827-1835 (2017).
335. Weiss, J., Schwechheimer, K., Cavenee, W.K., Herlyn, M. & Arden, K.C. Mutation and expression of the p53 gene in malignant melanoma cell lines. *Int J Cancer* **54**, 693-699 (1993).
336. Christodoulou, I., Kolisis, F.N., Papaevangelidou, D. & Zoumpourlis, V. Comparative evaluation of human mesenchymal stem cells of fetal (Wharton's jelly) and adult (adipose tissue) origin during prolonged in vitro expansion: Considerations for cytotherapy. *Stem Cells Int* **2013**, 246134 (2013).
337. Niesen, F.H., Berglund, H. & Vedadi, M. The use of differential scanning fluorimetry to detect ligand interactions that promote protein stability. *Nat Protoc* **2**, 2212-2221 (2007).
338. Wilcken, R., Wang, G., Boeckler, F.M. & Fersht, A.R. Kinetic mechanism of p53 oncogenic mutant aggregation and its inhibition. *Proc Natl Acad Sci U S A* **109**, 13584-13589 (2012).
339. Cohen, S.I., et al. Proliferation of amyloid- β 42 aggregates occurs through a secondary nucleation mechanism. *Proc Natl Acad Sci U S A* **110**, 9758-9763 (2013).

340. Hellstrand, E., Boland, B., Walsh, D.M. & Linse, S. Amyloid beta-protein aggregation produces highly reproducible kinetic data and occurs by a two-phase process. *ACS Chem Neurosci* **1**, 13-18 (2010).
341. McColl, G., et al. Utility of an improved model of amyloid-beta (A β ₁₋₄₂) toxicity in *Caenorhabditis elegans* for drug screening for Alzheimer's disease. *Mol Neurodegener* **57**(2012).
342. Drake, J., Link, C.D. & Butterfield, D.A. Oxidative stress precedes fibrillar deposition of Alzheimer's disease amyloid beta-peptide (1-42) in a transgenic *Caenorhabditis elegans* model. *Neurobiol Aging* **24**, 415-420 (2003).
343. 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 aminothiols unnatural amino acids. *ACS Chem Biol* **10**, 1805-1816 (2015).
344. Maksimov, M.O., Pan, S.J. & James Link, A. Lasso peptides: Structure, function, biosynthesis, and engineering. *Nat Prod Rep* **29**, 996-1006 (2012).
345. Bi, T., Li, Y., Shekhtman, A. & Camarero, J.A. In-cell production of a genetically-encoded library based on the theta-defensin RTD-1 using a bacterial expression system. *Bioorg Med Chem* **26**, 1212-1219 (2018).
346. 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).
347. 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).
348. Liu, C.C. & Schultz, P.G. Adding new chemistries to the genetic code. *Annu Rev Biochem* **79**, 413-444 (2010).

349. Joung, J.K., Ramm, E.I. & Pabo, C.O. A bacterial two-hybrid selection system for studying protein–DNA and protein–protein interactions. *Proc Natl Acad Sci U S A* **97**, 7382-7387 (2000).
350. Dhanda, S.K., et al. Development of a novel clustering tool for linear peptide sequences. *Immunology* **155**, 331-345 (2018).
351. Bastian, M., Heymann, S. & Jacomy, M. Gephi: An open source software for exploring and manipulating networks. *Third Int AAAI Conf Weblogs Soc Media*, 361-362 (2009).
352. Girvan, M. & Newman, M.E. Community structure in social and biological networks. *Proc Natl Acad Sci U S A* **99**, 7821-7826 (2002).
353. Chia, S., et al. Monomeric and fibrillar α -synuclein exert opposite effects on the catalytic cycle that promotes the proliferation of A β 42 aggregates. *Proc Natl Acad Sci U S A* **114**, 8005-8010 (2017).
354. Brenner, S. The genetics of *Caenorhabditis elegans*. *Genetics* **77**, 71–94 (1974).

Appendices

Appendix A – Graphical representation of fitting the data obtained from DSF experiments involving T-p53C(Y220C), to the Boltzmann equation using the OriginPro software.

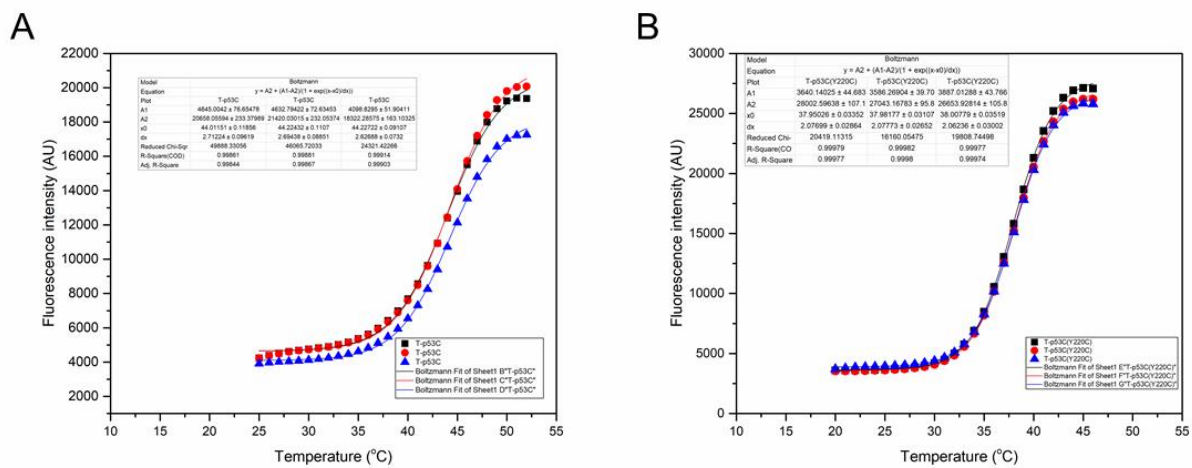


Figure A. 1. Graphical representation of the fluorescence data acquired by DSF for T-p53C and T-p53C(Y220C) as in Figure 5.11, after fitting to the Boltzmann equation using OriginPro.

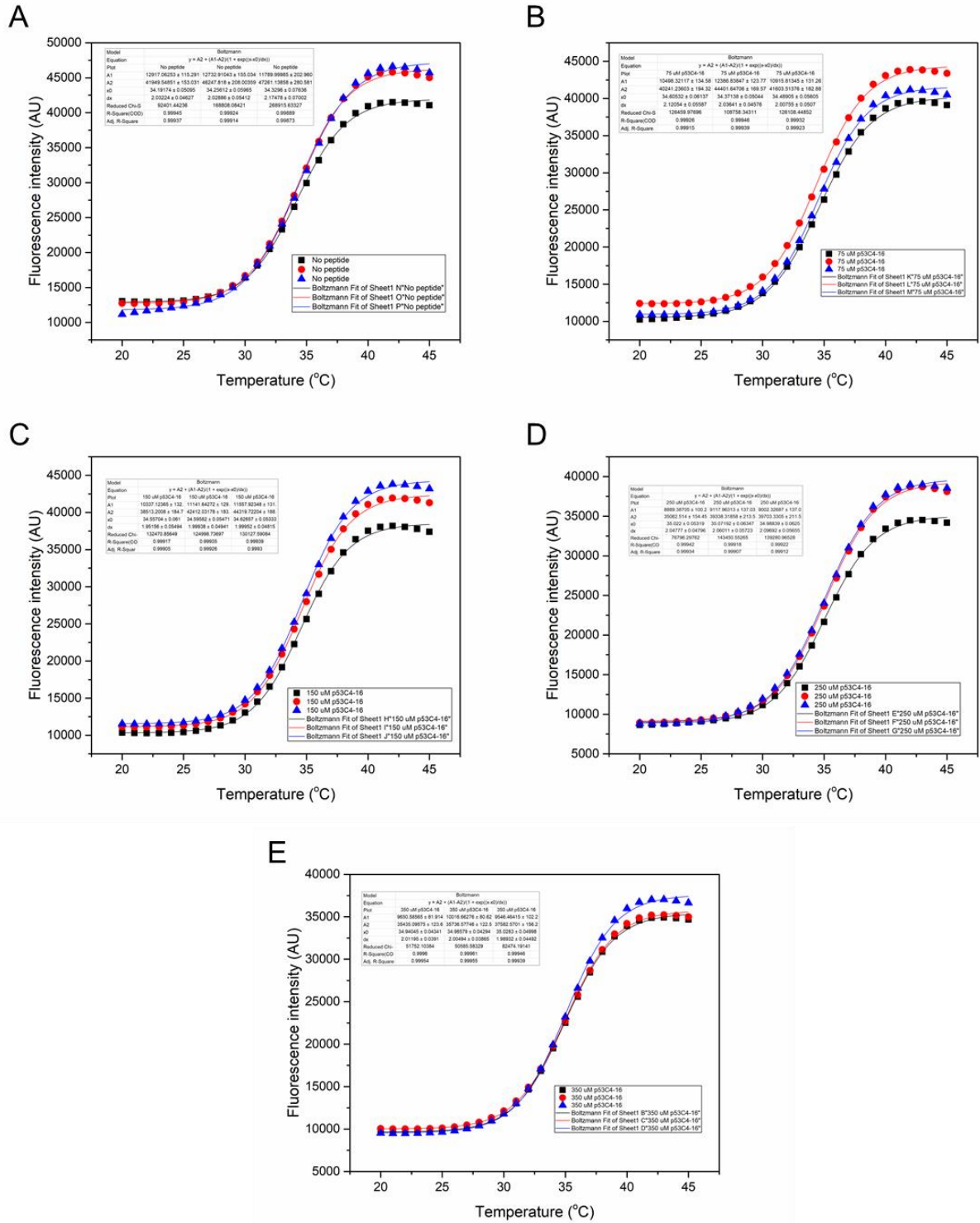


Figure A. 2. Graphical representation of the fluorescence data acquired by DSF for T-p53C(Y220C) in the presence and absence of p53C4-16 as in Figure 5.12, after fitting to the Boltzmann equation using OriginPro.

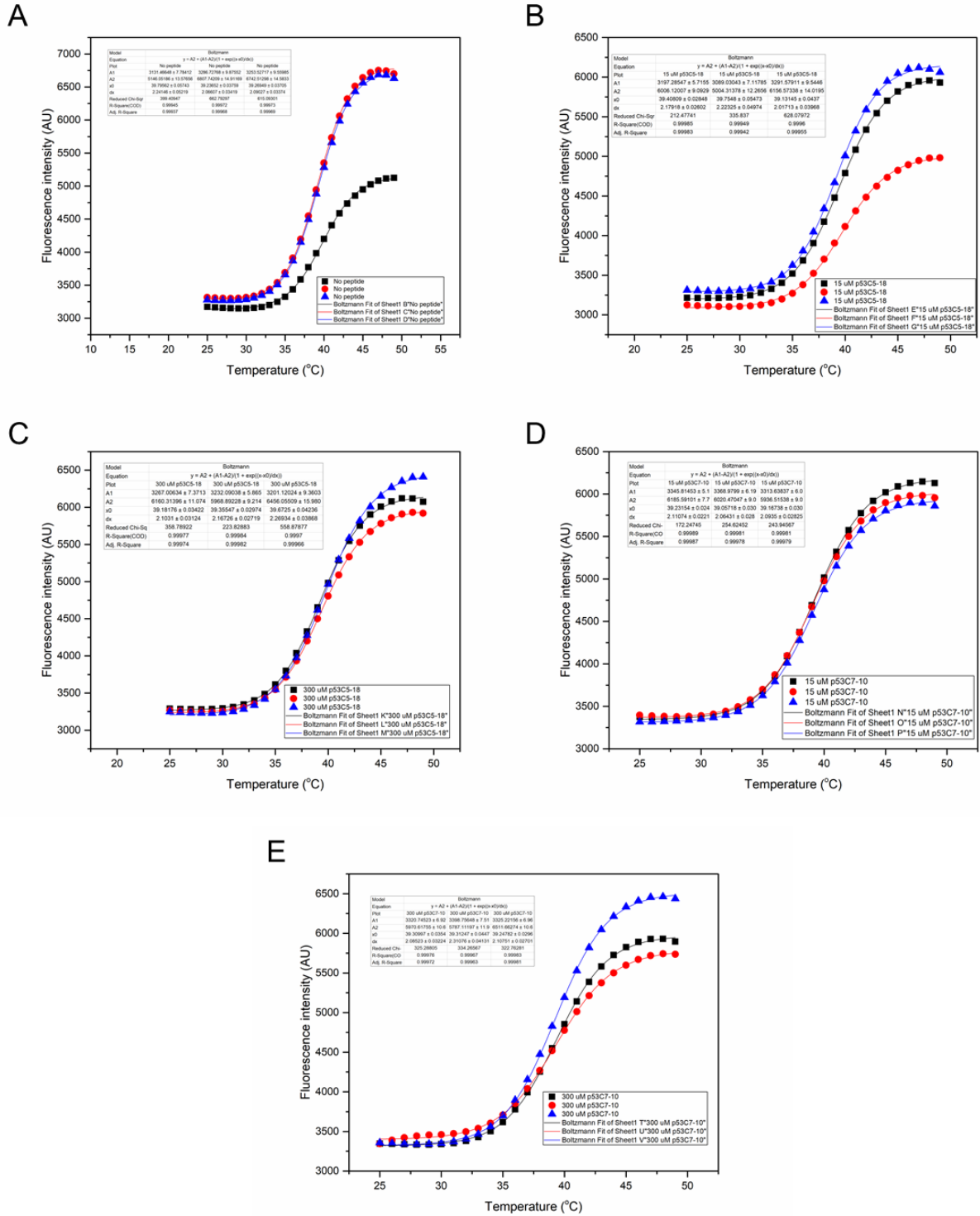


Figure A. 3. Graphical representation of the fluorescence data acquired by DSF for T-p53C(Y220C) in the presence and absence of p53C5-18 and p53C7-10 as in Figure 5.13 and Figure 5.14, after fitting to the Boltzmann equation using OriginPro.

Appendix B – Sequences and frequency of appearance of the A β -targeting heptapeptide sequences appearing at least twenty times within the sorted population, as determined by high-throughput sequencing of the enriched library after the 7th round of sorting.

Number	Peptide name	Cluster	Amino acid sequence							Number of reads	Reads / Total peptide reads (%)
1	A β C7-1	I	C	K	V	W	Q	L	L	64118	15.86
2	A β C7-2	I	C	K	V	W	M	P	L	42464	10.504
3	A β C7-3	I	C	R	V	W	T	E	L	28675	7.093
4	A β C7-4	-	C	R	L	T	S	K	A	23144	5.725
5	A β C7-5	I	C	R	V	W	M	V	P	20170	4.989
6	A β C7-6	I	C	R	V	W	C	A	L	18397	4.551
7	A β C7-7	I	C	R	V	W	Q	T	V	14267	3.529
8	A β C7-8	I	C	S	V	W	M	E	L	12638	3.126
9	A β C7-9	I	C	R	V	W	Q	A	L	11718	2.898
10	A β C7-10	I	C	K	V	W	Q	V	L	9249	2.288
11	A β C7-11	I	C	R	V	W	S	L	L	9124	2.257
12	A β C7-12	I	C	K	V	W	M	A	L	7513	1.858
13	A β C7-13	I	C	K	V	W	S	Q	L	7166	1.773
14	A β C7-14	II	C	R	I	V	P	S	L	6435	1.592
15	A β C7-15	I	C	R	V	W	Q	L	L	6030	1.492
16	A β C7-16	II	C	I	V	V	P	S	I	5871	1.452
17	A β C7-17	-	C	C	Y	V	K	H	T	5853	1.448
18	A β C7-18	III	C	G	S	P	I	V	L	5792	1.433
19	A β C7-19	I	C	R	V	W	C	E	L	5223	1.292
20	A β C7-20	I	C	K	V	W	M	E	V	5043	1.247
21	A β C7-21	-	C	V	Q	T	V	P	F	5041	1.247
22	A β C7-22	I	C	R	V	W	S	P	L	4843	1.198
23	A β C7-23	I	C	R	V	W	M	G	L	4212	1.042
24	A β C7-24	I	C	T	V	W	M	A	I	3654	0.904
25	A β C7-25	III	C	G	S	P	I	F	L	2618	0.648
26	A β C7-26	I	C	V	V	W	Q	P	L	2611	0.646
27	A β C7-27	I	C	R	V	W	Q	V	V	2595	0.642
28	A β C7-28	II	C	R	V	V	P	A	I	2408	0.596
29	A β C7-29	I	C	R	V	W	S	A	L	2327	0.576
30	A β C7-30	I	C	A	V	W	Q	A	L	2280	0.564
31	A β C7-31	II	C	Q	V	V	P	S	V	2004	0.496
32	A β C7-32	I	C	R	V	W	C	A	V	1971	0.488

Number	Peptide name	Cluster	Amino acid sequence							Number of reads	Reads / Total peptide reads (%)
33	AβC7-33	VII	C	E	V	R	L	A	I	1767	0.437
34	AβC7-34	IV	C	Q	V	V	W	R	R	1629	0.403
35	AβC7-35	IX	C	Y	Q	I	A	Y	I	1510	0.374
36	AβC7-36	VIII	C	R	P	V	A	W	W	1299	0.321
37	AβC7-37	V	C	A	V	R	S	Y	L	1257	0.311
38	AβC7-38	XI	C	T	V	P	N	V	I	1123	0.278
39	AβC7-39	X	C	R	W	M	W	C	R	1106	0.274
40	AβC7-40	I	C	K	V	W	C	V	M	1059	0.262
41	AβC7-41	I	C	R	V	W	Q	C	V	1028	0.254
42	AβC7-42	III	C	G	N	P	V	F	L	959	0.237
43	AβC7-43	I	C	A	V	W	M	Q	L	953	0.236
44	AβC7-44	III	C	G	C	T	I	F	L	937	0.232
45	AβC7-45	III	C	G	S	P	L	Y	L	894	0.221
46	AβC7-46	IV	C	S	V	V	L	F	R	870	0.215
47	AβC7-47	V	C	K	V	R	N	F	L	858	0.212
48	AβC7-48	III	C	G	N	P	V	M	L	839	0.208
49	AβC7-49	I	C	R	V	W	M	M	L	815	0.202
50	AβC7-50	I	C	R	V	W	S	V	V	803	0.199
51	AβC7-51	I	C	R	V	W	Q	E	V	796	0.197
52	AβC7-52	I	C	V	V	W	Q	Q	I	790	0.195
53	AβC7-53	VI	C	R	V	L	Y	A	Q	668	0.165
54	AβC7-54	II	C	V	V	V	P	S	I	655	0.162
55	AβC7-55	-	T	Y	Y	L	G	F	L	653	0.162
56	AβC7-56	III	C	G	S	V	I	F	L	652	0.161
57	AβC7-57	V	C	E	M	R	S	Y	L	612	0.151
58	AβC7-58	I	C	Q	V	W	M	D	L	608	0.15
59	AβC7-59	II	C	V	V	V	P	S	L	582	0.144
60	AβC7-60	I	C	R	V	W	Q	D	P	575	0.142
61	AβC7-61	VI	C	R	L	L	V	V	Q	490	0.121
62	AβC7-62	I	C	R	V	W	M	L	L	486	0.12
63	AβC7-63	I	C	V	V	W	Q	L	L	473	0.117
64	AβC7-64	VI	C	R	Y	L	I	A	Q	464	0.115
65	AβC7-65	XII	C	E	V	S	L	N	L	449	0.111
66	AβC7-66	XIV	C	R	W	G	L	D	L	446	0.11
67	AβC7-67	I	C	R	V	W	Y	G	I	435	0.108
68	AβC7-68	IV	C	V	V	V	W	R	K	427	0.106
69	AβC7-69	V	C	S	L	R	S	Y	L	421	0.104
70	AβC7-70	I	C	A	V	W	Q	L	L	376	0.093
71	AβC7-71	-	C	M	V	N	G	D	M	373	0.092
72	AβC7-72	I	C	N	V	W	Q	V	V	361	0.089
73	AβC7-73	VI	C	R	V	I	W	A	Q	356	0.088
74	AβC7-74	-	S	E	C	T	L	N	L	350	0.087

Number	Peptide name	Cluster	Amino acid sequence							Number of reads	Reads / Total peptide reads (%)
75	AβC7-75	I	C	K	V	W	T	V	V	335	0.083
76	AβC7-76	I	C	V	V	W	C	Q	P	326	0.081
77	AβC7-77	I	C	R	V	W	M	A	A	318	0.079
78	AβC7-78	I	C	R	V	W	M	T	V	316	0.078
79	AβC7-79	VIII	C	R	L	V	S	W	W	316	0.078
80	AβC7-80	I	C	R	V	W	Q	T	R	296	0.073
81	AβC7-81	II	C	L	V	V	P	S	V	291	0.072
82	AβC7-82	IV	C	M	L	V	W	R	R	288	0.071
83	AβC7-83	I	C	H	V	W	S	V	V	287	0.071
84	AβC7-84	V	C	T	L	R	T	F	L	284	0.07
85	AβC7-85	I	C	Y	V	W	C	P	L	273	0.068
86	AβC7-86	V	C	R	M	R	S	Y	L	267	0.066
87	AβC7-87	V	C	A	L	R	S	Y	L	257	0.064
88	AβC7-88	VI	C	R	I	L	V	A	Q	240	0.059
89	AβC7-89	II	C	D	I	I	P	S	L	228	0.056
90	AβC7-90	I	C	R	V	W	E	L	L	227	0.056
91	AβC7-91	I	C	R	V	R	Q	E	L	220	0.054
92	AβC7-92	III	C	G	T	V	V	Y	L	219	0.054
93	AβC7-93	I	C	R	V	W	C	F	C	218	0.054
94	AβC7-94	I	C	A	V	W	C	W	P	212	0.052
95	AβC7-95	I	C	K	V	W	M	W	K	211	0.052
96	AβC7-96	IV	C	Q	V	V	L	W	R	209	0.052
97	AβC7-97	I	C	R	V	W	S	V	P	207	0.051
98	AβC7-98	VIII	C	R	V	V	A	W	W	205	0.051
99	AβC7-99	IV	C	K	L	V	V	F	R	204	0.05
100	AβC7-100	V	C	S	L	R	T	F	L	200	0.049
101	AβC7-101	II	C	Y	Y	V	P	S	L	198	0.049
102	AβC7-102	I	C	R	V	W	T	Q	V	191	0.047
103	AβC7-103	III	C	G	S	V	L	W	L	190	0.047
104	AβC7-104	IX	C	Y	E	I	A	N	I	190	0.047
105	AβC7-105	X	C	R	W	V	W	C	R	187	0.046
106	AβC7-106	II	C	T	Y	V	P	S	L	179	0.044
107	AβC7-107	VI	C	R	I	L	Y	A	Q	176	0.044
108	AβC7-108	I	C	K	V	F	Q	V	L	173	0.043
109	AβC7-109	III	C	G	V	V	L	Y	L	171	0.042
110	AβC7-110	IV	C	T	M	V	V	W	R	167	0.041
111	AβC7-111	IV	C	R	T	V	V	W	R	166	0.041
112	AβC7-112	I	C	R	V	A	A	V	L	165	0.041
113	AβC7-113	I	C	R	V	W	C	L	P	165	0.041
114	AβC7-114	II	C	R	V	V	P	S	L	164	0.041
115	AβC7-115	I	C	R	V	W	S	V	R	163	0.04
116	AβC7-116	III	C	G	S	V	L	F	L	163	0.04

Number	Peptide name	Cluster	Amino acid sequence							Number of reads	Reads / Total peptide reads (%)
			C	E	V	A	I	N	L		
117	AβC7-117	XII	C	E	V	A	I	N	L	162	0.04
118	AβC7-118	I	C	V	V	W	C	T	R	161	0.04
119	AβC7-119	III	C	G	T	P	L	I	L	160	0.04
120	AβC7-120	IV	C	S	V	V	Y	W	R	159	0.039
121	AβC7-121	VII	C	S	V	M	V	A	I	159	0.039
122	AβC7-122	V	C	R	V	R	S	Y	L	159	0.039
123	AβC7-123	XV	C	Y	I	I	A	G	L	158	0.039
124	AβC7-124	III	C	G	R	P	V	L	L	158	0.039
125	AβC7-125	III	C	G	E	M	L	Y	L	154	0.038
126	AβC7-126	I	C	L	V	W	M	G	L	153	0.038
127	AβC7-127	V	C	R	Y	R	T	Y	L	151	0.037
128	AβC7-128	IV	C	S	V	V	L	W	R	150	0.037
129	AβC7-129	V	C	V	L	R	N	Y	L	149	0.037
130	AβC7-130	VIII	C	K	V	V	S	W	W	142	0.035
131	AβC7-131	VI	C	R	I	V	V	A	Q	139	0.034
132	AβC7-132	V	C	V	V	R	T	F	L	139	0.034
133	AβC7-133	XVI	C	V	V	P	N	M	V	138	0.034
134	AβC7-134	I	C	S	V	W	Q	S	L	137	0.034
135	AβC7-135	-	C	G	I	I	L	F	R	135	0.033
136	AβC7-136	IV	C	E	V	V	W	R	R	135	0.033
137	AβC7-137	I	C	R	V	W	S	P	C	134	0.033
138	AβC7-138	-	C	P	L	M	L	A	I	132	0.033
139	AβC7-139	IV	C	A	V	V	L	F	R	128	0.032
140	AβC7-140	III	C	G	R	P	I	V	L	127	0.031
141	AβC7-141	I	C	F	V	W	Q	C	R	125	0.031
142	AβC7-142	IV	C	S	V	L	V	F	R	124	0.031
143	AβC7-143	V	C	R	M	R	T	Y	L	122	0.03
144	AβC7-144	III	C	G	Y	V	V	Y	L	122	0.03
145	AβC7-145	I	C	R	V	W	C	I	I	122	0.03
146	AβC7-146	I	C	R	V	W	V	P	L	121	0.03
147	AβC7-147	-	C	R	V	F	I	M	Q	120	0.03
148	AβC7-148	III	C	G	S	P	V	Y	L	119	0.029
149	AβC7-149	V	C	K	W	R	T	F	L	118	0.029
150	AβC7-150	III	C	G	A	T	V	Y	L	115	0.028
151	AβC7-151	XIII	C	E	F	V	S	W	I	108	0.027
152	AβC7-152	VI	C	R	L	V	V	A	Q	108	0.027
153	AβC7-153	V	C	D	L	R	S	Y	L	107	0.026
154	AβC7-154	X	C	A	W	L	W	C	R	107	0.026
155	AβC7-155	I	C	R	V	S	Q	A	L	107	0.026
156	AβC7-156	VII	C	S	L	P	G	A	I	105	0.026
157	AβC7-157	-	C	M	W	I	E	C	V	105	0.026
158	AβC7-158	III	C	G	R	T	L	V	L	103	0.025

Number	Peptide name	Cluster	Amino acid sequence							Number of reads	Reads / Total peptide reads (%)
159	AβC7-159	IV	C	E	V	V	V	L	R	103	0.025
160	AβC7-160	II	C	K	V	V	P	T	L	101	0.025
161	AβC7-161	-	C	R	P	Y	M	P	G	101	0.025
162	AβC7-162	III	C	G	H	P	I	V	L	101	0.025
163	AβC7-163	-	C	L	M	D	G	W	V	101	0.025
164	AβC7-164	I	C	R	V	W	S	H	P	98	0.024
165	AβC7-165	XVI	C	S	V	P	N	L	V	97	0.024
166	AβC7-166	I	C	R	W	W	G	G	I	97	0.024
167	AβC7-167	III	C	G	S	V	I	W	L	97	0.024
168	AβC7-168	III	C	G	R	G	I	S	L	97	0.024
169	AβC7-169	III	C	G	D	V	L	W	L	97	0.024
170	AβC7-170	V	C	S	V	R	T	F	L	96	0.024
171	AβC7-171	III	C	G	E	I	V	V	L	94	0.023
172	AβC7-172	-	C	W	E	I	G	Y	I	92	0.023
173	AβC7-173	II	C	E	V	V	P	S	L	91	0.023
174	AβC7-174	I	C	I	V	W	Q	C	L	88	0.022
175	AβC7-175	IV	C	R	V	M	V	F	R	87	0.022
176	AβC7-176	-	C	G	P	L	I	E	L	87	0.022
177	AβC7-177	I	C	R	V	W	A	L	L	85	0.021
178	AβC7-178	XIII	C	E	F	V	S	W	H	84	0.021
179	AβC7-179	I	C	R	V	S	C	P	L	84	0.021
180	AβC7-180	IV	C	Q	L	L	V	F	R	84	0.021
181	AβC7-181	I	C	V	V	W	Q	G	L	84	0.021
182	AβC7-182	VII	C	S	V	R	L	A	I	84	0.021
183	AβC7-183	III	C	G	P	T	F	Y	L	83	0.021
184	AβC7-184	X	C	R	I	A	W	C	R	82	0.02
185	AβC7-185	-	C	Y	L	V	A	C	I	81	0.02
186	AβC7-186	XI	C	R	V	Q	A	V	I	80	0.02
187	AβC7-187	XIII	C	R	F	V	S	S	V	80	0.02
188	AβC7-188	IV	C	T	V	V	V	F	R	79	0.02
189	AβC7-189	IV	C	R	I	V	V	A	R	79	0.02
190	AβC7-190	III	C	G	S	P	V	L	L	79	0.02
191	AβC7-191	-	C	M	I	R	T	E	V	78	0.019
192	AβC7-192	I	C	V	V	W	S	P	L	77	0.019
193	AβC7-193	XV	C	W	R	I	A	E	L	77	0.019
194	AβC7-194	III	C	G	S	T	F	F	L	75	0.019
195	AβC7-195	-	C	E	I	S	W	C	F	74	0.018
196	AβC7-196	V	C	P	V	R	N	Y	L	74	0.018
197	AβC7-197	-	C	A	V	P	E	A	K	73	0.018
198	AβC7-198	III	C	G	G	T	L	Y	L	73	0.018
199	AβC7-199	VI	C	A	I	V	W	A	Q	72	0.018
200	AβC7-200	-	C	Y	E	I	R	Q	L	72	0.018

Number	Peptide name	Cluster	Amino acid sequence							Number of reads	Reads / Total peptide reads (%)
201	AβC7-201	IV	C	G	V	V	V	Y	R	72	0.018
202	AβC7-202	III	C	G	E	M	L	W	L	71	0.018
203	AβC7-203	II	C	W	T	V	G	T	I	71	0.018
204	AβC7-204	I	C	R	V	W	G	A	T	71	0.018
205	AβC7-205	-	C	K	I	V	P	A	F	70	0.017
206	AβC7-206	I	C	V	V	W	C	A	P	70	0.017
207	AβC7-207	-	C	V	V	Q	W	C	R	69	0.017
208	AβC7-208	-	C	C	G	T	R	M	R	69	0.017
209	AβC7-209	I	C	V	V	W	S	S	L	69	0.017
210	AβC7-210	IV	C	V	M	V	V	W	R	69	0.017
211	AβC7-211	-	C	Y	M	I	G	Y	K	68	0.017
212	AβC7-212	-	C	M	M	L	Y	F	R	68	0.017
213	AβC7-213	XIII	C	Y	F	V	S	W	V	66	0.016
214	AβC7-214	I	C	L	V	W	C	P	L	66	0.016
215	AβC7-215	VI	C	R	V	L	N	V	Q	65	0.016
216	AβC7-216	IV	C	F	M	V	V	W	R	64	0.016
217	AβC7-217	XVII	C	L	F	G	L	N	L	63	0.016
218	AβC7-218	III	C	G	D	P	L	M	L	63	0.016
219	AβC7-219	II	C	T	Y	V	P	S	I	63	0.016
220	AβC7-220	I	C	R	V	W	S	G	L	60	0.015
221	AβC7-221	XIII	C	G	W	V	T	E	L	60	0.015
222	AβC7-222	III	C	G	E	V	V	W	L	59	0.015
223	AβC7-223	-	C	P	L	V	V	R	V	59	0.015
224	AβC7-224	-	C	M	D	A	W	A	L	59	0.015
225	AβC7-225	-	C	Y	Q	V	S	G	L	57	0.014
226	AβC7-226	-	C	S	E	R	I	A	L	57	0.014
227	AβC7-227	I	C	R	V	A	Q	A	L	56	0.014
228	AβC7-228	II	C	R	V	V	P	P	I	55	0.014
229	AβC7-229	-	C	I	G	R	D	V	T	55	0.014
230	AβC7-230	I	C	R	V	S	A	A	L	54	0.013
231	AβC7-231	I	C	K	V	W	C	G	L	54	0.013
232	AβC7-232	I	C	V	V	T	P	V	V	54	0.013
233	AβC7-233	III	C	G	E	T	L	V	L	54	0.013
234	AβC7-234	I	C	R	V	T	A	A	L	54	0.013
235	AβC7-235	-	S	R	F	G	G	C	Q	54	0.013
236	AβC7-236	-	C	G	V	Y	M	R	V	53	0.013
237	AβC7-237	VII	C	E	I	R	L	A	L	53	0.013
238	AβC7-238	XVIII	C	R	P	C	G	W	V	53	0.013
239	AβC7-239	III	C	P	I	L	L	C	L	52	0.013
240	AβC7-240	-	C	C	A	Q	L	C	Q	52	0.013
241	AβC7-241	I	C	A	V	W	Q	A	V	51	0.013
242	AβC7-242	-	S	F	A	G	G	W	G	51	0.013

Number	Peptide name	Cluster	Amino acid sequence							Number of reads	Reads / Total peptide reads (%)
243	AβC7-243	III	C	G	E	P	I	L	L	51	0.013
244	AβC7-244	-	C	P	M	L	G	S	T	50	0.012
245	AβC7-245	I	C	R	V	W	S	A	V	50	0.012
246	AβC7-246	III	C	G	D	I	V	Y	L	50	0.012
247	AβC7-247	III	C	G	R	H	L	W	L	49	0.012
248	AβC7-248	XI	C	R	V	P	G	V	I	49	0.012
249	AβC7-249	-	C	E	V	A	W	P	F	48	0.012
250	AβC7-250	III	C	G	T	V	V	C	L	48	0.012
251	AβC7-251	I	C	V	V	W	S	T	I	48	0.012
252	AβC7-252	IV	C	Y	R	V	V	W	R	48	0.012
253	AβC7-253	III	C	G	V	P	L	F	L	47	0.012
254	AβC7-254	III	C	G	Q	P	V	V	L	46	0.011
255	AβC7-255	I	C	R	V	Y	C	V	P	46	0.011
256	AβC7-256	V	C	V	V	K	S	Y	L	46	0.011
257	AβC7-257	-	C	A	L	F	M	R	V	45	0.011
258	AβC7-258	XIII	C	E	F	V	S	W	L	45	0.011
259	AβC7-259	II	C	W	V	V	G	S	I	44	0.011
260	AβC7-260	III	C	G	P	I	V	H	L	44	0.011
261	AβC7-261	II	C	I	I	V	P	S	L	44	0.011
262	AβC7-262	I	C	R	V	W	C	A	F	44	0.011
263	AβC7-263	III	C	G	F	P	I	L	L	44	0.011
264	AβC7-264	III	C	G	Q	V	V	F	L	44	0.011
265	AβC7-265	-	C	G	Y	L	L	S	S	44	0.011
266	AβC7-266	IV	C	E	M	V	Y	V	R	44	0.011
267	AβC7-267	-	C	W	R	G	R	P	E	43	0.011
268	AβC7-268	-	C	N	G	G	L	E	R	43	0.011
269	AβC7-269	III	C	G	D	P	V	V	L	43	0.011
270	AβC7-270	III	C	G	R	S	L	V	V	43	0.011
271	AβC7-271	-	C	C	R	G	R	A	P	43	0.011
272	AβC7-272	-	C	V	T	L	G	G	V	42	0.01
273	AβC7-273	XIII	C	R	L	V	S	E	I	42	0.01
274	AβC7-274	-	C	N	W	N	Q	W	W	42	0.01
275	AβC7-275	IV	C	R	V	V	L	Y	R	41	0.01
276	AβC7-276	II	C	W	M	V	G	S	I	41	0.01
277	AβC7-277	IV	C	G	I	V	V	V	R	40	0.01
278	AβC7-278	IV	C	T	V	V	F	F	R	40	0.01
279	AβC7-279	I	S	A	V	R	C	V	W	39	0.01
280	AβC7-280	III	C	G	Q	K	V	Y	L	39	0.01
281	AβC7-281	XIII	C	G	W	V	S	E	V	39	0.01
282	AβC7-282	IV	C	K	E	V	V	W	R	39	0.01
283	AβC7-283	IV	C	R	V	V	V	W	R	39	0.01
284	AβC7-284	IV	C	A	V	V	L	R	L	39	0.01

Number	Peptide name	Cluster	Amino acid sequence							Number of reads	Reads / Total peptide reads (%)
285	AβC7-285	XI	C	R	I	P	G	V	I	38	0.009
286	AβC7-286	-	C	G	P	Y	Y	W	L	38	0.009
287	AβC7-287	I	C	R	V	L	C	L	G	38	0.009
288	AβC7-288	IV	C	R	S	V	V	W	R	38	0.009
289	AβC7-289	-	S	G	D	G	S	N	A	37	0.009
290	AβC7-290	-	C	S	R	T	I	I	K	37	0.009
291	AβC7-291	XIV	C	P	W	G	L	D	L	36	0.009
292	AβC7-292	IV	C	K	V	V	L	F	R	36	0.009
293	AβC7-293	III	C	G	T	P	V	V	L	36	0.009
294	AβC7-294	VI	C	R	V	I	Y	V	Q	35	0.009
295	AβC7-295	IV	C	R	I	V	L	W	R	35	0.009
296	AβC7-296	III	C	G	A	V	M	F	L	35	0.009
297	AβC7-297	-	C	E	G	S	V	S	W	35	0.009
298	AβC7-298	X	C	R	W	A	W	C	R	35	0.009
299	AβC7-299	IV	C	A	V	V	V	W	R	35	0.009
300	AβC7-300	-	S	R	E	W	L	D	V	35	0.009
301	AβC7-301	-	C	L	N	S	F	E	I	35	0.009
302	AβC7-302	II	C	Y	W	V	P	S	L	34	0.008
303	AβC7-303	I	C	K	A	W	Q	S	L	34	0.008
304	AβC7-304	-	C	V	W	R	R	L	W	34	0.008
305	AβC7-305	III	C	G	K	H	L	Y	L	34	0.008
306	AβC7-306	I	C	R	V	W	Q	W	V	34	0.008
307	AβC7-307	-	C	L	E	S	R	G	G	33	0.008
308	AβC7-308	XIX	C	T	V	I	L	V	R	33	0.008
309	AβC7-309	III	C	E	A	T	V	N	L	33	0.008
310	AβC7-310	XVIII	C	R	W	C	S	W	V	33	0.008
311	AβC7-311	XIII	C	R	P	V	S	E	V	33	0.008
312	AβC7-312	II	C	E	F	V	P	T	L	33	0.008
313	AβC7-313	-	S	T	A	G	R	G	T	32	0.008
314	AβC7-314	XVII	C	L	R	G	L	N	L	32	0.008
315	AβC7-315	-	C	L	V	V	H	A	T	32	0.008
316	AβC7-316	-	C	V	I	W	V	P	G	31	0.008
317	AβC7-317	III	C	R	A	E	L	W	L	31	0.008
318	AβC7-318	III	C	G	V	E	L	W	L	31	0.008
319	AβC7-319	III	C	G	A	P	L	L	L	30	0.007
320	AβC7-320	-	C	T	V	R	L	W	G	30	0.007
321	AβC7-321	IV	C	T	I	V	V	C	R	30	0.007
322	AβC7-322	III	C	G	R	I	I	D	L	30	0.007
323	AβC7-323	IV	C	N	L	I	V	V	R	30	0.007
324	AβC7-324	III	C	G	E	V	V	V	L	30	0.007
325	AβC7-325	IV	C	E	V	V	V	V	R	30	0.007
326	AβC7-326	III	C	G	L	E	I	V	L	29	0.007

Number	Peptide name	Cluster	Amino acid sequence							Number of reads	Reads / Total peptide reads (%)
327	AβC7-327	-	C	Q	V	G	K	G	G	29	0.007
328	AβC7-328	-	C	D	E	C	G	L	V	29	0.007
329	AβC7-329	I	C	F	V	W	E	A	L	29	0.007
330	AβC7-330	-	C	R	R	S	T	G	R	29	0.007
331	AβC7-331	I	C	R	V	W	S	E	V	28	0.007
332	AβC7-332	-	C	M	D	V	Y	A	V	28	0.007
333	AβC7-333	X	C	S	W	V	W	C	R	28	0.007
334	AβC7-334	-	C	E	S	A	I	I	V	28	0.007
335	AβC7-335	II	C	M	F	V	P	T	L	28	0.007
336	AβC7-336	-	S	E	E	N	G	Q	A	27	0.007
337	AβC7-337	-	C	V	M	H	K	C	G	27	0.007
338	AβC7-338	XIII	C	G	W	T	T	E	V	27	0.007
339	AβC7-339	-	S	I	R	R	T	P	V	27	0.007
340	AβC7-340	I	C	R	V	R	Q	D	L	27	0.007
341	AβC7-341	-	C	V	A	Q	S	R	N	27	0.007
342	AβC7-342	III	C	G	Q	T	L	V	L	27	0.007
343	AβC7-343	-	C	S	V	I	V	R	V	27	0.007
344	AβC7-344	I	C	A	V	W	M	M	L	27	0.007
345	AβC7-345	IV	C	R	M	A	V	W	R	27	0.007
346	AβC7-346	-	C	P	V	V	W	W	S	26	0.006
347	AβC7-347	IV	C	R	I	V	V	F	R	26	0.006
348	AβC7-348	-	C	R	E	R	G	S	W	26	0.006
349	AβC7-349	I	C	R	V	W	Q	H	V	26	0.006
350	AβC7-350	III	C	G	A	P	L	Y	V	26	0.006
351	AβC7-351	II	C	W	L	V	G	T	I	26	0.006
352	AβC7-352	-	C	T	A	L	I	W	R	26	0.006
353	AβC7-353	III	C	G	V	P	I	V	L	26	0.006
354	AβC7-354	I	C	Y	V	W	Q	S	R	26	0.006
355	AβC7-355	I	C	R	V	W	S	M	I	26	0.006
356	AβC7-356	I	C	S	V	W	C	P	L	26	0.006
357	AβC7-357	IV	C	S	L	V	V	V	R	26	0.006
358	AβC7-358	-	C	F	D	R	W	A	F	25	0.006
359	AβC7-359	VIII	C	R	L	D	S	T	W	25	0.006
360	AβC7-360	-	C	G	W	R	L	W	Q	25	0.006
361	AβC7-361	-	C	V	A	G	V	R	L	25	0.006
362	AβC7-362	-	C	L	A	W	V	G	W	25	0.006
363	AβC7-363	-	C	R	M	I	I	L	Q	25	0.006
364	AβC7-364	I	C	R	V	Y	M	E	L	25	0.006
365	AβC7-365	IV	C	A	V	V	L	I	R	24	0.006
366	AβC7-366	I	C	R	V	W	S	W	R	24	0.006
367	AβC7-367	VI	C	R	L	M	V	L	Q	24	0.006
368	AβC7-368	IV	C	D	V	V	V	I	R	24	0.006

Number	Peptide name	Cluster	Amino acid sequence							Number of reads	Reads / Total peptide reads (%)
369	AβC7-369	III	C	G	R	V	I	N	L	24	0.006
370	AβC7-370	II	C	T	V	V	P	S	V	24	0.006
371	AβC7-371	V	C	L	V	R	S	Y	L	24	0.006
372	AβC7-372	VII	C	S	L	W	V	A	I	24	0.006
373	AβC7-373	IV	C	G	L	V	I	F	R	23	0.006
374	AβC7-374	XV	C	Y	R	I	A	G	L	23	0.006
375	AβC7-375	III	C	G	A	T	V	N	L	23	0.006
376	AβC7-376	-	S	T	E	S	V	W	A	23	0.006
377	AβC7-377	-	C	S	E	P	P	R	A	23	0.006
378	AβC7-378	I	C	L	V	R	Q	E	L	23	0.006
379	AβC7-379	-	C	R	F	R	A	W	V	23	0.006
380	AβC7-380	-	S	A	V	G	T	V	R	23	0.006
381	AβC7-381	-	C	G	S	R	S	G	V	23	0.006
382	AβC7-382	-	C	D	P	E	A	V	S	23	0.006
383	AβC7-383	III	C	E	R	G	I	S	L	23	0.006
384	AβC7-384	III	C	G	K	V	L	Y	V	22	0.005
385	AβC7-385	-	C	V	R	A	N	R	A	22	0.005
386	AβC7-386	-	C	V	Q	T	A	A	V	22	0.005
387	AβC7-387	-	S	G	G	V	S	V	A	22	0.005
388	AβC7-388	XIII	C	R	F	V	S	E	V	22	0.005
389	AβC7-389	-	C	L	C	V	Q	D	M	22	0.005
390	AβC7-390	-	S	C	G	R	A	G	V	22	0.005
391	AβC7-391	XX	C	S	W	D	S	W	G	22	0.005
392	AβC7-392	III	C	G	P	T	V	V	L	22	0.005
393	AβC7-393	-	C	M	L	L	L	P	V	21	0.005
394	AβC7-394	XIX	C	T	L	I	L	L	R	21	0.005
395	AβC7-395	-	C	R	A	W	T	V	G	21	0.005
396	AβC7-396	I	C	K	V	W	R	L	L	21	0.005
397	AβC7-397	V	C	R	Y	R	T	C	L	21	0.005
398	AβC7-398	I	C	W	V	W	Q	S	L	21	0.005
399	AβC7-399	I	C	F	V	W	T	T	L	21	0.005
400	AβC7-400	-	C	H	A	T	R	S	W	21	0.005
401	AβC7-401	I	C	V	V	W	H	V	V	21	0.005
402	AβC7-402	-	S	P	V	C	F	A	G	21	0.005
403	AβC7-403	-	C	R	D	A	I	A	T	21	0.005
404	AβC7-404	III	C	G	V	M	V	M	L	20	0.005
405	AβC7-405	II	C	R	V	V	C	S	I	20	0.005
406	AβC7-406	IV	C	A	V	V	V	V	R	20	0.005
407	AβC7-407	I	C	V	P	G	A	V	R	20	0.005
408	AβC7-408	IV	C	C	K	V	V	L	V	20	0.005
409	AβC7-409	I	C	R	V	W	R	T	L	20	0.005
410	AβC7-410	XX	S	Q	G	L	S	W	D	20	0.005

Number	Peptide name	Cluster	Amino acid sequence								Number of reads	Reads / Total peptide reads (%)
411	AβC7-411	-	C	L	A	I	G	Q	I	20	0.005	
412	AβC7-412	-	C	L	R	V	L	V	G	20	0.005	
413	AβC7-413	I	C	V	V	W	T	T	R	20	0.005	
414	AβC7-414	-	S	R	W	G	R	S	R	20	0.005	
415	AβC7-415	-	C	S	G	M	T	P	R	20	0.005	
416	AβC7-416	-	C	H	C	G	A	L	R	20	0.005	
Sum										404,280	100	

Appendix C – Sequences and frequency of appearance of the A β -targeting Cluster I heptapeptide sequences, as determined by high-throughput sequencing of the enriched library after the 7th round of sorting.

Number	Peptide name	Amino acid sequence							Number of reads	Reads/Total Cluster I reads (%)	Reads/Total heptapeptide reads (%)
1	A β C7-1	C	K	V	W	Q	L	L	64118	21.144	15.86
2	A β C7-2	C	K	V	W	M	P	L	42464	14.003	10.504
3	A β C7-3	C	R	V	W	T	E	L	28675	9.456	7.093
4	A β C7-5	C	R	V	W	M	V	P	20170	6.651	4.989
5	A β C7-6	C	R	V	W	C	A	L	18397	6.067	4.551
6	A β C7-7	C	R	V	W	Q	T	V	14267	4.705	3.529
7	A β C7-8	C	S	V	W	M	E	L	12638	4.168	3.126
8	A β C7-9	C	R	V	W	Q	A	L	11718	3.864	2.898
9	A β C7-10	C	K	V	W	Q	V	L	9249	3.05	2.288
10	A β C7-11	C	R	V	W	S	L	L	9124	3.009	2.257
11	A β C7-12	C	K	V	W	M	A	L	7513	2.478	1.858
12	A β C7-13	C	K	V	W	S	Q	L	7166	2.363	1.773
13	A β C7-15	C	R	V	W	Q	L	L	6030	1.988	1.492
14	A β C7-19	C	R	V	W	C	E	L	5223	1.722	1.292
15	A β C7-20	C	K	V	W	M	E	V	5043	1.663	1.247
16	A β C7-22	C	R	V	W	S	P	L	4843	1.597	1.198
17	A β C7-23	C	R	V	W	M	G	L	4212	1.389	1.042
18	A β C7-24	C	T	V	W	M	A	I	3654	1.205	0.904
19	A β C7-26	C	V	V	W	Q	P	L	2611	0.861	0.646
20	A β C7-27	C	R	V	W	Q	V	V	2595	0.856	0.642
21	A β C7-29	C	R	V	W	S	A	L	2327	0.767	0.576
22	A β C7-30	C	A	V	W	Q	A	L	2280	0.752	0.564
23	A β C7-32	C	R	V	W	C	A	V	1971	0.65	0.488
24	A β C7-40	C	K	V	W	C	V	M	1059	0.349	0.262
25	A β C7-41	C	R	V	W	Q	C	V	1028	0.339	0.254
26	A β C7-43	C	A	V	W	M	Q	L	953	0.314	0.236
27	A β C7-49	C	R	V	W	M	M	L	815	0.269	0.202
28	A β C7-50	C	R	V	W	S	V	V	803	0.265	0.199
29	A β C7-51	C	R	V	W	Q	E	V	796	0.262	0.197
30	A β C7-52	C	V	V	W	Q	Q	I	790	0.261	0.195
31	A β C7-58	C	Q	V	W	M	D	L	608	0.2	0.15
32	A β C7-60	C	R	V	W	Q	D	P	575	0.19	0.142
33	A β C7-62	C	R	V	W	M	L	L	486	0.16	0.12
34	A β C7-63	C	V	V	W	Q	L	L	473	0.156	0.117
35	A β C7-67	C	R	V	W	Y	G	I	435	0.143	0.108
36	A β C7-70	C	A	V	W	Q	L	L	376	0.124	0.093
37	A β C7-72	C	N	V	W	Q	V	V	361	0.119	0.089
38	A β C7-75	C	K	V	W	T	V	V	335	0.11	0.083

Number	Peptide name	Amino acid sequence							Number of reads	Reads/Total Cluster I reads (%)	Reads/Total heptapeptide reads (%)
39	AβC7-76	C	V	V	W	C	Q	P	326	0.108	0.081
40	AβC7-77	C	R	V	W	M	A	A	318	0.105	0.079
41	AβC7-78	C	R	V	W	M	T	V	316	0.104	0.078
42	AβC7-80	C	R	V	W	Q	T	R	296	0.098	0.073
43	AβC7-83	C	H	V	W	S	V	V	287	0.095	0.071
44	AβC7-85	C	Y	V	W	C	P	L	273	0.09	0.068
45	AβC7-90	C	R	V	W	E	L	L	227	0.075	0.056
46	AβC7-91	C	R	V	R	Q	E	L	220	0.073	0.054
47	AβC7-93	C	R	V	W	C	F	C	218	0.072	0.054
48	AβC7-94	C	A	V	W	C	W	P	212	0.07	0.052
49	AβC7-95	C	K	V	W	M	W	K	211	0.07	0.052
50	AβC7-97	C	R	V	W	S	V	P	207	0.068	0.051
51	AβC7-102	C	R	V	W	T	Q	V	191	0.063	0.047
52	AβC7-108	C	K	V	F	Q	V	L	173	0.057	0.043
53	AβC7-112	C	R	V	A	A	V	L	165	0.054	0.041
54	AβC7-113	C	R	V	W	C	L	P	165	0.054	0.041
55	AβC7-115	C	R	V	W	S	V	R	163	0.054	0.04
56	AβC7-118	C	V	V	W	C	T	R	161	0.053	0.04
57	AβC7-126	C	L	V	W	M	G	L	153	0.05	0.038
58	AβC7-134	C	S	V	W	Q	S	L	137	0.045	0.034
59	AβC7-137	C	R	V	W	S	P	C	134	0.044	0.033
60	AβC7-141	C	F	V	W	Q	C	R	125	0.041	0.031
61	AβC7-145	C	R	V	W	C	I	I	122	0.04	0.03
62	AβC7-146	C	R	V	W	V	P	L	121	0.04	0.03
63	AβC7-155	C	R	V	S	Q	A	L	107	0.035	0.026
64	AβC7-164	C	R	V	W	S	H	P	98	0.032	0.024
65	AβC7-166	C	R	W	W	G	G	I	97	0.032	0.024
66	AβC7-174	C	I	V	W	Q	C	L	88	0.029	0.022
67	AβC7-177	C	R	V	W	A	L	L	85	0.028	0.021
68	AβC7-179	C	R	V	S	C	P	L	84	0.028	0.021
69	AβC7-181	C	V	V	W	Q	G	L	84	0.028	0.021
70	AβC7-192	C	V	V	W	S	P	L	77	0.025	0.019
71	AβC7-204	C	R	V	W	G	A	T	71	0.023	0.018
72	AβC7-206	C	V	V	W	C	A	P	70	0.023	0.017
73	AβC7-209	C	V	V	W	S	S	L	69	0.023	0.017
74	AβC7-214	C	L	V	W	C	P	L	66	0.022	0.016
75	AβC7-220	C	R	V	W	S	G	L	60	0.02	0.015
76	AβC7-227	C	R	V	A	Q	A	L	56	0.018	0.014
77	AβC7-230	C	R	V	S	A	A	L	54	0.018	0.013
78	AβC7-231	C	K	V	W	C	G	L	54	0.018	0.013
79	AβC7-232	C	V	V	T	P	V	V	54	0.018	0.013
80	AβC7-234	C	R	V	T	A	A	L	54	0.018	0.013
81	AβC7-241	C	A	V	W	Q	A	V	51	0.017	0.013
82	AβC7-245	C	R	V	W	S	A	V	50	0.016	0.012
83	AβC7-251	C	V	V	W	S	T	I	48	0.016	0.012
84	AβC7-255	C	R	V	Y	C	V	P	46	0.015	0.011
85	AβC7-262	C	R	V	W	C	A	F	44	0.015	0.011
86	AβC7-279	S	A	V	R	C	V	W	39	0.013	0.01
87	AβC7-287	C	R	V	L	C	L	G	38	0.013	0.009
88	AβC7-303	C	K	A	W	Q	S	L	34	0.011	0.008

Number	Peptide name	Amino acid sequence							Number of reads	Reads/Total Cluster I reads (%)	Reads/Total heptapeptide reads (%)
		C	R	V	W	Q	W	V			
89	AβC7-306	C	R	V	W	Q	W	V	34	0.011	0.008
90	AβC7-329	C	F	V	W	E	A	L	29	0.01	0.007
91	AβC7-331	C	R	V	W	S	E	V	28	0.009	0.007
92	AβC7-340	C	R	V	R	Q	D	L	27	0.009	0.007
93	AβC7-344	C	A	V	W	M	M	L	27	0.009	0.007
94	AβC7-349	C	R	V	W	Q	H	V	26	0.009	0.006
95	AβC7-354	C	Y	V	W	Q	S	R	26	0.009	0.006
96	AβC7-355	C	R	V	W	S	M	I	26	0.009	0.006
97	AβC7-356	C	S	V	W	C	P	L	26	0.009	0.006
98	AβC7-364	C	R	V	Y	M	E	L	25	0.008	0.006
99	AβC7-366	C	R	V	W	S	W	R	24	0.008	0.006
100	AβC7-378	C	L	V	R	Q	E	L	23	0.008	0.006
101	AβC7-396	C	K	V	W	R	L	L	21	0.007	0.005
102	AβC7-398	C	W	V	W	Q	S	L	21	0.007	0.005
103	AβC7-399	C	F	V	W	T	T	L	21	0.007	0.005
104	AβC7-401	C	V	V	W	H	V	V	21	0.007	0.005
105	AβC7-407	C	V	P	G	A	V	R	20	0.007	0.005
106	AβC7-409	C	R	V	W	R	T	L	20	0.007	0.005
107	AβC7-416	C	V	V	W	T	T	R	20	0.007	0.005
Sum									303,245	100	75.01

Appendix D – Sequences and frequency of appearance of the A β -targeting Cluster II heptapeptide sequences as determined by high-throughput sequencing of the enriched library after the 7th round of sorting.

Number	Peptide name	Aminoacid sequence							Number of reads	Reads/Total Cluster II reads (%)	Reads/Total heptapeptide reads (%)
		C	R	I	V	P	S	L			
1	A β C7-14	C	R	I	V	P	S	L	6435	32.682	1.592
2	A β C7-16	C	I	V	V	P	S	I	5871	29.817	1.452
3	A β C7-28	C	R	V	V	P	A	I	2408	12.23	0.596
4	A β C7-31	C	Q	V	V	P	S	V	2004	10.178	0.496
5	A β C7-54	C	V	V	V	P	S	I	655	3.327	0.162
6	A β C7-59	C	V	V	V	P	S	L	582	2.956	0.144
7	A β C7-81	C	L	V	V	P	S	V	291	1.478	0.072
8	A β C7-89	C	D	I	I	P	S	L	228	1.158	0.056
9	A β C7-101	C	Y	Y	V	P	S	L	198	1.006	0.049
10	A β C7-106	C	T	Y	V	P	S	L	179	0.909	0.044
11	A β C7-114	C	R	V	V	P	S	L	164	0.833	0.041
12	A β C7-160	C	K	V	V	P	T	L	101	0.513	0.025
13	A β C7-173	C	E	V	V	P	S	L	91	0.462	0.023
14	A β C7-203	C	W	T	V	G	T	I	71	0.361	0.018
15	A β C7-219	C	T	Y	V	P	S	I	63	0.32	0.016
16	A β C7-228	C	R	V	V	P	P	I	55	0.279	0.014
17	A β C7-259	C	W	V	V	G	S	I	44	0.223	0.011
18	A β C7-261	C	I	I	V	P	S	L	44	0.223	0.011
19	A β C7-276	C	W	M	V	G	S	I	41	0.208	0.01
20	A β C7-302	C	Y	W	V	P	S	L	34	0.173	0.008
21	A β C7-312	C	E	F	V	P	T	L	33	0.168	0.008
22	A β C7-335	C	M	F	V	P	T	L	28	0.142	0.007
23	A β C7-351	C	W	L	V	G	T	I	26	0.132	0.006
24	A β C7-370	C	T	V	V	P	S	V	24	0.122	0.006
25	A β C7-405	C	R	V	V	C	S	I	20	0.102	0.005
		Sum							19,690	100	4.87

Appendix E – Plasmids used herein

Table A. 1. Plasmids that encode SICLOPPS libraries and individual members

Plasmid	Encoded Protein	Marker	Origin of replication	Source
pSICLOPPS	I _C -SGGYLPPL-I _N -CBD	Cm ^R	ACYC	Prof. S. Benkovic
pSICLOPPSKanR	I _C -KanR	Cm ^R	ACYC	This work
pSICLOPPS-CysX ₁ X ₂ X ₃ sub-library	I _C -CysX ₁ X ₂ X ₃ -I _N -CBD sub-library	Cm ^R	ACYC	This work
pSICLOPPS-SerX ₁ X ₂ X ₃ sub-library	I _C -SerX ₁ X ₂ X ₃ -I _N -CBD sub-library	Cm ^R	ACYC	This work
pSICLOPPS-ThrX ₁ X ₂ X ₃ sub-library	I _C -ThrX ₁ X ₂ X ₃ -I _N -CBD sub-library	Cm ^R	ACYC	This work
pSICLOPPS-CysX ₁ X ₂ X ₃ X ₄ sub-library	I _C -CysX ₁ X ₂ X ₃ X ₄ -I _N -CBD sub-library	Cm ^R	ACYC	This work
pSICLOPPS-SerX ₁ X ₂ X ₃ X ₄ sub-library	I _C -SerX ₁ X ₂ X ₃ X ₄ -I _N -CBD sub-library	Cm ^R	ACYC	This work
pSICLOPPS-ThrX ₁ X ₂ X ₃ X ₄ sub-library	I _C -ThrX ₁ X ₂ X ₃ X ₄ -I _N -CBD sub-library	Cm ^R	ACYC	This work
pSICLOPPS-CysX ₁ X ₂ X ₃ X ₄ X ₅ sub-library	I _C -CysX ₁ X ₂ X ₃ X ₄ X ₅ -I _N -CBD sub-library	Cm ^R	ACYC	This work
pSICLOPPS-SerX ₁ X ₂ X ₃ X ₄ X ₅ sub-library	I _C -SerX ₁ X ₂ X ₃ X ₄ X ₅ -I _N -CBD sub-library	Cm ^R	ACYC	This work
pSICLOPPS-ThrX ₁ X ₂ X ₃ X ₄ X ₅ sub-library	I _C -ThrX ₁ X ₂ X ₃ X ₄ X ₅ -I _N -CBD sub-library	Cm ^R	ACYC	This work
pSICLOPPS-CysX ₁ X ₂ X ₃ X ₄ X ₅ X ₆ sub-library	I _C -CysX ₁ X ₂ X ₃ X ₄ X ₅ X ₆ -I _N -CBD sub-library	Cm ^R	ACYC	This work
pSICLOPPS-SerX ₁ X ₂ X ₃ X ₄ X ₅ X ₆ sub-library	I _C -SerX ₁ X ₂ X ₃ X ₄ X ₅ X ₆ -I _N -CBD sub-library	Cm ^R	ACYC	This work
pSICLOPPS-ThrX ₁ X ₂ X ₃ X ₄ X ₅ X ₆ sub-library	I _C -ThrX ₁ X ₂ X ₃ X ₄ X ₅ X ₆ -I _N -CBD sub-library	Cm ^R	ACYC	This work
pSICLOPPS-Random1	I _C -unknown peptide sequence1-I _N -CBD	Cm ^R	ACYC	This work
pSICLOPPS-Random2	I _C -unknown peptide sequence2-I _N -CBD	Cm ^R	ACYC	This work
pSICLOPPS(H24L/F26A)KanR	I _C (H24L/F26A)KanR	Cm ^R	ACYC	This work
pSICLOPPS(H24L/F26A)-Random1	I _C (H24L/F26A)-unknown peptide sequence1-I _N -CBD	Cm ^R	ACYC	This work
pSICLOPPS(H24L/F26A)-Random2	I _C (H24L/F26A)- unknown peptide sequence2-I _N -CBD	Cm ^R	ACYC	This work
pSICLOPPS-AβC7-1	I _C -CKVWQLL-I _N -CBD	Cm ^R	ACYC	This work
pSICLOPPS-AβC7-2	I _C -CKVWMPL-I _N -CBD	Cm ^R	ACYC	This work
pSICLOPPS-AβC7-3	I _C -CKVWTEL-I _N -CBD	Cm ^R	ACYC	This work
pSICLOPPS-AβC7-7	I _C -CRVWQTV-I _N -CBD	Cm ^R	ACYC	This work
pSICLOPPS-AβC7-14	I _C -CRIVPSL-I _N -CBD	Cm ^R	ACYC	This work
pSICLOPPS(H24L/F26A)-AβC7-1	I _C (H24L/F26A)-CKVWQLL-I _N -CBD	Cm ^R	ACYC	This work
pSICLOPPS(H24L/F26A)-AβC7-2	I _C (H24L/F26A)-CKVWMPL-I _N -CBD	Cm ^R	ACYC	This work
pSICLOPPS(H24L/F26A)-AβC7-3	I _C (H24L/F26A)-CKVWTEL-I _N -CBD	Cm ^R	ACYC	This work

Plasmid	Encoded Protein	Marker	Origin of replication	Source
pSICLOPPS(H24L/F26A)-AβC7-7	I _C (H24L/F26A)-CRVWQTV-I _N -CBD	Cm ^R	ACYC	This work
pSICLOPPS(H24L/F26A)-AβC7-14	I _C (H24L/F26A)-CRIVPSL-I _N -CBD	Cm ^R	ACYC	This work
pSICLOPPS-AβC7-371	I _C -CLVRSYL-I _N -CBD	Cm ^R	ACYC	This work
pSICLOPPS-AβC7-405	I _C -CRVVCSI-I _N -CBD	Cm ^R	ACYC	This work
pSICLOPPS-AβC7-416	I _C -CVVWTTR-I _N -CBD	Cm ^R	ACYC	This work
pSICLOPPS-AβC7-1(C1S)	I _C -SKVWQLL-I _N -CBD	Cm ^R	ACYC	This work
pSICLOPPS-AβC7-1(C1T)	I _C -TKVWQLL-I _N -CBD	Cm ^R	ACYC	This work
pSICLOPPS-AβC7-1(K2A)	I _C -CAVWQLL-I _N -CBD	Cm ^R	ACYC	This work
pSICLOPPS-AβC7-1(V3A)	I _C -CKAWQLL-I _N -CBD	Cm ^R	ACYC	This work
pSICLOPPS-AβC7-1(W4A)	I _C -CKVAQLL-I _N -CBD	Cm ^R	ACYC	This work
pSICLOPPS-AβC7-1(Q5A)	I _C -CKVWALL-I _N -CBD	Cm ^R	ACYC	This work
pSICLOPPS-AβC7-1(L6A)	I _C -CKVWQAL-I _N -CBD	Cm ^R	ACYC	This work
pSICLOPPS-AβC7-1(L7A)	I _C -CKVWQLA-I _N -CBD	Cm ^R	ACYC	This work
pSICLOPPS-AβC7-14(C1S)	I _C -SRIVPSL-I _N -CBD	Cm ^R	ACYC	This work
pSICLOPPS-AβC7-14(C1T)	I _C -TRIVPSL-I _N -CBD	Cm ^R	ACYC	This work
pSICLOPPS-AβC7-14(R2A)	I _C -CAIVPSL-I _N -CBD	Cm ^R	ACYC	This work
pSICLOPPS-AβC7-14(I3A)	I _C -CRAVPSL-I _N -CBD	Cm ^R	ACYC	This work
pSICLOPPS-AβC7-14(V4A)	I _C -CRIAPSL-I _N -CBD	Cm ^R	ACYC	This work
pSICLOPPS-AβC7-14(P5A)	I _C -CRIVASL-I _N -CBD	Cm ^R	ACYC	This work
pSICLOPPS-AβC7-14(S6A)	I _C -CRIVPAL-I _N -CBD	Cm ^R	ACYC	This work
pSICLOPPS-AβC7-14(L7A)	I _C -CRIVPSA-I _N -CBD	Cm ^R	ACYC	This work
pSICLOPPS-p53C4-4	I _C -CFSS-I _N -CBD	Cm ^R	ACYC	This work
pSICLOPPS-p53C4-16	I _C -CLEQ-I _N -CBD	Cm ^R	ACYC	This work
pSICLOPPS-p53C4-19	I _C -CLRR-I _N -CBD	Cm ^R	ACYC	This work
pSICLOPPS-p53C4-21	I _C -CAAQ-I _N -CBD	Cm ^R	ACYC	This work
pSICLOPPS-p53C5-18	I _C -TRGGC-I _N -CBD	Cm ^R	ACYC	This work
pSICLOPPS-p53C7-10	I _C -SKRSGMQ-I _N -CBD	Cm ^R	ACYC	This work
pSICLOPPS(H24L/F26A)-p53C4-4	I _C (H24L/F26A)-CFSS-I _N -CBD	Cm ^R	ACYC	This work
pSICLOPPS(H24L/F26A)-p53C4-16	I _C (H24L/F26A)-CLEQ-I _N -CBD	Cm ^R	ACYC	This work
pSICLOPPS(H24L/F26A)-p53C4-19	I _C (H24L/F26A)-CLRR-I _N -CBD	Cm ^R	ACYC	This work
pSICLOPPS(H24L/F26A)-p53C4-21	I _C (H24L/F26A)-CAAQ-I _N -CBD	Cm ^R	ACYC	This work
pSICLOPPS(H24L/F26A)-p53C5-18	I _C (H24L/F26A)-TRGGC-I _N -CBD	Cm ^R	ACYC	This work
pSICLOPPS(H24L/F26A)-p53C7-10	I _C (H24L/F26A)-SKRSGMQ-I _N -CBD	Cm ^R	ACYC	This work
pSICLOPPS-His ₆ -Random1	His ₆ -I _C -unknown peptide sequence1-I _N -CBD	Cm ^R	ACYC	This work
pSICLOPPS-His ₆ -p53C4-4	His ₆ -I _C -CFSS-I _N -CBD	Cm ^R	ACYC	This work
pSICLOPPS-His ₆ -p53C4-16	His ₆ -I _C -CLEQ-I _N -CBD	Cm ^R	ACYC	This work

Plasmid	Encoded Protein	Marker	Origin of replication	Source
pSICLOPPS-His ₆ -p53C4-19	His ₆ -I _C -CLRR-I _N -CBD	Cm ^R	ACYC	This work
pSICLOPPS-His ₆ -p53C4-21	His ₆ -I _C -CAAQ-I _N -CBD	Cm ^R	ACYC	This work
pSICLOPPS-His ₆ -p53C5-18	His ₆ -I _C -TRGGC-I _N -CBD	Cm ^R	ACYC	This work
pSICLOPPS-His ₆ -p53C7-10	His ₆ -I _C -SKRSGMQ-I _N -CBD	Cm ^R	ACYC	This work
pET-Random1	I _C -unknown peptide sequence1-I _N -CBD	Kan ^R	ColE1	This work

Table A. 2. Plasmids that encode MisPs

Plasmid	Encoded Protein	Marker	Origin of replication	Source
pETA β 42-GFP	A β 42-GFP	Kan ^R	ColE1	Prof. M. H. Hecht
pET-Sac-Abeta(M1-42)	Met-A β 42 (for in vitro experiments)	Amp ^R	ColE1	Prof. S. Linse
pETA β 42	Met-A β 42 (for ThS experiments)	Kan ^R	ColE1	This work
pETp53wt-GFP	p53Cwt-GFP	Kan ^R	ColE1	This work
pETp53C(Y220C)-GFP	p53C(Y220C)-GFP	Kan ^R	ColE1	This work
pETp53C(V143A)-GFP	p53C(V143A)-GFP	Kan ^R	ColE1	This work
pETp53C(F270L)-GFP	p53C(F270L)-GFP	Kan ^R	ColE1	This work
pETT-p53C-GFP	T-p53C-GFP	Kan ^R	ColE1	This work
pETT-p53C(Y220C)-GFP	T-p53C(Y220C)-GFP	Kan ^R	ColE1	This work
pETT-p53C(V143A)-GFP	T-p53C(V143A)-GFP	Kan ^R	ColE1	This work
pETT-p53C(F270L)-GFP	T-p53C(F270L)-GFP	Kan ^R	ColE1	This work
pASKT-p53C-GFP	T-p53C-GFP-His ₆	Amp ^R	ColE1	This work
pASKT-p53C(Y220C)-GFP	T-p53C(Y220C)-GFP-His ₆	Amp ^R	ColE1	This work
pBADT-p53C-GFP	T-p53C-GFP-His ₆	Amp ^R	ACYC	This work
pBADT-p53C(Y220C)-GFP	T-p53C(Y220C)-GFP-His ₆	Amp ^R	ACYC	This work
pTrcT-p53C-GFP	T-p53C-GFP-His ₆	Amp ^R	ColE1	This work
pTrcT-p53C(Y220C)-GFP	T-p53C(Y220C)-GFP-His ₆	Amp ^R	ColE1	This work
pETT-p53C-BFP	T-p53C-BFP	Kan ^R	ColE1	This work
pETT-p53C(Y220C)-BFP	T-p53C(Y220C)-BFP	Kan ^R	ColE1	This work
pETT-p53C-RFP	T-p53C-RFP	Kan ^R	ColE1	This work
pETT-p53C(Y220C)-RFP	T-p53C(Y220C)-RFP	Kan ^R	ColE1	This work
pCDFT-p53C-BFP	T-p53C-BFP	Sm ^R	CDF	This work
pCDFT-p53C(Y220C)-BFP	T-p53C(Y220C)-BFP	Sm ^R	CDF	This work
pETSOD1wt-GFP	SOD1wt-GFP	Kan ^R	ColE1	This work
pETSOD1(A4V)-GFP	SOD1(A4V)-GFP	Kan ^R	ColE1	This work
pETSOD1(G37R)-GFP	SOD1(G37R)-GFP	Kan ^R	ColE1	This work
pETSOD1(G85R)-GFP	SOD1(G85R)-GFP	Kan ^R	ColE1	This work
pETSOD1(G93A)-GFP	SOD1(G93A)-GFP	Kan ^R	ColE1	This work
pETSOD1wt	SOD1wt	Kan ^R	ColE1	S. Panoutsou
pETSOD1(A4V)	SOD1(A4V)	Kan ^R	ColE1	S. Panoutsou
pETSOD1(G37R)	SOD1(G37R)	Kan ^R	ColE1	S. Panoutsou
pETSOD1(G85R)	SOD1(G85R)	Kan ^R	ColE1	S. Panoutsou

Plasmid	Encoded Protein	Marker	Origin of replication	Source
pETSOD1(G93A)	SOD1(G93A)	Kan ^R	ColE1	S. Panoutsou
pETHTT _{ex1} -25Q-GFP	HTT _{ex1} -25Q-GFP	Kan ^R	ColE1	This work
pETHTT _{ex1} -46Q-GFP	HTT _{ex1} -46Q-GFP	Kan ^R	ColE1	This work
pETHTT _{ex1} -97Q-GFP	HTT _{ex1} -97Q-GFP	Kan ^R	ColE1	This work
pETT-p53C-His ₆	T-p53C-His ₆	Kan ^R	ColE1	This work
pETT-p53C(Y220C)-His ₆	T-p53C(Y220C)-His ₆	Kan ^R	ColE1	This work

Appendix F – Primers used herein

Table A. 3. Primers used for the construction of pSICLOPPS libraries and individual members

Name	Short Description	Primer sequence (5'- 3')	Use
DD015	DnaE(H24L/F26A) (BglI)rev	TTTTTTGCCCATTTGGCTAGCA GAGCATTAAGGTCTTGGGGAA GACCAATAT	Reverse primer for the H24L/F26A mutagenesis of the C-terminal domain of the Ssp DnaE intein (I _C) containing a BglI site (underlined)
DD052	HisIcDnaE(NcoI) for	AAAAACCATGGTTCACCACCAC CACCACCACAAAGTTATCGGTC GTCGTTTC	Forward primer for the introduction of a His ₆ -tag at the N-terminus of the C-terminal domain of the Ssp DnaE intein (I _C) containing a NcoI site (underlined)
DD129	SKVWQLL(BglI) for	AAAAAGCCAATGGGGCGATCG CCCACAATAGCAAGGTGTGGC AGTTG	Forward primer for the construction of pSICLOPPS-AβC7-1(C1S) containing a BglI site (underlined)
DD130	TKVWQLL(BglI) for	AAAAAGCCAATGGGGCGATCG CCCACAATACCAAGGTGTGGCA GTTG	Forward primer for the construction of pSICLOPPS-AβC7-1(C1T) containing a BglI site (underlined)
DD131	CAVWQLL(BglI) for	AAAAAGCCAATGGGGCGATCG CCCACAATTGCGGGTGTGGCA GTTGTTG	Forward primer for the construction of pSICLOPPS-AβC7-1(K2A) containing a BglI site (underlined)
DD132	CKAWQLL(BglI) for	AAAAAGCCAATGGGGCGATCG CCCACAATTGCAAGGCGTGGCA GTTGTTGTGC	Forward primer for the construction of pSICLOPPS-AβC7-1(V3A) containing a BglI site (underlined)
DD133	CKVAQLL(BglI) for	AAAAAGCCAATGGGGCGATCG CCCACAATTGCAAGGTGGCGCA GTTGTTGTGCTTAAG	Forward primer for the construction of pSICLOPPS-AβC7-1(W4A) containing a BglI site (underlined)
DD134	CKVWALL(BglI) for	AAAAAGCCAATGGGGCGATCG CCCACAATTGCAAGGTGTGGC GTTGTTGTGCTTAAG	Forward primer for the construction of pSICLOPPS-AβC7-1(Q5A) containing a BglI site (underlined)
DD135	CKVWQAL(BglI) for	AAAAAGCCAATGGGGCGATCG CCCACAATTGCAAGGTGTGGCA GGCGTTGTGCTTAAGTTTTG	Forward primer for the construction of pSICLOPPS-AβC7-1(L6A) containing a BglI site (underlined)
DD136	CKVWQLA(BglI) for	AAAAAGCCAATGGGGCGATCG CCCACAATTGCAAGGTGTGGCA GTTGGCGTGCTTAAGTTTTGG	Forward primer for the construction of pSICLOPPS-AβC7-1(L7A) containing a BglI site (underlined)
DD137	SRIVPSL(BglI) for	AAAAAGCCAATGGGGCGATCG CCCACAATAGCCGCATCGTCCC CAG	Forward primer for the construction of pSICLOPPS-AβC7-14(C1S) containing a BglI site (underlined)
DD138	TRIVPSL(BglI) for	AAAAAGCCAATGGGGCGATCG CCCACAATACCCGCATCGTCCC CAG	Forward primer for the construction of pSICLOPPS-AβC7-14(C1T) containing a BglI site (underlined)
DD139	CAIVPSL(BglI) for	AAAAAGCCAATGGGGCGATCG CCCACAATTGCGCCATCGTCCC CAGCTTG	Forward primer for the construction of pSICLOPPS-AβC7-14(R2A) containing a BglI site (underlined)
DD140	CRAVPSL(BglI) for	AAAAAGCCAATGGGGCGATCG CCCACAATTGCCGCGCCGTCCC CAGCTTGTGC	Forward primer for the construction of pSICLOPPS-AβC7-14(I3A) containing a BglI site (underlined)
DD141	CRIAPSL(BglI) for	AAAAAGCCAATGGGGCGATCG CCCACAATTGCCGCATCGCCCC CAGCTTGTGCTTAAG	Forward primer for the construction of pSICLOPPS-AβC7-14(V4A) containing a BglI site (underlined)
DD142	CRIVASL(BglI) for	AAAAAGCCAATGGGGCGATCG CCCACAATTGCCGCATCGTCGC CAGCTTGTGCTTAAG	Forward primer for the construction of pSICLOPPS-AβC7-14(P5A) containing a BglI site (underlined)

Name	Short Description	Primer sequence (5'- 3')	Use
DD143	CRIVPAL(BglI) for	AAAAAGCCAATGGGGCGATCG CCCACAATTGCCGCATCGTCCC CGCCTTGTGCTTAAGTTTTG	Forward primer for the construction of pSICLOPPS-AβC7-14(L6A) containing a BglI site (underlined)
DD144	CRIVPSA(BglI) for	AAAAAGCCAATGGGGCGATCG CCCACAATTGCCGCATCGTCCC CAGCGCGTGCTTAAGTTTTGG	Forward primer for the construction of pSICLOPPS-AβC7-14(L7A) containing a BglI site (underlined)
DD175	CVVWTTR(BglI) for	AAAAAGCCAATGGGGCGATCG CCCACAATTGCGTGGTCTGGAC GACCCGGTGCTTAAGTTTTG	Forward primer for the construction of pSICLOPPS-AβC7-416 containing a BglI site (underlined)
DD176	CRVCSI(BglI) for	AAAAAGCCAATGGGGCGATCG CCCACAATTGCCGGTGGTGTG CAGCATCTGCTTAAGTTTTG	Forward primer for the construction of pSICLOPPS-AβC7-405 containing a BglI site (underlined)
DD178	CLVRSYL(BglI) for	AAAAAGCCAATGGGGCGATCG CCCACAATTGCCTGGTGAGGTC CTACCTGTGCTTAAGTTTTG	Forward primer for the construction of pSICLOPPS-AβC7-371 containing a BglI site (underlined)
DG002	Kan(HindIII) rev	TTTTTTAAGCTTTTAGAAAAAC TCATCGAGC	Reverse primer annealing to KanR containing a HindIII site (underlined)
GS032	TetraCysfor	GGAATTCGCCAATGGGGCGATC GCCCACAATTGC(NNS) ₃ TGCTTA AGTTTTGGC	Degenerate forward primer for the construction of the CysX ₁ X ₂ X ₃ sub-library containing a BglI site (underlined)
GS033	TetraSerfor	GGAATTCGCCAATGGGGCGATC GCCCACAATAGC(NNS) ₃ TGCTTA AGTTTTGGC	Degenerate forward primer for the construction of the SerX ₁ X ₂ X ₃ sub-library containing a BglI site (underlined)
GS034	TetraThrfor	GGAATTCGCCAATGGGGCGATC GCCCACAATACC(NNS) ₃ TGCTTA AGTTTTGGC	Degenerate forward primer for the construction of the ThrX ₁ X ₂ X ₃ sub-library containing a BglI site (underlined)
GS035	CBDrev	AAAAAAAAGCTTTCATTGAAGC TGCCACAAGG	Reverse primer annealing to CBD containing a HindIII site (underlined)
GS037	pARCBDseqfor	CTATAACTATGGCTGGAATG	Forward primer annealing to the pSICLOPPS backbone, before the 5'-end of the C-terminal domain of the Ssp DnaE intein.
GS043	KanBglIfor	AAAAAAGCCAATGGGGCGATGA GCCATATCAACGGGAAAC	Forward primer annealing to KanR containing a BglI site (underlined)
GS069	SspCys Zipper	AAAAAAGCCAATGGGGCGATC GCCCACAATTGC	Forward zipper primer for the construction of the Cys sub-libraries containing a BglI site (underlined)
GS070	SspSer Zipper	AAAAAAGCCAATGGGGCGATC GCCCACAATAGC	Forward zipper primer for the construction of the Ser sub-libraries containing a BglI site (underlined)
GS071	SspThr Zipper	AAAAAAGCCAATGGGGCGATC GCCCACAATACC	Forward zipper primer for the construction of the Thr sub-libraries containing a BglI site (underlined)
GS072	PentaCysfor	GGAATTCGCCAATGGGGCGATC GCCCACAATTGC(NNS) ₄ TGCTTA AGTTTTGGC	Degenerate forward primer for the construction of the CysX ₁ X ₂ X ₃ X ₄ sub-library containing a BglI site (underlined)
GS073	PentaSerfor	GGAATTCGCCAATGGGGCGATC GCCCACAATAGC(NNS) ₄ TGCTTA AGTTTTGGC	Degenerate forward primer for the construction of the SerX ₁ X ₂ X ₃ X ₄ sub-library containing a BglI site (underlined)

Name	Short Description	Primer sequence (5' - 3')	Use
GS074	PentaThrfor	GGAATTC <u>GCCCAATGGGGCGATC</u> GCCACAATACC(NNS) ₄ TGCTTA AGTTTTGGC	Degenerate forward primer for the construction of the ThrX ₁ X ₂ X ₃ X ₄ sub-library containing a BglI site (underlined)
GS075	HexaCysfor	GGAATTC <u>GCCCAATGGGGCGATC</u> GCCACAATTGC(NNS) ₅ TGCTTA AGTTTTGGC	Degenerate forward primer for the construction of the CysX ₁ X ₂ X ₃ X ₄ X ₅ sub-library containing a BglI site (underlined)
GS076	HexaSerfor	GGAATTC <u>GCCCAATGGGGCGATC</u> GCCACAATAGC(NNS) ₅ TGCTTA AGTTTTGGC	Degenerate forward primer for the construction of the SerX ₁ X ₂ X ₃ X ₄ X ₅ sub-library containing a BglI site (underlined)
GS077	HexaThrfor	GGAATTC <u>GCCCAATGGGGCGATC</u> GCCACAATACC(NNS) ₅ TGCTTA AGTTTTGGC	Degenerate forward primer for the construction of the ThrX ₁ X ₂ X ₃ X ₄ X ₅ sub-library containing a BglI site (underlined)
GS078	HeptaCysfor	GGAATTC <u>GCCCAATGGGGCGATC</u> GCCACAATTGC(NNS) ₆ NNSNNS NNSNNSNSTGCTTAAGTTTTG GC	Degenerate forward primer for the construction of the CysX ₁ X ₂ X ₃ X ₄ X ₅ X ₆ sub-library containing a BglI site (underlined)
GS079	HeptaSerfor	GGAATTC <u>GCCCAATGGGGCGATC</u> GCCACAATAGC(NNS) ₆ NNSNN SNNSNNSNSTGCTTAAGTTTTG GC	Degenerate forward primer for the construction of the SerX ₁ X ₂ X ₃ X ₄ X ₅ X ₆ sub-library containing a BglI site (underlined)
GS080	HeptaThrfor	GGAATTC <u>GCCCAATGGGGCGATC</u> GCCACAATACC(NNS) ₆ NNSNNS NNSNNSNSTGCTTAAGTTTTG GC	Degenerate forward primer for the construction of the ThrX ₁ X ₂ X ₃ X ₄ X ₅ X ₆ sub-library containing a BglI site (underlined)

Table A. 4. Primers used for the construction of plasmids encoding MisPs

Name	Short Description	Primer sequence (5' - 3')	Use
DD002	p53HisXhoIRev	AAAA <u>ACTCGAGT</u> TAGTGGTGGT GGTGGTGGTGGGTGTTGTTGGA CAGTGCTCG	Reverse primer for the construction of pETp53-His ₆ variants containing a XhoI site (underlined)
DD004	Abeta(NcoI)For	AAAA <u>ACCATGGATG</u> CGGAATTT CGCCATG	Forward primer for the construction of Met-Aβ ₄₂ containing a NcoI site (underlined)
DD006	BFPlinker(BamHI) for	AAAA <u>AGGATCCGGT</u> TCTGGTTC TATGAGCGAAGAACTGATCAA AG	Forward primer for the introduction of BFP downstream of p53 variants and into the pET28 vector containing a BamHI site (underlined)
DD007	BFP(XhoI) (HindIII) rev	AAAA <u>ACTCGAGAAGCT</u> TTTAGT TCAGTTTGTGACCCAGTTTAG	Reverse primer for the introduction of RFP downstream of p53 variants and into the pET28 vector containing a HindIII and a XhoI site (underlined)
DD008	p53(NcoI)For	AAAA <u>ACCATGGT</u> TTTCATCTTCT GTCCCTTCCCAG	Forward primer for the introduction of p53 into the pCDF-1b vector containing a NcoI site (underlined)
DD009	T-F270Lfor	CTGGGACGGGACAGCTTAGAG GTGCGTGTTTGTG	Forward point mutagenesis primer for the introduction of the F270L mutation in the sequence of T-p53C
DD010	T-F270Lrev	CACAAACACGCACCTCTAAGCT GTCCCGTCCCAG	Reverse point mutagenesis primer for the introduction of the F270L mutation in the sequence of T-p53C
DD011	V143Afor	GCCAAGACCTGCCCTGCGCAGC TGTGGGTTGATTC	Forward point mutagenesis primer for the introduction of the V143A mutation in the sequence of p53C
DD012	V143Arev	GAATCAACCCACAGCTGCGCA GGGCAGGTCTTGCC	Reverse point mutagenesis primer for the introduction of the V143A mutation in the sequence of p53C
DD070	Tp53(ΔNcoI)for	CACCCGCGTCCGCGCAATGGCC ATCTACAAG	Forward point mutagenesis primer for the elimination of the NcoI site in the sequence of T-p53C
DD071	Tp53(ΔNcoI)rev	CTTGTAGATGGCCATTGCGCGG ACGCGGGTG	Reverse point mutagenesis primer for the elimination of the NcoI site in the sequence of T-p53C
DD079	RFPlinker(BamHI) for	AAAA <u>AGGATCCGGT</u> TCTGGTTC TGCTTCTCCGAAGACGTTATC	Forward primer for the introduction of RFP downstream of p53 variants and into the pET28 vector containing a BamHI site (underlined)
DD083	RFP(XhoI)rev	TTTT <u>TCTCGAGT</u> TAAAGCACCGG TGGAGTGACGAC	Reverse primer for the introduction of RFP downstream of p53 variants and into the pET28 vector containing a XhoI site (underlined)
DD084	HD46(PstI)rev	GGGGGTGGAGGGGGAGGTGGC <u>TGCAGT</u> TGTTGCTG	Reverse primer for the construction of pETHHTT _{ex1} -46Q-GFP containing a PstI site (underlined)
GS002	EGFPprev	AAAAAA <u>AAGCTT</u> TCTCGAGTTAG TGGTGGTGGTGGTGGTGTGGT AGAGTTCATCCATGCC	Reverse primer for the construction of GFP fused protein constructs containing a HindIII site (underlined)
GS003	p53for	AAAAAAT <u>CTAGA</u> AGGAGGAAA CGCATATGTCATCTTCTGTCCCT TCCCAG	Forward primer for the construction of p53 protein constructs containing a XbaI site (underlined)

Name	Short Description	Primer sequence (5'- 3')	Use
GS004	p53rev	AAAAAAGGATCCCTGCAGGGT GTTGTTGGACAGTGCTCG	Reverse primer for the construction of the p53 protein fusion with GFP containing a BamHI site (underlined)
GS007	p53Y220Cfor	AGTGTGGTGGTGCCCTGTGAGC CGCCTGAGGTTG	Forward point mutagenesis primer for the introduction of the Y220C mutation in the sequence of p53C
GS008	p53Y220Crev	CAACCTCAGGCGGCTCACAGG GCACCACCACACT	Reverse point mutagenesis primer for the introduction of the Y220C mutation in the sequence of p53C
GS011	p53M133Lfor	CCTGCCCTCAACAAGCTGTTTT GCCAACTGGCC	Forward point mutagenesis primer for the introduction of the M133L mutation in the sequence of p53C
GS012	p53M133Lrev	GGCCAGTTGGCAAACAGCTTG TTGAGGGCAGG	Reverse point mutagenesis primer for the introduction of the M133L mutation in the sequence of p53C
GS013	p53V203Afor	GGAAATTTGCGTGC GGGAGTAT TTGGATGAC	Forward point mutagenesis primer for the introduction of the V203A mutation in the sequence of p53C
GS014	p53V203Arev	GTCATCCAAATACTCCC GCACG CAAATTTCC	Reverse point mutagenesis primer for the introduction of the V203A mutation in the sequence of p53C
GS015	p53N239Yfor	CAACTACATGTGTTACAGTTCC TGCATGGGC	Forward point mutagenesis primer for the introduction of the N239Y mutation in the sequence of p53C
GS016	p53N239Yrev	GCCCATGCAGGA ACTGTAACAC ATGTAGTTG	Reverse point mutagenesis primer for the introduction of the N239Y mutation in the sequence of p53C
GS041	F270Lfornew	CTGGGACGGAACAGCTTAGAG GTGCGTGT TTTGTG	Forward point mutagenesis primer for the introduction of the F270L mutation in the sequence of p53C
GS042a	F270Lrevnew	CACAAACACGCACCTCTAAGCT GTTCCGTC C CAG	Reverse point mutagenesis primer for the introduction of the F270L mutation in the sequence of p53C
GS042b	N268Dfornew	GTAATCTACTGGGACGGGACA GCTTTGAGGTGCGTG	Forward point mutagenesis primer for the introduction of the N268D mutation in the sequence of p53C
GS042c	N268Drevnew	CACGCACCTCAAAGCTGTCCCG TCCCAGTAGATTAC	Reverse point mutagenesis primer for the introduction of the N268D mutation in the sequence of p53C
GS058	SOD1for	AAAAAACATATGGCGACGAAG GCCGTGTGCGTG	Forward primer for the construction of pETSOD1-GFP variants containing a NdeI site (underlined)
GS059	SOD1rev	AAAAAAGGATCCACTAGTTTGG GCGATCCCAATTACACC	Reverse primer for the construction of pETSOD1-GFP variants containing a BamHI site (underlined)
GS060	SOD1(A4V) for	AAAAAACATATGGCGACGAAG GTGGTGTGCGTGCTG	Forward primer for the construction of pETSOD1(A4V)-GFP containing a NdeI site (underlined)
GS100	SODwt assembly For1	ATGGCGACGAAGGCCGTGTGC GTGCTGAAGGGCGACGGCCCA GTGCAGGGGCATCATC	Forward primer for the <i>SOD1</i> gene assembly (segment 1).
GS101	SODwt assembly Rev1	CACACCTTCACTGGTCCATTAC TTTCCTTCTGCTCGAAATTGAT GATGCCCTGCACTGGG	Reverse primer for the <i>SOD1</i> gene assembly (segment 1).

Name	Short Description	Primer sequence (5'- 3')	Use
GS102	SODwt assembly For2	GGACCAGTGAAGGTGTGGGA AGCATTAAGGACTGACTGAA GGCCTGCATGGATTCC	Forward primer for the <i>SOD1</i> gene assembly (segment 2).
GS103	SODwt assembly Rev2	CTGGTACAGCCTGCTGTATTAT CTCCAAACTCATGAACATGGAA TCCATGCAGGCC	Reverse primer for the <i>SOD1</i> gene assembly (segment 2).
GS104	SODwt assembly For3	CAGCAGGCTGTACCAGTGCAG GTCCTCACTTTAATCCTCTATCC AGAAAACACGG	Forward primer for the <i>SOD1</i> gene assembly (segment 3).
GS105	SODwt assembly Rev3	GTCTCCAACATGCCTCTCTTCA TCCTTTGGCCACCGTGTTTTCT GGATAGAGG	Reverse primer for the <i>SOD1</i> gene assembly (segment 3).
GS106	SODwt assembly For4	GAGAGGCATGTTGGAGACTTG GGCAATGTGACTGCTGACAAA GATGGTGTGGCCG	Forward primer for the <i>SOD1</i> gene assembly (segment 4).
GS107	SODwt assembly Rev4	CCTGAGAGTGAGATCACAGAA TCTTCAATAGACACATCGGCCA CACCATCTTTGTC	Reverse primer for the <i>SOD1</i> gene assembly (segment 4).
GS108	SODwt assembly For5	CTGTGATCTCACTCTCAGGAGA CCATTGCATCATTGGCCGCAC	Forward primer for the <i>SOD1</i> gene assembly (segment 5).
GS109	SODwt assembly Rev5	GCCCAAGTCATCTGCTTTTTCA TGGACCACCAGTGTGCGGCCAA TGATGC	Reverse primer for the <i>SOD1</i> gene assembly (segment 5).
GS110	SODwt assembly For6	GCAGATGACTTGGGCAAAGGT GGAAATGAAGAAAGTACAAAG ACAGGAAACGC	Forward primer for the <i>SOD1</i> gene assembly (segment 6).
GS111	SODwt assembly Rev 6	TTGGGCGATCCCAATTACACCA CAAGCCAAACGACTTCCAGCGT TTCCTGTCTTTGTAC	Reverse primer for the <i>SOD1</i> gene assembly (segment 6).
GS112	SOD(G37R)for	GTGGGGAAGCATTAAACGACT GACTGAAGGCC	Forward point mutagenesis primer for the introduction of the G37R mutation in the sequence of <i>SOD1</i>
GS113	SOD(G37R)rev	GGCCTTCAGTCAGTCGTTTAAT GCTTCCCCAC	Reverse point mutagenesis primer for the introduction of the G37R mutation in the sequence of <i>SOD1</i>
GS114	SOD(G85R)for	CATGTTGGAGACTTGCGCAATG TGACTGCTG	Forward point mutagenesis primer for the introduction of the G85R mutation in the sequence of <i>SOD1</i>
GS115	SOD(G85R)rev	CAGCAGTCACATTGCGCAAGTC TCCAACATG	Reverse point mutagenesis primer for the introduction of the G85R mutation in the sequence of <i>SOD1</i>
GS116	SOD(G93A)for	CTGCTGACAAAGATGCTGTGGC CGATGTGTC	Forward point mutagenesis primer for the introduction of the G93A mutation in the sequence of <i>SOD1</i>
GS117	SOD(G93A)rev	GACACATCGGCCACAGCATCTT TGTCAGCAG	Reverse point mutagenesis primer for the introduction of the G93A mutation in the sequence of <i>SOD1</i>
GS118	p53wt assembly For1	ATGTCATCTTCTGTCCCTTCCA GAAAACCTACCAGGGCAGCTA CGGTTTCCGTCTGGGC	Forward primer for <i>TP53</i> gene assembly (segment 1).
GS119	p53wt assembly Rev1	GGAGTACGTGCAAGTCACAGA CTTGGCTGTCCCAGAATGCAAG AAGCCAGACGGAAACC	Reverse primer for <i>TP53</i> gene assembly (segment 1).
GS120	p53wt assembly For2	GACTTGCACGTACTCCCCTGCC CTCAACAAGATGTTTTGCCAAC TGGCCAAGACC	Forward primer for <i>TP53</i> gene assembly (segment 2).

Name	Short Description	Primer sequence (5'- 3')	Use
GS121	p53wt assembly Rev2	CCGGCGGGGGTGTGGAATCA ACCCACAGCTGCACAGGGCAG GTCTTGCCAGTTGGC	Reverse primer for <i>TP53</i> gene assembly (segment 2).
GS122	p53wt assembly For3	GATTCCACACCCCGCCCGCA CCC CGTCCGCGCCATGGCCAT CTACAAGCAGTCACAG	Forward primer for <i>TP53</i> gene assembly (segment 3).
GS123	p53wt assembly Rev3	CAGCGCTCATGGTGGGGCAG CGCCTCACAACCTCCGTCATGT GCTGTGACTGCTTGTAG	Reverse primer for <i>TP53</i> gene assembly (segment 3).
GS124	p53wt assembly For4	CCCACCATGAGCGCTGCTCAGA TAGCGATGGTCTGGCCCCTCT CAGCATCTTATC	Forward primer for <i>TP53</i> gene assembly (segment 4).
GS125	p53wt assembly Rev4	CCAAATACTCCACACGCAAATT TCCTTCCACTCGGATAAGATGC TGAGGAGGG	Reverse primer for <i>TP53</i> gene assembly (segment 4).
GS126	p53wt assembly For5	GCGTGTGGAGTATTTGGATGAC AGAAACACTTTTCGACATAGTG TGGTGGTGCC	Forward primer for <i>TP53</i> gene assembly (segment 5).
GS127	p53wt assembly Rev5	GTGGTACAGTCAGAGCCAACCT CAGGCGGCTCACAGGGCACCA CCACACTATG	Reverse primer for <i>TP53</i> gene assembly (segment 5).
GS128	p53wt assembly For6	GGCTCTGACTGTACCACCATCC ACTACAACACTACATGTGTAACAG TTCCTGCATG	Forward primer for <i>TP53</i> gene assembly (segment 6).
GS129	p53wt assembly Rev6	GTGTGATGATGGTGAGGATGGG CCTCCGGTTCATGCCGCCCATG CAGGAACGTGTTAC	Reverse primer for <i>TP53</i> gene assembly (segment 6).
GS130	p53wt assembly For7	CCTCACCATCATCACACTGGAA GACTCCAGTGGTAATCTACTGG GACGGAACAGCTTTG	Forward primer for <i>TP53</i> gene assembly (segment 7).
GS131	p53wt assembly Rev7	GTGCGCCGGTCTCTCCAGGAC AGGCACAAACACGCACCTCAA AGCTGTTCCGTCCCAG	Reverse primer for <i>TP53</i> gene assembly (segment 7).
GS132	p53wt assembly For8	GAGAGACCGCGCACAGAGGA AGAGAATCTCCGCAAGAAAGG GGAGCCTCACCAG	Forward primer for <i>TP53</i> gene assembly (segment 8).
GS133	p53wt assembly Rev8	GGTGTGTTGGACAGTGCTCGC TTAGTGCTCCCTGGGGCAGCT CGTGGTGAGGCTCCCC	Reverse primer for <i>TP53</i> gene assembly (segment 8).
IM022	Aβ42(XhoI) rev	<u>CCGCTCGAGTTACGCAATCACC</u> ACGCCGCCAC	Reverse primer for the construction of Met-Aβ42 containing a XhoI site (underlined)