

LAWRENCE LIVERMORE NATIONAL LABORATORY

Developing New Tools for the in vivo Generation/Screening of Cyclic Peptide Libraries. A New Combinatorial Approach for the Detection of Bacterial Toxin Inhibitors

J. A. Camarero

January 31, 2007

Disclaimer

This document was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor the University of California nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or the University of California. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or the University of California, and shall not be used for advertising or product endorsement purposes.

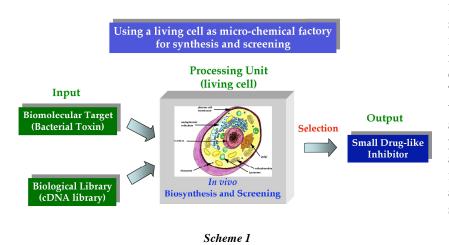
This work was performed under the auspices of the U.S. Department of Energy by University of California, Lawrence Livermore National Laboratory under Contract W-7405-Eng-48.

FY06 LDRD Final Report Developing New Tools for the *in vivo* Generation/Screening of Cyclic Peptide Libraries. A New Combinatorial Approach for the Detection of Bacterial Toxin Inhibitors LDRD Project Tracking Code: 04-ERD-040 Julio A. Camarero, Principal Investigator

Abstract

A new combinatorial approach for the biosynthesis and screening of small drug-like toxin inhibitors inside living cells is presented. This approach has been initially used as proof of principle for finding inhibitors against the LF factor from *Bacillus anthracis*.

Key to our 'living combinatorial' approach is the use of a living cell as a micro-chemical factory



for both synthesis and screening of potential inhibitors for a given molecular recognition event (see Scheme 1). This powerful technique posses the advantage that both processes synthesis and screening happen inside the cell thus accelerating the whole screening/selection process.

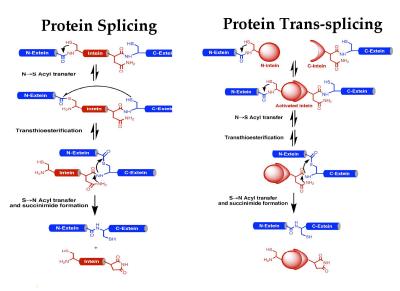
Introduction and motivation

Deliberate exposure of the civilian population of the United States to *B. anthracis* spores by means of the United States Postal Service revealed a gap in the nation's overall preparedness against bioterrorism. Unfortunately, anthrax is just one of many potential weapons of biochemical terrorism, there are so many other toxins like ricin or botulin that could also be used against the civilian population. This alarming situation uncovers an urgent need for tests to rapidly diagnose, effective antidotes or vaccines to prevent, and drugs to cure disease (like new antiviral/antibiotics against new pathogenic bacterial or viral strains). Thus, the development of new methodologies for fast and efficient synthesis, and screening of small drug-like molecules with high affinity against toxins should be one of the top priorities in the fight against bioterrorism. These molecules could be used as specific antidotes but also, when attached to the appropriate platform, as reliable biosensors.

In order to be useful, these new methodologies should be fast -allowing high throughput analysis- and efficient, i.e. provide specific binders in a short period of time. The available methods for solving this daunting problem are either based on rational or combinatorial approaches. The rational approach uses the molecular structure of the target to be knocked out, and then using docking software is able to find potential binders within a virtual library of small organic molecules. Despite the big advances in computing technology, however, this is still a slow process yielding around one ligand per year. Furthermore, most of the docking software available is based on static docking programs, which reduces considerably the efficiency of this technique.¹ Combinatorial approaches, on the other hand, use a random approach to generate as many compounds as possible with the hope that some of the members in these huge mixtures of compounds (also called libraries) will have activity against the molecular target (this is a common approach in nature, active antibodies are created and selected in that way). Combinatorial libraries can be generated using chemical² or biological³ tools. Chemical generation of libraries is usually generated in solid-phase and consequently the libraries are limited to contain 10⁶ different compounds as maximun.² Biological libraries can reach up to 10⁹ members per library (as in phage display technology) however are limited to the generation of peptide or proteins libraries only. In both cases, however, the main limitation is the screening process, which is carried in vitro and is very time consuming.

As solution we have developed totally new and revolutionary approach, a *living combinatorial approach*. This approach uses living cells for the generation of libraries of small drug-like biomolecules, which are then screened inside the cell for activity (see Scheme 1). On of the main advantages of this novel approach is that all the processes (i.e. biosynthesis of the library and screening) happen inside a living cell and therefore no *in vitro* screening is required, thus speeding the screening process. Moreover, the fact that the screening process takes place in a complex media composed by thousand of proteins (i.e. inside the cytoplasma's cell) favors the selection of members of the library with high specificity for the target will be selected. This will minimize the selection of universal binders, a real problem when in vitro screening methods are employed.

We have used this approach for finding inhibitors agains LF factor from anthrax. This protein is a Zn-protease that, once inside the cell, degrades some key host proteins thus disabling the host targeted cell.^{4,5} The fact that this toxin, as well as the microorganism that produce it, can be used for bioterrorism or biochemical warfare constitute a major threat to our national security.



For this reason there is also an urgent need to produce small, chemically stable and readily available inhibitors for this or similar toxins. **Technical plan**

Key to our combinatorial approach is the use of a library of living cells. Each individual cell within this living library will produce a unique small drug-like biomolecule. This compound will be then screened inside the same cell for activity against the corresponding molecular target, in this case anthrax LF. Therefore, this approach requires the development of in vivo methods for the

Figure 1. Protein splicing is new tool that allows to do protein chemistry inside living cells.

biosynthesis of libraries as well as ways to screen the activity of those molecules in vivo and report the activity.

Biosynthesis of circular polypeptides. Circular peptides are one the most common scaffolds used in nature to produce high affinity drug-like effectors (eg. antibiotics, immuno-supressants, etc).

Indeed, peptide cyclization is commonly used in medicinal chemistry for modifying the properties of bioactive peptides.^{6,7} In particular, backbone (head-to-tail) cyclization has been widely used to rigidify peptide structure, therefore minimizing the entropic cost of receptor binding, and to improve the *in vivo* stability of peptides.

Circular peptides will be biosynthesized inside the cell by using recombinant DNA expression techniques, which will allow us to create vast libraries of these compounds. The circularization of the peptide libraries will be accomplished by making use of protein splicing,⁸⁻¹⁰ a naturally

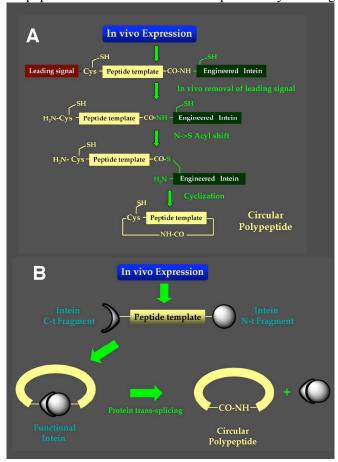


Figure 2. Biosynthesis of circular peptides can be achieved by engineering protein splicing (**A**) or by using protein transsplicing (**B**).

share a disulfide-stabilized core (3 disulfide bonds) with an unusual knotted structure (Figure 3).¹⁵ Cyclotides contrast with other circular poylpeptides in that they have highly defined threedimensional structures, and despite their small size, can be considered as miniproteins. Their unique circular backbone topology and knotted arrangement of the 3 disulfide bonds makes them exceptionally stable to thermal and enzymatic degradation.¹⁵ Furthermore, their well defined structures have been also associated with a range of biological activities, including uterotonic activity, inhibition of neurotension binding, hemolytic, anti-HIV, insecticidal as well as trypsin inhibitory activity. Altogether, these characteristics makes cyclotides ideal candidates to be used as molecular scaffolds for new ligand/drug design and discovery.¹⁵

In vivo screening. The screening process for finding which cyclotides within the library are able to inhibit LF factor toxin has been carried out using a fluorescence-based reporter¹⁶ (see Scheme

occurring process which involves the splicing of polypeptide sequences after the translation step (see Fig.1). This process is mediated by an intein domain (red sequence in Fig. 1) and it can happen in $cis^{8,10-12}$ or $trans^{9,13,14}$ (see Fig. 2).

The first approach uses an engineered intein unit and a leading polypeptide sequence to introduce a C-terminal α thioester and a N-terminal Cvs residue at the peptide sequence to be cyclized, respectively. These two chemical moieties can react with each other to vield the corresponding circular 2A). peptide (Fig. The second approach is based on the use of protein trans-splicing (Fig. 2B). In this case the N-terminal and C-terminal intein fragments are fused to the C- and Nterminus of the peptide to be cyclized. respectively. Thus, when the corresponding chimeric precursor is expressed, the two intein fragments are able to associate spontaneously allowing protein splicing to occur. This event produces the circularization of the corresponding peptide sequence. *Library design*. The molecular library has designed using the cyclotide scaffold as template. Cyclotides are a new emerging family of large cyclic polypeptides (\approx 30 residues long) that

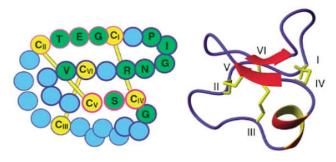
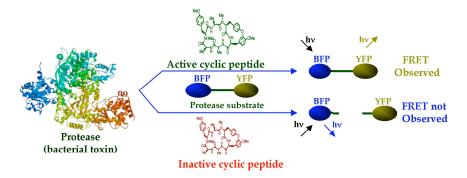


Figure 3. Schematic of a cyclotide showing conserved residues (green), cysteine residues (yellow) and variable residues (blue) along with the three dimensional structure of a cyclotide. The variable residues will be randomized in the library.

2). LF is a Zn-protease that cleave specific proteins crucial for the survival of the targeted cell.^{4,5} Our fluorescence-based screening reporter uses the catalytic activity of the enzyme in combination with a fluorogenic substrate (see Scheme 2). The fluorogenic substrate is composed by a short peptide containing the natural substrate of the protease and flanked by two fluorescent proteins. Cyan fluorescent protein (CFP) and yellow fluorescent protein (YFP) are both engineered mutants of the natural occurring green fluorescent protein (GFP) and are designed to form a FRET (fluorescence resonance energy transfer) pair.¹⁶ Thus, when the protease and the fluorogenic substrate are expressed

in a cell that has biosynthesized an active (i.e. inhibitor) cyclic peptide, the protease is inhibited and therefore unable to cleave the fluorogenic substrate thus allowing the FRET signal to be observed, i.e. emission of the YFP fluorophore when exciting the CFP fluorophore. On the other hand if the cyclic peptide is inactive against LF the still active protease cleaves the fluorogenic substrate leading to the loss of the FRET signal. This in vivo FRET-based inhibition assay will allow the selection of the individual cells that are expressing active (i.e. inhibitor) compounds from the ones that are not. For this purpose we will use a flow cytometric screening technique ¹⁷. Combination of both techniques will provide a powerful living combinatorial tool for finding inhibitors against bacterial toxins. Both approaches, i.e. biosynthesis of libraries with in vivo screening methods, will be combined as depicted in Scheme 3. The first step involves the creation of a combinatorial library of plasmids (in excess of 10⁹ members). All the plasmids of the library contain the DNA coding for the proteolytic domain of the protein toxin and the corresponding fluorogenic substrate for the corresponding toxin. In addition, any individual plasmid contains a variable DNA sequence codifying for the different cyclotides present on the library. This variable sequence is fused to the appropriate engineered intein element in order to allow its cyclization once expressed inside the cell.

It is important to note that although this approach has been initially used for finding inhibitors of bacterial toxins with enzymatic activity, such as LF, the in vivo biosynthesis/screening approach is general. Thus, it is expected that the introduction of slight modifications on the screening scheme would be able to provide a truly generic method based on darwinian selection.¹⁸⁻²⁰ This



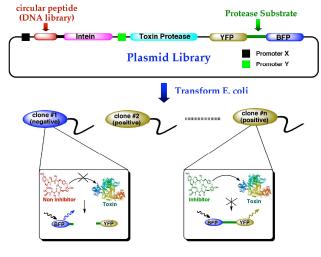
Scheme 2. Principle of a chameleon reporter for finding Anthrax Lethal Factor inhibitors.

slightly modified approach could be easily used for finding small cyclic peptides that could act as effective inhibitors on any protein-protein interaction, thereby providing a general approach for discovering new pharmaceutical compound leads.

Results

During the duration of this project we have been able to accomplish the biosynthesis of several cyclic peptides (cyclotides Kalata B1 and MCoTI-II and sunflower trypsin inhibitor SFTI-I) inside *E. coli* cells as well as a FRET-based screening for reporting LF activity inside living cells. *Biosynthesis of cyclic peptide based libraries in vivo.* Our approach for the biosynthesis of

circular polypeptides is based on the use of an intramolecular version of the native chemical ligation combined with the use of a modified protein splicing unit.^{10,21} In order to accomplish the cyclization, the different linear precursors were fused at the N-terminus with a Met residue, and at the C-terminus with either a VMA or Gyrase engineered inteins (available in the pTYB and pTXB expression vectors family from New England Biolabs). In all the cases the Met residue was efficiently removed in vivo by the endogenous Met amino peptidase of E. coli. This in vivo proteolytic event unmasked the required N-terminal Cys residue that can then react in an





intramolecular fashion with the α -thioester generated by the engineered intein at the *C*-terminus of the linear polypeptide precursor. In the case of cyclotide MCoTI-II and SFTI-1 peptide, the corresponding circular reduced polypeptides were able to fold spontaneously in the cytoplasm to adopt the native structure (Fig. 4A and 4B). We have recently shown that totally reduced cyclic Kalata B1 (KB1) can also fold efficiently into its native form under conditions similar to those found in the cytoplasm.²² Using this biosynthetic approach we have also produced a small library based on the KB1 and SFTI-1 scaffolds (Fig. 5C). Structural characterization for the cyclotides KB1 and MCoTI-II was accomplished using standard 2D-homonuclear NMR techniques.²² The biological activity of MCoTI-II and SFTI-1 was tested using a trypsin inhibitory assay. In both cases the obtained K_i values were similar to those reported for the corresponding natural products,^{23,24} around 20 nM for SFTI-I and 25 pM for MCoTI-II

In vivo FRET-based reporter for LF activity. Our in vivo FRET-based reporter for LF was designed to contain the natural substrate of LF flanked by two fluorescent proteins, cyan fluorescent protein (CFP) and yellow fluorescent protein (YFP)²⁵ (Figure 4D). This reporter protein was readily expressed in *E. coli* showing high levels of FRET in vivo the absence of LF. On the other hand, the FRET signal was significantly reduced when LF and the FRET-based reporter protein were sequentially expressed in *E. coli* (Fig. 4D).

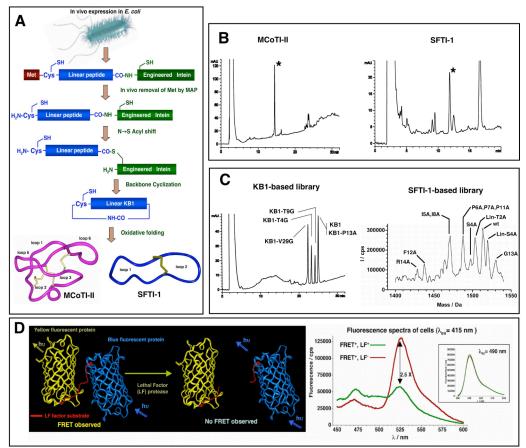


Figure 4. **A.** *In vivo* biosynthesis of backbone cyclized polypeptides using engineered protein splicing units. **B.** In vivo biosynthesis of cyclotide MCoTI-II and trypsin inhibitor SFTI-1. Analytical HPLC traces of different cellular lysates after being purified on a trypsin-column. The asterisk indicates the corresponding oxidized cyclic product. **C.** *In vivo* biosynthesis of cyclotide- and SFTI-based libraries. Left panel, analytical HPLC trace of a KB1-based library. Right panel, ES-MS trace of a SFTI-1 library encoding different Ala mutants. **D.** FRET-based reporter to screen for LF activity in side living *E. coli* cells.

In vivo biosynthesis of ramdonly generated cyclotide-based libraries. Randomly generated cyclotide-based libraries were created at the DNA level using double stranded DNA inserts with degenerated sequences for some of the different loops of the cyclotide scaffold (i.e. loops 2, 3 and/or 5, see figure 4A).

Among the different strategies that have been developed to produce clonable degenerated DNA sequences ²⁶⁻²⁸ we have used method described by Scott and Smith ²⁹ (Fig. 5). This approach involves the generation of a double-stranded degenerated DNA by PCR. Briefly, a long degenerated synthetic oligonucleotide (which codes the whole cyclotide, \approx 100 nt long) template was PCR amplified using 5'- and 3'-primers corresponding to the non degenerated flanking regions. The resulting double stranded degenerated DNA was double digested and a then ligated to a linearized intein expression vector (pTXB or pTYB family) to produce a library of plasmids. These libraries were then transformed into electrocompetent *E. coli* cells to finally obtain a library of cells containing typically up to $\approx 10^9$ different clones (i.e. cyclotide sequences). The degenerated regions. This scheme uses 32 codons to encode all 20 amino acids and encodes only 1 stop codon. Alternatively, the degenerated template was also synthesized from mixtures of trinucleotide codons representing all 20 amino acids and no stop codons ³⁰.

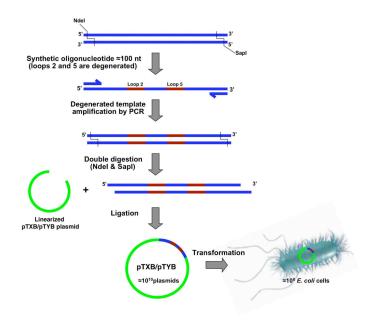


Figure 5. Schematic strategy for assembling a double stranded DNA coding for a full-length cyclotide with degenerated inserts for loops 2 and 5.

The first cyclotide-based libraries generated using this protocols were based on loops 2 and/or 5 using the cyclotides Kalata B1 and MCoTI-II as scaffolds. The complexity of these libraries was around $\approx 10^6$ and $\approx 10^9$ members for combinations involving only one of the loops or two loops, respectively.

Strategic Aligment

This project introduces a totally new approach for the rapid production and screening of high affinity ligands against any molecular target thus revolutionazing the efforts in national biosecurity and drug-discovery at the LLNL. These ligands could be used as powerful antidotes against any particular toxin, but also and more importantly, as biosensors when attached to appropriate platforms. Thus, the whole project aligns extremely well with the National

Institute of Allergy and Infectious Diseases (NAIAD) strategic plan for biodefense research ³¹ and the LLNL responsibilities in homeland security.

Exit Plan

We have submitted during FY06 two research proposals requesting funding to continue this exciting project:

1) Department of Defense, Congressionally Directed Medical Research Programs, Breast Cancer Program, Idea Award.

Title: A new combinatorial approach for the generation of small cyclic peptide

antagonists of the MDM2 oncoprotein as anticancer agents

Funds requested: 600K in 2 years.

2) JSTO FY2007 call basic research program in chemical, biological, radiological and nuclear defense of the joint science and technology office physical science and technology, Division Defense Threat Reduction Agency (DTRA)

Title: A cell-based approach for the rapid biosynthesis/screening of cyclic peptide libraries against bacterial toxins.

Funds requested: 862K for 3 years.

Summary

In summary we report here the first biosynthesis, using DNA recombinant techniques, of different circular Cys-containing peptides in *E. coli*, including the cyclotide MCoTI-II and the Bowman-Birk sunflower trypsin inhibitor (SFTI-1). We have also shown that our biosynthetic approach can be used to generate cyclotide- and SFTI-based libraries that could be screened *in vitro* or *in vivo*. Finally, we have also developed a new cell-based FRET reporter for the *in vivo* screening of LF inhibitors. Future work will focus in the screening of cyclotide- and SFTI-based libraries against LF using our FRET-based reporter *in vivo*.

The development of this novel living combinatorial approach introduces a generic technology that combines chemistry and biology for fast and efficient identification of high-specific ligands to biological toxins. These ligands can be used as powerful antidotes against toxins. It is also anticipated that due to the high stability and high specificity of cylotide-based antidotes they could also be used as reliable and resistant biosensors when attached to the appropriate platforms, thus providing cheap and reliable detector systems that can be used to improve response capability to bioterrorist attacks on the military as well as on the civilian population.

Peer-reviewed publications generated during this project

- 1) R. Kimura, A. T. Tuyet and J. A. Camarero (2006) Biosynthesis of the cyclotide Kalata B1 using protein splicing, *Angew. Chem. Int. Ed.*, **45**, 973-976.
- R. Kimura, K. Krishnan and J. A. Camarero (2006) Biosynthesis of the cyclotide Kalata B1 using protein splicing tools in Understanding Biology using Peptides (Ed.; Blondelle, S. E.), Springer, New York, pp. 287-288.
- 3) J. A. Camarero (2005) Editorial Hot Topic: Polypeptide Chemical Ligation Tools in Protein Engineering, *Protein Pept. Lett.* **12**, 721-722.
- 4) R. Kimura and J. A. Camarero (2005) Expressed Protein Ligation. A new tool for the biosynthesis of cyclic polypeptides, *Protein Pept. Lett.* **12**, 789-794.
- 5) Y. Kwon, K. Welsh, A. R. Mitchell and J. A. Camarero (2004) Preparation of peptide *p*-nitroanilides using an aryl hydrazine resin, *Org. Lett.* **6**, 4145-4151.

Presentations in scientific conferences

- 1) Poster presentation at the 20th Symposium of the Protein Society, August 5-9, 2006, San Diego, CA.
- 2) Oral presentation at the 29th European Peptide Symposium, September 3-8, 2006, Gdansk, Poland.
- **3)** Oral presentation at the 232nd American Chemical Society National Meeting & Exposition, September 10-14, 2006, San Francisco, CA.
- 4) Invited oral presentation at the CHI PepTalk 2007, 3rd Annual Protein & Antibody Therapeutics Meeting, January, 9-10, 2007, San Diego, CA.

Patents and Records of Invention

1) A new Fmoc-compatible method for the solid-phase synthesis of peptide C-terminal athioesters, J. A. Camarero, A. R. Mitchell, J. J. de Yoreo, invention case No. IL-11175 (DOE case #S-101056).

Acknowledgements

This work was performed under the auspices of the U.S. Department of Energy by University of California, Lawrence Livermore National Laboratory under contract W-7405-Eng-48. The project 04-ERD-040 was funded by the Laboratory Directed Research and Development Program at LLNL.

References

(1) Arkin, M. R.; Randal, M.; DeLano, W. L.; Hyde, J.; Luong, T. N.; Oslob, J. D.; Raphael, D. R.; Taylor, L.; Wang, J.; MacDowell, R. S.; Wells, J. A.; Braisted, A. C. *Proc. Natl. Acad. Sci. U S A* **2003**, *100*, 1603-1608.

- (2) Lam, K. S.; Lebl, M.; Krchnak, V. Chem. Rev. 1997, 97.
- (3) Clackson, T.; Hoogenboom, H. R.; Griffiths, A. D.; Winter, G. *Nature* **1991**, *352*, 624-8.
- (4) Vitale, G.; Bernardi, L.; Napolitani, G.; Mock, M.; Montecucco, C. *Biochem. J.* **2000**, *352*, 739-745.
- (5) Schmidt, J. J.; Stafford, R. G.; Millard, C. B. Anal. Biochem. 2001, 296, 130-137.
- (6) Kessler, H. Angew. Chem. Int. Ed. Engl. **1982**, 21, 512-523.
- (7) Hruby, V. J.; Al-Obeidi, F. J. Biochem. 1990, 268, 249-262.
- (8) Camarero, J. A.; Muir, T. W. J. Am. Chem. Soc. 1999, 121, 5597-5598.

(9) Scott, C. P.; Abel-Santos, E.; Wall, M.; Wahnon, D.; Benkovic, S. J. *Proc. Natl. Acad. Sci. USA* **1999**, *96*, 13638-13643.

- (10) Camarero, J. A.; Fushman, D.; Cowburn, D.; Muir, T. W. *Bioorg Med Chem* **2001**, *9*, 2479-84.
- (11) Camarero, J. A.; Muir, T. W. J. Chem. Soc., Chem. Comm. 1997, 1997, 1369-1370.
- (12) Camarero, J. A.; Pavel, J.; Muir, T. W. Angew. Chem. Int. Ed. 1998, 37, 347-349.
- (13) Scott, C. P.; Abel-Santos, E.; Jones, A. D.; Benkovic, S. J. Chem. Biol. 2001, 8, 801-815.
- (14) Abel-Santos, E.; Scott, C. P.; Benkovic, S. J. *Methods Mol. Biol.* **2003**, 205, 281-294.
- (15) Craik, D. J.; Simonsen, S.; Daly, N. L. Curr. Opin. Drug Discov. Devel. 2002, 5, 251-260.
- (16) Miyawaki, A.; Llopis, J.; Heim, R.; McCaffery, J. M.; Adams, J. A.; Ikura, M.; Tsien, R. Y. *Nature* **1997**, *388*, 882-7.
- (17) Daugherty, P. S.; Iverson, B. L.; Georgiu, G. J. Immunol. Meth. 2000, 243, 211-227.
- (18) Wang, L.; Brock, A.; Herberich, B.; Schultz, P. G. Science 2001, 292, 498-500.
- (19) Chin, J. W.; Martin, A. B.; King, D. S.; Wang, L.; Schultz, P. G. *Proc. Natl. Acad. Sci. U S A* **2002**, 99, 11020-11024.
- (20) presentation, S. E.-L. o.
- (21) Kimura, R.; Camarero, J. A. Protein Pept Lett 2005, 12, 789-94.
- (22) Kimura, R. H.; Tran, A. T.; Camarero, J. A. Angew Chem Int Ed Engl 2006, 45, 973-6.
- (23) Avrutina, O.; Schmoldt, H. U.; Gabrijelcic-Geiger, D.; Le Nguyen, D.; Sommerhoff, C. P.;
- Diederichsen, U.; Kolmar, H. Biol Chem 2005, 386, 1301-6.
- (24) Korsinczky, M. L.; Schirra, H. J.; Craik, D. J. Curr Protein Pept Sci 2004, 5, 351-64.
- (25) Nguyen, A. W.; Daugherty, P. S. Nat Biotechnol 2005, 23, 355-60.
- (26) Sparks, A. B.; Rider, J. E.; Hoffman, N. G.; Fowlkes, D. M.; Quilliam, L. A.; Kay, B. K. *Proc. Natl. Acad. Sci. USA* **1996**, *93*, 1540-1544.
- (27) Kay, B. K.; Adey, N. B.; He, Y. S.; Manfredi, J. P.; Mataragnon, A. H.; Fowlkes, D. M. *Gene* **1993**, *128*, 59-65.
- (28) Cwirla, S. E.; Peters, E. A.; Barrett, R. W.; Dower, W. J. *Proc Natl Acad Sci U S A* **1990**, 87, 6378-82.
- (29) Scott, J. K.; Smith, G. P. Science **1990**, 249, 386-90.
- (30) Virnekas, B.; Ge, L.; Pluckthun, A.; Schneider, K. C.; Wellnhofer, G.; Moroney, S. E. *Nucleic Acids Res* **1994**, 22, 5600-7.
- (31) <u>http://www.niaid.nih.gov/biodefense/research/strategic.pdf</u>, s.