Supplementary Material (ESI) for Chemical Communications
This journal is (c) The Royal Society of Chemistry 2010

# Supporting Information for

# Examining the Evolvability of Lead Peptides from Small Library Screens

Jinglei Zhang, Berea A.R. Williams, Mikael Nilsson, and John C. Chaput*

Center for Evolutionary Medicine and Informatics, The Biodesign Institute, and Department of Chemistry and Biochemistry, Arizona State University, Tempe, Arizona 85287-5301.

[*] To whom correspondence should be addressed.
John C. Chaput. Telephone: (480) 727-0392. E-mail: john.chaput@asu.edu

## Methods

*General Supplies.* Oligonucleotides were purchased from the W.M. Keck Facility at Yale University. Peptides were purchased from Genscript Corp. DNA sequencing was performed at the DNA sequencing facility at Arizona State University. Chemical and biological reagents were purchased from the following vendors: Klenow DNA polymerase (New England BioLabs) Superscript II reverse transcriptase (Invitrogen), pJET1.2 kit (Fermentas), carboxy-derivatized MagnaBind™ beads (Pierce), rabbit reticulocyte lysate (Nova Red, Novagen), $^{35}$S-methionine (Amersham Biosciences), oligo(dT) cellulose (New England BioLabs).

*Biased Library Design and Synthesis.* Two DNA libraries were designed based on the lead peptides, TRF23 (FRGWAHIFFGPHVIYRG) and TRF26 (AHKVVPQRQIRHAYNRY). The two libraries consisted of a degenerate open reading frame encoding 17 contiguous amino acid residues flanked on both sides with constant regions that could be modified by overlap PCR. The libraries were designed using a genetic algorithm to control the levels of mutagenesis for each peptide (4% for TRF 26 and 12% for TRF23) (1,2). Based on the level of nucleotide doping, we expected the TRF26 library to sample all possible combinations of one and two amino acid mutations and the TRF23 library to sample a large fraction of five amino acid mutations (see Figure S1). The libraries were ordered from the Keck facility at Yale University, and constructed by solid-phase synthesis on an automated DNA synthesizer. The TRF26 library 5'-TTC TAA TAC GAC TCA CTA TAG GGA CAA TTA CTA TTT ACA ATT ACA ATG 126 246 445 135 135 226 245 216 245 436 216 246 126 346 446 216 346 ATG GGA ATG TCT GGA TC-3' was assembled using the following mixtures of nucleotide phosphoramidites: **1**=97%G+1%C+1%T+1%A; **2**=97%C+1%G+1%T+1%A; **3**=97%T+1%G+1%C+1%A; **4**=97%A+1%G+1%C+1%T; **5**=98%G+2%C; and **6**=98%C+2%G. The TRF23 library 5'-TTC TAA TAC GAC TCA TAT AGG GAC AAT TAC TAT TTA CAA TTA CA ATG 226 135 336 235 346 416 126 226 226 336 446 416 325 126 216 135 336 ATG GGA ATG TCT GGA TCT-3', was constructed using the following mixtures of nucleotide phosphoramidites: (**1**=91%A+3%T+3%G+3%C; **2**=3%A+91%T+3%G+3%C; **3**=3%A+3%T+91%G+3%C; **4**=3%A+3%T+3%G+91%C, **5**=94%G+6%C; and **6**=6%G+94%C). These nucleotide mixtures were designed to control the level of doping and maximize the occurrence of all 20 natural amino

acid residues.

The DNA libraries were made double-stranded by overlap primer extension with Klenow DNA polymerase. The unpaired region on the DNA primer encoded the sequence information necessary to crosslink the DNA-puromycin oligonucleotide to the RNA pool after transcription of the DNA library into messenger RNA. A solution containing 45 µl of 10X Klenow buffer, 5 µl of BSA (10 mg/ml), 90 µl of oligonucleotide library (50 µM), 90 µl of DNA primer (50 µM) (5'-ATA GCC GGT GCT ACC GCT CAG GGC CTG ATA AGA TCC AGA CAT TCC CAT-3'), and 220 µl of water was heated at 94 °C for 5 minutes and cooled on ice. Following the annealing step, 5 µl of Klenow polymerase (5,000 units/ml), 40 µl of dNTPs (20 mM) and 5 µl of water was added to the annealed solution and the reaction was incubated at 37 °C for two hrs. The DNA was phenol/chloroform/isoamyl alcohol (25:24:1, pH 8.0) extracted, ethanol precipitated, dissolved in 300 µl of water and quantified.

*Unbiased Random Library.* The random-sequence library was previously designed and pre-selected to contain an intact open reading frame devoid of stop codons, frame shifts, and internal initiation sites (2). The library encoded 18 contiguous random amino acid positions with a complexity of $4 \times 10^{12}$ unique sequences. The random library was amplified by PCR using the forward primer (5'-TTC TAA TAC GAC TCA CTA TAG GGA CAA TTA CTA TTT ACA ATT ACA-3') and the reverse primer (5'-ATA GCC GGT GCT ACC GCT CAG GGC CTG ATA AGA TCC AGA CAT TCC CAT ATG ATG-3') to modify the upstream region with a T7 transcription promoter and a TMV translation enhancer sequence and the downstream region with a photocrossslinking site. The final library (222 bp) had the following composition: 5'-TTC TAA TAC GAC TCA CTA TAG GGA CAA TTA CTA TTT ACA ATT ACA ATG GAC TAC AAA GAC GAC GAC GAT AAG AAG ACT NAC TGN $(NNN)_{18}$ NAC TGG TCA GCG AGC TGC CAT CAT CAT CAT CAT CAT ATG GGA ATG TCT GGA TCT TAT CAG GCC CTG AGC GGT AGC ACC GGC TAT-3', where N represents a distribution of nucleotides that encodes an equal distribution of all 20 natural amino acid residues. The PCR products were phenol/chloroform/isoamyl alcohol (25:24:1, pH 8.0) extracted, ethanol precipitated, re-suspended in water and quantified.

*Construction of mRNA-Peptide Fusion Molecules.* The use of mRNA display to construct mRNA-peptide fusion molecules has been described previously (2,3). In brief, the initial libraries were amplified by PCR and transcribed with T7 RNA polymerase. The mRNA library was purified by denaturing urea-PAGE, recovered by electroelution and ethanol precipitated. Purified mRNA was photo-crosslinked to a psoralen-DNA-puromycin linker (5'-psoralen-TAG CCG GTG-$(PEG_9)_2$-$A_{15}$-ACC-puromycin, underlined positions denote 2'-methoxy nucleosides) by annealing the crosslinker to the RNA and irradiating the mixture for 15 min at 366 nm in a 96-well plate (50 µl per well). The crosslinked material was purified by denaturing urea-PAGE, recovered by electroelution, and ethanol precipitated. The crosslinked mRNA (1 nmol) was translated *in vitro* by incubating with rabbit reticulocyte lysate and $^{35}$S-methionine for 1 hr at 30 °C. Fusion formation was promoted by incubating overnight at -20 °C in the presence of 600 mM of KCl and 75 mM of $MgCl_2$. mRNA-peptide fusions were purified from the crude lysate by oligo(dT) cellulose chromatography, and reverse transcribed with Superscript II.

*Immobilization of Transferrin on Magnetic Beads.* Carboxy-derivatized MagnaBind™ (0.5 ml) beads were washed 3× with 1 ml of PBS buffer (100 mM $NaH_2PO_4$, 150 mM NaCl, pH 7.2). During each wash step, the beads were gently agitated, magnetically precipitated and the supernatant was discarded. The beads were combined with 1 ml of transferrin (5.4 mg/ml) dissolved in 50 mM MES (N-morpholinoethane sulfonic acid). EDC (1-ethyl-3-(3-dimethylaminopropyl carbodiimide hydrochloride) was added to the mixture with a concentration of 1 mg/ml to activate the coupling of primary amines. The mixture was incubated for 2.5 hour at 4 ºC with shaking. After incubation, the beads were magnetically precipitated and the supernatant was removed. The beads were washed 3× with 1 ml PBS buffer. The amount of protein immobilized on the beads was determined by subtracting the amount of protein remaining in the supernatant and wash fractions from the initial amount of protein added to the beads.

*Evolution of Transferrin-binding Peptides by mRNA Display.* Purified mRNA-peptide fusion libraries were exchanged into 500 µl of selection buffer (50 mM $NaH_2PO_4$, pH 8.0, 0.3 M NaCl, 0.1% NP-40, 0.1 mM EDTA) and incubated with transferrin-coated magnetic beads (100 µl of beads containing ~1 mg of transferrin) for 1 hr with mild shaking (1,000 rpm) at 4 ºC. The beads

were placed on ice, washed six times with 500 µl of selection buffer, and the cDNA was amplified by PCR to generate a new pool of DNA molecules that was enriched in sequences that encoded peptides with affinity to transferrin. The volume of beads was reduced to 10 µl in round 2 and 5 µl in round 3. After a total of three rounds of in vitro selection and amplification, the pool of DNA was inserted into the pJET1.2 sequencing plasmid, transformed into E. coli, and cloned. Fifteen clones from each biased library were randomly chosen and sequenced (Table S1), and the peptide sequences were aligned to attain a consensus motif.

The transferrin selection using the random-sequence library was performed as described above with the following changes. In round one, the magnetic beads contained a mixture of transferrin, streptavidin, α1-antitrypsin, and FV to enrich for peptides with general protein-binding affinity. In round two, four independent selections were performed by incubating the mRNA-peptide fusions produced from the output of round one with magnetic beads that contained transferrin, streptavidin, α1-antitrypsin, and FV in four separate tubes. In round three, the transferrin selection was carried forward by incubating mRNA-peptide fusions produced from the output of round two with trasferrin-coated magnetic beads and the decoy proteins streptavidin, α1-antitrypsin, and FV were present in the wash steps to retain peptides with high specificity for transferrin. After a total of three rounds of in vitro selection and amplification, the pool of DNA was inserted into the pJET1.2 sequencing plasmid, transformed into E. coli, and cloned. Fifteen clones from each biased library were randomly chosen and sequenced. A peptide sequence alignment was not possible due to the diversity of sequences that remained in the pool.

*Affinity Determination by SPR.* Four peptides from each of the doped library selections and three peptides from the unbiased library were randomly selected for characterization using a Biacore T100 surface plasmon resonance instrument. Transferrin protein was immobilized on a CM5 chip using standard amine coupling chemistry, yielding 6,962 response units of immobilized protein. Individual binding assays were performed at multiple concentrations in standard PBS-Tween buffer (10 mM $NaH_2PO_4$, 150 mM NaCl, 0.05% Tween-20, pH 7.4) with a flow rate of 30 µl/min. Each assay consisted of a 100 sec contact time followed by a 300 sec dissociation time. All sensograms were double referenced using buffer injections and the reference cell to subtract nonspecific background binding. Solution binding affinity values were determined using

the affinity fits in the Biacore software package using a 1:1 binding model and represented the average of at least two independent trials.

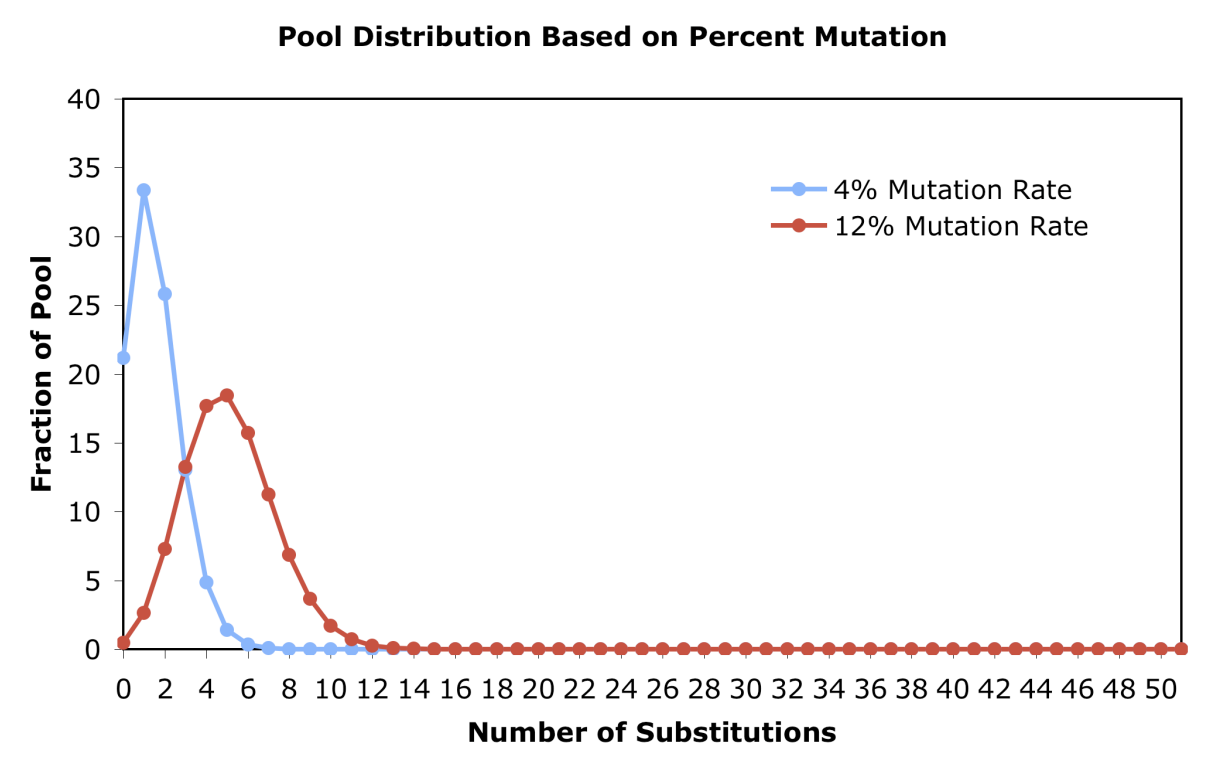


Figure S1. Predicted Distribution of Amino Acid Substitutions for TRF23 and TRF26 Libraries. Doping the TRF23 library with a mutagensis rate of 12% will produce a library of variants that contain on average 5-6 amino acid substitutions per sequence. Likewise, doping the TRF26 library with a mutagensis rate of 4% will produce a library of variants that contain on average 1-2 amino acid substitutions per sequence. The distribution was calculated using the model: $P(N)=f^N(L!/N!(L-N)!)$ (N is the number of substitution, L is the length of the coding region, and f is the mutation rate) (see reference 2).

| TRF23 Selection Peptides | TRF26 Selection Peptides |
|---|---|
| LRAWAQMLLGPQVRSRG | GHKVVPQRQIRHAYNRY |
| LRAWAQMLLGPQVLSRR | PTRRCPSARSATPTTAT |
| LRGWEPMLLGRRCCRGG | GAPAPDPPRLQPLHGNV |
| LRGWAAMLLASQVMSRW | ADKVVPHARSATRTTAT |
| LWGWALMLLRPQVMSRG | AHKVVPQPRLQPLHGNV |
| LKVWAHMLLGPQAMSRG | GHKSCPSARSATPSTAT |
| LKVWAHMWLGPQVLSSG | DHKVVPQRQIRPAYNRY |
| LRWWSTCCWGRRRCRAG | ATRWCPSARSATPTTAT |
| SRVLTQMLLWPKLMSSV | GHKVVPQRQMRHAYNRN |
| LRGWAQLLFGPQAMSRW | PTGWCPAPDPPRLHPLH |
| LRRWAQMLLGPPVMLRG | GHKVVPQLEMRHAYNLY |
| LRRWVQMLLGQKVMSRG | AYKVVPQRQRRYAYNRY |
| LRGWSKRLLGPQVLSRS | DHKVVHQRQLRHAYNRY |
| LRVWAQMMLVPHVMSRE | AHKVVPQRQMRHAYSRY |
| LRGWAQMLLVPQVMSRG | GHKVVPQRQIRHAYNRY |

Table S1. Transferrin-binding peptides obtained after three rounds of mRNA display selection.

Reference:

1. J. Pollard, S. D. Bell, and A. D. Ellington, *Curr. Prot. Nucleic Acid Chem*. 2000, **9.2**, 1-23.

2. G. Cho, A. D. Keefe, R. Liu, D. S. Wilson, J. W. Szostak, *J. Mol. Biol*. 2000, **297**, 309-319.

3. R. Liu, J. E. Barrick, J. W. Szostak, R. W. Roberts, *Methods Enzymol.,* 2000, **318**, 268-293.