A Phenotypic Screen Identifies Cyclic Peptide Compounds that Induce Repeat Expansion in Yeast

An honors thesis for the Department of Biology

Nicholas G. Economos

Tufts University, 2013

ACKNOWLEDGMENTS

We'd like to thank the Kritzer Laboratory of the Tufts University chemistry department, specifically Dr. Joshua A Kritzer and Jason Gavenonis, for the *CFX6* cyclic peptide library and plasmids, and for their close collaboration on this project. Thank you to Alexander Shishkin for the SMY223/1 strain. We'd also like to thank the entire Mirkin Laboratory, especially Dr. Sergei M. Mirkin and Anna Aksenova, for their pivotal role in this project and for sharing their ideas at lab meetings.

Table of Contents

Abstract	**1**
Introduction	**2**
Materials and Methods	**8**
Results	**14**
Discussion	**25**
Conclusion	**29**
Appendices and Supplementary Information	**31**
References	**46**

Abstract

Expansion of repeat sequences in the eukaryotic genome can result in dangerous consequences often implicated in the pathology of severe human diseases. Out-of-context telomeric extension in immortalized cancer cell lines and expanded GAA repeat tracts in sufferers of the disease Friedreich's ataxia are two such examples. The mechanisms related to repeat expansion have been widely investigated, but are poorly understood and likely complex. To identify small molecules that may induce repeat expansion and to further understand these mechanisms, we developed a phenotypic screening system to identify small cyclic peptides that induce the expansion of tandem repeat sequences modeled in yeast (*Saccharomyces cerevisiae*) using a large intracellularly constructed cyclic peptide library. This paper presents the isolation and characterization of three cyclic peptides that demonstrate inverted human telomeric repeat expansion activity after screening an initial pool of 125,000 transformants. Further analyses reveal that all three of the isolated peptides additionally induce expansion in GAA repeats modeled in a similar system. This work showcases the efficient ability of our screening system to identify agents with repeat expanding properties and highlights three interesting molecules that induce instability in at least two different repeat sequences.

INTRODUCTION

Cyclic peptides represent a diverse class of compounds that feature a cyclized protein backbone as a result of linked N and C termini. They are prevalent in nature and feature a wide variety of structures, often with the aid of disulphide bridges and other side chain linkages, and functions, such as antimicrobial activity in plants (Craik, 2006). Due to their circular construct, cyclic peptides also demonstrate resistance to protease degradation and demonstrate remarkable stability *in vivo* (Satoh et al. 1996). Recently, biochemical methods have been developed for the synthesis and engineering of small species of these interesting structures to harness their inherent bioactive potency (Scott et al., 1999; Clark et al., 2005). The variety of function, small size, and remarkable stability of chemically synthesized cyclic peptides further identify them as potential structures for large combinatorial library production and coupling with functional screens for biological activity (Scott et al., 1999).

A previously used method for cyclic peptide production called SICLOPPS, or split-intein-mediated circular ligation of proteins and peptides, features *in vivo* synthesis of large peptide libraries (Scott et al., 1999). Cyclic peptides can be generated using a single genetic construct allowing for efficient intracellular production of large-scale combinatorial libraries. The initially translated large peptide product features C-terminal and N-terminal regions that spontaneously associate to form an active ligase domain. The folded peptide then mediates the cyclization of a small variable target region forming a head-to-tail ligated cyclic peptide that is excised within the cell. Initially developed for bacterial hosts, screening strategies have previously been able to select for dam methyltransferase and ClpXP protease inhibition activity, among others, from

large SICLOPPS libraries expressed in *Escherichia coli* (Naumann, Tavassoli, and Benkovic, 2008; Cheng et al., 2007). Recent advances have been able to adapt SICLOPPS methods to produce cyclic peptides within eukaryotic systems as well. Yeast-compatible SICLOPPS constructed libraries have been created and optimized for phenotypic screening, and have been used previously to efficiently select for α-synuclein toxicity reduction activity in a yeast model (Kritzer et al., 2009). Thus, screening SICLOPPS libraries in eukaryotic systems has the potential to rapidly identify cyclic peptides that induce a phenotype of interest. To explore the capabilities of these small active agents further, our laboratory is particularly interested in those cyclic peptides that may influence the stability of repetitive elements in the genome of yeast (*S. cerevisiae*) models.

Telomeric Repeat Sequences

Telomeres are dynamic and vital complexes of tandem noncoding repetitive DNA sequences, proteins, and RNA that play an essential role in the eukaryotic cell. Present in all eukaryotic organisms, telomeres lie at the ends of linear chromosomes protecting them from degradation. These terminal sequences further protect themselves with a complicated network of proteins and form t-loops to evade repair machinery recognition and minimize damage, such as gross chromosomal rearrangements (de Lange, 2005; reviewed in Greider, 1999). In humans, telomeres are composed of repeating TTAGGG hexanucleotide sequences and can span several kilobases in length (Moyzis et al., 1988). Lately, these structures have been subject to intense research, especially on their apparent relevance to cancer and cellular aging (Reviewed in Donate and Blasco, 2011). The proper function of telomeres is critical in all cells. Should a telomere shorten, become unprotected, or excessively lengthen, dangerous chromosomal instabilities may

result (Shay and Wright, 2011). Thus, by identifying factors related to the expansion and shortening of telomeres, researchers can gain useful insight into the mechanisms that affect these vital structures.

A particular factor of interest in telomere research is the associated ribonucleoprotein telomerase. Telomerase adds telomeric repeats to the ends of chromosomes using a short internal RNA template and maintains the length of telomeres by reverse transcriptase activity (Greider and Blackburn, 1985). The presence of this ribozyme has evolved out of an inherent problem that stems from the semi-conservative 5' to 3' replication mechanism during the replication of linear DNA molecules. In what is known as the "end-replication problem," DNA replication machinery cannot fully complete nascent lagging strand synthesis and a shortening event at the ends of chromosomes characterizes each round of replication (reviewed in Wynford-Thomas and Kipling, 1997). However, significant telomerase activity is only reserved for a small population of cells in humans. Normal somatic cells undergo gradual telomere shortening, a natural process that results in cell senescence and gives cells a characteristic lifespan (reviewed in Kenyon, 2010).

Telomerase creates significant problems when it becomes inappropriately activated in somatic cells. In a majority of cancerous tumor cells telomerase is active, thus the natural path to senescence is interrupted and cell lifespan is extended indefinitely (Kim et. al., 1994). The immortalization of a cell line through telomere length regulation via telomerase results in boundless division and has been attributed to approximately 85% of human cancers (Reviewed in Cesare and Reddel, 2010). The remaining 15% of cancers similarly increase cell life by maintaining telomere length.

This feat is accomplished using a different process known as alternative lengthening of telomeres (ALT). While the exact nature of ALT processes is not well understood, homologous recombination between telomeric sequences within the cell likely plays a major role in telomere extension in the absence of telomerase (Dunham et al., 2000). Ultimately, regardless of the mechanisms involved, telomeric sequence expansions, when out of context, play an instrumental role in oncogenesis.

While the role of telomeres in cancer is apparent, their applications can extend in a different direction, giving insight into the process of organism aging and aging-related disorders. Studies support the hypothesis that human cells have a lifetime, and this time is determined by the gradual shortening of their telomeres (Alssopp et. al., 1992). The average length of telomeres in cells decreases with age and strongly correlates with organism lifespan (reviewed in Kenyon, 2010). The disease Hutchinson-Gilford progeria syndrome shows increased attrition of telomeres, resulting in symptoms that mirror what appears to be an expedited lifetime (reviewed in Kenyon, 2010). Interestingly, fibroblasts from a Hutchinson-Gilford progeria syndrome cell line exhibiting these accelerated telomere attrition symptoms can be rescued by expressing the hTERT catalytic subunit of telomerase (Kudlow et al., 2008). These surprising results indicate an advantageous application of telomeric expansion, and a future direction for potential research on telomere related disease. Another telomere-related disease called Dyskeratosis congenita (DKC) is a progressive congenital disorder caused by inherited defects in the telomerase complex, and demonstrates symptoms that also strongly resemble aging. Autosomal dominant DKC is associated with mutations in an RNA

component of telomerase, hTERC, while X-linked DKC is caused by mutations in the gene for dyskerin, a protein implicated in telomerase function (Vulliami et. al., 2001).

Remarkably, telomeric repeat sequences are not limited to the ends of chromosomes. Interstitial telomeric sequences (ITSs) are prevalent throughout the human genome (Ruiz-Herrera, Nergadz, Sanragostino, and Guilotto, 2008). Short and long tracts of these internal repeat sequences have been identified, and lengths are highly polymorphic across individuals (Lin and Yan, 2008). Preliminary research shows that genomic instability may characterize these regions, though the implications of this have yet to be revealed (Mondello, Pirzio, Azzalin, and Giulotto, 2000). Interstitial telomeric sequences have not been thoroughly investigated, but their variable lengths and inherent instability make them an interesting candidate for potential future research on diseases characterized by chromosomal rearrangement, such as cancer and genetic disorders.

GAA Repeat Sequences

GAA trinucleotide repeats are common repeating structures that are prevalent throughout prokaryotic and eukaryotic genomes, especially in the case of the class Mammalia (Kassai-Jáger et al., 2008; Clark et al., 2004). These simple repeating sequences are inherently instable, and thus highly polymorphic within genomes, and demonstrate expansion activity in both proliferating and nonreplicating cells (Clark et al., 2004; De Baise et al., 2007; Ditch et al., 2009). The mechanisms responsible for this instability have been widely investigated. DNA replication, DNA repair, homologous recombination, and transcription processes are all potential mechanisms implicated in repeat expansion (reviewed in Mirkin, 2006; Ditch et al., 2009). Further, evidence supports that interplay between these processes and the propensity for GAA repeat

tracts to form unusual secondary structures play a central role in their instability (reviewed in McMurray, 1999; reviewed in Mirkin, 2006).

GAA repeat expansion is implicated in the hereditary autosomal recessive neurodegenerative disease Friedreich's ataxia (FRDA) (Campuzano et al., 1996). Afflicting an estimated 1 in 50,000 individuals, FRDA is the most common hereditary ataxia in humans (Harding, 1983). FRDA belongs to a much larger class of nucleotide repeat disorders, all of which can be characterized by trinucleotide repeat expansion events within the genome resulting in neurodegenerative symptoms (Bates and Lehrach, 1994). In the case of FRDA, GAA repeats are located within the intron of the FXN gene that codes for frataxin, an essential mitochondrial protein in humans (Campuzano et al., 1997). Intronic expansion and inactivation of the FXN gene results in a significant reduction in the expression frataxin, presumably inducing oxidative damage within the cell and causing cell death (Campuzano et al., 1997). There are currently very few effective therapies available for FRDA sufferers, and symptomatic relief is usually the focus of these approaches, rather than the genetic nature of the disease. Thus, there remains a limitation in our ability to treat FRDA due to the lack of a detailed understanding of the GAA expanding pathology of this disease in humans. To investigate potential mechanisms implicated in diseases characterized by repeat instability, our laboratory is interested in using repeats modeled in yeast to explore small molecules that may induce repeat expansion.

In this study, we developed a phenotypic screen for interstitial telomeric repeat expansion activity using a large yeast compatible cyclic peptide library expressed in *Saccharomyces cerevisiae*. We successfully isolated three peptides that stimulated

telomeric repeat expansions in yeast cells. Additionally, we located those amino acids within the eight-reside cyclic peptide candidates that were essential for activity. Remarkably, we also observed that all three peptide isolates from our initial screen induced GAA repeat expansions as well. These results showcase the impressive screening capabilities of large yeast-optimized cyclic peptide libraries, and an efficient means to identify factors that may induce repeat instability in yeast models. Further, we report three cyclic peptides confirmed to induce instability in at least two repeat sequences.

MATERIALS AND METHODS

Strains

The *S. cerevisiae* SMY735/2 strain (*MATa leu2-1 trp1-63 ura3-52 his3-200 ADE2-Int-(CCCTAA)14 ChrIII-(75423-75715)::URA3-Int-(CCCTAA)25*) was obtained from Anna Aksenova and used throughout the initial screen. This strain carries two selectable cassettes: an *ADE2* cassette, and a *URA3* cassette, with 14 and 25 inverted human telomeric repeats respectively located within an artificial intron. The *URA3* cassette is integrated near the *ARS306* locus on chromosome 3, while the *ADE2* cassette resides in the natural location of the *ADE2* gene on the fifteenth chromosome.

Further investigation of cyclic peptide effects on telomeric repeats was conducted using the *S. cerevisiae* SMY717/1 strain (*MATa leu2-1 trp1-63 ura3-52 his3-200 ChrIII-(75423-75715)::URA3-Int-(CCCTAA)25*) obtained from Anna Aksenova. Only the *URA3* cassette, artificially intronated with 25 inverted telomeric repeats, was incorporated in this strain.

Analysis of cyclic peptide effects on GAA repeats was conducted using *S. cerevisiae* strain SMY223/1 (*MATa leu2-1 trp1-63 ura3-52 his3-200 Chr1585 III-(75423-75715)::URA3-Int+tet269-GAA100d*) obtained from Alexander Shishkin. This strain features a *URA3*, T269 balanced cassette integrated at the same location described above, and features a $(GAA)_{100}$ artificial intron (Shishkin et al., 2009).

Media

Yeast complete minus Leu media (Leu-) was prepared using U.S. Biological Leu-, Ura- dropout mix and was supplied with powdered uracil (Sigma-Aldrich) before autoclaving (Amberg, Burke, & Strathern, 2005). 5-FOA Leu- media contains 0.1% 5-FOA (5-fluoroocitic acid) and was prepared using Leu-, Ura- dropout mix supplied with powdered uracil (Sigma-Aldrich) (Amberg, Burke, & Strathern, 2005).

Cyclic Peptide Plasmid Library

A yeast-compatible SICLOPPS-based cyclic peptide library was used in this experiment (Kritzer et al., 2009). We used SICLOPPS constructs expressed from a high-copy 2μ pRS425 plasmid under a constitutive *GPD* promoter. Our plasmid library, entitled *CFX6,* was acquired from Joshua A Kritzer of the Tufts University chemistry department. The cyclic peptide library was generated from different amino acid combinations composing an 8-unit peptide, where 6 positions are variable with conserved C1 and F2 residues. Thus, a total of $6.4x10^7$ are possible. All cyclic peptide plasmids described contain a yeast-selectable *LEU2* marker for transformant selection.

Alanine Substituted Cyclic Peptide Plasmids

Plasmids featuring point-mutated versions of CP283, CP339, and CP382 were provided by Jason Gavenonis from the Kritzer labortory. Eight plasmids were provided for each

peptide, each corresponding to a single amino acid residue changed to alanine (with the exception of the A8 of CP339 and A3 of CP382, which were changed to glycine). Additionally, a plasmid was provided for each peptide that featured the conserved C1 residue replaced with a similarly nucelophilic serine. By convention, alanine-scanning mutant peptides are described by a letter-number-letter motif that represents the wild type amino acid, its position, and the residue that is used to replace it (e.g. F2A, A8G).

Uncyclized Mutant Peptides

Uncyclized "TH" intein-disabled peptide plasmids for CP283, CP339, and CP382 were provided by Jason Gavenonis from the Kritzer labortory. Mutations of two amino acid residues to alanine in the intein C-terminal region (T69A, and H72A) inhibit the cyclization step in the SICLOPPS process (Kritzer et al, 2009). Intein-disabled versions of cyclic peptides still feature the association of the intein halves, but fail to complete the head-to-tail ligation event required to fully process the peptide.

Yeast Transformation

The protocol used to transform yeast cells was a modified Gietz protocol (See Appendix A in supplement for full protocol).

DNA Isolation

Frozen stock transformants expressing cyclic peptides constructs were obtained from Anna Aksenova. For DNA isolation protocol refer to Appendix B in supplement.

Bacterial Transformation

The standard bacterial transformation procedure as described in *Molecular Cloning* (1989) was used (Sambrok, Fritsch, & Maniatis, 1989). Chemically competent bacterial cells of the XL10-GOLD *E. coli* strain were obtained from Anna Aksenova.

Plasmid Extraction
In order to extract cloned plasmids from bacteria (*E. coli*) we used a Miniprep DNA isolation kit (Zymo Research) to extract DNA from transformed bacterial colonies.

Phenotype Analysis for Screen
After yeast transformation onto Leu-, colony color was visually analyzed and recorded. At least six regular transformants from the plates were streaked onto a 5-FOA Leu- media plate. Plates were analyzed for elevated colony frequencies after three or four days of growth.

Spotting Assays
To easily observe and compare the relative frequencies of 5-FOA resistant SMY717/1 colonies transformed with alanine-scanning mutants, colonies were resuspended in a total volume of 200μL and 10-15μL of the dilutions were plated onto 5-FOA Leu- and Leu- media. Single spots were evaluated after three days of growth.

Fluctuation Testing
To measure expansion and inactivation rates, transformants were plated on Leu- selective media and were grown for 4 days at 30°C. Six separate randomly chosen colonies, and in the case of GAA repeat expansion analysis 12 colonies, were diluted into 200μL of water and then serially diluted 10-fold to yield 10, 100, 1000, 10000, and 100000-fold dilutions. 100000-fold colony dilutions were plated on Leu- media as an initial assessment of growth and to estimate the total number of cells in the initial dilution. SMY717/1 colonies transformed with alanine-scanning mutants featured varied dilutions ideal for plating on 5-FOA Leu- media, depending on the nature of the replaced amino acid and its role in peptide function. Thus, dilutions were adjusted for each

mutant peptide to give an ideal number of colonies for analysis on 5-FOA Leu- media. 1000-fold dilutions were plated on Leu-5-FOA media for SMY223/1 colonies transformed with wild-type cyclic peptide plasmids, while 10-fold dilutions were plated for those colonies transformed with TH intein-disabled constructs.

Plated dilutions were allowed to grow at 30°C for 5 days and colony counts on each plate were recorded. Four randomly chosen colonies (or as many as could be taken from the plate in some cases with plates yielding less than four colonies) were then selected from each 5-FOA Leu- plate and subjected to PCR analysis to determine if any expansion events had occurred.

Mutation rates for expansion and gene inactivation events were calculated using the MSS Maximum Likelihood Method with the Fluctuation AnaLysis CalculatOR (FALCOR) web tool (Sarkar, Ma, and Sandri, 1992; Ma, Sandri, and Sakar, 1992; Hall et al., 2009). 95% confidence intervals were determined by FALCOR using methods described by Rosche and Foster (2000). Final rates represented in data were calculated and adjusted using an unpublished method by Anna Aksenova.

PCR

Artificial introns within our cassettes were cloned and amplified using appropriate primers that flanked the region of interest and run with an optimized PCR program (See Fig. 1; see Tables 1and 2). The PCR mix for programs C, D and E contained 5x Green Buffer with 7.5mM $MgCl_2$, pH8.5 (Promega), 25 pmols of each of the two primers used, and 1.25 units (0.25µL) of Taq polymerase (SibEnzyme). The PCR mix for programs A and B contained 10x Taq Buffer with $(NH_4)_2SO_4$ and 20mM $MgCl_2$, pH8.8 (Fermentas),

25 pmols of each of the two primers used, and 1.25 units (0.25μL) of Taq polymerase (SibEnzyme)

Table 1. Primers used with corresponding PCR programs, strains, and cassettes.

Primer Name	**Sequence**	**PCR Program**	**Strain/ Cassette**
ADIN3	CTTATTAACGAAATTGCCCCAAGGGTATGT	A	SMY735/2 *ADE2*
ADINSPY3	ATCTCTCGAGCAATTCTAGAT	A	SMY735/2 *ADE2*
BHTUNI	CTCGACCATGATCATCG	B	SMY735/2 *URA3*
BHT1OR2	GAGAAATCTCTCGCCTCG	B	SMY735/2 *URA3*
UIRL1	GAAGTAACAAAGGAACCTAGAGGGTA	C	SMY735/2 *URA3*
UIRL2	TACAGATCAGTCAATATAGGAGGTT	C	SMY735/2 *URA3*
A2	CTCGATGTGCAGAACCTGAAGCTTGATCT	D	SMY223/1 *URA3*
B2	GCTCGAGTGCAGACCTCAAATTCGATGA	D	SMY223/1 *URA3*
108F	(XXXX)	E	SMY717/1 *URA3*
108R	(XXXX)	E	SMY717/1 *URA3*

Table 2. List of PCR programs used.

PCR Program	Initial Step:	Step 1:	Step 2:	Step 3:	Number of Cycles (1-3)
A	95°C - 15”	94°C - 20”	58°C - 15”	72°C - 30”	30
B	95°C - 15”	94°C - 15”	57°C - 15”	72°C - 45”	30
C	95°C - 15”	94°C - 15”	64°C - 15”	72°C - 45”	30
D	95°C - 10”	94°C - 10”	62°C - 10”	72°C - 45”	32
E	95°C - 10”	94°C - 10”	70°C - 10”	72°C - 50”	32

Gel Electrophoresis

To separate the DNA fragments obtained by PCR, a 2.0% agarose gel in TAE buffer was run at approximately 5V/cm. Fragments were visualized by post-staining with ethidium bromide and UV (255nm). The relative distances traveled by fragments

compared to the unexpanded original strain fragments were used to determine the occurrence of any expansion events within the repeat introns in our cassettes.

RESULTS

A Large Cyclic Peptide Library Expressed in Yeast

The purpose of this experiment was to efficiently screen a large yeast-optimized cyclic peptide library for species that induce repeat expansion in a *S. cerevisiae* model. Using the eight-residue combinatorial *CFX6* SICLOPPS library, we sought to identify cyclic peptides that demonstrate this target activity by expressing constructs *in vivo* and using a phenotypic screening system.

The Selective System

To identify repeat expansion activity in yeast we utilized two selective cassettes (*URA3* and *ADE2*) with artificial introns featuring tandem repeat sequences integrated into the genome of the *S. cerevisiae* strain SMY735/2 (See Fig. 1a,b). Significant repeat expansion in either intron results in a gene inactivation event. In the case of repeat expansion within *ADE2*, gene inactivation results in the accumulation of red pigment. Intronic repeat expansion within the *URA3* cassette renders affected colonies resistant to the drug 5-FOA, and thus 5-FOA resistance rates can be used as a measure of expansion rate. By using this dual-selection system, we could effectively identify and quantify repeat expansion events in our model system.

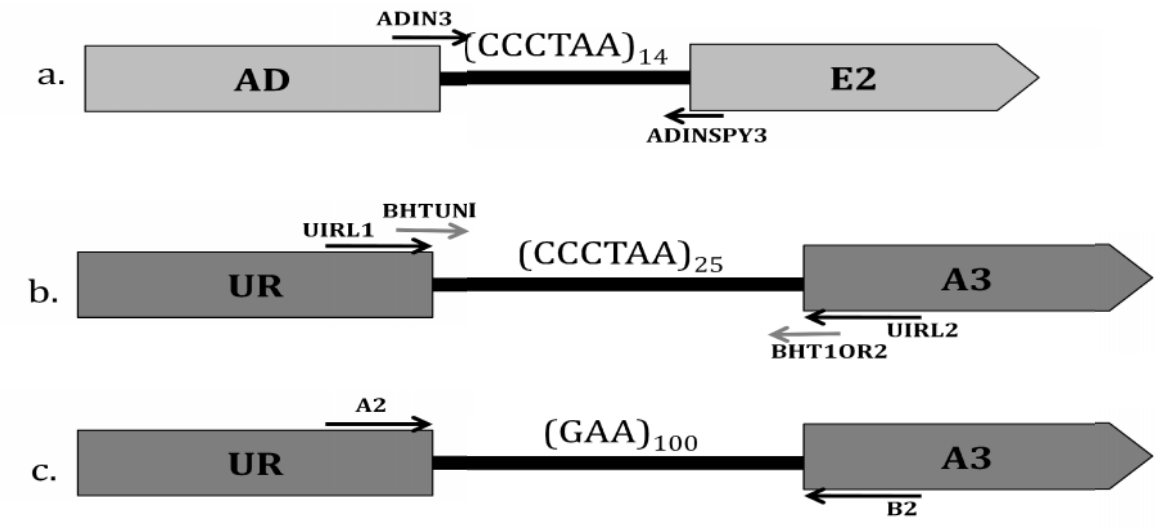


Figure 1a-c. Selectable cassettes used to screen for repeat expansion in *S. cerevisiae*. Repeat sequences are labeled and are located within an artificial intron in the cassette. Primer locations used for PCR analysis are shown.

Library Screen for Cyclic Peptides That Affect Human Telomeric Repeat Stability

Our screening experiment focused on the use of the yeast-optimized *CFX6* SICLOPPS cyclic peptide library and our phenotypic selectable system to investigate the potential effects of cyclic peptides on inverted telomeric sequence expansion. Initially, Anna Aksenova obtained *CFX6* cyclic peptide transformants of the SMY735/2 strain. Cyclic peptide plasmids featured a selectable *LEU2* marker for selection of transformants of our *leu2-* strains on media lacking leucine. In total, ~1.2×10^5 transformants were yielded on Leu- media. As a control, ~2.4×10^4 transformants of CPU-Q120, a control cyclic peptide featuring an inactive HPQ motif, were analyzed. Anna Aksenova isolated approximately 400 transformants that showed a color change from white to pink or red. Among the 400 selected transformants, 73 demonstrated elevated frequencies of 5-FOA resistant colonies on 5-FOA Leu- media. These 73 plasmids were subjected to further analysis by extracting plasmid DNA from transformants expressing selected cyclic peptides, retransformation into the original

SMY735/2, and stringent selection for color and elevated colony frequency on 5-FOA containing media (See scheme in Fig 2.) Control CPU-120 transformants did not express the selective phenotype.

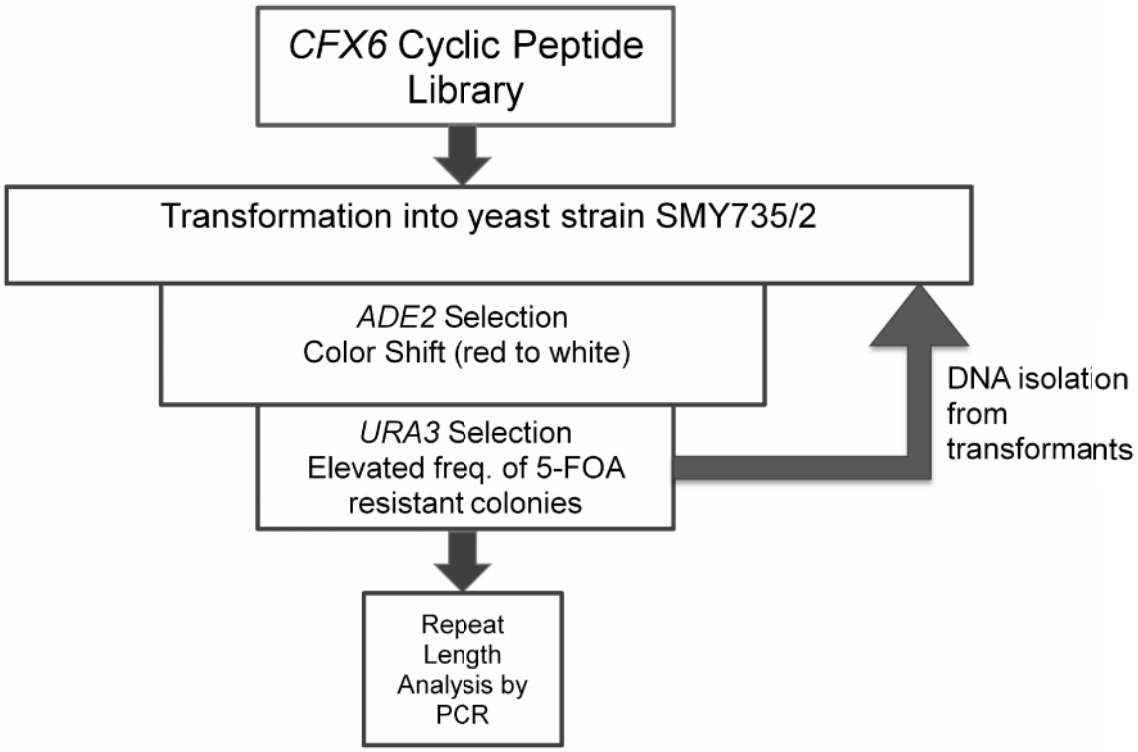
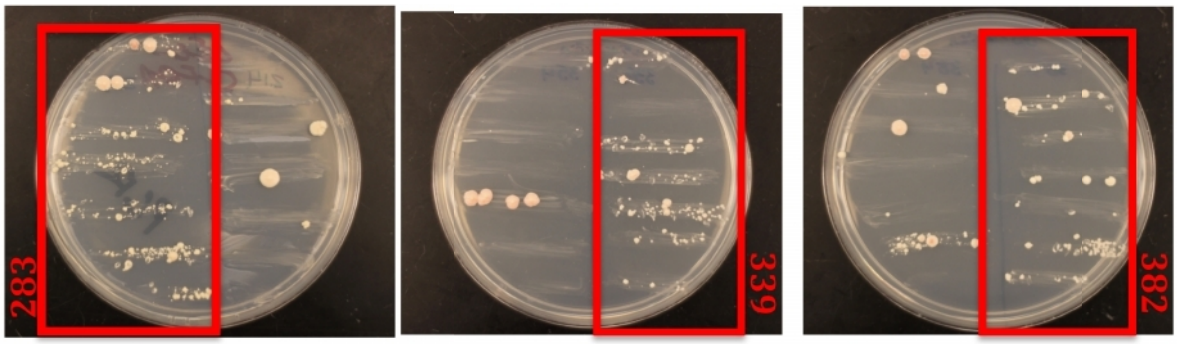


Figure 2. Scheme of the cyclic peptide library screen. Flow chart describes the scheme of our screen used to find peptides with telomeric repeat expanding properties from the *CFX6* library.

Three Cyclic Peptides Identified in Screen Affect Telomeric Repeat Stability

Ultimately, of the 73 transformants analyzed, 70 exhibited a reddened color, one (CP5) demonstrated growth on 5-FOA Leu- media, and 3 (CP283, CP339, and CP382) showed consistent elevated frequencies of 5-FOA resistant colonies on 5-FOA Leu- media (See Fig. 3). CP5, CP283, CP339, and CP382 transformants were all pink or red on Leu- media.

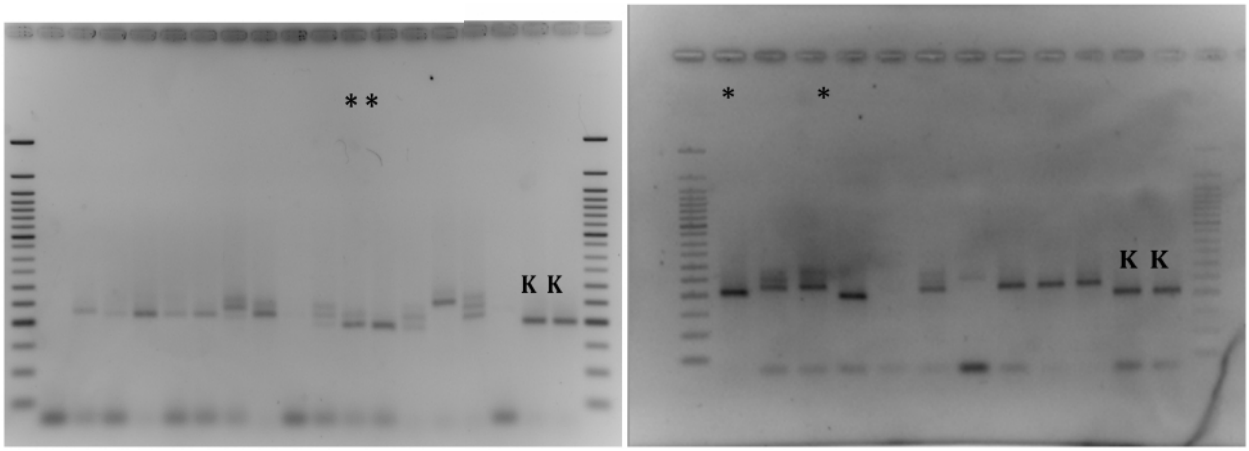


a. b. c.

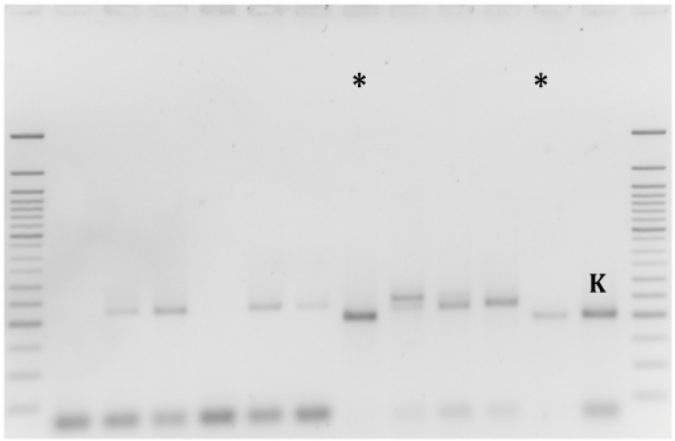
Figures 3a-c. Cyclic peptide CP283, CP339, and CP382 transformant phenotypes on Leu-5-FOA

PCR Analysis of Expansions:

Results from initial screening revealed that the *URA3* cassette was more selective for cyclic peptide activity than *ADE2*, and thus we chose to analyze repeats within the intron of the *URA3* cassette as a means for comparing repeat lengths in transformants. Anna Aksenova's earlier PCR analysis of the CP5 transformant suggested that the observed phenotypes were not the result of repeat expansion. PCR analysis of the *URA3* intron in CP283, CP339, and CP382 transformants confirmed the presence of $(CCCATT)_{25}$ repeat sequence expansions (see Figures 4a-c). Of the 5-FOA resistant colonies analyzed, 86% of CP283 transformants, 80% of CP339 transformants, and 77.7% of CP382 transformants demonstrated expanded $(CCCTAA)_{25}$ repeat introns. The remaining analyzed colonies demonstrated 5-FOA resistance in the absence of an expansion event. Plasmids for CP283, CP339, and CP382 were sequenced and their amino acid sequences were determined (See Table 3).



5a. 5b.



5c.

Figures 4a-c. PCR analysis of *URA3* introns in CP283, CP339, and CP382 transformants. (a) *URA3* repeat length analysis of CP283 transformants (b) *URA3* repeat length analysis of CP339 transformants, (c) *URA3* repeat length analysis of CP382 transformants. Lanes denoted with a "K" are control unexpanded introns from the original SMY735/2 strain. Unexpanded introns observed in 5-FOA resistant transformants are denoted by an asterisk. Primers BHTUNI and BHT1OR2 were used.

Table 3. Sequencing results for plasmids for CP283, CP339, and CP382. Nucleotide sequences represent the region of the plasmid that codes for the processed cyclic peptide. Amino acid sequences were determined from the nucleotide sequence. Information was provided by the Kritzer laboratory.

Cyclic Peptide	Nucleotide Sequence	Amino Acid Sequence
283	TGTTTCTACGTGTGGGAGACCGGG	CFYVWETG
339	TGTTTCCACCTCGTCTACTGGGCG	CFHLVYWA
382	TGTTTCGCGACCTACTGCTTCTAC	CFATYCFY

Investigation of Cyclic Peptide Structure-Activity Relationship by Point Mutagenesis

All analyses of cyclic peptide structure and function after the initial screen focused exclusively on the 5-FOA resistant phenotype as a means for identification and quantification of repeat expansion due to the more selective nature of the *URA3* cassette. To elucidate which amino acids within the eight-unit constructs were required for peptide function, alanine-scanning mutant peptide transformants were spotted on 5-FOA Leu- selective media to quickly identify transformants that lost the ability to induce a resistant phenotype (Fig. 5). Such a phenotype conversion event could presumably be attributed to the alteration of a cyclic peptide property that is critical to its function. Methodic substitution of individual peptide residues allowed each side chain to be analyzed. Uncyclized, intein-disabled TH mutant transformants were analyzed as a control. TH mutant peptides fail to complete the head-to-tail ligation step of SICLOPPS mediated peptide construction. Thus, those peptides carrying this detrimental mutation should no longer render transformants 5-FOA resistant by expansion activity.

A reduction in 5-FOA resistant colony frequency, as compared to the wild type peptide, was observed for mutants F2A, Y3A, W5A, and E6A of CP283. F2A, H3A, L4A, Y6A, and W7A of CP339, and F2A, Y5A, F7A, and Y8A of CP382 also demonstrate a similar reduction in 5-FOA resistance. These results suggest that these substituted residues are critical for the full functioning of their respective peptides. In some cases, however, comparison between spot phenotypes were difficult to analyze and thus served as a rough estimate for the relative expansion abilities of mutant peptides. TH intein-disabled mutants behaved as expected and resulted in a strong reduction in 5-FOA resistance suggesting the spliced cyclic nature of the cyclic peptides was

essential. Notably, the *CFX6* cyclic peptide library featured a conserved nucleophilic C1 residue that was mechanistically required for the *in vivo* cyclization of SICLOPPS constructs. The observed elevated frequencies of 5-FOA resistant colonies in C1A mutants of all three peptides were a totally unexpected result.

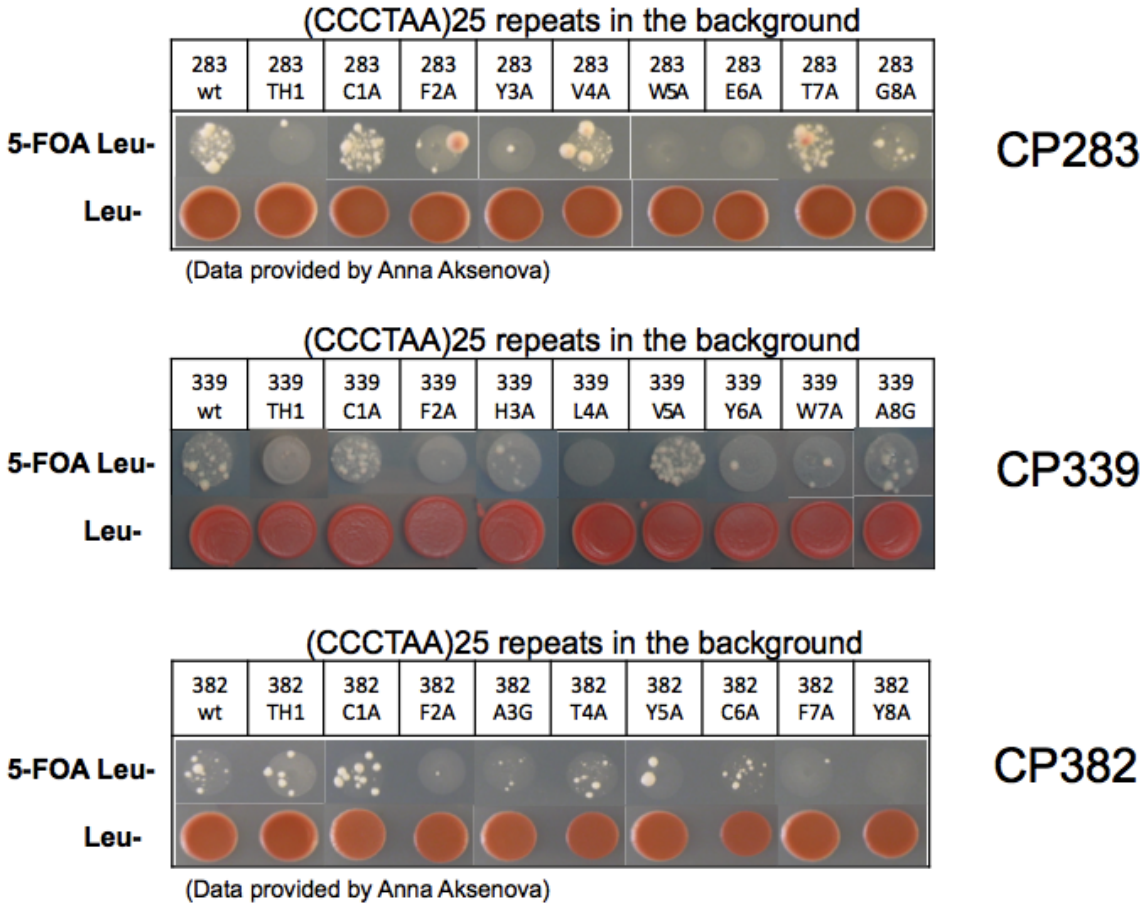


Figure 5. Spot assays of alanine-scanning peptide transformants for CP283, CP339, and CP382. Wild-type and intein-disabled (TH) transformant assays are provided for comparison. CP283 and CP382 assays were conducted by Anna Aksenova.

In order to further understand and quantify the relative expansion inducing activities of mutant peptides, fluctuation testing was conducted on the alanine-scanning peptide transformants for CP283 and CP339 (Fig. 6, 7). Fluctuation tests were not full scale, but rather six colony assays to quickly reveal the nature of the role of side chains

in the function of cyclic peptide isolates. We predicted activity abolishment in C1A mutants, a control replacement of the mechanistically essential C1 residue. Further, C1S mutants, which feature a similar nucleophilic residue to the native side chain, were also analyzed to further asses the required nucleophilic nature of the first position one side chain.

Expansion rate comparisons for CP283 mutants revealed that residues Y3, W5, and E6 demonstrated expansion rates similar to the empty vector control (PRS425) when substituted with alanine, suggesting that these residues were essential for CP283 repeat expansion activity. Further, the CP283 F2A mutant transformant demonstrated an order of magnitude reduction in expansion rate compared to the wild-type peptide. Thus, while the F2 residue clearly played an important role in peptide function, it was not absolutely required for activity. Unexpectedly, and similar to the result observed in the spot assays, mutation of the essential position one nucleophilic residue did not abolish peptide activity, and expansion rates similar to the wild-type peptide were observed. High C1S mutant expansion rates suggested that the C1 residue of the CP283 may be readily replaced by a similarly nucleophilic serine, but the unexpected results of the C1A mutant analyses make the structure-activity-relationship of the position one residue unclear.

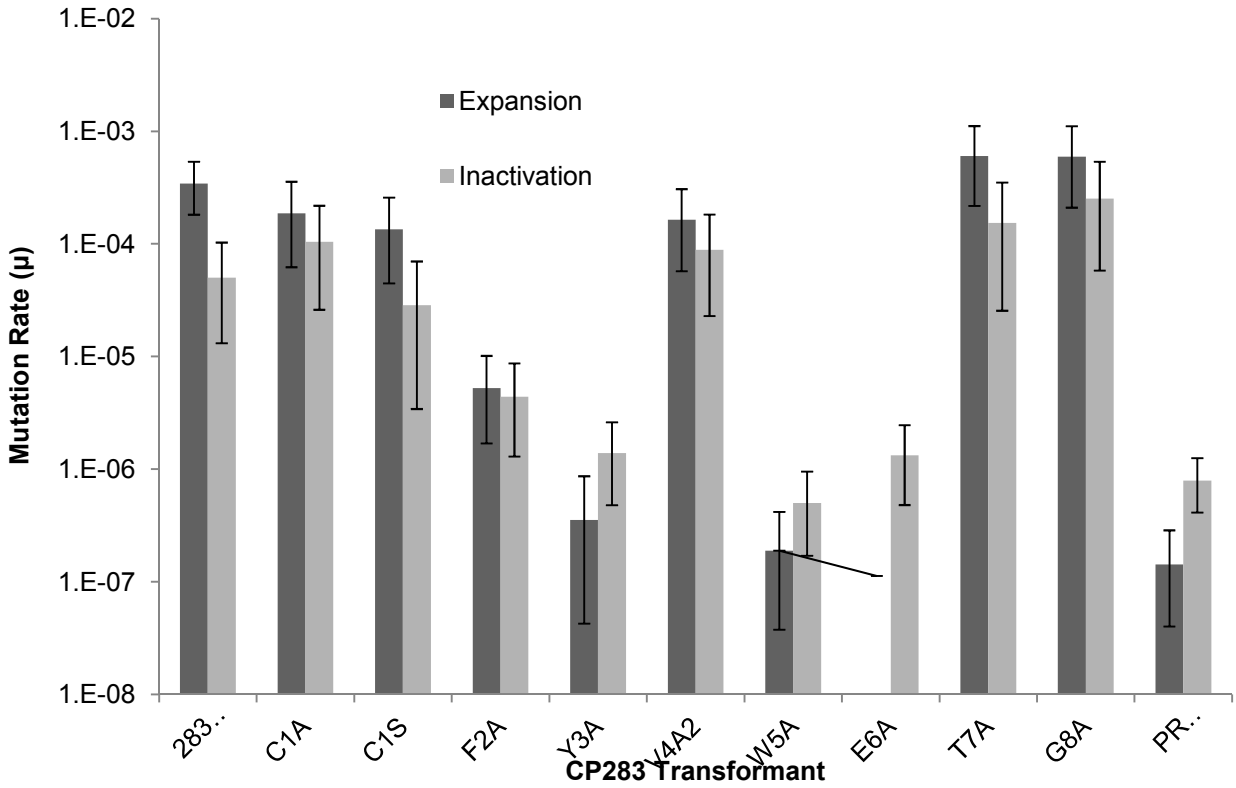


Figure 6. Rates of $(CCCTAA)_{25}$ expansion, and *URA3* inactivation in the absence of expansion, for CP283 alanine-scanning mutant peptides. PRS425 is the empty vector used as a negative control. Values are the median mutation rates determined by fluctuation tests of six cultures. Error bars represent 95% confidence intervals. 283wt data was provided by Anna Aksenova. See also Tables S1-S11, S30, and S31 in the supplement.

In the case of CP339, only the W7 side chain was essential to the expansion inducing function of the cyclic peptide. Mutation of residues F2, H3, L4, and Y6 resulted in expansion rate drops over an order of magnitude less that the wild-type peptide. These residues, however, did not completely eliminate expansion-inducing activity, and thus played an important, yet individually nonessential role in peptide function. Finally, mutant transformants C1A, V5A, and A8G all showed significantly lower expansion rates than the wild type peptide on a scale less than an order of magnitude, indicating a

real, but weak role for these side chains. C1 analyses yielded similar results to CP283 described above.

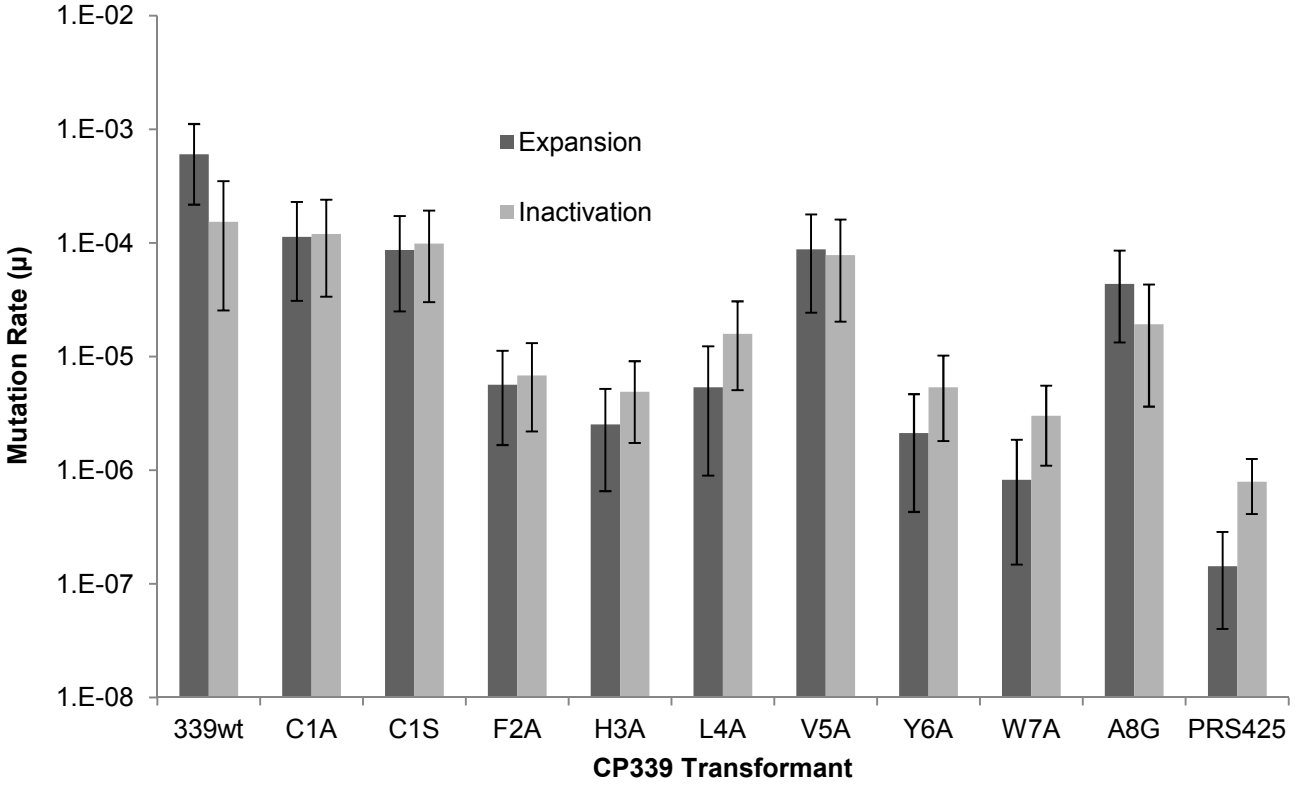


Figure 7. Rates of $(CCCTAA)_{25}$ expansion, and *URA3* inactivation in the absence of expansion, for CP339 alanine-scanning mutant peptides and wild-type peptide. PRS425 is the empty vector used as a negative control. Values are the median mutation rates determined by fluctuation tests of six cultures. Error bars represent 95% confidence intervals. See also Tables S12-S22, S30, and S31 in the supplement

Cyclic Peptides Induce GAA Repeat Instability

With three cyclic peptides in hand, we turned our sights to GAA repeats to further investigate and characterize the repeat expanding nature of these molecules. To observe any potential effects of the cyclic peptide isolates on $(GAA)_{100}$ repeats within the *URA3* cassette (See Fig. 1c), fluctuation testing and expansion analyses by PCR were conducted on CP283, CP339, and CP382 transformants as well as their

respective TH mutants (described above) in the *S. cerevisiae* SMY223/1 strain. The selective cassette used in this part of the experiment was identical to the *URA3* system used in earlier analyses, except a $(GAA)_{100}$ artificial intron was present. Wild-type CP283, CP339, and CP382 transformants demonstrated $(GAA)_{100}$ expansion rates significantly higher than their intein-disabled counterparts (Fig. 8). The TH intein-disabled mutant transformants did not show expansion rates significantly elevated beyond the values for the empty vector, as expected. Thus, the cyclized, spliced nature of the expressed cyclic peptides is imperative for GAA repeat expansion to occur.

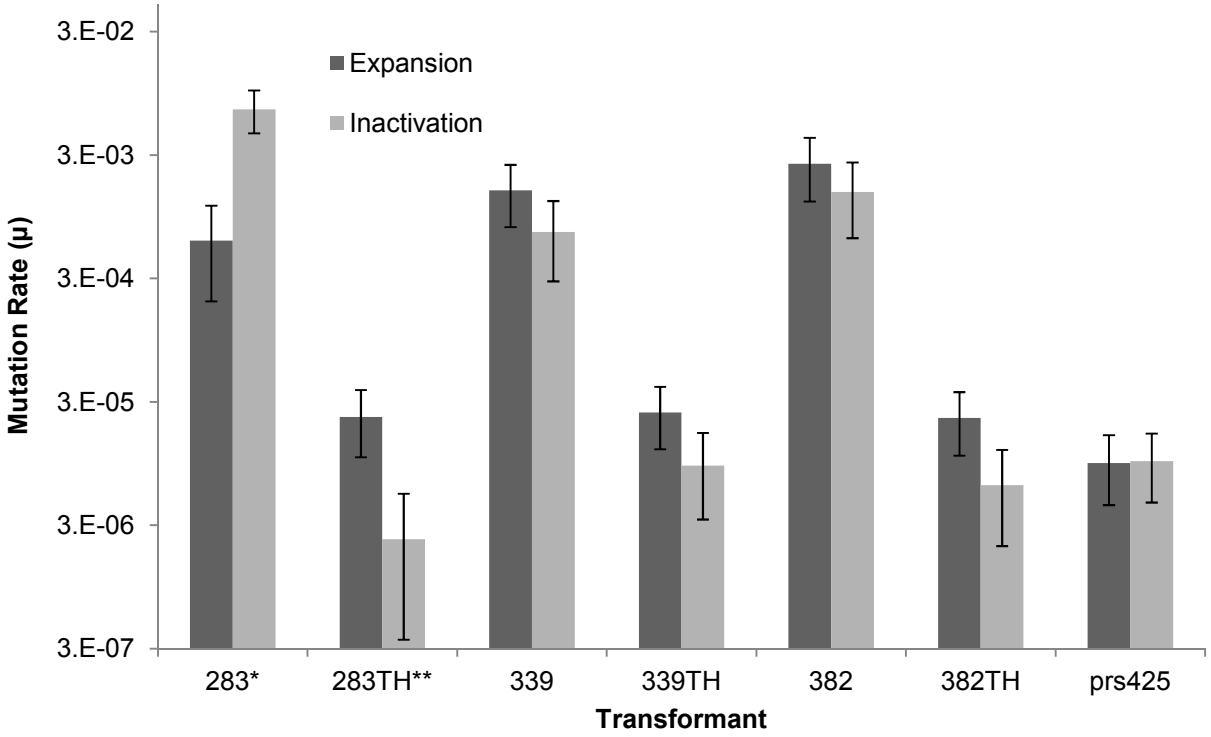


Figure 8. Rates of $(GAA)_{100}$ expansion and *URA3* inactivation in the absence of expansion for cyclic peptides and TH mutants. prs425 is the empty vector used as a negative control. Values are the median mutation rates determined by fluctuation tests of twelve cultures. Error bars represent 95% confidence intervals. 283 represents combined data with Anna Aksenova. 283TH data was provided by Anna Aksenova. See also Tables S23-29, S32 and S33.

Ultimately, using a large intracellularly produced cyclic peptide library expressed in yeast and a phenotypic dual-selection system, we successfully isolated three cyclic peptides capable of expanding inverted human telomeric repeat sequenced in our model. We have determined that the hydrophobic aromatic nature is very important to peptide function, and in the case of CP283 and CP339 so is the presence of a charged side chain. We further demonstrated that all three of these cyclic peptides are capable of expanding GAA tandem repeats in a yeast model as well. It should also be noted that *URA3* inactivation rates in the absence of repeat expansion were also measured and appear to correlate closely with expansion rates.

DISCUSSION

In this work, we developed an efficient phenotypic screen for interstitial telomeric repeat expansion activity using a yeast-expressed (*S. cerevisiae*) cyclic peptide library and cassettes that selectively expressed a target phenotype as a result of repeat expansion within an artificial intron. Visual assessment of colony phenotypes allowed for a time-efficient and unambiguous screening process. Ultimately, we isolated three peptides that cause inverted human telomeric sequences expansion.

Additionally, we took advantage of point mutated alanine-scanning peptides to efficiently assay for functionally important residues within the eight-unit constructs and determined the structure-activity-relationship of the interaction. We determined that aromatic side chains are important to the function of all three peptide candidates, and in the case of CP283 and CP339, demonstrated that some aromatic residues are absolutely essential for peptide function. A hydrophobic side chain, L4 of CP339, was

also determined to be important. Additionally, the charged E6 residue of CP283 was required for activity while the charged H3 side chain of CP339 played a significant role in peptide function. No polar residues (with the exception of the conserved C1 residue) were featured in any of the isolated cyclic peptide constructs.

We employed two control methods to ensure that the observed effects exclusively resulted from cyclic peptide activity, one of which yielded an unexpected result. TH intein-disabling mutations in cyclic peptides resulted in activity abolishment as expected. C1A mutants, however, induced a repeat expansion phenotype similar to that of the wild-type peptide for all three peptides. Mutation of the essential position one nucleophilic residue should have resulted in a phenotype similar to that of transformants expressing TH disabled and empty vector constructs. Sequence analyses of these plasmids did not reveal any errors in the C1A genetic constructs. It should be noted that this result was not observed in screens conducted by Kritzer et al. (2009) using a similar yeast-compatible SICLOPPS library in *S. cerevisiae*. The only explanation we have, albeit highly speculative, is that another nucleophilic residue within the full SICLOPPS construct can be used to complete the cyclization and splicing processes, resulting in a larger active cyclic peptide.

Finally, we also observed significant GAA repeat expansion activity as a result of all three isolates in an adapted version of the model system. We determined that this effect was dependent on the presence of the spliced cyclized peptide and not due to other elements in the *CFX6* library SICLOPPS constructs.

Interestingly, although the cyclic isolates are responsible for significant repeat expansion events, in many cases they demonstrated *URA3* gene inactivation activity as

well. This secondary effect can be seen in 5-FOA resistant colonies that express a cyclic peptide in the absence of repeat expansion. Notably, the rates of these inactivation events appear to correlate with expansion rates and thus may be related to the mode of action of the cyclic peptides. These events may be the result of mutation or epigenetic inactivation, perhaps due to peptide function.

To further understand the means by which these peptides function we hope to identify a functional target. While we cannot exclude the possibility of a peptide-nucleic acid interaction, results from previous studies screening for cyclic peptide activities suggest that protein-protein interactions are the most likely explanation for these observations (Naumann, Tavassoli, and Benkovic, 2008; Cheng et al., 2007). The most efficient means to reveal peptide-protein interactions taking place in the system would be affinity-based techniques.

All three cyclic peptides isolated from the screening experiment demonstrated the remarkable ability to induce instability within inverted telomeric and GAA repeat sequences in the system. These interesting findings suggest certain factors as promising target candidates for investigation. For example, in a genome-wide screen for pathways that cause GAA repeat instability in yeast conducted by Zhang et al. (2012), 22 genes were identified to result in large-scale GAA repeat expansion after deletion, doxycycline regulation, or mRNA perturbation. Surprisingly, two mutant genes implicated in telomeric maintenance, TET-*TEN1* and TET-*CDC13*, were indentified in this screen and induced large-scale GAA repeat expansions. Along with Stn1, Cdc13 and Ten1 are sequence specific factors that recognize and stabilize telomeres, and associate with the single stranded 3’ overhang that characterizes the ends of linear

chromosomes in yeast (Paschini, Mandell and Lundblad, 2010; Lin and Zakian, 1996). The roles of Ten1 and Cdc13 association or interaction with interstitial repeats have not yet been investigated. Interestingly, both telomeric and GAA repeat sequences form characteristic secondary DNA structures *in vivo*. GAA tracts have been shown to form triplexes, while telomeres form t-loops that feature a single stranded 3' overhang that folds back to invade the double stranded telomere structure (reviewed in McMurray 1999; reviewed in Greider 1999). T-loops, however, have yet to be identified in yeast or explored in the context of interstitial telomeric repeats. Although any proposed mechanisms of cyclic peptide action remain speculative, interruption of the Ten1-Stn1-Cdc13 complex and its potential endogenous interactions with GAA and telomeric repeats, or their related secondary structures, remains an interesting possibility that warrants further investigation. We hope to investigate any effects the peptides may have on the telomeres of yeast cells to further explore the possibility of telomere implicated factors as potential targets. We also hope to further evaluate the screening strains described by Zhang et al. (2012), particularly the TET-*TEN1* and TET-*CDC13* strains, in hopes of identifying a target.

Additionally, the prevalence of aromatic residues in the cyclic peptides and the presence of the positively charged histidine side chain in CP339 suggest a potential DNA-peptide interaction may be taking place. Although such an interaction has yet to be observed as a result of cyclic peptide activity screens, we cannot exclude this possibility.

Ultimately, this work presents a novel tool for identifying factors that can affect the stability of tandem repeat sequences within a yeast model. The system design

overcomes difficulties that can plague researchers attempting to investigate repeat expansion *in vivo*. Intracellular production of peptides (SICLOPPS) features an easy method for introduction into a yeast model by simple plasmid transformation. The effects of these potent molecules are then easily discriminated by the selective cassette phenotypes, a feature that results in remarkable efficiency. Finally, target elucidation can quickly identify proteins and pathways, novel or otherwise, implicated in repeat expansion induction activity. This approach to expansion identification also lends itself well to alteration. The discriminatory phenotypes in this screen depend exclusively on repeat expansion within introns, and thus the potential expanding effects of small agents on any repeat sequence can be investigated using this system design. We demonstrated this ability by examining two different repeat sequences.

Notably, this novel biochemical genetic technology has the ability to reveal proteins and pathways that are perturbed specifically by interactions with small molecules. This information has promise to identify future drug targets in disease models, and could be of particular interest to biopharmocologists.

CONCLUSION

A phenotypic screen that depends on expansion of an artificial repeat intron within a selective cassette has the ability to efficiently identify molecules that induce repeat expansion in yeast (*S. cerevisiae*) models. Intracellularly produced cyclic peptides are easily screened using such a system and can demonstrate repeat expanding properties when expressed in yeast cells. Surprisingly, small cyclic peptide

molecules can influence the stability of more than one type of repeat nucleotide sequence.

Our results present for the first time, cyclic peptide species with repeat expansion-inducing properties. Our easily adaptable screening method allows for the identification of cyclic peptides that cause the expansion of any repeat sequence in a yeast model. We anticipate that identifying the targets of cyclic peptides isolated using our screening design may shed light on proteins and pathways implicated in mechanisms of repeat expansion.

Appendix A. Yeast Transformation Protocol

1. A single colony was inoculated into 10mL of YPD broth and grown overnight at 30°C
2. The following day, 2mL of the inoculum was diluted into 50mL of YPD media (~5x10^6cells/mL) and incubated for 4 hours at 30°C
3. The cells were spun down at 3,000rpm for 2 minutes, and supernatant was discarded
4. Pellet was resuspeded in 25mL of sterile water, spun down at 3,000rpm for 2 minutes, and supernatant was discarded
5. The pellet was resuspended in 1mL of 0.1M LiAc and spun down at 3,000rpm for 1 minute
6. Supernatant was removed, cells were resuspended in 400µL of 0.1M LiAc and aliquoted as 50µL samples into fresh tubes
7. Cells were spun down for 1 minute at 3,000rpm, supertanant was removed
8. 240µL of PEG (MW≈3350, 50% w/v), 36µL of 1M LiAc, 10 µL of ss-carrier DNA (10mg/mL), 2 µL of plasmid DNA, and 72 µL of water w ere added, in order, to the tubes
9. Tubes were briefly vortexing to homogenize their contents and incubated at room temperature for 20 minutes
10. Heat shocked at 42°C for 30 minutes and spun down at 3,000rpm for 1 minute
11. Supernatant was removed and resulting pellet was resuspended in 200µL of water and plated onto Leu- media.

Appendix B. DNA Isolation Protocol

1. Yeast colonies were resuspended in 290 µL of 0.9M Sorbitol, 0.1M EDTA pH7.5 solution, and 10µL of 2.5mg/ml Lyticase solution
2. Incubated for 30 minutes at 37°C and spun down for 2 minutes at 8000rpm
3. Remaining supernatant was discarded and cells were resuspended in 270µL of 50mM Tris pH7.5, 20mM EDTA solution and 20µL of 10% SDS

4. Incubated at 65°C for 20 minutes, 150µL of chilled 5M Potassium Acetate was added, and tubes were inverted to mix
5. Incubated at 4°C for 30 minutes and spun down for 10 minutes at 13,000rpm
6. Supernatant was transferred into fresh tubes and 900µL of cold EtOH was added and mixed by inverting
7. Incubated at -20°C for 1 hour and spun down for 20 minutes at 13,200rpm
8. Supernatant was discarded and resulting pellet was washed twice with 70% EtOH and fully dried
9. Pellet was resuspended in 75µL of TE pH8.0

Plate	Leu- Colonies	Leu-5-FOA colonies	Colonies Analyzed	Expansion Events	Inactivation Events	Failed PCR
1	52	10	4	3	1	0
2	41	22	4	2	1	1
3	30	23	4	2	2	0
4	38	17	4	4	0	0
5	29	14	4	1	1	2
6	39	10	4	2	1	1

Table S1. Fluctuation test data for CP283C1A SMY717/1 transformants

Plate	Leu- Colonies	Leu-5-FOA colonies	Colonies Analyzed	Expansion Events	Inactivation Events	Failed PCR
1	108	88	4	2	2	0
2	107	77	4	4	0	0
3	106	1252	4	3	1	0
4	148	74	4	4	0	0
5	134	90	4	4	0	0
6	56	0	0	0	0	0

Table S2. Fluctuation test data for CP283C1S SMY717/1 transformants

Plate	Leu- Colonies	Leu-5-FOA colonies	Colonies Analyzed	Expansion Events	Inactivation Events	Failed PCR
1	118	1408	4	1	3	0
2	104	796	4	4	0	0
3	164	1940	4	2	2	0
4	104	700	4	3	1	0
5	90	60	4	1	3	0
6	150	1756	4	2	2	0

Table S3. Fluctuation test data for CP283F2A SMY717/1 transformants

Plate	Leu- Colonies	Leu-5-FOA colonies	Colonies Analyzed	Expansion Events	Inactivation Events	Failed PCR
1	GROWTH	172	4	1	2	1
2	201	379	4	0	4	0
3	116	155	4	1	3	0
4	180	314	4	2	1	1
5	169	184	4	0	3	1
6	164	253	4	0	3	1

Table S4. Fluctuation test data for CP283Y3A SMY717/1 transformants

Plate	Leu- Colonies	Leu-5-FOA colonies	Colonies Analyzed	Expansion Events	Inactivation Events	Failed PCR
1	123	332	4	2	1	1
2	112	304	4	2	2	0
3	132	498	4	4	0	0
4	101	468	4	3	1	0
5	152	618	4	2	2	0
6	108	426	4	3	1	0

Table S5. Fluctuation test data for CP283V4A2 SMY717/1 transformants

Plate	Leu- Colonies	Leu-5-FOA colonies	Colonies Analyzed	Expansion Events	Inactivation Events	Failed PCR
1	77	45	4	3	1	0
2	76	33	4	0	4	0
3	96	15	4	1	2	1
4	103	41	4	0	3	1
5	88	18	4	0	2	2
6	107	44	4	2	3	1

Table S6. Fluctuation test data for CP283W5A SMY717/1 transformants

Plate	Leu- Colonies	Leu-5-FOA colonies	Colonies Analyzed	Expansion Events	Inactivation Events	Failed PCR
1	113	96	4	0	1	3
2	134	155	4	0	3	1
3	146	189	4	0	4	0
4	127	106	4	0	3	1
5	135	421	4	0	4	0
6	114	4	4	0	4	0

Table S7. Fluctuation test data for CP283E6A SMY717/1 transformants

Plate	Leu- Colonies	Leu-5-FOA colonies	Colonies Analyzed	Expansion Events	Inactivation Events	Failed PCR
1	96	768	4	3	1	0
2	67	1112	4	4	0	0
3	61	864	4	3	1	0
4	43	568	4	4	0	0
5	41	788	4	4	0	0
6	52	1	1	0	1	0

Table S8. Fluctuation test data for CP283T7A SMY717/1 transformants

Plate	Leu- Colonies	Leu-5-FOA colonies	Colonies Analyzed	Expansion Events	Inactivation Events	Failed PCR
1	62	306	4	2	2	0
2	90	2104	4	4	0	0
3	58	386	4	3	1	0
4	80	2364	4	4	0	0
5	65	470	4	2	2	0
6	83	578	4	2	1	1

Table S9. Fluctuation test data for CP283G8A SMY717/1 transformants

Plate	Leu- Colonies	Leu-5-FOA colonies	Colonies Analyzed	Expansion Events	Inactivation Events	Failed PCR
1	22	5	4	0	4	0
2	72	8	4	1	3	0
3	73	7	4	2	2	0
4	52	29	4	0	4	0
5	81	14	4	0	4	0
6	38	18	4	1	3	0
7	104	66	4	0	4	0
8	78	58	4	3	1	0
9	40	23	4	0	4	0
10	57	38	4	0	4	0
11	58	54	4	0	4	0
12	66	10	4	0	4	0

Table S10. Fluctuation test data for PRS425 empty vector SMY717/1 transformants

Transformant	Dilution
283C1A	1000
283C1S	100
283F2A	1
283Y3A	1
283V4A2	100
283W5A	1
283E6A	1
283T7A	100
283G8A	100
PRS425	1

Table S11. Dilution info for fluctuation tests on CP283 alanine mutants in SMY717/1. Dilution values are the fold values for dilutions of the initial transformant suspended in 200µL of water

Plate	Leu- Colonies	Leu-5-FOA colonies	Colonies Analyzed	Expansion Events	Inactivation Events	Failed PCR
1	32	158	4	2	2	0
2	22	92	4	4	0	0
3	34	67	4	2	2	0
4	34	15	4	2	1	1
5	25	130	4	2	2	0
6	22	0	0	0	0	0

Table S12. Fluctuation test data for CP339 (wild-type) SMY717/1 transformants

Plate	Leu- Colonies	Leu-5-FOA colonies	Colonies Analyzed	Expansion Events	Inactivation Events	Failed PCR
1	28	276	4	3	1	0
2	39	2016	4	2	2	0
3	37	1436	4	2	2	0
4	30	956	4	4	0	0
5	23	848	4	0	4	0
6	15	1	1	0	1	0

Table S13. Fluctuation test data for CP339C1A SMY717/1 transformants

Plate	Leu-	Leu-5-	Colonies	Expansion	Inactivation	Failed

	Colonies	FOA colonies	Analyzed	Events	Events	PCR
1	53	1148	4	1	3	0
2	56	2120	4	2	2	0
3	74	1276	4	1	3	0
4	46	1472	4	4	0	0
5	23	46	4	0	4	0
6	70	2400	4	3	1	0

Table S14. Fluctuation test data for CP339C1S SMY717/1 transformants

Plate	Leu- Colonies	Leu-5-FOA colonies	Colonies Analyzed	Expansion Events	Inactivation Events	Failed PCR
1	23	268	4	4	0	0
2	18	310	4	1	3	0
3	17	105	4	2	2	0
4	20	57	4	1	3	0
5	30	226	4	1	3	0
6	12	47	4	1	3	0

Table S15. Fluctuation test data for CP339F2A SMY717/1 transformants

Plate	Leu- Colonies	Leu-5-FOA colonies	Colonies Analyzed	Expansion Events	Inactivation Events	Failed PCR
1	48	322	4	1	3	0
2	57	164	4	1	3	0
3	65	306	4	0	4	0
4	85	358	4	2	2	0
5	82	848	4	2	2	0
6	75	291	4	1	3	0

Table S16. Fluctuation test data for CP339H3A1 SMY717/1 transformants

Plate	Leu-	Leu-5-	Colonies	Expansion	Inactivation	Failed

	Colonies	FOA colonies	Analyzed	Events	Events	PCR
1	31	85	4	2	2	0
2	37	96	4	1	3	0
3	25	6	4	0	4	0
4	25	75	4	1	3	0
5	21	0	0	0	0	0
6	26	10	4	0	3	1

Table S17. Fluctuation test data for CP339L4A SMY717/1 transformants

Plate	Leu- Colonies	Leu-5-FOA colonies	Colonies Analyzed	Expansion Events	Inactivation Events	Failed PCR
1	22	764	4	3	1	0
2	22	696	4	3	1	0
3	26	870	4	3	1	0
4	24	534	4	2	2	0
5	27	1	1	0	0	1
6	17	2	2	0	2	0

Table S18. Fluctuation test data for CP339V5A SMY717/1 transformants

Plate	Leu- Colonies	Leu-5-FOA colonies	Colonies Analyzed	Expansion Events	Inactivation Events	Failed PCR
1	16	93	4	0	4	0
2	15	118	4	4	0	0
3	27	77	4	1	3	0
4	17	36	4	0	4	0
5	12	48	4	0	4	0
6	16	156	4	2	2	0

Table S19. Fluctuation test data for CP339Y6A SMY717/1 transformants

Plate	Leu- Colonies	Leu-5-FOA colonies	Colonies Analyzed	Expansion Events	Inactivation Events	Failed PCR
1	35	72	4	1	2	1
2	36	88	4	1	3	0
3	20	10	4	0	4	0
4	29	53	4	2	2	0
5	25	127	4	0	4	0
6	43	72	4	0	4	0

Table S20. Fluctuation test data for CP339W7A SMY717/1 transformants

Plate	Leu- Colonies	Leu-5-FOA colonies	Colonies Analyzed	Expansion Events	Inactivation Events	Failed PCR
1	25	133	4	1	3	0
2	54	508	4	2	2	0
3	49	132	4	4	0	0
4	40	150	4	4	0	0
5	36	568	4	3	1	0
6	25	0	0	0	0	0

Table S21. Fluctuation test data for CP339A8G SMY717/1 transformants

Transformant	Dilution
339wt	100
339C1A	10
339C1S	10
339F2A	1
339H3A1	1
339L4A	10
339V5A	10
339Y6A	1
339W7A	1
339A8G	10

Table S22. Dilution info for fluctuation tests on CP339 alanine mutants in SMY717/1. Dilution values are the fold values for dilutions of the initial transformant suspended in 200µL of water

Plate	Leu- Colonies	Leu-5-FOA colonies	Colonies Analyzed	Expansion Events	Inactivation Events	Failed PCR
1	114	356	4	0	4	0
2	93	2208	4	1	3	0
3	81	1564	4	0	4	0
4	101	1908	4	0	4	0
5	89	1964	4	0	4	0
6	92	1268	4	0	4	0
7	36	1352	4	2	2	0
8	70	1568	4	1	3	0
9	101	1287	4	0	4	0
10	68	1220	4	0	4	0
11	80	2220	4	0	4	0
12	138	1936	4	0	4	0

Table S23. Fluctuation test data for CP283 SMY223/1 transformants, Data provided by Anna aksenova

Plate	Leu- Colonies	Leu-5-FOA colonies	Colonies Analyzed	Expansion Events	Inactivation Events	Failed PCR
1	247	669	4	2	2	0
2	193	1033	4	4	0	0
3	222	1265	4	3	1	0
4	185	50	4	1	3	0
5	136	840	4	4	0	0
6	252	998	4	4	0	0
7	239	1142	4	3	1	0
8	249	1208	4	3	1	0
9	250	1064	4	4	0	0
10	185	1129	4	2	2	0
11	206	1036	4	2	2	0
12	238	1103	4	2	2	0

Table S24. Fluctuation test data for CP339 SMY223/1 transformants

Plate	Leu- Colonies	Leu-5-FOA colonies	Colonies Analyzed	Expansion Events	Inactivation Events	Failed PCR
1	203	815	4	2	2	0
2	185	598	4	2	1	1
3	220	1043	4	1	3	0
4	147	868	4	3	0	1
5	256	1116	4	4	0	0
6	160	837	4	3	1	0
7	213	1734	4	2	1	1
8	231	811	4	4	0	0
9	304	788	4	4	0	0
10	160	794	4	2	2	0
11	253	820	4	4	0	0
12	239	681	4	3	1	0

Table S25. Fluctuation test data for 339TH SMY223/1 transformants

Plate	Leu- Colonies	Leu-5-FOA colonies	Colonies Analyzed	Expansion Events	Inactivation Events	Failed PCR
1	74	650	4	2	2	0
2	103	671	4	3	1	0
3	112	855	4	2	2	0
4	118	820	4	1	2	1
5	187	902	4	3	1	0
6	92	708	4	2	2	0
7	77	717	4	3	1	0
8	124	1050	4	3	0	1
9	164	1348	4	3	1	0
10	152	1266	4	2	1	1
11	131	830	4	2	2	0
12	139	858	4	4	0	0

Table S26. Fluctuation test data for CP382 SMY223/1 transformants

Plate	Leu- Colonies	Leu-5-FOA colonies	Colonies Analyzed	Expansion Events	Inactivation Events	Failed PCR
1	391	1303	4	3	0	1
2	350	1128	4	1	3	0
3	630	3690	4	4	0	0
4	351	1772	4	2	0	2
5	508	1349	4	3	0	1
6	409	2283	4	3	0	1
7	348	1067	4	1	2	1
8	458	1625	4	3	0	1
9	398	1236	4	3	1	0
10	352	1008	4	3	0	1
11	454	1483	4	2	1	1
12	390	1012	4	2	2	0

Table S27. Fluctuation test data for 382TH SMY223/1 transformants

Plate	Leu- Colonies	Leu-5-FOA colonies	Colonies Analyzed	Expansion Events	Inactivation Events	Failed PCR
1	303	735	4	3	1	0
2	321	635	4	1	2	1
3	375	1001	4	1	3	0
4	314	750	4	2	2	0
5	309	1592	4	2	2	0
6	771	1126	4	2	2	0
7	852	1528	4	2	2	0
8	402	992	4	1	1	2
9	762	838	4	2	1	1
10	390	946	4	2	1	1
11	348	1300	4	3	1	0
12	434	940	4	0	3	1

Table S28. Fluctuation test data forPRS425 (empty vector) SMY223/1 transformants

Transformant	Dilution
283	1000
283TH	10
339	1000
339TH	10
382	1000
382TH	10
pRS425GDP	10

Table S29. Dilution info for fluctuation tests on wild-type cyclic peptides and TH mutants in SMY717/1. Dilution values are the fold values for dilutions of the initial transformant suspended in 200µL of water

Transformant	Expansion Rate (µ)	Upper CI	Lower CI
283C1A	1.87E-04	3.57E-04	6.22E-05
283C1S	1.35E-04	2.58E-04	4.46E-05
283F2A	5.25E-06	1.01E-05	1.69E-06
283Y3A	3.54E-07	8.66E-07	4.21E-08
283V4A2	1.64E-04	3.07E-04	5.72E-05
283W5A	1.89E-07	4.17E-07	3.75E-08
283E6A	0.00E+00	1.13E-07	0.00E+00
283T7A	6.01E-04	1.11E-03	2.17E-04
283G8A	5.93E-04	1.11E-03	5.15E-04
339wt	6.01E-04	1.11E-03	2.17E-04
339C1A	1.13E-04	2.30E-04	3.10E-05
339C1S	8.65E-05	1.73E-04	2.50E-05
339F2A	5.66E-06	1.12E-05	1.67E-06
339H3A1	2.52E-06	5.20E-06	6.54E-07
339L4A	5.37E-06	1.23E-05	8.95E-07
339V5A	8.80E-05	1.78E-04	2.43E-05
339Y6A	2.13E-06	4.67E-06	4.30E-07
339W7A	8.21E-07	1.86E-06	1.47E-07
339A8G	4.36E-05	8.56E-05	1.33E-05
PRS425	1.42E-07	2.87E-07	4.01E-08

Table S30. Expansion rates calculated from fluctuation tests on SMY717/1 transformants with 95% upper and lower confidence intervals. No expansion events were observed for analyzed transformants of 283E6A, upper confidence interval represents the hypothetical occurrence of a single expansion event.

Transformant	Inactivation Rate (μ)	Upper CI	Lower CI
283C1A	1.05E-04	2.18E-04	2.60E-05
283C1S	2.86E-05	6.98E-05	3.39E-06
283F2A	4.39E-06	8.70E-06	1.30E-06
283Y3A	1.38E-06	2.61E-06	4.79E-07
283V4A2	8.83E-05	1.82E-04	2.29E-05
283W5A	5.01E-07	9.50E-07	1.70E-07
283E6A	1.33E-06	2.46E-06	4.81E-07
283T7A	1.53E-04	3.51E-04	2.55E-05
283G8A	2.52E-04	5.37E-04	5.79E-05
339wt	1.53E-04	3.51E-04	2.55E-05
339C1A	1.19E-04	2.40E-04	3.36E-05
339C1S	9.84E-05	1.93E-04	3.03E-05
339F2A	6.82E-06	1.32E-05	2.19E-06
339H3A1	4.89E-06	9.12E-06	1.74E-06
339L4A	1.58E-05	3.07E-05	5.07E-06
339V5A	7.82E-05	1.61E-04	2.03E-05
339Y6A	5.37E-06	1.02E-05	1.80E-06
339W7A	3.01E-06	5.56E-06	1.10E-06
339A8G	1.92E-05	4.30E-05	3.63E-06
PRS425	7.91E-07	1.26E-06	4.12E-07

Table S31. *URA3* gene inactivation rates calculated from fluctuation tests on SMY717/1 transformants with 95% upper and lower confidence intervals.

Transformant	Expansion Rate (μ)	Upper CI	Lower CI
283	6.07E-04	1.17E-03	1.97E-04
283TH	2.26E-05	3.75E-05	1.07E-05
339	1.55E-03	2.50E-03	7.81E-04
339TH	2.46E-05	3.97E-05	1.24E-05
382	2.55E-03	4.14E-03	1.26E-03
382TH	2.22E-05	3.61E-05	1.10E-05
PRS425	9.56E-06	1.61E-05	4.38E-06

Table S32. Expansion rates calculated from fluctuation tests on SMY223/1 transformants with 95% upper and lower confidence intervals.

Transformant	Inactivation Rate (μ)	Upper CI	Lower CI
283	7.05E-03	1.00E-02	4.50E-03
283TH	2.31E-06	5.40E-06	3.55E-07
339	7.13E-04	1.27E-03	2.84E-04
339TH	9.11E-06	1.68E-05	3.36E-06
382	1.50E-03	2.62E-03	6.37E-04
382TH	6.34E-06	1.22E-05	2.03E-06
PRS425	9.92E-06	1.67E-05	4.59E-06

Table S33. *URA3* gene inactivation rates calculated from fluctuation tests on SMY223/1 transformants with 95% upper and lower confidence intervals.

REFERENCES

Allsopp, R.C., Vaziri, H., Patterson, C., Goldstein, S., Younglai, E.V., Futcher, A.B., Greider, C.W., and Harley, C.B. (1992) Telomere length predicts replicative capacity of human fibroblasts. *Proc Natl Acad Sci* USA 89(21): 10114-10118

Amberg, D. C., Burke, D. and Strathern, J. N. (2005) *Methods in Yeast Genetics: a Cold Spring Harbor Laboratory Course Manual*. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY

Bates, G. and Lehrach, H. (1994) Trinucleotide repeat expansions and human genetic disease. *Bioessays* 16(4): 277-284

Campuzano, V., Montermini, L., Moltò, M.D., Pianese, L., Cossée, M., Cavalcanti, F., Monros, E., Rodius, F., Duclos, F., Monticelli, A., et al. (1996) Friedreich's ataxia: Autosomal recessive disease caused by intronic GAA triplet repeat expansion. *Science* 271: 1423-1427

Campuzano, V., Montermini, L.,. Lutz, Y., Cova, L., Hindelang, C., Jiralerspong, S., Trottier, Y., Kish, S.J., Faucheux, B., Trouillas, P., et al. (1997) Frataxin is reduced in Friedreich's ataxia patients and is associated with mitochondrial membranes. *Hum Mol Genet* 6(11): 1771-1780

Cesare, A.J., and Reddel, R.R. (2010) Alternative lengthening of telomeres: Models, mechanisms and implications. *Nature Reviews* 11: 319-330

Cheng, L., Naumann, T.A., Horswill, A.R., Hong, S., Venters, B.J., Tomsho, J.W., Benkovic, S.J., and Keiler, K.C. (2007) Discovery of antibacterial cyclic peptides that inhibit the ClpXP protease. *Protein Science* 16: 1536-1541

Clark, R.J., Fscher, H., Dempster, L., Daly, N.L., Rosengran, K. J., Nevin, S.T., Meunier, F.A., Adams, D.J., and Craik, D.J. (2005) Engineering stable peptide toxins by means of backbone cylization: Stabilization of the α-conotoxin MII. *Proc Natl Acad Sci* USA 102(39): 13767-13772

Clark, R.M., Dalgliesh, G.L., Endres, D., Gomez, M., Taylor, J., and Bidichandani, S.I. (2004) Expansion of GAA triplet repeats in the human genome: unique origin of the FRDA mutation at the center of an Alu. *Genomics* 83: 373-383

Craik, D.J. (2006) Seamless proteins tie up their loose ends. *Science* 311(5767): 1563-1564

De Baise, I., Rasmussen, A., Monticelli, A., Al-Mahdawi, S., Pook, M., Cocozza, S., and Bisichandani, S.I. (2007) Somatic instability of the expanded GAA triplet-repeat sequence in Friedreich's ataxia progresses throughout life. *Genomics* 90:1-5

de Lange, T. (2005) Shelterin: the protein complex that shapes and safeguards human telomeres. *Genes Dev.* 19: 2100-2110

Ditch, S., Sammarco, M.C., Banerjee, A., and Grabczyk, E. (2009) Progressive GAA.TTC repeat expansion in human cell lines. *PLoS Genet* 5(10):e1000704

Donate, L.E. and Blasco, M.A. (2010) Telomeres in cancer and aging. *Phil Trans R Soc B* 366(1561): 76-84

Dunham, M.A., Neumann, A.A., Fasching, C.L. and Reddel, R.R. (2000) Telomere maintenance by recombination in human cells. *Nature Genetics* 26: 447-450.

Greider, C.W. (1999) Telomeres do D-loop-T-loop. *Cell* 97: 419-422

Greider, C.W. and Blackburn, E.H. (1985) Identification of a specific telomere terminal transferase activity in Tetrahymena extracts. *Cell* 43: 405-413

Hall, B.M., Ma, C., Liang, P. & Singh, K.K. (2009) Fluctuation AnaLysis CalculatOR (FALCOR): a web tool for the determination of mutation rate using Luria-Delbruck fluctuation analysis. *Bioinformatics*, 25(12): 1564-1565

Harding, A.E. (1983) Classification of the hereditary Ataxias and paraplegias. *Lancet* 1:1151-1155

Kassai-Jáger, E., Ortutay, C., Tóth, G., Vellai, T., and Gáspári, Z. (2008) Distribution and evolution of short tandem repeats in closely related bacterial genomes. *Gene* 410: 18-25

Kenyon, C.J. (2010) The genetics of aging. *Nature* 464: 504-512

Kim, N.W., Piatyszek, M.A., Prowse, K.R., Harley, C.B., West, M.D., Ho, P.L.C., Coviello, G.M., Wright, W.E., Weinrich, S.L., and Shay, J.W. (1994) Specific association of human telomerase activity with immortal cells and cancer. *Science* 266: 2011-2015

Kritzer, J.A., Hamamichi, S., McCaffery, J.M., Santagata, S., Naumann, T.A., Caldwell, K., Caldwell, G.A., and Lindquist, S. (2009) Rapid selection of cyclic peptides that reduce α- synuclein toxicity in yeast and animal models. *Nature Chemical Biology* 5(9): 655-662

Kudlow, B.A., Stanfel, M.N., Burtner, C.R., Johnston, E.D., Kennedy, B.K. (2008) Suppression of proliferative defects associated with processing-defective Lamin A mutants in hTERT or inactivation of p53. *Molecular Biology of the Cell* 19: 5238-5248

Lin, K.W., and Yan, J. (2008) Ending in the middle: Current knowledge of interstitial telomeric sequences. *Mutation Research* 658: 95-110

Ma,W.T., Sandri,G.H. and Sarkar,S. (1992) Analysis of the Luria-Delbruck distribution using discrete convolution powers. *J Appl Prob*, 29: 255-267.

McMurray, C.T. (1999) DNA secondary structure: A common and causative factor for expansion in human disease. *Proc Natl Acad Sci* USA 96: 1823-1825

Mirkin, S.M. (2006) DNA structures, repeat expansions and human hereditary disorders. *Curr Opin Struct Biol* 16:1-8

Mondello, C., Pirzio, L., Azzalin, C.M., Giulotto, E. (2000) Instability of interstitial telomeric sequences in the human genome. *Genomics* 68: 111-117

Moyzis, R.K., Buckingham, J.M., Cram, L.S., Dani, M., Deaven, L.L., Jones, M.D., Meyne, J.M., Ratliff, R.L., and Wu, J. (1988) A highly conserved repetitive DNA sequence, $(TTAGGG)_n$, present at the telomeres of human chromosomes. *Proc Natl Acad Sci* USA 85: 6622-6626

Naumann, T.A., Tavassoli, A., and Benkovic, S.J. (2008) Genetic selection of cyclic peptide dam methyltransferase inhibitors. *ChemBioChem* 9:194-197

Paschini, M., Mandell, E.K., and Lundblad, V. (2010) Structure prediction-driven genetics in *Saccharomyces* cerevisiae identifies an interface between the t-RPA proteins Stn1 and Ten1. *Genetics* 185: 11-21

Ruiz-Herrera, A., Nergadze, S.G., Santagostino, and M., Giulotto, E. (2008) Telomeric repeats far from the ends: Mechanisms of origin and role in evolution. *Cytogenet Genome Res* 122: 219-228

Sambrook, J., Fritsch, E.F., and Maniatis, T. (1987) Molecular Cloning: A Laboratory Manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.

Sarkar,S., Ma,W.T. and Sandri,G.v.H. (1992) On fluctuation analysis: a new, simple and efficient method for computing the expected number of mutants. *Genetica*, 85: 173-179

Satoh, T., Li, S., Friedman, T.M., Wiaderkiewicz, R., Korngold, R., and Huang, Z. (1996) Synthetic peptide derivatives from the fourth domain of CD4 Antagonize CD4 function and inhibit T cell activation. *Biochemical and Biophysical Research Communications* 224(2): 438-443

Scott, C.P., Abel-Santos, E., Wall, M., Wahnon, D.C., and Benkovic, S.J. (1999) Production of cyclic peptides and proteins *in vivo*. *Proc. Natl. Acad. Sci.* USA 96(24): 13638-13643

Shay, J.W. and Wright, W.E. (2011) Role of telomeres and telomerase in cancer. *Semin Cancer Biol* 21(6): 349-353

Shishkin, A.A., Voineagu, I., Matera, R., Cherng, N., Chernet, B.T., Krasilnikova, M.M., Narayanan, V., Lobachev, K.S., Mirkin, S.M. (2009) Large-scale expansions of Friedenreich's ataxia GAA repeats in yeast. *Molecular Cell* 35: 82-92

Tam, J.P., Lu, Y., Yang, J., Chiu, K. (1999) An unusual structural motif of antimicrobial peptides containing end-to-ed macrocycle and cystine-knot disulfides. *Proc. Natl. Acad. Sci.* USA 96: 8913-8918

Vulliami, T., Marrone, A., Goldman, F., Dearlove, A., Bessler, M., Mason, P., and Dokal, I. (2001) The RNA component of telomerase is mutated in autosomal dominant dyskeratosis congenital. *Nature* 413: 432-435

Wynford-Thomas, D. and Kipling, D. (1997) Cancer and the knockout mouse. *Nature* 389: 551-552

Zhang, Y., Shishkin, A., Nishida, Y., Marcinkowski-Desmond, D., Saini, N., Volkov, K.V., Mirkin, S.M., and Lobachev, K.S. (2012) Genome-wide screen identifies pathways that govern GAA/TTC repeat fragility and expansions in dividing and nondividing yeast cells. *Molecular Cell* 48: 254-265