

DISEASES AND DISORDERS

Bacterial production and direct functional screening of expanded molecular libraries for discovering inhibitors of protein aggregation

Dafni C. Delivoria^{1,2}, Sean Chia³, Johnny Habchi³, Michele Perni³, Ilias Matis¹, Nikoleta Papaevgeniou^{1,4}, Martin Reczko⁵, Niki Chondrogianni¹, Christopher M. Dobson³, Michele Vendruscolo³, Georgios Skretas^{1*}

Protein misfolding and aggregation are associated with a many human disorders, including Alzheimer's and Parkinson's diseases. Toward increasing the effectiveness of early-stage drug discovery for these conditions, we report a bacterial platform that enables the biosynthesis of molecular libraries with expanded diversities and their direct functional screening for discovering protein aggregation inhibitors. We illustrate this approach by performing, what is to our knowledge, the largest functional screen of small-size molecular entities described to date. We generated a combinatorial library of ~200 million drug-like, cyclic peptides and rapidly screened it for aggregation inhibitors against the amyloid- β peptide (A β 42), linked to Alzheimer's disease. Through this procedure, we identified more than 400 macrocyclic compounds that efficiently reduce A β 42 aggregation and toxicity in vitro and in vivo. Finally, we applied a combination of deep sequencing and mutagenesis analyses to demonstrate how this system can rapidly determine structure-activity relationships and define consensus motifs required for bioactivity.

INTRODUCTION

The phenomenon of protein misfolding and aggregation is a defining feature of a wide range of human diseases with very high socio-economic impact, including neurodegenerative disorders, type 2 diabetes, and cancer (1). Since aggregated proteins can cause disease, either because they can no longer efficiently perform their physiological function (loss of function) or because they form harmful aggregated species with cytotoxic properties (toxic gain of function) (1), compounds that prevent, delay, or reverse protein aggregation constitute valuable leads for the development of potential therapeutics. Many such molecules are currently in preclinical and clinical development (2). As a proof of concept for the therapeutic value of this approach, tafamidis, a small molecule that prevents the misfolding and aggregation of the carrier protein transthyretin by binding and stabilizing its tetrameric native form, has been approved for the treatment of familial amyloid polyneuropathy in Europe and Japan (Vyndaqel, Pfizer) (3). More recently, migalastat, a chemical rescuer of the misfolding of the lysosomal enzyme α -galactosidase (4), has been approved for the treatment of the lysosomal storage disorder Fabry disease in Europe and the United States (Galafold, Amicus Therapeutics). Despite these encouraging results, the vast majority of protein-misfolding diseases remain incurable, as no disease-modifying drug has reached the clinic in most cases. Among the reasons for the failure of current clinical trials, we mention an incomplete understanding of the exact molecular mechanism of action of the anti-amyloid- β peptide (A β) candidates and the late treatment of the

patients (5). Thus, it is imperative to develop systematic and robust approaches to discover previously unidentified and effective disease-modifying agents, which are urgently required for this type of disorders.

Advances in key scientific and technological areas are needed to increase the success rate with which effective drugs against these complex diseases are discovered. One such area is chemical library construction. The availability of molecular libraries with expanded diversities is expected to markedly increase the chances for identifying compounds with the desired properties (6, 7). Because of current limitations in organic synthesis and the isolation of natural products, however, the diversity of currently tested small-molecule libraries is typically not higher than 10^3 to 10^6 (8). Considering that the size of the chemical space for small molecules, i.e., the number of all possible low-molecular weight structures has been estimated to be $\sim 10^{60}$ (9), it is clear that drug screening efforts will benefit from increased diversity. In addition, even when chemical libraries with larger sizes are available, the majority of screening methodologies for the identification of drug-like compounds are not sufficiently high throughput to efficiently handle very large libraries. Functional screening assays in multiwell plate format, for example, become impractical for libraries with more than 10^6 to 10^7 members.

Genetically encoded combinatorial libraries can enable a marked expansion in the number and chemical complexity of lower-molecular weight compounds, which can be generated and subsequently tested for bioactivity (7, 10, 11). By using approaches of this type, molecular libraries with diversities ranging from many millions to even tens of trillions of test compounds have already been generated (6, 7, 10, 12, 13), and molecules with valuable biological activities have been discovered. These bioactivities include modulation of the aggregation process of misfolding-prone and disease-associated proteins, such as the A β and huntingtin (11, 14–17).

One important shortcoming when investigating DNA-encoded libraries for protein misfolding and aggregation diseases, however, is that they can only be screened for binding against immobilized

Copyright © 2019
The Authors, some
rights reserved;
exclusive licensee
American Association
for the Advancement
of Science. No claim to
original U.S. Government
Works. Distributed
under a Creative
Commons Attribution
License 4.0 (CC BY).

¹Institute of Chemical Biology, National Hellenic Research Foundation, Athens 11635, Greece. ²School of Chemical Engineering, National Technical University of Athens, Athens 15780, Greece. ³Centre for Misfolding Diseases, Department of Chemistry, University of Cambridge, Cambridge CB2 1EW, UK. ⁴Faculty of Biology and Pharmacy, Institute of Nutrition, Friedrich Schiller University of Jena, Jena 07743, Germany. ⁵Institute for Fundamental Biomedical Science, Biomedical Sciences Research Center "Alexander Fleming," Athens 16672, Greece.

*Corresponding author. Email: gskretas@eie.gr

protein targets (10). Despite their efficiency in identifying strong binders, these affinity-based selections cannot readily provide functional information regarding the aggregation inhibition activity of the identified hits (10). As a result, the selected binders need to be resynthesized chemically and evaluated again for aggregation-inhibitory activity in secondary assays. This procedure adds substantial time, complexity, and cost to the overall screening process and is regarded as a major bottleneck by the pharmaceutical industry (12). Furthermore, in many cases, the outcome of the selection process results in the identification of a large fraction of hits that are either completely inactive (12) or have opposite effects on protein misfolding and aggregation than the ones intended originally (15).

In an effort to generate new and efficient systems for discovering previously unidentified inhibitors of pathogenic protein aggregation, we have recently reported the development of a synthetic biology platform that enables the discovery of chemical rescuers of disease-associated protein misfolding (18). In this system, combinatorial libraries of lower-molecular weight peptide macrocycles are biosynthesized in *Escherichia coli* cells and are simultaneously screened for their ability to correct the problematic folding of misfolding-prone, disease-associated proteins using a flow cytometric ultrahigh-throughput genetic screen.

In the present work, we demonstrate how this bacterial discovery platform can be expanded to enable the production and direct functional screening of molecular libraries with greatly increased diversities, thus considerably surpassing the capabilities of other systems reported to date. We used this system to generate a combinatorial library of ~200 million peptide macrocycles and to perform simultaneous functional screening for aggregation inhibition activity against the 42-residue form of A β (A β 42), which is associated with Alzheimer's disease. Within a matter of only a few days, our bacterial platform enabled the production and screening of the complete library and identified hundreds of hits. Analysis of the selected macrocycles revealed that they form different clusters with distinct sequence characteristics. Selected macrocycles derived from the most dominant clusters were subjected to *in vitro* biochemical and biophysical testing and were found to be highly potent inhibitors of A β 42 aggregation at substoichiometric ratios. *In vivo* testing in established models of Alzheimer's disease in the nematode *Caenorhabditis elegans* demonstrated that the selected macrocycles were effective in decreasing the deposition of A β 42 aggregates and in markedly reversing A β 42-induced pathogenic effects. We then used a combination of high-throughput sequencing and site-directed mutagenesis analyses to determine structure-activity relationships for the selected macrocycles and to define consensus motifs required for high bioactivity in these molecules. Overall, our discovery platform enables the simultaneous production and functional screening of molecular libraries with markedly expanded diversities for the identification of compounds with therapeutic potential for inhibiting the aggregation of disease-associated polypeptides.

RESULTS

Construction and characterization of a low-weight molecular library with expanded diversity

The molecular libraries that we have chosen to use for the discovery of protein aggregation inhibitors are combinatorial libraries of head-to-tail cyclic heptapeptides, with an average molecular mass of about 770 Da. These macrocycles fall within the class of small

molecules (molecular mass, <900 Da) but occupy an area of chemical space beyond the classical Lipinski's rule of five (bRo5 space; molecular mass, 500 to 1000 Da), where different rules for drug-likeness compared to classical small-molecule therapeutics apply (19, 20). The very large number of possible amino acid combinations comprising a peptide sequence (of seven amino acids in our case) enables greatly expanded levels of molecular diversity compared to available synthetic and natural small-molecule libraries (8). Furthermore, the cyclic nature of these molecules affords higher binding affinities for other proteins, enhanced ability to penetrate biological barriers, and enhanced resistance to proteolysis compared to their linear analogs (21).

Libraries of head-to-tail cyclic peptides can be conveniently produced in *E. coli* cells by the "split intein-mediated circular ligation of peptides and proteins" (SICLOPPS) method, where a circularly permuted split intein catalyzes the formation of a peptide bond between the termini of the target protein or peptide (22). SICLOPPS is a well-established technique, which has been previously used to identify cyclic peptides with different bioactivities (23). The only external requirement for the intein splicing reaction and peptide cyclization to take place is the presence of a nucleophilic amino acid (Cys, Ser, or Thr) as the first amino acid of the to-be-cyclized peptide (18). Thus, to maximize the diversity of our macrocycle library, we constructed a combinatorial heptapeptide library with the general formula cyclo-NuX₁X₂X₃X₄X₅X₆, where Nu is any one of the nucleophilic amino acids Cys, Ser, or Thr and X is any one of the 20 natural amino acids. The maximum theoretical diversity of this library is $3 \times 20^6 = 192$ million different sequences. The libraries of genes encoding these cyclic heptapeptide libraries were constructed using degenerate polymerase chain reaction (PCR) primers, in which the randomized amino acids (X) were encoded using randomized NNS codons, where N is A, T, G, or C and S is G or C (see Materials and Methods). The generated peptide-encoding gene libraries were cloned into the vector pSICLOPPS (18) to form the combined pSICLOPPS-NuX₁X₂X₃X₄X₅X₆ vector library (Fig. 1A). These vectors express a combinatorial library of tetrapartite fusion proteins comprising the following: (i) the C-terminal domain of the Ssp DnaE intein (I_C), (ii) a NuX₁X₂X₃X₄X₅X₆ heptapeptide sequence, (iii) the N-terminal domain of the Ssp DnaE intein (I_N), 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 (Fig. 1A). Cloning of the resulting gene libraries into the pSICLOPPS plasmid yielded a total of 1.2×10^9 independent transformants, as judged by plating experiments after serial dilutions.

To assess the quality of our constructed library, we initially chose 150 randomly selected clones and performed colony PCR and SDS-polyacrylamide gel electrophoresis (SDS-PAGE)/Western blot. This analysis revealed that approximately 45% of the analyzed clones contained a DNA insert of the correct size and produced full-length I_C-peptide-I_N-CBD precursor fusion protein (molecular mass, ~25 kDa), which could undergo processing (appearance of a band with a molecular mass of ~20 kDa) (Fig. 1B). This processing is a prerequisite for successful intein splicing and indicates possible formation of a cyclic product. According to these results, the generated library contains approximately 5.6×10^8 clones that apparently produce cyclic heptapeptides, a number that covers the theoretical diversity of our combined library by almost threefold.

To characterize the constructed library further, we performed deep sequencing analysis of the peptide-encoding region of the pSICLOPPS-NuX₁X₂X₃X₄X₅X₆ vector library. Of the ~3.4 million plasmid sequences

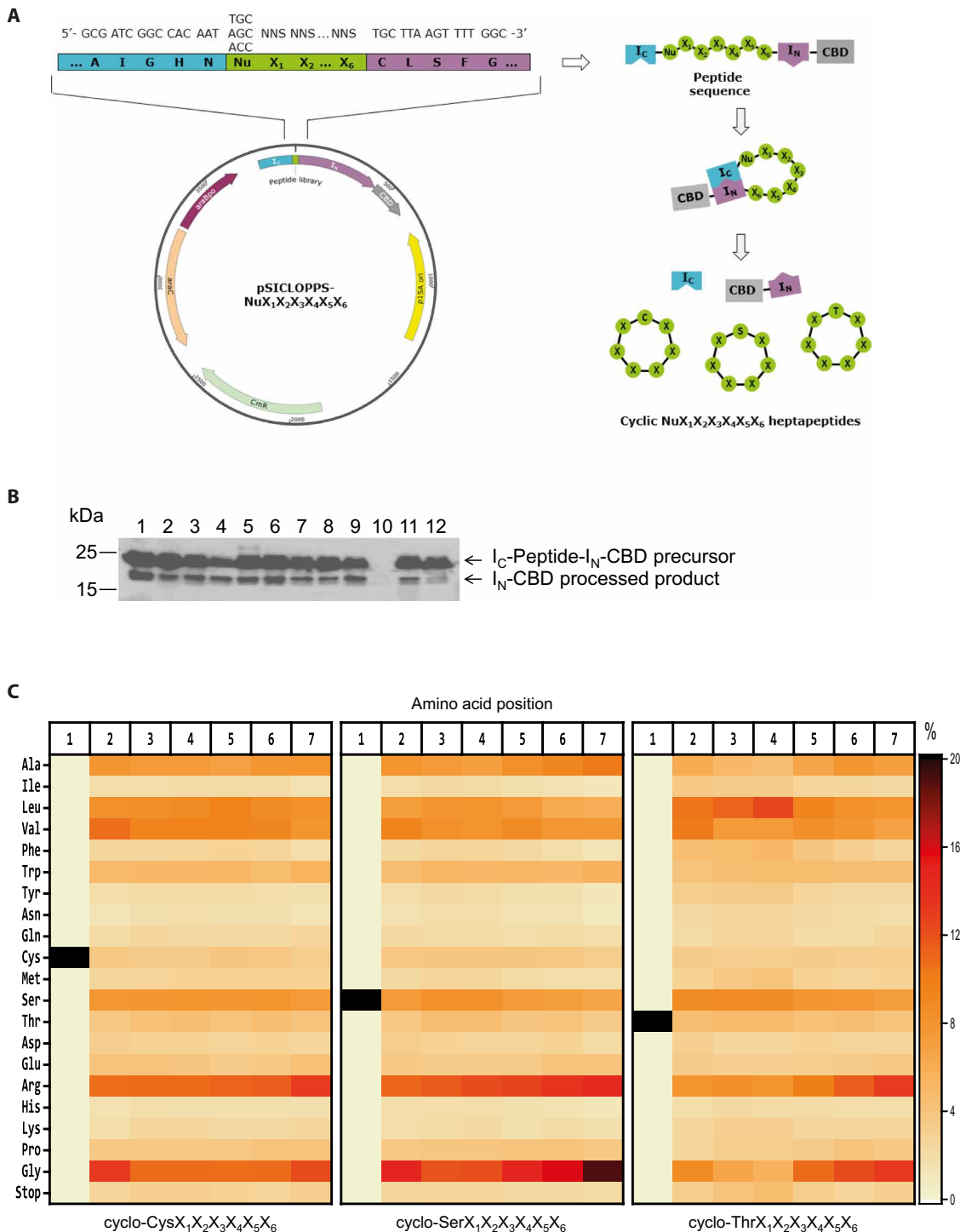


Fig. 1. Construction and characterization of a combinatorial cyclic heptapeptide library with expanded diversity. (A) Left: Representation of the pSICLOPPS-NuX₁X₂X₃X₄X₅X₆ vector library encoding the combinatorial heptapeptide library cyclo-NuX₁X₂X₃X₄X₅X₆. Nu: Cys, Ser, or Thr; X: any of the 20 natural amino acids; NNS: randomized codons, where N = A, T, C, or G and S = G or C; I_C: C-terminal domain of the Ssp DnaE split intein; I_N: N-terminal domain of the Ssp DnaE split intein. Right: Peptide cyclization using the SICLOPPS construct. Upon interaction between the two intein domains I_C and I_N, the encoded I_C-NuX₁X₂X₃X₄X₅X₆-I_N-CBD fusions undergo intein splicing and peptide cyclization, leading to the production of the cyclo-NuX₁X₂X₃X₄X₅X₆ library. (B) Western blot analysis of 12 randomly picked individual clones from the combinatorial heptapeptide library cyclo-NuX₁X₂X₃X₄X₅X₆, showing the expression and processing of the precursor fusion protein I_C-peptide-I_N-CBD. The 25-kDa band corresponds to the unprocessed precursor and the 20-kDa band to the processed I_N-CBD construct, and indicates, wherever present, successful intein splicing and peptide cyclization. Clone 10, for which the precursor is not expressed, was to contain a stop codon in its peptide-encoding region. (C) Heatmap representation of the amino acid distribution at each position of the constructed cyclo-CysX₁X₂X₃X₄X₅X₆ (left), cyclo-SerX₁X₂X₃X₄X₅X₆ (middle), and cyclo-ThrX₁X₂X₃X₄X₅X₆ (right) sublibraries, as demonstrated by the deep sequencing analysis results.

that we analyzed, ~75% were unique at the DNA level and ~95% of those were found to encode unique peptide sequences (table S1). All amino acids were found to be encoded at every position of the generated library, albeit with an overrepresentation of residues corresponding to Gly and Arg (Fig. 1C). Together, these results indicate that we have constructed a very high-diversity library encoding the vast majority, if not all, of the theoretically possible ~192 million cyclo-NuX₁X₂X₃X₄X₅X₆ heptapeptide sequences.

Ultrahigh-throughput functional screening for Aβ42 aggregation inhibitors

To perform direct functional screening of our vast library of cyclic heptapeptides and readily identify bioactive macrocyclic inhibitors of pathogenic protein aggregation, we used an ultrahigh-throughput system that we previously developed (18). Because of the high aggregation propensity of Aβ, *E. coli* cells overexpressing Aβ42–green fluorescent protein (GFP) produce a misfolded fusion that eventually accumulates into insoluble inclusion bodies lacking fluorescence (24). Conditions that inhibit Aβ aggregation, however, result in the formation of soluble and fluorescent Aβ42–GFP, and bacterial cells expressing this fusion acquire a fluorescent phenotype (18, 24). On the basis of this, production of the macrocyclic peptide libraries under investigation and their screening for misfolding-rescuing activity in this system are carried out simultaneously in *E. coli* cells in an integrated fashion, by selecting and isolating the bacterial clones biosynthesizing the molecules that enhance the fluorescence of chimeric fusions of misfolding-prone proteins with the GFP (Fig. 2A).

Electrocompetent *E. coli* Tuner (DE3) cells carrying the expression vector pETAβ42–GFP (24), which produces Aβ42–GFP under the control of the strong bacteriophage T7 promoter, were cotransformed with the pSICLOPPS–NuX₁X₂X₃X₄X₅X₆ vector library. Approximately 3×10^9 transformants carrying both vectors were harvested, pooled together, and grown in Luria–Bertani (LB) liquid medium containing 0.005% 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 reached a level of about 0.5, 0.1 mM isopropyl-β-D-thiogalactoside (IPTG) was added to the medium so as to induce overexpression of the Aβ42–GFP reporter. After about 2 hours at 37°C, $\sim 3 \times 10^9$ cells were screened, and the population exhibiting the top 1 to 3% fluorescence was isolated using fluorescence-activated cell sorting (FACS) (fig. S1A). The isolated cells were regrown and screened for a total of seven rounds, at which point the mean fluorescence of the population displayed an almost sixfold increase compared to the starting library (Fig. 2B). No further substantial increase in fluorescence was observed after additional rounds of sorting.

After the seventh round of FACS screening, 10 individual clones were randomly chosen from the sorted population, and their peptide-encoding vectors were isolated and then retransformed into fresh *E. coli* Tuner (DE3) cells carrying pET Aβ42–GFP. Protein production was induced from both plasmids, and the levels of Aβ42–GFP fluorescence of these cultures were measured. Aβ42–GFP fluorescence of the isolated peptide-expressing clones was found to be markedly increased compared to cells expressing the same Aβ42–GFP fusion in the presence of two random cyclic peptide sequences previously found to have no effect on Aβ42–GFP fluorescence and aggregation (Fig. 2C) (18). 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 (fig. S1B). Furthermore, the observed phenotypic effects were dependent on the ability of the Ssp DnaE intein to perform protein splicing, as the double amino acid substitution H24L/F26A in the C-terminal half 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 (25), was found to reduce Aβ42–GFP fluorescence back to wild-type levels (Fig. 2C and fig. S1B). Last, the observed increases in fluorescence were found to be Aβ42 specific, as the isolated pSICLOPPS–NuX₁X₂X₃X₄X₅X₆ vectors did not enhance the levels of cellular green fluorescence when the sequence of Aβ42 was replaced in the same vector with that of the DNA-binding (core) domain of human p53 containing a tyrosine to cysteine substitution at position 220 [p53C (Y220C)], a protein whose misfolding and aggregation is associated with certain forms of cancer (fig. S1C) (26).

Analysis of the expressed Aβ42–GFP fusions by SDS–PAGE and Western blotting revealed that the bacterial clones expressing the selected cyclic heptapeptides produce markedly increased levels of soluble Aβ42–GFP compared to random cyclic peptides, despite the fact that accumulation of total Aβ42–GFP protein remained at similar levels (Fig. 2D, top, and fig. S1D). Furthermore, when the same cell lysates were analyzed by native PAGE and Western blotting, we observed that coexpression of the selected cyclic peptides reduced the accumulation of higher-order Aβ42–GFP aggregates, which could not enter the gel, and increased the amounts of species with higher electrophoretic mobility (Fig. 2D, bottom left). These higher electrophoretic mobility species correspond to the fraction of the total Aβ42–GFP that exhibits fluorescence (Fig. 2D, bottom right). Since the solubility and fluorescence of bacterially expressed Aβ42–GFP has been found to be inversely proportional to the aggregation propensity of Aβ42 (18, 24, 27), the results described above suggest that Aβ42 aggregation is markedly decreased in the presence of the selected cyclic heptapeptides. Similar results were acquired when Aβ42 was produced in an unfused, GFP-free form. When we tested the effects of the selected cyclic heptapeptides on Aβ42 aggregation with an *in vivo* assay using whole-cell staining of intracellular formation of Aβ42 aggregates with thioflavin S (ThS) (28), we observed that coproduction of the selected peptides resulted in decreased levels of ThS fluorescence, further indicating a reduced aggregate formation (Fig. 2E).

DNA sequencing of the 10 selected clones revealed five distinct cyclic heptapeptide sequences: cyclo-CKVWQLL (present six times among the sequenced clones), cyclo-CRVWTEL, cyclo-CKVWMPL, cyclo-CIVVPSI, and cyclo-CRIVPSL (fig. S1E).

High-throughput analysis of the isolated hits

We previously found that low-molecular weight peptide macrocycles are a rich source of chemical rescuers of disease-associated protein misfolding and aggregation (18). On the basis of that initial observation, and in combination with the fact that multiple distinct cyclic heptapeptide sequences were identified among the 10 selected clones initially tested (fig. S1E), we hypothesized that numerous Aβ42-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 (Fig. 2B). This analysis revealed 416 distinct cyclic heptapeptide sequences appearing at

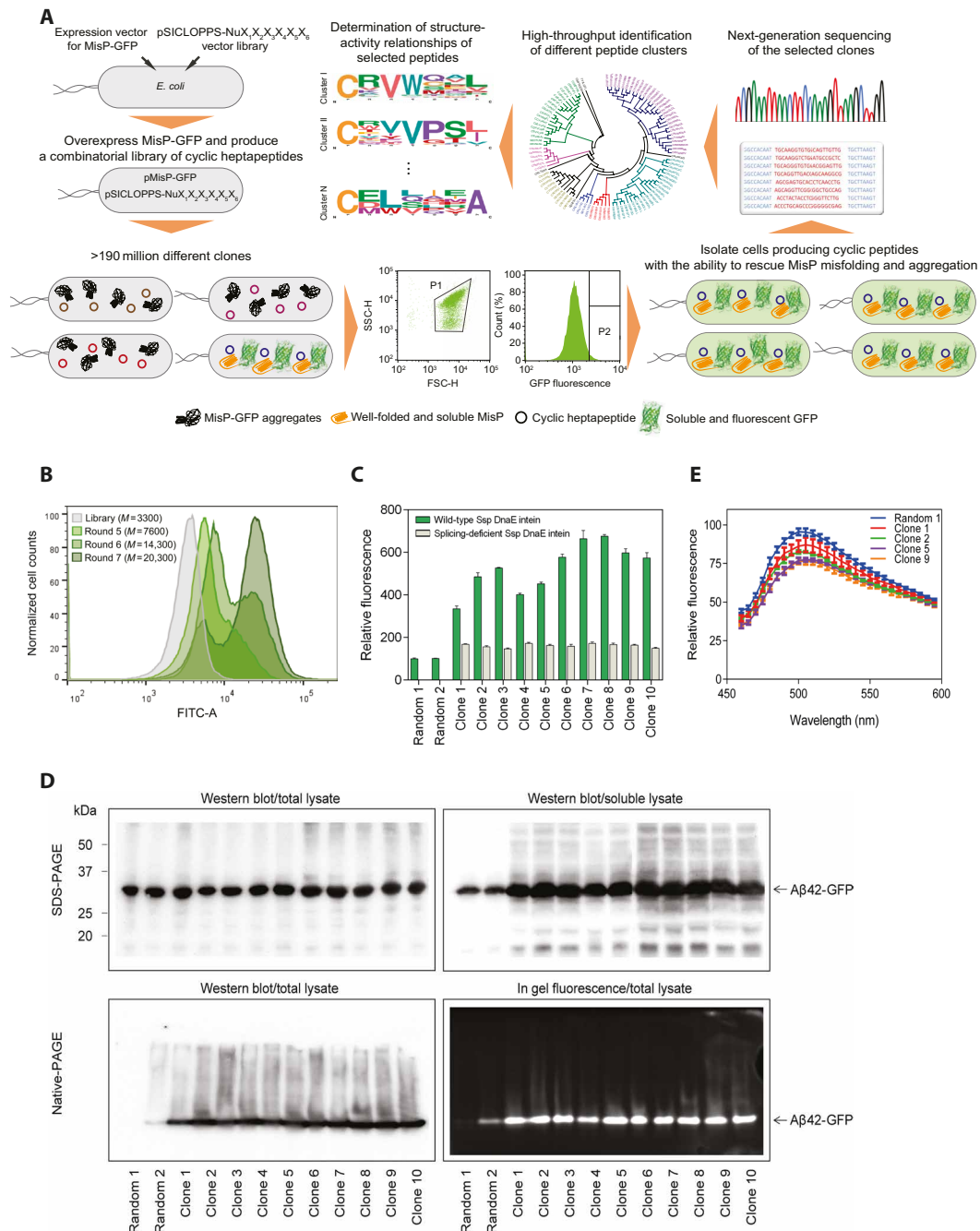


Fig. 2. Biosynthesis and ultrahigh-throughput screening of a combinatorial cyclic heptapeptide library with expanded diversity for discovering inhibitors of protein aggregation. (A) Schematic of the used bacterial platform for discovering inhibitors of protein aggregation and for the high-throughput analysis of the selected hits. pMisP-GFP: plasmid encoding a misfolded protein-GFP fusion; pSICLOPPS-NuX₁X₂X₃X₄X₅X₆: vector library encoding the combinatorial heptapeptide library cyclo-NuX₁X₂X₃X₄X₅X₆; Nu: Cys, Ser, or Thr; X_i: any of the 20 natural amino acids; FSC-H: forward scatter; SSC-H: side scatter; P: sorting gate. (B) FACS of *E. coli* Tuner (DE3) cells overexpressing Aβ42-GFP and the combined cyclic heptapeptide library. *M*: mean GFP fluorescence in arbitrary units. FITC-A: filter for fluorescein isothiocyanate. (C) Relative fluorescence of *E. coli* Tuner (DE3) cells overexpressing Aβ42-GFP and 10 randomly selected cyclic heptapeptide clones isolated after the seventh round of FACS shown in (B) and using either the wild-type split Ssp DnaE intein (green bars) or the splicing-deficient variant H24L/F26A (white bars) (25). Two randomly picked cyclic peptide sequences (random 1 and 2) previously shown to have no effect on Aβ42-GFP fluorescence and aggregation (18) were used as a negative control. The fluorescence of the bacterial population producing cyclic peptide random 1 was arbitrarily set to 100. Mean values ± SEM are presented (*n* = 3 independent experiments, each performed in three replicates). (D) Top: Western blot analysis of total (left) and soluble (right) lysates of *E. coli* Tuner (DE3) cells overexpressing Aβ42-GFP and the 10 individual cyclic peptide sequences tested in (C). The predicted molecular mass of the Aβ42-GFP fusion is ~32 kDa. Bottom: Western blotting using the anti-Aβ antibody 6E10 (left) and in-gel fluorescence (right) analyses of total lysates following native PAGE of *E. coli* Tuner (DE3) cells coexpressing Aβ42-GFP and the 10 individual cyclic peptide sequences tested in (C). (E) Emission spectra of *E. coli* Tuner (DE3) cells overexpressing Aβ42 along with four of the selected cyclic heptapeptide sequences tested in (C) and stained with ThS. The maximum fluorescence of the bacterial population producing cyclic peptide random 1 was arbitrarily set to 100. Mean values ± SEM are presented (*n* = 1 experiment performed in three replicates).

least 20 times within the sorted population, thus indicating that their presence in the selected pool is not coincidental. 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 (fig. S1F).

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) (Fig. 3A, left). 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 (Fig. 3A, right, and table S2). 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 (Fig. 3A, right, and table S2), thus indicating a highly efficient selection process.

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 (fig. S2A). From the 416 cyclic heptapeptides selected, 323 of them formed 1467 unique pairs with more than 70% sequence identity and formed 20 distinct clusters with similar sequence characteristics (Fig. 3B and fig. S2B). Clusters I and II were the most dominant, comprising 75.0 and 4.9% of the selected bacterial clones, respectively, as well as 25.7 and 6% of the unique cyclic heptapeptide sequences selected (Fig. 3B and table S3). 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 (fig. S1E).

The selected cyclic heptapeptides inhibit A β 42 aggregation in vitro

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 (Fig. 4A and table S4), were chosen for subsequent analysis and were synthesized chemically in milligram quantities (fig. S3A). These cyclic peptides were selected because they were both encountered in the postselection pool investigated initially (fig. S1E) and, more importantly, they were the most frequently encountered members among the two most dominant clusters (clusters I and II) (table S4).

A β C7-1 and A β C7-14 were initially evaluated by monitoring their effects on the kinetics of A β 42 aggregation by thioflavin T (ThT) staining using a highly reproducible approach previously described (29, 30). Monomeric A β 42 was purified after recombinant production in *E. coli*, and aggregation kinetic experiments were initiated using 2 μ M A β 42 in the absence and presence of A β C7-1 and A β C7-14. Both A β C7-1 and A β C7-14 inhibited A β 42 aggregation very effectively at substoichiometric ratios as low as 0.5 molar equivalents for A β C7-1 and 0.1 molar equivalents for A β C7-14 (Fig. 4, B and C). 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, albeit to a different

extent (Fig. 4, B and C, right). Furthermore, we found that the 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 (fig. S3B). Thus, it is likely that these selected macrocycles are not binding irreversibly to A β 42 species and redirecting the aggregation process toward off-pathway aggregates. 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 (fig. S3C).

The selected cyclic heptapeptides inhibit A β 42 aggregation and toxicity in vivo

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 A β 42-induced pathogenicity 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 (31). Upon temperature upshift, these nematodes (hereafter referred to as A β worms) exhibit muscle-localized A β 42 aggregation and eventually the emergence of a paralysis phenotype (31). 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 (Fig. 5, A and B). Both peptides were able to restore the total fitness of the A β worms to approximately the levels of the wild-type animals (Fig. 5C). Furthermore, A β worms treated with either one of the selected cyclic peptides produced 50 to 60% fewer A β 42 aggregates, as determined by imaging of the worms using the amyloid-specific dye 2-[[5'-(4-hydroxyphenyl)(2,2'-bithiophen)-5-yl]-methylene]-propanedinitrile (NIAD-4) (Fig. 5, D and E).

To exclude the possibility of promoter- or strain-specific effects, we also 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 (32). Consistent with our previous observations, the administration of both cyclic peptides resulted in a significant delay in the emergence of its characteristic paralysis phenotype (fig. S4). 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.

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

To identify the functionally important residues within the selected peptides, we performed nucleophile substitutions at position 1 and Ala-scanning mutagenesis at positions 2 to 7 for both A β C7-1 and A β C7-14. Then, we compared the effects 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 (Fig. 6, A and B).

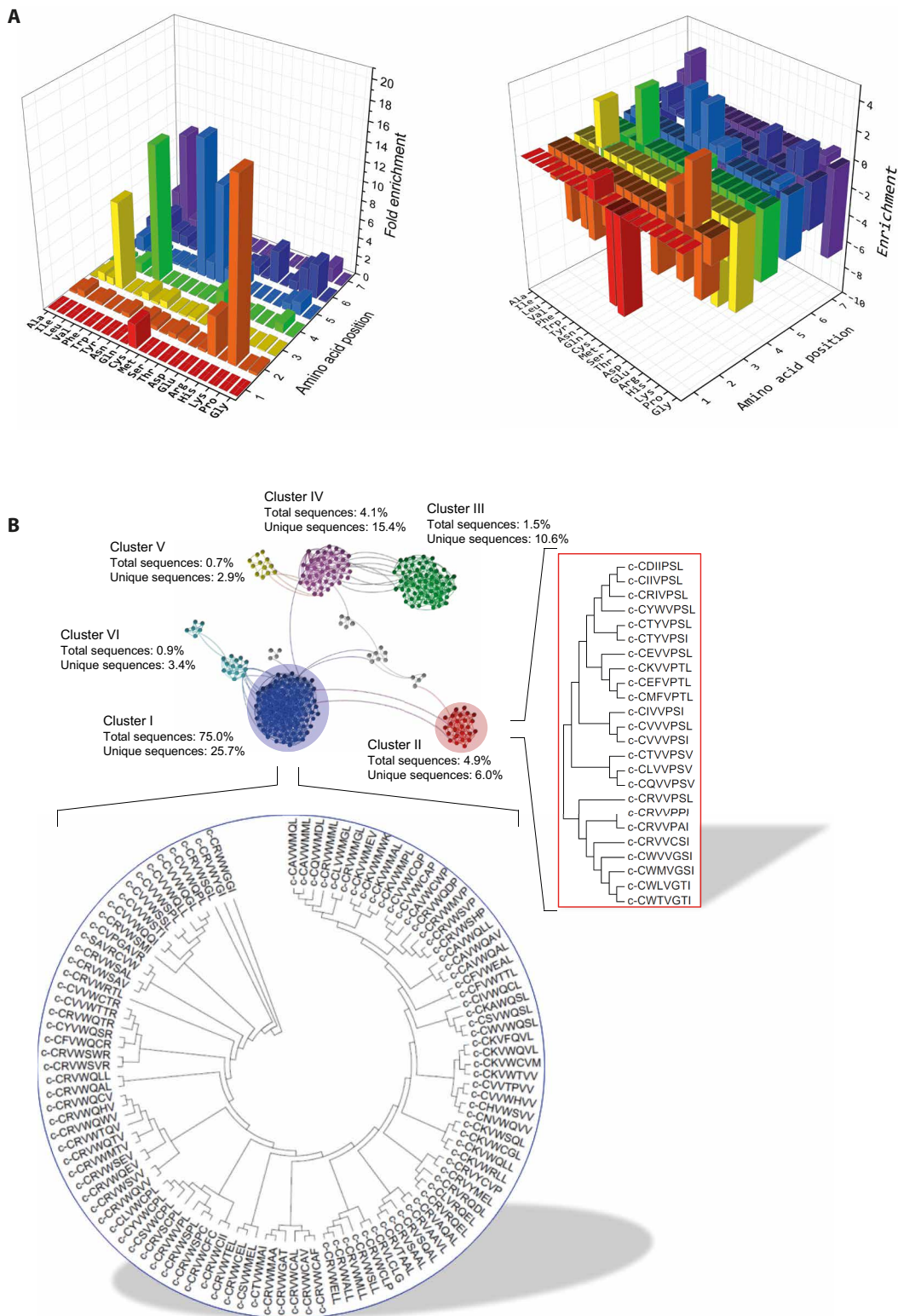


Fig. 3. Sequence analysis of the selected cyclic heptapeptide pool. (A) Left: Frequency of appearance of the 20 natural amino acids at each position of the heptapeptide sequences selected after the seventh round of sorting (Fig. 2B). Right: Enrichment of the 20 natural amino acids at each position of the heptapeptide sequences selected after the seventh round of sorting (Fig. 2B). Values represent the \log_2 -fold change of the amino acid frequency of appearance of the peptides from the sorted pool compared to the initial library. (B) Visualization of the main clusters formed by the selected cyclic heptapeptides according to their sequence similarities. Nodes represent different cyclic peptide sequences, and solid lines connect pairs of peptides that share at least 70% sequence identity. The sequences of the members of the two most dominant clusters (clusters I and II) are shown in the corresponding dendrograms.

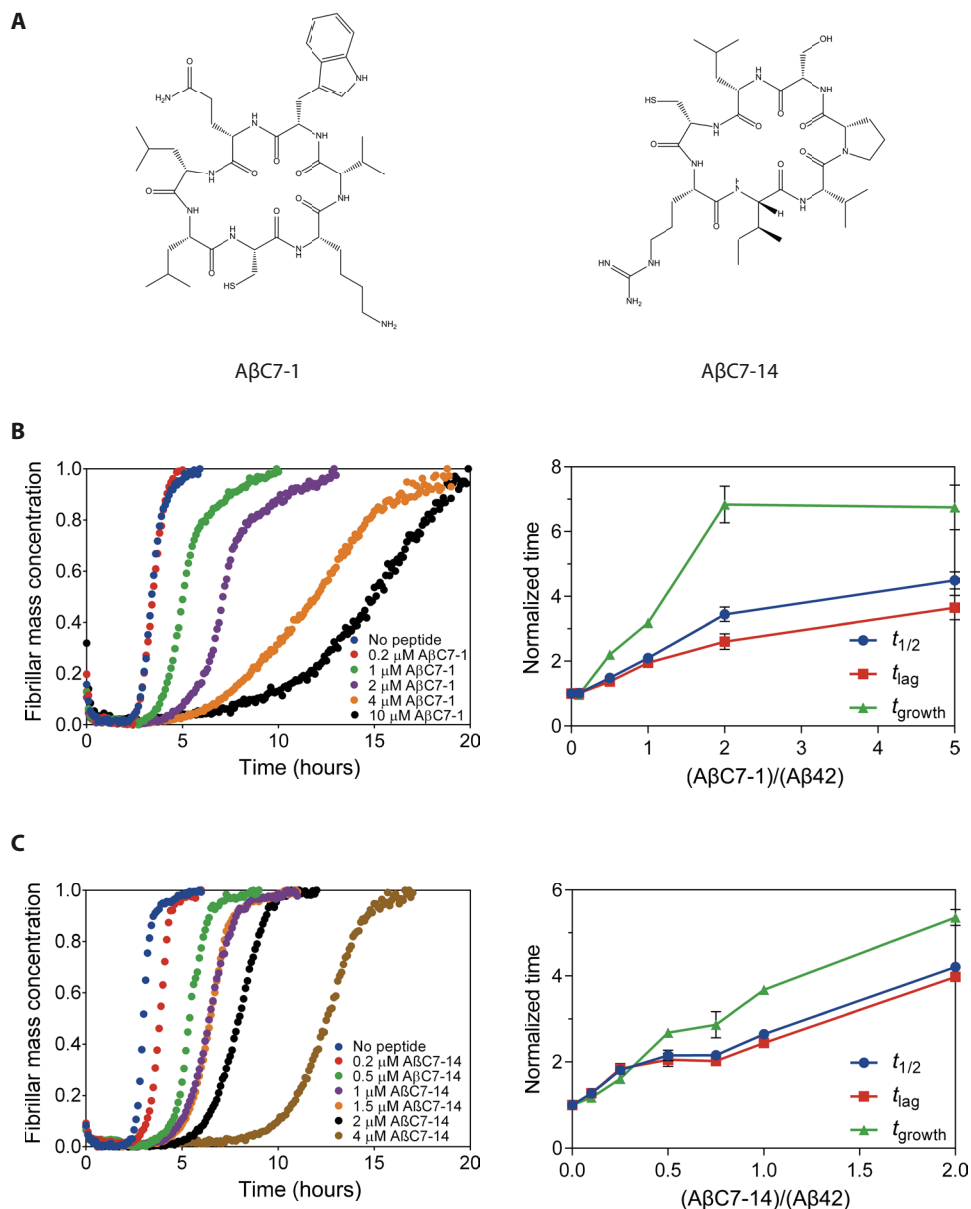


Fig. 4. AβC7-1 and AβC7-14 inhibit Aβ42 aggregation in vitro. (A) Chemical structures of the selected cyclic heptapeptides AβC7-1 and AβC7-14. (B) Kinetic profiles of the aggregation of $2 \mu\text{M}$ Aβ42 in the absence and presence of AβC7-1 at different molar ratios (left) and the normalized $t_{1/2}$, t_{lag} , and t_{growth} values of the corresponding aggregation reactions (right). (C) As in (B) for AβC7-14. In (B) and (C), mean values \pm SEM are presented ($n = 1$ experiment performed in three replicates).

The latter observation is somewhat unexpected, considering the dominant appearance of Cys¹ sequences among the selected cyclic heptapeptide pool (Fig. 3A, left), but it may be related to our previous results, where Thr played a crucial role in the identified cyclic peptides against Aβ42 aggregation (18). Since the isolation of the bioactive sequences in our system requires repeated rounds of bacterial culturing, protein overexpression and FACS, the scarcity of Thr¹-containing sequences in the isolated cyclic heptapeptide pool may be occurring because of a toxicity effect of these sequences on bacterial growth, which can result in de-enrichment of the clones that produce them, despite their efficiency in preventing protein aggregation.

Furthermore, for both peptides, Ala-scanning mutagenesis at the majority of the positions 2 to 7 resulted in markedly Aβ42-GFP fluorescence decrease and concomitant increase in aggregation

(Fig. 6, A and B). Specifically, for AβC7-1, substitutions at positions 2, 3, 4, and 7 resulted in a ~30 to 70% decrease in Aβ42-GFP fluorescence, while for AβC7-14, substitutions at all positions except Ser⁶ resulted in a ~45 to 80% decrease (Fig. 6, A and B). These observations indicate that a number of residues in both selected cyclic heptapeptides are important for optimal aggregation inhibition activity. 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 (Fig. 6C and table S4). Similarly, for cluster II,

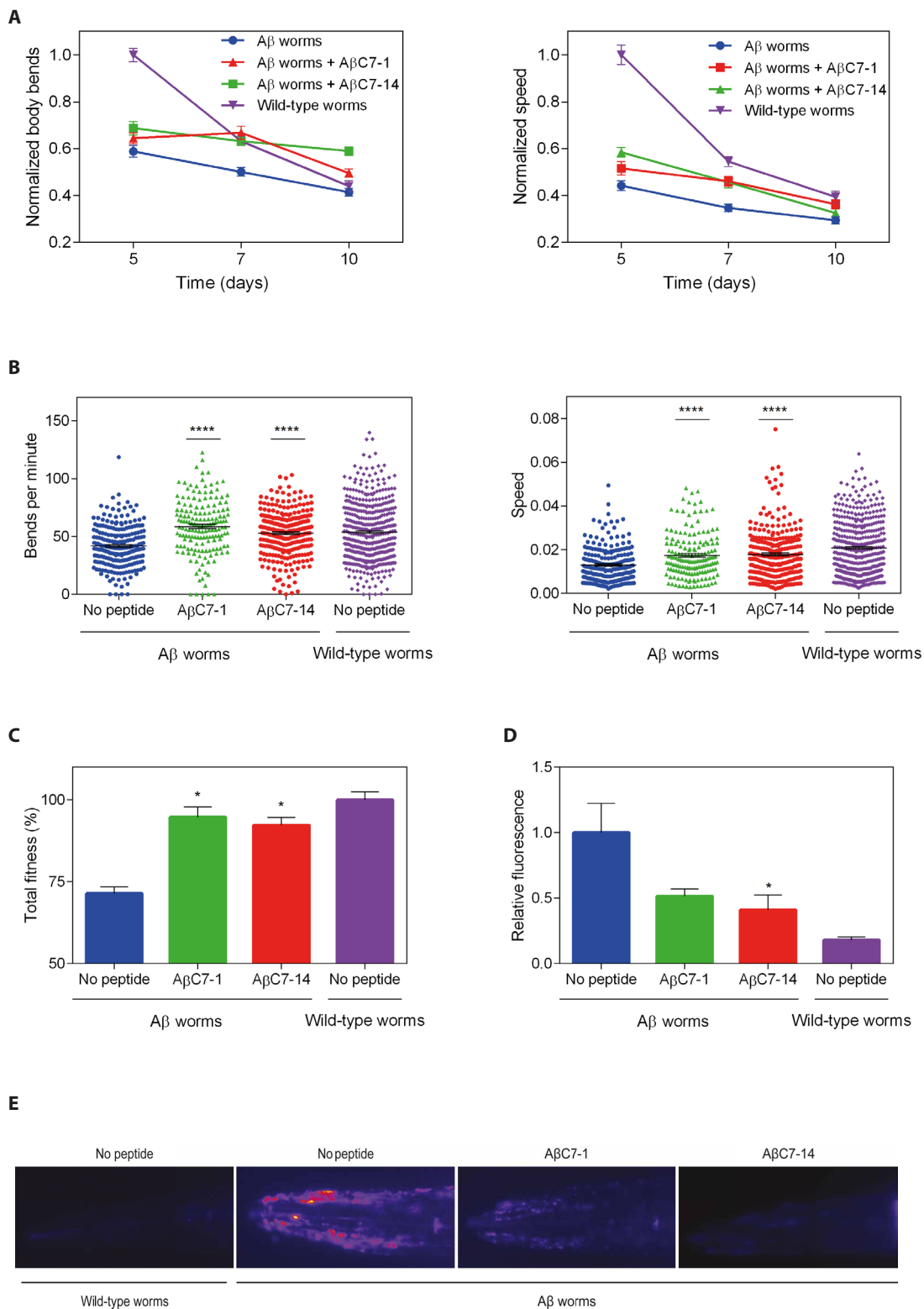


Fig. 5. AβC7-1 and AβC7-14 inhibit Aβ42 aggregation in vivo. (A) Normalized motility (left) and normalized speed of movement (right) of Aβ42 and wild-type worms in the absence and presence of 40 μM AβC7-1 and 5 μM AβC7-14 during days 5 to 10 of adulthood. (B) Motility (left) and speed (right) of individual Aβ and wild-type worms in the absence and presence of AβC7-1 and AβC7-14 at day 7 of adulthood. (C) Total fitness (*S*₁) of the worms as in (B). (D) Relative fluorescence of Aβ42 and wild-type worms at day 7 of adulthood showing a 50 to 60% decrease in Aβ42 aggregate formation in the presence of AβC7-1 and AβC7-14. (E) Representative images from (D). In (A) to (C), ~200 worms were analyzed on average, while in (D), 25 worms were analyzed in total. In all panels, mean values ± SEM are presented (*n* = number of worms tested in one experiment). Statistical significance is denoted by **P* ≤ 0.05 and *****P* ≤ 0.0001, for differences to the “No peptide Aβ worms” sample.

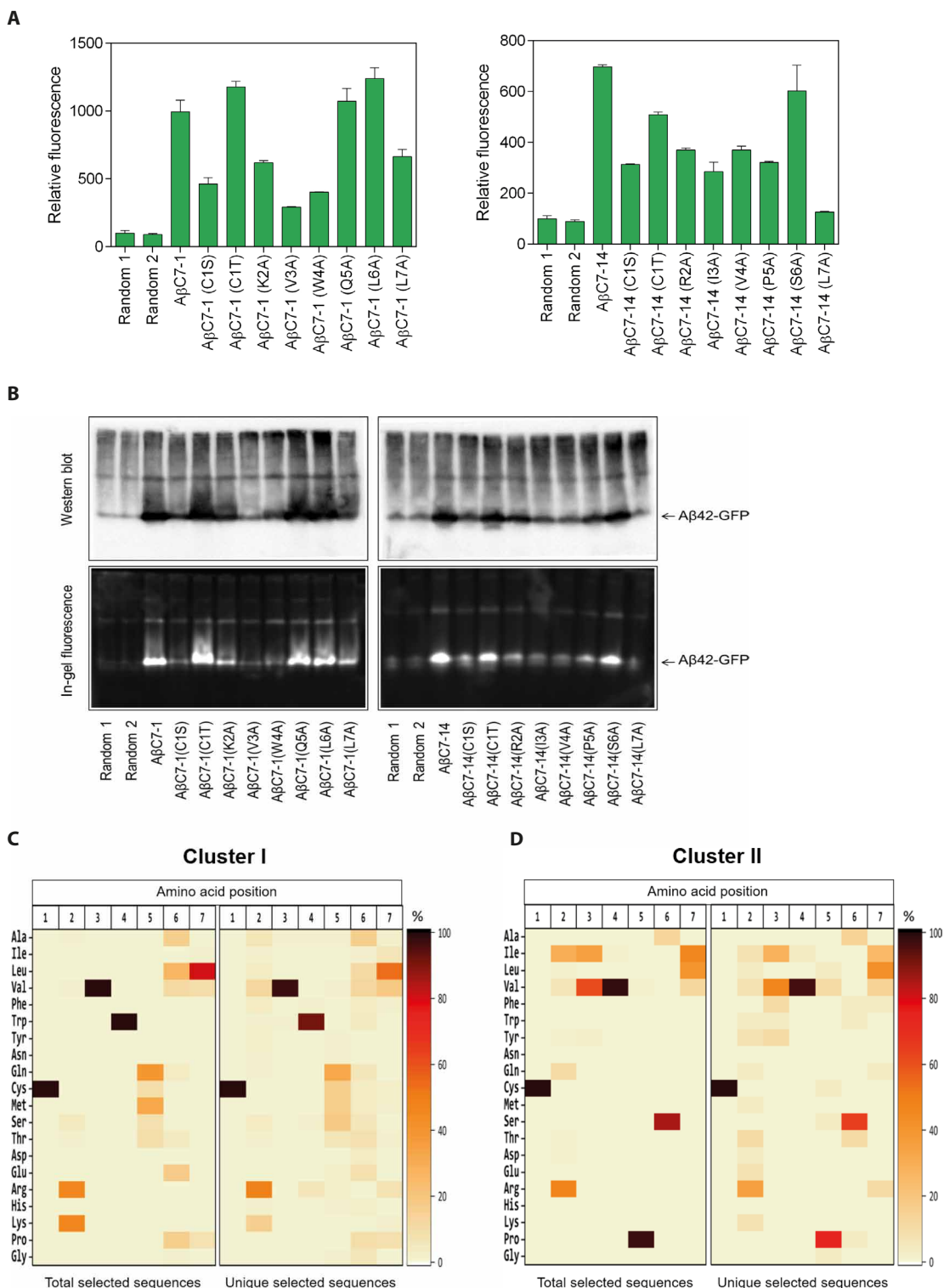


Fig. 6. Structure-activity analysis of the selected heptapeptides A β C7-1 and A β C7-14. (A) Relative fluorescence of *E. coli* Tuner (DE3) cells overexpressing A β 42-GFP and A β C7-1 (left) or A β C7-14 (right) or the indicated variants thereof as measured by flow cytometry. The fluorescence of the bacterial population coproducing the random cyclic peptide was arbitrarily set to 100. Experiments were carried out in triplicate ($n = 1$ experiment), and the reported values correspond to the mean value \pm SEM. (B) Western blotting using the anti-A β antibody 6E10 (top) and in-gel fluorescence (bottom) analyses following native PAGE of total lysates of *E. coli* Tuner (DE3) cells coexpressing A β 42-GFP and A β C7-1 (left) or A β C7-14 (right) along with the indicated variants thereof. (C) Heatmap representation of the amino acid distribution at each position of the peptide sequences corresponding to cluster I (Fig. 3B), as demonstrated by the deep sequencing analysis results. The total (left) or the unique (right) heptapeptide sequences were included in the analysis. (D) As in (C) 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, the frequency of appearance was >97% (Fig. 6D and table S4). Together, our results indicate that the most bioactive motifs against A β misfolding and aggregation in the investigated macrocyclic 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 any one of the 20 natural amino acids; Π is any one of the polar amino acids Q, C, S, and T; Δ is R, I, V, or Q; and Ψ is any one of the aliphatic amino acids L, V, and I.

DISCUSSION

We have reported how a previously developed bacterial platform can be expanded to enable the simultaneous production and functional screening of molecular libraries with greatly increased diversities for the discovery of inhibitors of disease-associated protein aggregation. We have generated a complete combinatorial library of nearly 200 million head-to-tail cyclic heptapeptides in the cytoplasm of *E. coli* cells and have rapidly screened them to discover inhibitors of the pathogenic misfolding and aggregation of A β 42. We thus found head-to-tail cyclic heptapeptides that efficiently reduce A β 42 aggregation and toxicity both in vitro and in vivo. Our highly effective screening methodology, coupled with high-throughput sequencing analysis of the isolated hits, enabled the identification of >400 cyclic heptapeptide putative inhibitors of A β 42 aggregation. In addition, these results provide further support to our previous observations that low-molecular weight peptide macrocycles are a very rich source of chemical rescuers of protein misfolding (18) and that they may constitute a promising class of potential therapeutics (33).

Our unbiased selection process yielded distinct groups of bioactive macrocyclic peptides with different sequence characteristics. For the two most dominant clusters, we used a combination of site-directed mutagenesis and deep sequencing analyses to rapidly define the sequence motifs providing optimal bioactivity. These were found to be cyclo-(C,T) (R,K)VW (Π ,A,M)X (Ψ ,P) for cluster I and cyclo-(C,T) Δ (I,V)VP (S,A) Ψ for cluster II, where X is any one of the 20 natural amino acids; Π is any one of the polar amino acids Q, C, S, or T; Δ is R, I, V, or Q; and Ψ is any one of the aliphatic amino acids L, V, or I. Our in vitro validation indicated that these macrocyclic peptides likely exert protective effects by interfering with microscopic reaction steps underlying the aggregation of A β , which affect the generation of oligomers over time. In the context of an in vivo system, as observed in *C. elegans*, where aggregation proceeds on a far longer time scale, this delay in aggregation is much more pronounced and can be considered as effective as an overall arrest of the entire process (34).

To our knowledge, the present work describes the largest screen of small molecule-like molecular entities with the ability to perform direct functional screening beyond simple detection of binding to the target protein described to date. Compared to other reported functional compound screens for misfolding rescuing or other bioactivities in vitro or in vivo (8, 18, 25), we have demonstrated that the system that we described has the ability to generate and evaluate molecular libraries with 20 to 1000 higher diversity than what can be currently achieved. Furthermore, as the diversity of the generated peptide macrocycle libraries are limited only by the theoretical diversity of

the library design and the transformation efficiency of *E. coli* cells, our system can allow the evaluation of libraries with tens or even hundreds of billions of members. Notably, *E. coli* can support the biosynthesis of not only head-to-tail cyclic peptides, as investigated here, but also side chain-to-tail cyclic peptides (35), bicyclic peptides (36), lasso peptides (37), θ -defensins (38), cyclotides (39), and other macrocyclic structures (40) that include both natural and noncanonical amino acids (41). Contrary to other approaches that allow the investigation of even wider areas of molecular space, such as mRNA display (10) and DNA-encoded libraries (7, 42), our technology goes beyond simple detection of binding to the target protein and, instead, selects directly for compounds rescuing aggregation. This is an important advantage, since compound resynthesis and testing for the desired bioactivity following affinity-based selections of DNA- and genetically encoded libraries is time consuming, expensive, and results in a high discovery rate of binders that do not exhibit the desired biological activity (42).

It is noteworthy that the sequences of the A β 42-targeting cyclic heptapeptide discovered here diverge completely from those isolated from our previous screen that included combinatorial libraries of shorter cyclopeptides (18). This result suggests that, apart from the specific amino acid residues in the primary sequence of the macrocyclic peptide interacting directly with the target protein and are necessary for bioactivity (18), there is probably a conformational component that is also important for molecular recognition between these macrocycles and their targets and that larger cyclopeptide scaffolds are not mere extensions of shorter bioactive sequences. Furthermore, the selected macrocycles bear no resemblance with the sequence of A β 42, and thus, their discovery would have been very challenging using rational or computationally guided design as, for example, in the case of classical β sheet breaker peptides (43) and other designed peptide-based inhibitors of A β aggregation (44, 45). Last, also note that the selected cyclopeptides have drug-like molecular characteristics, when compared to those of existing macrocyclic drugs and, in some aspects, to those of conventional drugs as well (table S5).

Our biotechnological approach for producing and evaluating molecular libraries with expanded diversities is not restricted to A β 42 but is highly versatile and can be applied broadly for targeting a variety of misfolding-prone proteins of both globular and intrinsically disordered nature, as we have shown previously (18). We are currently using this system to screen molecular libraries with expanded diversities, such as the ones described here, and have identified candidate macrocyclic rescuers of the misfolding and aggregation of variants of human Cu/Zn superoxide dismutase and p53, as well as of huntingtin, whose misfolding and aggregation are associated with amyotrophic lateral sclerosis, cancer, and Huntington's disease, respectively (1).

The biosynthetic production of the lower-weight molecular libraries under investigation and their simultaneous screening for bioactivity in a simple bacterium like *E. coli* offer great simplicity and speed and reduces the overall cost of the discovery process markedly (7, 12). Once the peptide macrocycle library has been constructed, one can identify the entire repertoire of aggregation inhibitors for a target protein and, at the same time, acquire an initial understanding of structure-activity relationships for the acquired hits in less than a month. The simplicity, speed, and wide applicability of this approach could permit academic and industrial laboratories to simultaneously perform parallel screenings against multiple targets and to prioritize further compound development according to the number and nature

of the hits uncovered by the screen. Overall, our approach represents a highly adaptable strategy for investigating molecular libraries with expanded diversities, which enables the discovery of New molecular entities that effectively target peptides and proteins associated with protein misfolding diseases.

MATERIALS AND METHODS

Construction of the combinatorial cyclic heptapeptide library

The vector sublibraries pSICLOPPS-CysX₁X₂X₃X₄X₅X₆, pSICLOPPS-SerX₁X₂X₃X₄X₅X₆, and pSICLOPPS-ThrX₁X₂X₃X₄X₅X₆ (table S6) were generated as described previously (18). Briefly, the degenerate forward primers GS078, GS079, and GS080 were used together with the reverse primer GS035 and pSICLOPPS as a template (table S6). Cys, Ser, and Thr were encoded in these primers by the codons TGC, AGC, and ACC, respectively, while the randomized amino acids (X) were encoded using random NNS codons, where N = A, T, G, or C and S = G or C. 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 of the peptide sublibraries starting with Cys, Ser, or Thr, respectively, together with the reverse primer GS035. The resulting PCR products were then digested with Bgl I and Hind III for 5 hours and inserted into a similarly digested and dephosphorylated pSICLOPPSKanR vector (18). The ligation reactions were optimized at a 12:1 insert:vector ratio and performed at 16°C for 4 hours. Approximately 10 µg of the pSICLOPPSKanR vector was used for each sublibrary. The ligated DNA was then purified using spin columns, transformed into electrocompetent MC1061 cells, plated onto LB agar plates containing chloramphenicol (25 µg/ml), and incubated at 37°C for 14 to 16 hours. This process resulted in approximately 1.2 billion independent transformants, as judged by plating experiments after serial dilutions.

Cyclic heptapeptide library screening

Electrocompetent *E. coli* Tuner (DE3) cells (Novagen, USA) carrying the expression vector pETAβ42-GFP (24) were cotransformed with the combined pSICLOPPS-NuX₁X₂X₃X₄X₅X₆ vector library. Approximately 10⁹ transformants carrying both vectors 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 to 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 2 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, ~3 × 10⁹ cells were gated on a side-scatter (SSC-H) versus forward-scatter (FSC-H) plot to eliminate non-cellular events and were subjected to FACS for the isolation of the bacterial population exhibiting the top ~2% fluorescence. The isolated cells were regrown 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.

High-throughput sequencing analysis

High-throughput sequencing 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. Briefly, the combined pSICLOPPS-NuX₁X₂X₃X₄X₅X₆ vector library and the enriched peptide library after the seventh round of sorting were digested with Nco I and BsrG I, and the resulting ~250 base pair (bp) products that contained the variable peptide-encoding region were isolated and analyzed. Ion proton reads were aligned to a reference sequence using Bowtie2 (v2.2.8). The alignment information stored in the CIGAR string of the resulting Sequence Alignment Map file was parsed and mapped to matching and mismatching sequences using the tool Biostar59647 of the JVarkit utilities. From the resulting XML file, a custom awk script extracted the mismatching insert sequences, which were then clustered using the CD-HIT tool (v4.6.1) (46), together with their read counts. From the obtained data, only the 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 clustering tool (<http://tools.iedb.org/cluster2/>) and the fully interconnected clusters (cliques) method (47). This approach allows all peptides in a clique to share a minimal level of identity, while at the same time, one peptide can be part of multiple cliques (47). As sequence similarity analysis was 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 identity were identified, and after reintegration 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 identity. The remaining 93 cyclic peptides did not share a minimal level of 70% identity with any other of the peptides. The results were then presented in an undirected network graph using the Gephi graph visualization software (48), and cluster identification was performed using the Girvan-Newman Algorithm (49).

Aβ42 aggregation kinetic experiments

Kinetic experiments were performed as described previously (30). Briefly, appropriate amounts of the synthetic cyclic peptides were added to 2 µM of monomeric Aβ42 to obtain the desired cyclic peptide:Aβ42 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 AβC7-1 and AβC7-14, respectively. Under these conditions, both AβC7-1 and AβC7-14 remained stable in a monomeric state for the duration of the in vitro experiments, as judged by dynamic light scattering analyses. 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-coated plate (Corning 3881), at 80 µl 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.

C. elegans motility

Strains

The following strains were used for this experiment: (i) GMC101, herein referred to as A β worms; genotype *dvIs100* [*unc-54p::A-beta-1-42::unc-54 3'-UTR + mtl-2p::GFP*]; *mtl-2p::GFP* constitutively expresses the GFP in intestinal cells; *unc-54p::A-beta-1-42* expresses A β 42 in body wall muscle cells, resulting in A β 42 aggregation and worm paralysis after temperature upshift from 20° to 25°C (31). (ii) N2, wild-type *C. elegans* var Bristol, herein referred to as wild-type worms (50).

Propagation procedures

C. elegans worms were propagated using standard conditions and as described previously (30, 50). Briefly, the worms were treated with hypochlorite bleach, and eggs were hatched overnight in M9 buffer [KH_2PO_4 (3 g/liter), Na_2HPO_4 (6 g/liter), NaCl (5 g/liter), and 1 mM MgSO_4] and then distributed on nematode growth medium (NGM) [1 mM CaCl_2 , 1 mM MgSO_4 , cholesterol (5 mg/ml), 250 mM KH_2PO_4 (pH 6), agar (17 g/liter), NaCl (3 g/liter), and casein (7.5 g/liter)] 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% (v/v) acetonitrile. Synthetic cyclic peptides were provided to the nematodes as is, without any additional steps to enhance their permeability. At that point, 75 μM 5-fluoro-2'-deoxyuridine was also added to the plates to inhibit growth of offspring. The plates were then transferred to 24°C to promote A β 42 expression and aggregation.

Motility assay

On days 5 to 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/s for 1 min using a homemade microscopic setup, and the body bends were quantified using a custom-tracking algorithm as described previously (30, 51). In total, ~2300 worms were analyzed per peptide 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 (30). Briefly, live animals were stained by incubating with 1 μM NIAD-4 [0.1% (v/v) dimethyl sulfoxide 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 anesthetized by adding 40 mM NaN_3 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 \times 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 because of the high background signal in the intestine.

Statistical analyses

Statistical analyses were performed using Prism (GraphPad Software Inc., La Jolla, CA, USA), and mean values were compared using unpaired two-tailed *t* tests. For animal experiments, group sizes were chosen on the basis of 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.

SUPPLEMENTARY MATERIALS

Supplementary material for this article is available at <http://advances.sciencemag.org/cgi/content/full/5/10/eaax5108/DC1>

Section S1. Supplementary Materials and Methods

Fig. S1. Identification of potential A β 42 aggregation inhibitors using a bacterial genetic screen.

Fig. S2. Identification of different cyclic peptide clusters appearing in the sorted population.

Fig. S3. A β C7-1 and A β C7-14 inhibit the aggregation of A β 42 in vitro.

Fig. S4. A β C7-1 and A β C7-14 inhibit the aggregation of A β 42 in vivo.

Table S1. Deep sequencing analysis of the peptide-encoding regions of ~3.4 million clones from the constructed pSICLOPPS-NuX₁X₂X₃X₄X₅X₆ library.

Table S2. Enrichment (blue) and depletion (red) of the 20 amino acids in each position of the heptapeptide sequences.

Table S3. Distribution of the heptapeptide sequences in the different clusters identified.

Table S4. Sequences and frequency of appearance of cluster I and cluster II heptapeptide sequences as determined by high-throughput sequencing of the enriched library after the seventh round of sorting.

Table S5. Molecular properties of the selected cyclic heptapeptides A β C7-1 and A β C7-14 compared to those of conventional drugs, oral macrocyclic (MC) drugs, and nonoral MC drugs.

Table S6. Plasmids and PCR primers used in this study.

References (52, 53)

[View/request a protocol for this paper from Bio-protocol.](#)

REFERENCES AND NOTES

1. F. Chiti, C. M. Dobson, Protein misfolding, amyloid formation, and human disease: A summary of progress over the last decade. *Annu. Rev. Biochem.* **86**, 27–68 (2017).
2. J. Cummings, G. Lee, A. Ritter, K. Zhong, Alzheimer's disease drug development pipeline: 2018. *Alzheimers Dement.* **4**, 195–214 (2018).
3. T. Coelho, G. Merlini, C. E. Bulawa, J. A. Fleming, D. P. Judge, J. W. Kelly, M. S. Maurer, V. Plante-Bordeneuve, R. Labaudiniere, R. Mundayat, S. Riley, I. Lombardo, P. Huertas, Mechanism of action and clinical application of tafamidis in hereditary transthyretin amyloidosis. *Neurol. Ther.* **5**, 1–25 (2016).
4. D. G. Warnock, D. G. Bichet, M. Holida, O. Goker-Alpan, K. Nicholls, M. Thomas, F. Eyskens, S. Shankar, M. Adera, S. Sitaraman, R. Khanna, J. J. Flanagan, B. A. Wustman, J. Barth, C. Barlow, K. J. Valenzano, D. J. Lockhart, P. Boudes, F. K. Johnson, 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).
5. E. Karran, J. Hardy, A critique of the drug discovery and phase 3 clinical programs targeting the amyloid hypothesis for Alzheimer disease. *Ann. Neurol.* **76**, 185–205 (2014).
6. A. Mullard, DNA tags help the hunt for drugs. *Nature* **530**, 367–369 (2016).
7. R. A. Goodnow Jr., C. E. Dumelin, A. D. Keefe, DNA-encoded chemistry: Enabling the deeper sampling of chemical space. *Nat. Rev. Drug Discov.* **16**, 131–147 (2017).
8. T. Kogej, N. Blomberg, P. J. Greasley, S. Mundt, M. J. Vainio, J. Schamberger, G. Schmidt, J. Hüser, Big pharma screening collections: more of the same or unique libraries? The AstraZeneca–Bayer Pharma AG case. *Drug Discov. Today* **18**, 1014–1024 (2013).
9. J. Hert, J. J. Irwin, C. Laggner, M. J. Keiser, B. K. Shoichet, Quantifying biogenic bias in screening libraries. *Nat. Chem. Biol.* **5**, 479–483 (2009).
10. T. Passioura, T. Katoh, Y. Goto, H. Suga, Selection-based discovery of druglike macrocyclic peptides. *Annu. Rev. Biochem.* **83**, 727–752 (2014).
11. Y. Nagai, T. Tucker, H. Ren, D. J. Kenan, B. S. Henderson, J. D. Keene, W. J. Strittmatter, J. R. Burke, Inhibition of polyglutamine protein aggregation and cell death by novel peptides identified by phage display screening. *J. Biol. Chem.* **275**, 10437–10442 (2000).
12. B. Halford, How DNA-encoded libraries are revolutionizing drug discovery. *Chem. Eng. News* **95**, 28–33 (2017).
13. A. M. Davis, A. T. Plowright, E. Valeur, Directing evolution: The next revolution in drug discovery? *Nat. Rev. Drug Discov.* **16**, 681–698 (2017).
14. T. van Groen, K. Wiesehan, S. A. Funke, I. Kadish, L. Nagel-Steger, D. Willbold, Reduction of Alzheimer's disease amyloid plaque load in transgenic mice by D3, a D-enantiomeric peptide identified by mirror image phage display. *ChemMedChem* **3**, 1848–1852 (2008).
15. B. P. Orner, L. Liu, R. M. Murphy, L. L. Kiessling, Phage display affords peptides that modulate beta-amyloid aggregation. *J. Am. Chem. Soc.* **128**, 11882–11889 (2006).
16. D. Xue, M. Zhao, Y. J. Wang, L. Wang, Y. Yang, S. W. Wang, R. Zhang, Y. Zhao, R. T. Liu, A multifunctional peptide rescues memory deficits in Alzheimer's disease transgenic mice by inhibiting A β 42-induced cytotoxicity and increasing microglial phagocytosis. *Neurobiol. Dis.* **46**, 701–709 (2012).
17. T. Kawasaki, K. Onodera, S. Kamijo, Selection of peptide inhibitors of soluble A β (1–42) oligomer formation by phage display. *Biosci. Biotechnol. Biochem.* **74**, 2214–2219 (2010).
18. I. Matis, D. C. Delivoria, B. Mavroidi, N. Papaevgeniou, S. Panoutsou, S. Bellou, K. D. Papavasileiou, Z. I. Linardaki, A. V. Stavropoulou, K. Vekrellis, N. Boukos, F. N. Kolisis, E. S. Gonos, M. Margarity, M. G. Papadopoulos, S. Efthimiopoulos, M. Pelecanou,

- N. Chondrogianni, G. Skretas, An integrated bacterial system for the discovery of chemical rescuers of disease-associated protein misfolding. *Nat. Biomed. Eng.* **1**, 838–852 (2017).
19. E. A. Villar, D. Beglov, S. Chennamadhavuni, J. A. J. Porco, D. Kozakov, S. Vajda, A. Whitty, How proteins bind macrocycles. *Nat. Chem. Biol.* **10**, 723–731 (2014).
 20. V. Poongavanam, B. C. Doak, J. Kihlberg, Opportunities and guidelines for discovery of orally absorbed drugs in beyond rule of 5 space. *Curr. Opin. Chem. Biol.* **44**, 23–29 (2018).
 21. D. J. Craik, Chemistry. Seamless proteins tie up their loose ends. *Science* **311**, 1563–1564 (2006).
 22. C. P. Scott, E. Abel-Santos, M. Wall, D. C. Wahnou, S. J. Benkovic, Production of cyclic peptides and proteins in vivo. *Proc. Natl. Acad. Sci. U.S.A.* **96**, 13638–13643 (1999).
 23. A. Tavassoli, SICLOPPS cyclic peptide libraries in drug discovery. *Curr. Opin. Chem. Biol.* **38**, 30–35 (2017).
 24. C. Wurth, N. K. Guimard, M. H. Hecht, 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).
 25. J. A. Kritzer, S. Hamamichi, J. M. McCaffery, S. Santagata, T. A. Naumann, K. A. Caldwell, G. A. Caldwell, S. Lindquist, Rapid selection of cyclic peptides that reduce α -synuclein toxicity in yeast and animal models. *Nat. Chem. Biol.* **5**, 655–663 (2009).
 26. G. Wang, A. R. Fersht, First-order rate-determining aggregation mechanism of p53 and its implications. *Proc. Natl. Acad. Sci. U.S.A.* **109**, 13590–13595 (2012).
 27. W. Kim, Y. Kim, J. Min, D. J. Kim, Y. T. Chang, M. H. Hecht, A high-throughput screen for compounds that inhibit aggregation of the Alzheimer's peptide. *ACS Chem. Biol.* **1**, 461–469 (2006).
 28. A. Espargaró, R. Sabate, S. Ventura, Thioflavin-S staining coupled to flow cytometry. A screening tool to detect in vivo protein aggregation. *Mol. Biosyst.* **8**, 2839–2844 (2012).
 29. E. Hellstrand, B. Boland, D. M. Walsh, S. Linse, Amyloid β -protein aggregation produces highly reproducible kinetic data and occurs by a two-phase process. *ACS Chem. Neurosci.* **1**, 13–18 (2010).
 30. J. Habchi, P. Arosio, M. Perni, A. R. Costa, M. Yagi-Utsumi, P. Joshi, S. Chia, S. I. Cohen, M. B. Müller, S. Linse, E. A. Nollen, C. M. Dobson, T. P. J. Knowles, M. Vendruscolo, 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).
 31. G. McColl, B. R. Roberts, T. L. Pukala, V. B. Kenche, C. M. Roberts, C. D. Link, T. M. Ryan, C. L. Masters, K. J. Barnham, A. I. Bush, R. A. Cherny, 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.* **7**, 57 (2012).
 32. J. Drake, C. D. Link, D. A. Butterfield, 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).
 33. E. Marsault, M. L. Peterson, Macrocycles are great cycles: Applications, opportunities, and challenges of synthetic macrocycles in drug discovery. *J. Med. Chem.* **54**, 1961–2004 (2011).
 34. P. Arosio, M. Vendruscolo, C. M. Dobson, T. P. J. Knowles, Chemical kinetics for drug discovery to combat protein aggregation diseases. *Trends Pharmacol. Sci.* **35**, 127–135 (2014).
 35. J. R. Frost, N. T. Jacob, L. J. Papa, A. E. Owens, R. Fasan, Ribosomal synthesis of macrocyclic peptides in vitro and in vivo mediated by genetically encoded aminothiol unnatural amino acids. *ACS Chem. Biol.* **10**, 1805–1816 (2015).
 36. N. Bionda, R. Fasan, Ribosomal synthesis of natural-product-like bicyclic peptides in *Escherichia coli*. *Chembiochem* **16**, 2011–2016 (2015).
 37. M. O. Maksimov, I. Pelczar, A. J. Link, Precursor-centric genome-mining approach for lasso peptide discovery. *Proc. Natl. Acad. Sci. U.S.A.* **109**, 15223–15228 (2012).
 38. T. Bi, Y. Li, A. Shekhtman, J. A. Camarero, In-cell production of a genetically-encoded library based on the θ -defensin RTD-1 using a bacterial expression system. *Bioorg. Med. Chem.* **26**, 1212–1219 (2018).
 39. J. Austin, W. Wang, S. Puttamadappa, A. Shekhtman, J. A. Camarero, Biosynthesis and biological screening of a genetically encoded library based on the cyclotide MCoTI-I. *Chembiochem* **10**, 2663–2670 (2009).
 40. P. G. Arnison, M. J. Bibb, G. Bierbaum, A. A. Bowers, T. S. Bugni, G. Bulaj, J. A. Camarero, D. J. Campopiano, G. L. Challis, J. Clardy, P. D. Cotter, D. J. Craik, M. Dawson, E. Dittmann, S. Donadio, P. C. Dorrestein, K. D. Entian, M. A. Fischbach, J. S. Garavelli, U. Goransson, C. W. Gruber, D. H. Haft, T. K. Hemscheidt, C. Hertweck, C. Hill, A. R. Horswill, M. Jaspars, W. L. Kelly, J. P. Klinman, O. P. Kuipers, A. J. Link, W. Liu, M. A. Marahiel, D. A. Mitchell, G. N. Moll, B. S. Moore, R. Muller, S. K. Nair, I. F. Nes, G. E. Norris, B. M. Olivera, H. Onaka, M. L. Patchett, J. Piel, M. J. Reaney, S. Rebuffat, R. P. Ross, H. G. Sahl, E. W. Schmidt, M. E. Selsted, K. Severinov, B. Shen, K. Sivonen, L. Smith, T. Stein, R. D. Sussmuth, J. R. Tagg, G. L. Tang, A. W. Truman, J. C. Vederas, C. T. Walsh, J. D. Walton, S. C. Wenzel, J. M. Willey, W. A. van der Donk, Ribosomally synthesized and post-translationally modified peptide natural products: Overview and recommendations for a universal nomenclature. *Nat. Prod. Rep.* **30**, 108–160 (2013).
 41. C. C. Liu, P. G. Schultz, Adding new chemistries to the genetic code. *Annu. Rev. Biochem.* **79**, 413–444 (2010).
 42. B. Halford, Breakthroughs with bar codes DNA-encoded libraries help pharma find drug leads. *Chem. Eng. News* **95**, 28–33 (2017).
 43. C. Soto, E. M. Sigurdsson, L. Morelli, R. A. Kumar, E. M. Castaño, B. Frangione, β -sheet breaker peptides inhibit fibrillogenesis in a rat brain model of amyloidosis: Implications for Alzheimer's therapy. *Nat. Med.* **4**, 822–826 (1998).
 44. S. A. Sievers, J. Karanikolas, H. W. Chang, A. Zhao, L. Jiang, O. Zirafi, J. T. Stevens, J. Münch, D. Baker, D. Eisenberg, Structure-based design of non-natural amino-acid inhibitors of amyloid fibril formation. *Nature* **475**, 96–100 (2011).
 45. P.-N. Cheng, C. Liu, M. Zhao, D. Eisenberg, J. S. Nowick, Amyloid β -sheet mimics that antagonize protein aggregation and reduce amyloid toxicity. *Nat. Chem.* **4**, 927–933 (2012).
 46. L. Fu, B. Niu, Z. Zhu, S. Wu, W. Li, CD-HIT: Accelerated for clustering the next-generation sequencing data. *Bioinformatics* **28**, 3150–3152 (2012).
 47. S. K. Dhandu, K. Vaughan, V. Schulten, A. Grifoni, D. Weiskopf, J. Sidney, B. Peters, A. Sette, Development of a novel clustering tool for linear peptide sequences. *Immunology* **155**, 331–345 (2018).
 48. M. Bastian, S. Heymann, M. Jacomy, Gephi: An open source software for exploring and manipulating networks. *Third Int. AAAI. Conf. Weblogs. Soc. Media*, 361–362 (2009).
 49. M. Girvan, M. E. J. Newman, Community structure in social and biological networks. *Proc. Natl. Acad. Sci. U.S.A.* **99**, 7821–7826 (2002).
 50. S. Brenner, The genetics of *Caenorhabditis elegans*. *Genetics* **77**, 71–94 (1974).
 51. M. Perni, P. K. Challa, J. B. Kirkegaard, R. Limbocker, M. Koopman, M. C. Hardenberg, P. Sormanni, T. Müller, K. L. Saar, L. W. Y. Roode, J. Habchi, G. Vecchi, N. Fernando, S. Casford, E. A. A. Nollen, M. Vendruscolo, C. M. Dobson, T. P. J. Knowles, Massively parallel C. *elegans* tracking provides multi-dimensional fingerprints for phenotypic discovery. *J. Neurosci. Methods* **306**, 57–67 (2018).
 52. J. Habchi, S. Chia, R. Limbocker, B. Mannini, M. Ahn, M. Perni, O. Hansson, P. Arosio, J. R. Kumita, P. K. Challa, S. I. Cohen, S. Linse, C. M. Dobson, T. P. J. Knowles, M. Vendruscolo, 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).
 53. D. M. Walsh, E. Thulin, A. M. Minogue, N. Gustavsson, E. Pang, D. B. Teplow, S. Linse, A facile method for expression and purification of the Alzheimer's disease-associated amyloid beta-peptide. *FEBS J.* **276**, 1266–1281 (2009).

Acknowledgments: Nematode strains used in this study were provided by the *Caenorhabditis* Genetics Center, supported by the National Center for Research Resources of the U.S. National Institutes of Health. We thank S. Casford (University of Cambridge, UK) for technical assistance with the in vivo experiments, V. Harokopos (Biomedical Sciences Research Center "Alexander Fleming," Athens, Greece) for performing the deep sequencing, and E. Pappou (National Hellenic Research Foundation, Athens, Greece) for performing the flow cytometric phenotypic measurements. G. Georgiou (University of Texas at Austin, USA) is acknowledged for facilitating the flow cytometric sorting experiments. **Funding:** This work has received funding from the following: (i) the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (Project "ProMiDis"; grant agreement no. 8199934); (ii) the project "STHENOS-b" (MIS 5002398), which is funded by the Operational Programme "Competitiveness, Entrepreneurship and Innovation" (NSRF 2014-2020) and cofinanced by Greece and the EU (European Regional Development Fund); (iii) the project "NEUROOTHERAPY" in the framework of the research grant "Aristeia"; and (iv) the project "STHENOS" in the framework of action "KRIPIS," the last two financed by the Hellenic General Secretariat of Research and Technology (GSRT) and the National Strategic Reference Framework (NSRF 2007-2013). We also acknowledge support by a short-term scientific mission grant from COST Action BM1405 to D.C.D. **Author contributions:** G.S. conceived and coordinated the project. G.S., D.C.D., J.H., N.C., and M.P. designed the research. D.C.D., S.C., I.M., and N.P. performed the research. D.C.D., M.P., S.C., N.P., and G.S. analyzed the data. G.S., M.V., J.H., C.M.D., and N.C. supervised the research. G.S. and D.C.D. wrote the paper. All authors read and approved the final manuscript. **Competing interests:** G.S. is inventor on the patent application PCT/IB2018/000622 describing aspects of the herein described technology. G.S. and D.C.D. are inventors on a patent application for A β C7-1, A β C7-14, and other A β -targeting peptide macrocycles described in this article, which is currently in submission to the Hellenic Industrial Property Organisation. G.S. and D.C.D. are founders and equity holders of ResQ Biotech P.C. The authors declare no other competing interests. **Data and materials availability statement:** All data needed to evaluate the conclusions in the paper are present in the paper and/or the Supplementary Materials. Additional data related to this paper may be requested from the authors.

Submitted 30 March 2019

Accepted 25 September 2019

Published 16 October 2019

10.1126/sciadv.aax5108

Citation: D. C. Delivoria, S. Chia, J. Habchi, M. Perni, I. Matis, N. Papaevgeniou, M. Reczko, N. Chondrogianni, C. M. Dobson, M. Vendruscolo, G. Skretas, Bacterial production and direct functional screening of expanded molecular libraries for discovering inhibitors of protein aggregation. *Sci. Adv.* **5**, eaax5108 (2019).