mRNA-Display and Its Mechanism

Une victoire racontée en détail, on ne sait plus ce qui la distingue d'une défaite.
JP Sartre

Summary

mRNA display is a very potent technique which offers a systematic approach to discovery of functional peptide-ligands for protein targets. To date, however, the mechanism of this technique remains vaguely understood, which may result in obscure experimental design and the biased interpretation of results, and failure to produce strongly binding ligands. This paper provides a detailed analysis of a peptide-ligand selection using mRNA display, and suggests a thermodynamic model, which explains existing problems and may offer a hint on how to further develop and optimise mRNA display.

Keywords

mRNA display, PHD2, thermodynamic model, PCR bias

Introduction

DNA-encoded peptide libraries opened a new way to search for peptide ligands with useful properties — drug-like compounds [Goldflam et al.], protein inhibitors, protein activators, or co-crystallisation chaperons [Hipolito et al.]. Selection of peptide ligands from such libraries is based on iterative algorithm: (1) incubation of a DNA-tagged peptide-ligand library with a target protein, (2) recovery of binding ligands, and (3) regeneration of a less diverse library from the recovered DNA tags (Figure 1). Cycles of selection are repeated until library diversity reduces sufficiently to make its analysis possible. Identified peptide ligands are chemically synthesised and their binding to the target protein is verified. Selection is termed 'successful' if amongst thus recovered peptides, some demonstrate high affinity towards the target protein (*e.g.* $k_D < 10^{-7}$ M).

Based on this selection algorithm, phage display and mRNA display are of particular interest, because these methods enable selection from highly diverse libraries (~10^9 for phage display, and ~10^{12} for mRNA display) .Display techniques produced a number of useful peptide ligands. Unfortunately, many selections are failing to produce a ligand for a target protein [Menendez et al., Vodnik et al.], which has most commonly been attributed (1) to interference of the target protein with the phage, cDNA or mRNA part of a ligand or (2) to individual cDNA sequences amplification efficiencies.

A peptide-ligand recovery using display techniques is thought to be regulated by peptide's ability to form a complex with a target protein. Thus, it is often inferred that peptide-ligands recovered in higher proportion have higher affinity towards target protein; and ligand recovery is often taken inversely proportional to ligand-target complex dissociation constant (K_D value).

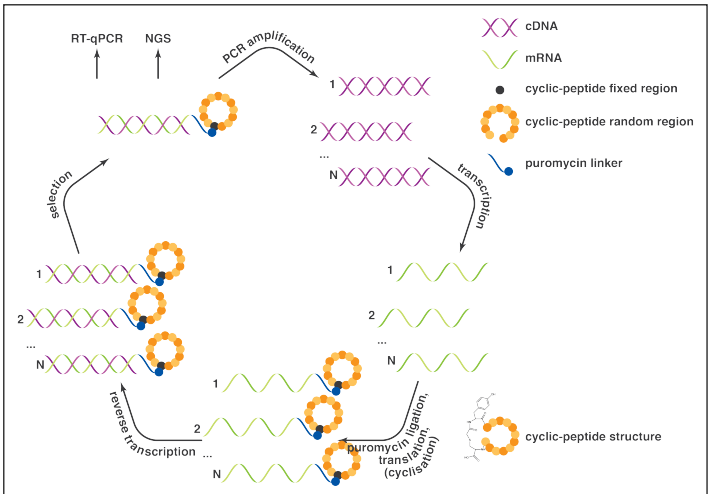

Figure 1. Selection Cycle.

Selection cycle begins with cDNA library, which is transcribed into mRNA library, and subsequently ligated to puromycin, translated, cyclised and reverse-transcribed *in vitro*, thus resulting in cDNA-mRNA-peptide library. This library is incubated with immobilised target protein.
The fraction of the library recovered after incubation is used
(1) to amplify cDNA library using Taq DNA polymerase, thus generating an input for the next cycle of selection
(2) to estimate the recovery using RT-qPCR, and
(3) to create an NGS 'snapshot' of the library at the end of a cycle.

Similarly, recovery of structurally analogous peptide-ligands is often viewed as evidence for a particular sequence (consensus sequence) being favoured in the selection [Takahashi et al.] (*i.e.* the complex between such consensus peptide-ligand and target-protein having lower K_D).

Thus, marketability of display techniques relies on two assumptions: (1) the recovery of a peptide-ligands in selection is directed by and is proportional to its affinity towards a protein-target, and (2) the amplification of DNA tags does not affect the outcome of selection process. To date, the former assumption has been invalidated for both phage displays [Derda et al.] and mRNA displays [Jalali-Yazdi et al.]. However, it remains unclear to which extent the results are biased by the DNA-tag amplification, and how the results of the selection are to be interpreted.

This paper provides detailed analysis of the successful selection of peptide-ligands of PHD2 protein-target. This work verifies the effect of the cDNA tags on the selection results and introduces a robust method to monitor and analyse display-based selections. This paper suggests a thermodynamic mechanism of selection that informs on to which extent the selection process is regulated by the ligand-to-target affinity, highlights potential limitations of mRNA display and indicates potential ways of protocol improvements.

(human recombinant PHD2) from a cyclic-peptide library using a version of mRNA display (so called RaPID system) based on (1) the flexible in vitro translation (FIT) of mRNA library, (2) mRNA-to-peptide ligation using puromycin, and (3) peptide cyclisation using the non-proteinogeinc amino acids incorporated into the peptide chain.

Selection against Human Recombinant PHD2

Selection was performed according to the previously published protocol [Hayashi et al.] with minor modifications using a version of mRNA display (so called RaPID system) based on (1) the flexible in vitro translation (FIT) of mRNA library, (2) mRNA-to-peptide ligation using puromycin, and (3) cyclisation triggered by a non-proteinogeinc amino acid incorporated into the peptide chain. Active (activity was confirmed using CODD peptide as a substrate) biotin-tagged human recombinant PHD2 immobilised on M-280 streptavidin magnetic beads was used as protein target. Selection was performed from a cDNA-mRNA-cyclic-peptide library. Original cDNA library was assembled from five libraries — $(NNK)_8$ to $(NNK)_{12}$ — combined at equimolar ratios (cDNA total length varying from 111 to 123 nucleotides, respectively). Original library diversity was estimated 10^{12}-10^{14} unique cDNA sequences, each present at 1 to 100 copies.

The cDNA library was transcribed into an mRNA library, which was *in vitro* ligated to puromycin, translated, reverse-transcribed and cyclised, thus resulting in cDNA-mRNA-peptide library. Resulting cDNA-mRNA-peptide library was incubated with immobilised PHD2 admixed at approximately 1:1 molar ratio. A fraction of the library recovered after incubation was used for cDNA library regeneration (using Taq DNA polymerase) to use as an input in the next cycle of selection. To eliminate the peptides interacting with magnetic beads, starting from the second cycle, three consecutive dummy incubations (without PHD2) were performed prior to incubation with immobilised PHD2.

Progression of selection was controlled in two ways: (1) real-time quantitative PCR (RT-qPCR) was used to estimate cDNA-library recovery relative to the input, and (2) next generation sequencing (NGS) was used to take a 'snapshot' of the library at the end of every cycle. Selection was judged complete when (1) the fraction of cDNA library recovered after incubation with immobilised target protein exceeded that recovered after dummy incubation, and when (2) the recovery reached approximately 1% of the input (Figure 2). Achieving this involved 6 cycles of selection.

Analysis of Selection Results

To estimate the fraction of individual peptides recovered after each cycle, a sample of cDNA library was subjected to NGS. As selection progresses, the library is being enriched in one peptide (vwdprtfylsri, and K_D = 3 nM; referred to as *the top peptide* throughout the paper), which constitutes over 75% of the library after the 6th cycle. After the final cycle, 206 unique peptide sequences were identified. These peptides are ranked based on their enrichment in final cycle (rank 1 being the most enriched peptide). 24 top-ranking peptides were selected and their fraction was plotted for every selection cycle (Figure 3).

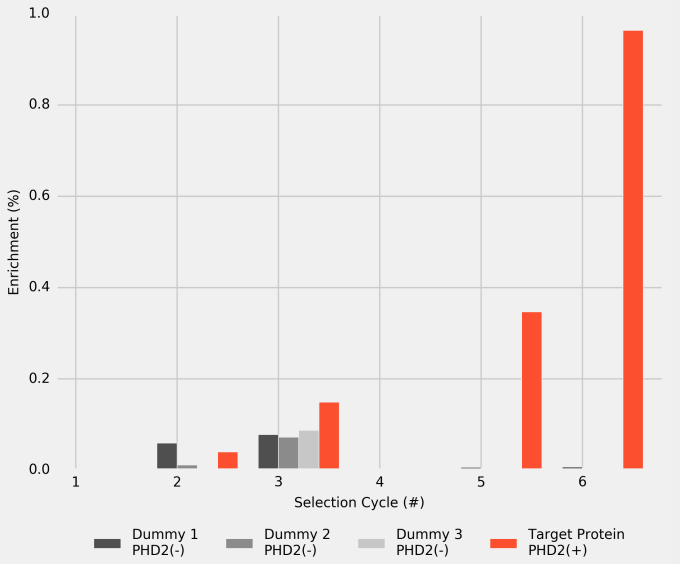

FIGURE 2. CDNA-LIBRARY RECOVERY BY CYCLE.

cDNA-library recovery after every cycle estimated using RT-qPCR and normalised by the input in every cycle.

[PHD2 −] after dummy incubation with M-280 streptavidin magnetic beads alone

[PHD2 +] after incubation with active human recombinant PHD2 immobilised on M-280 streptavidin magnetic beads

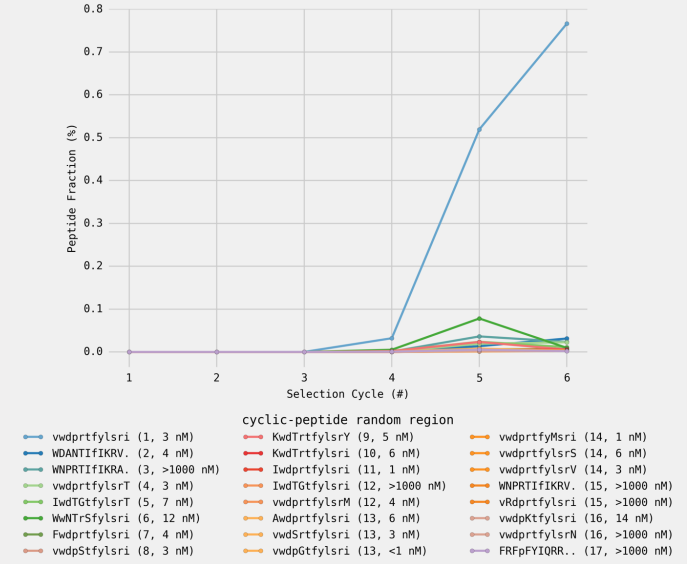

FIGURE 3. PEPTIDE FRACTIONS BY CYCLE.

Enrichment of the 24 most frequent peptides found after the final (6th) cycle of selection.
The peptide sequences are formatted so that amino acids in each position, if different from *the top peptide* (vwdprtfylsri), are capitalised; if the peptide sequence is shorter, the missing amino acid is indicated by the '.' at the end of the sequence.
Peptides rank and their K_D are provided in the brackets after the random-region peptide sequence.

To test whether peptides' enrichments in final cycle correlate with their K_D values, top 24 peptides were chemically synthesised and K_D values of their complexes with PHD2 were measured using surface plasmon resonance (SPR; Figure 3). Importantly, several of the less enriched peptides were found to have even higher affinity towards the target protein than the top peptide. No correlation was found between peptide fraction and its K_D with PHD2.

16 out of top 24 peptides differ from the top peptide and one another by no more than two amino acids, with K_D values differ by approximately 3 orders of magnitude. Noteworthy, peptide ranked 3 has K_D higher than 1000 nM and is followed in ranking by peptides with K_D in the lower nM region. These findings indicate, that the fraction of a peptide recovered in the final cycle is not exclusively regulated by its affinity towards a target protein *(i.e.* is not primarily driven by ligand's phenotype) but results from a more complex selection mechanism.

MRNA-DISPLAY PHYLOGENETICS

To look into genetic relations between structurally similar peptides found in the final cycle an open-source tool was developed. This tool converts NGS reads into a graph of cDNA- and corresponding peptide sequences. Two unique DNA sequences were joined together if these two sequences differ by one nucleotide. Thus, the unique cDNA sequences were arranged into a set of disjoint groups. cDNA sequences in each group were translated into peptide sequences *in silico* and the unique peptide sequences within such groups were joined together if two peptides differ by one amino acid. The resulting graph was visualised as a scatterplot, where each circle marker represents a unique peptide sequence and each line connecting any two circles indicates that two connected peptide sequences differ by one amino acid (Figure 4, and Supplementary___ showing peptide sequences and their characteristics).

Peptides which appear in earlier selection cycles are more frequent, and are encoded by a greater number of DNA mutants, some of which are not to be found in the original NNK library. Given the

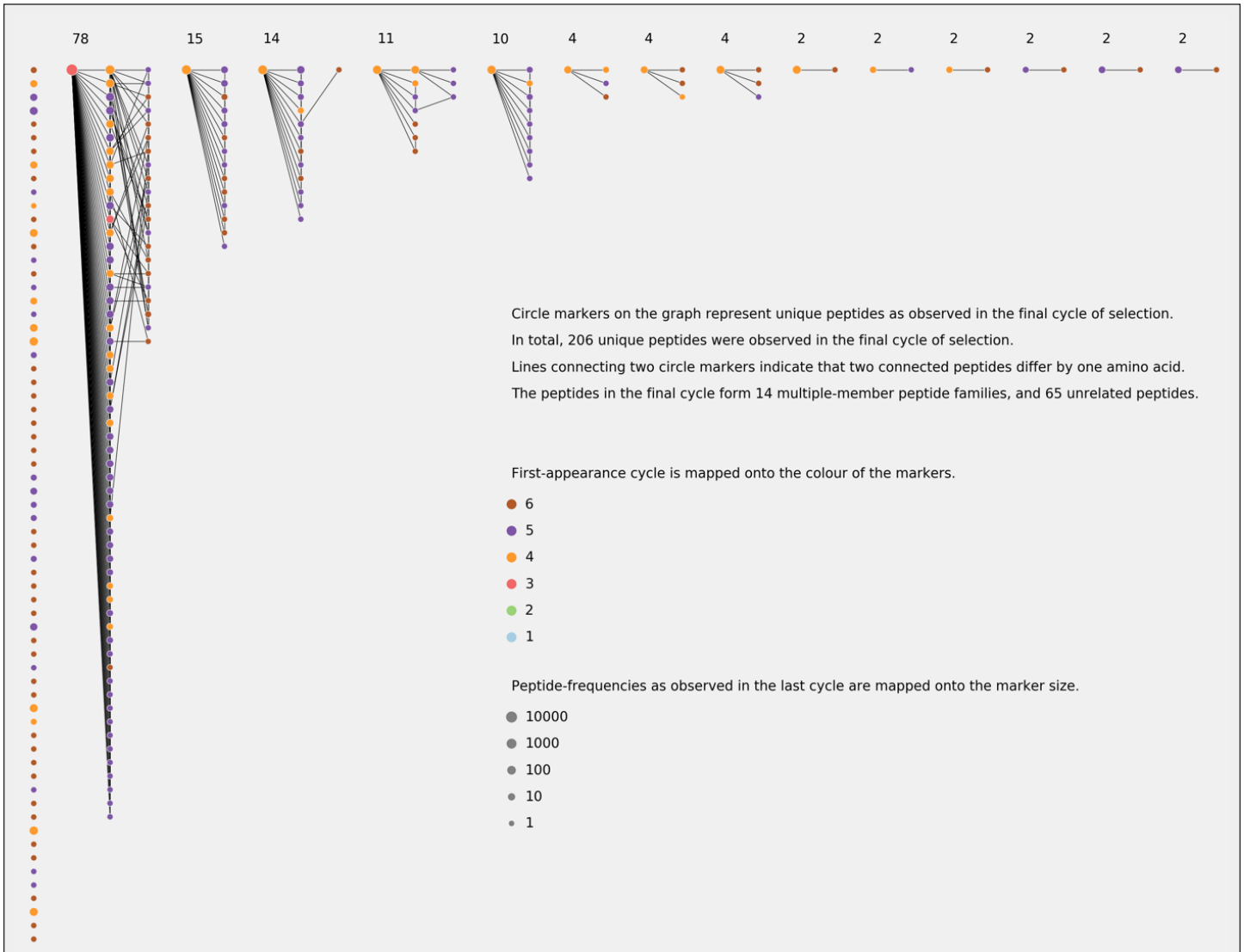

FIGURE 4. PEPTIDE SEQUENCES IN FINAL CYCLE

Phylogenetic trees of unique peptides observed in the final cycle of selection:

technique	modified mRNA display (so called RaPID system)
target	human recombinant PHD2
library	cDNA-mRNA-cyclic-peptide library ($NNK_{8\text{-}12}$)

constant rate of cDNA mutation in the process of PCR amplification, it is conceivable that the later appearing DNA sequences have fewer DNA mutants. These results suggest, that in a pair of peptides distinguishable by only one amino acid, the more frequent and earlier appearing peptide is likely to be genetically antecedent to the lest frequent later appearing peptide, thus, evincing genetic rather than accidental structural relatedness of the two peptides.

After the final cycle of selection, library sequencing produced 5601 reads, with 505 unique DNA sequences encoding for 206 unique peptide sequences. Of the 505 unique cDNA sequences found in final cycle, 361 (71%) have codons beyond original NNK library. The top peptide is encoded by 82 cDNA mutants, of which 78 have codons extrinsic to the original NNK library. These suggest that mutation of original sequence is a driving force of both sequence diversity and structural similarity observed in the final cycle.

The most prolific peptide family constitutes 35% (78 out of 206 unique peptide sequence) of all peptides. Crucially, some peptides in this family persist through selection and are highly enriched despite their low affinity towards target peptide (K_D >1000 nM). The probability of selecting a whole family of peptides from original library by chance is incomparably lower than the probability of mutation

when using Taq DNA polymerase. It is tempting to interpret structural similarity within this prolific peptide family as an indicator of selection favouring some consensus sequence. However such interpretation may bias the analysis in favour of non existent peptide sequence. Structural similarity between peptides in final cycle does not reflect the affinity of either individual peptides or some consensus peptide sequence towards target-protein, but results from mutation of original sequences.

Top Peptide Alanine Scan

To provide further insight into the mechanism of selection, 2 one-cycle selections were performed using different cDNA libraries :

(1) cDNA encoding top peptide (vwdprtfylsri), using cDN

12 cDNAs based on the most common E. coli codons encoding for 12 alanine-scan

analogues of the top peptide (Figure 5A)

(2) cDNA encoding for the top peptide based on the most common E. coli codons

12 cDNAs based on the most common E. coli codons encoding for 12 alanine-scan

analogues of the top peptide (Figure 5B)

Both selections have been performed with and without dummy incubations, each selection was repeated twice and all provided essentially identical results.

cDNAs based on the most frequent E. coli codons differ from one another by no more than 3 nucleotides and A/T count of their random region varies between 13 and 16. The original cDNA

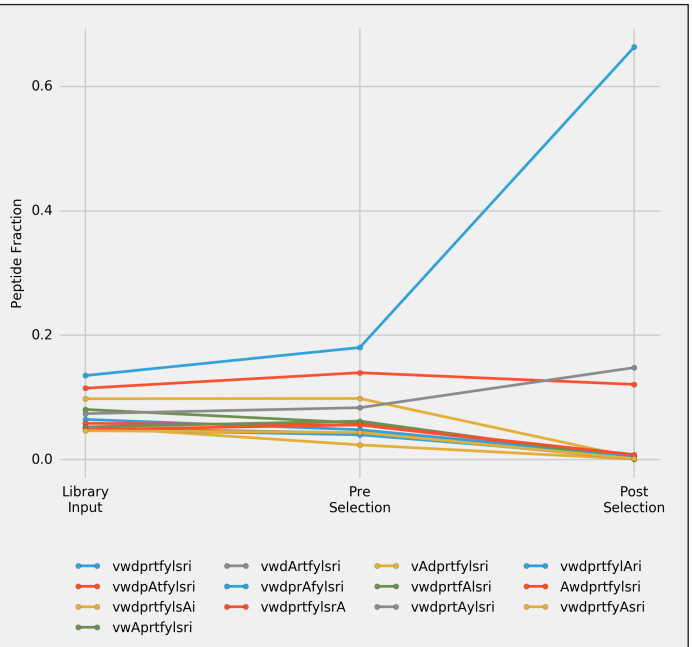

FIGURE 5A. ALA-SCAN **E.COLI CODONS**

Peptide-fractions monitoring in a single-cycle selection from a library of 13 cDNA-mRNA-cyclic-peptides, where:

- 1 cDNA encodes the top-peptide (vwdprtfylsri), **using cDNA from the original selection**;
- 12 cDNAs encode for Ala-scan analogue of the top peptide, using most frequent E. coli codons.

Input Library was transcribed, puromycine ligated, translated and reverse transcribed to convert it into *Pre-Selection Library*. *Post-Selection Library* was the library recovered after incubation of *Pre-Selection Library* with immobilised PHD2.
Ala-s in the peptide sequences are capitalised.

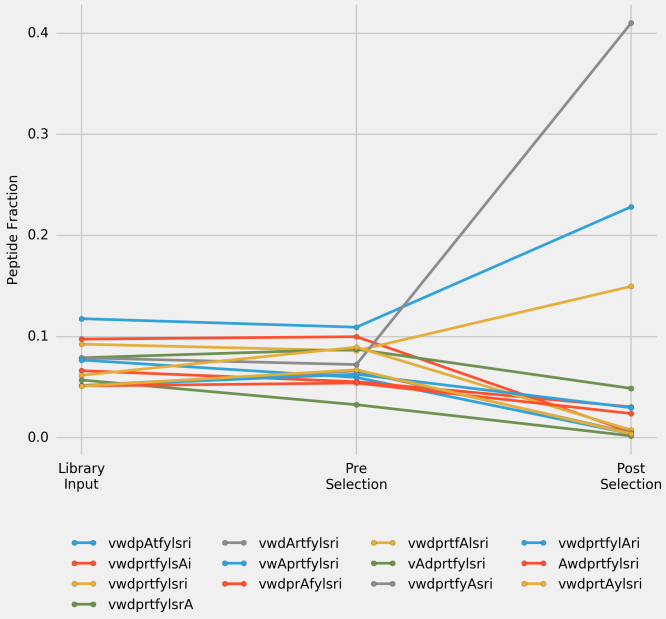

FIGURE 5B. ALA-SCAN **NNK CODONS**

Peptide-fractions monitoring in a single-cycle selection from a library of 13 cDNA-mRNA-cyclic-peptides, where:

- 1 cDNA encodes the top-peptide (vwdprtfylsri), **using most frequent E. coli codons**;
- 12 cDNAs encode for Ala-scan analogue of the top peptide, using most frequent E. coli codons.

Input Library was transcribed, puromycine ligated, translated and reverse transcribed to convert it into *Pre-Selection Library*. *Post-Selection Library* was the library recovered after incubation of *Pre-Selection Library* with immobilised PHD2.
Ala-s in the peptide sequences are capitalised.

sequence recovered from selection differs from the other cDNAs by up to 12 nucleotides, with A/T count equal 21.

Single-round selection using library (1) produced the same top peptide (enriched over 60%) as the original selection from random library. In selection using library (2) the same peptide was enriched only to 15% and ranked 3. Minimising the differences in cDNA not only affects the fraction of individual peptides, but also changes the ranks of peptides, and previously top peptide sequence becomes third. Changing the genotype of only one peptide affects not only the fraction of this peptide, but also the outcome of the whole selection. K_D values for chemically synthesised alanine-scan peptides were assayed using SPR, and compared with results of such selections (Table___).

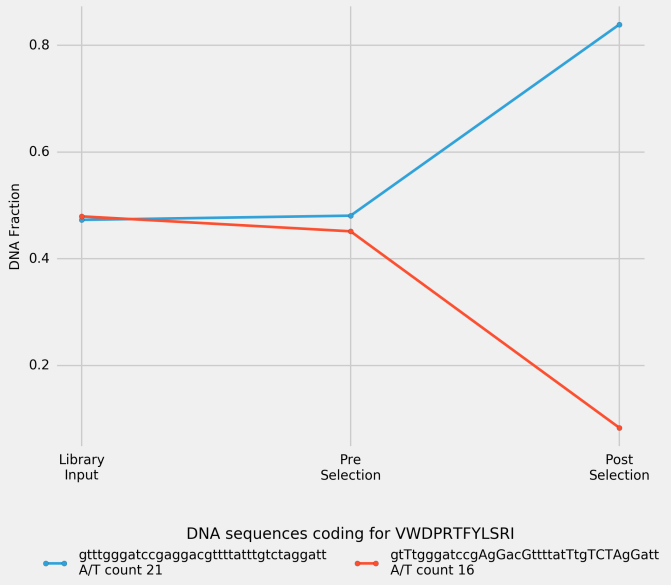

FIGURE 6. SAME PEPTIDE DIFFERENT CDNA

Peptide-fractions monitoring in a single-cycle selection from a library of 2 cDNA-mRNA-cyclic-peptides, where both cDNAs encode the top-peptide (vwdprtfylsri):
(1) **using cDNA from the original selection**;
(2) **using most frequent E. coli codons**
Input Library was transcribed, puromycine ligated, translated and reverse transcribed to convert it into *Pre-Selection Library*. *Post-Selection Library* was the library recovered after incubation of *Pre-Selection Library* with immobilised PHD2.
Nucleotides which differ from the original sequence are capitalised

TWO CDNAS ENCODING THE SAME PEPTIDE

To estimate the impact of cDNA composition on the selection outcome, one-cycle selection was performed with a library comprised of two cDNAs encoding for the most frequent peptide in original selection: (1) the original cDNA sequence recovered from selection, and (2) one composed from the most frequent E. coli codons.

PCR AMPLIFICATION EFFICIENCY

To verify whether PCR amplification efficiency is A/T content dependent, 8 $(NNK)_{12}$ sequences with A/T content varying from 16 to 23 were picked at random and used for the mock-selection test, in which the equimolar library was PCR-amplified for 120 cycles, and the samples were collected for the NGS sequencing after every 20 cycles of PCR (Figure 6).

The fraction in two sequences was achieved by performing PCR amplification alone (without selection) for as few as 40 cycles, for a very simple library. For a more complex library, it would take more cycles of PCR to produce this level of fraction. It appears that A/T content does not have significant influence on the PCR-amplification efficiency, and some more complex mechanism, such as forming intramolecular folds, may be involved.

DISCUSSION

In the process of selection, cDNA-mRNA-peptide ligands in the library undergo two distinct types of reactions, controlled by different parts of the fusion ligand: (1) complex formation between the ligand and the target protein (controlled by the peptide part of a ligand, and is characterised by K_D of ligand-target complex), and (2) PCR amplification (controlled by the cDNA part of a ligand, and characterised by cDNA amplification efficiency). The former follows linear law and is performed approximately 10 times during selection, while the latter follows exponential law and is performed approximately 150 times during selection. Close examination of the selection process suggests that the fraction of ligands recovered in selection is primarily depends on PCR amplification rather than complex formation with target protein, *i.e.* the process is overall driven by the genotype rather than phenotype of the cDNA-mRNA-peptide ligands. It remains clear however, that ligand-target binding does have influence on the selection.

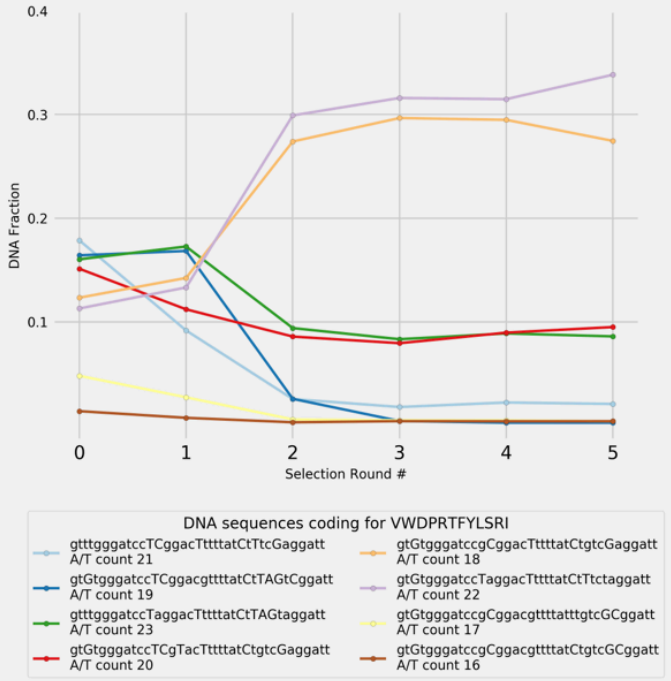

FIGURE 7. PCR BIAS

PCR-amplification of 8 randomly selected $(NNK)_{12}$ cDNAs encoding for the top peptide in the original selection (vwdprtfylsri). Every 20 cycles of PCR, part of the sample was used for the following amplification, and the rest was collected for NGS sequencing. 0^{th} point shows the composition of the original input library.

Unfortunately, currently existing methods do not allow one to closely monitor selection from the libraries with diversities above 10^6. Instead, to better understand to which extent the complex formation between a ligand and a target protein affects the selection outcome, a thermodynamic model was developed.

In the mixture of ligands and one target, the thermodynamic equilibrium of complex formation-dissociation between a particular ligand (n) and the target (L-T) is described by equations:

$$L_nT \rightleftharpoons L_n + T \quad (1)$$

$$K_{D,n} = \frac{[L_n][T]}{[L_nT]} \quad (2)$$

$$[L_nT] = \frac{[L_n][T]}{K_{D,n}} \quad (3)$$

The probability of a target being bound ($P_{b,T}$) is described by equation:

$$P_{b,T} = \frac{[LT]}{[T]+[LT]} = \frac{[L]}{K_D} : (1 + \frac{[L]}{K_D}) \quad (4)$$

Where L is an average ligand, characterised by total concentration of all ligands, and the average K_D of a complex with a target protein.

The amount of the unbound target protein ([T]) can be expressed as follows:

$$[T] = (1 - P_{b,T}) * T \quad (5)$$

Thus, the probability of a given ligand being bound ($P_{n,b}$) to a target protein is described by equation:

$$P_{b,n} = \frac{[L_nT]}{[L_n]+[L_nT]} = \frac{[T]}{K_{D,n}} : (1 + \frac{[T]}{K_{D,n}}) \quad (6)$$

Figure 8 shows the probability of a ligand being bound to a target protein depending on a given ligand $K_{D,n}$, and the average K_D of total ligand population. Primary diversity of the mRNA library (as opposed to secondary diversity resulting from PCR-induced mutations) is to a great extent regulated by the first cycle of selection, at which point each ligand is represented by 1 to 100 copies and can be eliminated completely from the selection, if not bound to a ligand.

The concentration of a target protein in during incubation with cDNA-mRNA-peptide fusion library is approximately 10^{-6} M, and the total library concentration is about 10^{-6} M (total-ligand to target ratio 1:1). Given that peptide ligands with 1 μM K_D values are found after selection, total library average K_D can be safely assumed to be at 10 μM or higher. It can be seen that a ligand which complex with the target protein K_D exceeds 100 μM, has close to 0 probability of being bound to the target protein. In the first cycle of selection, when only 1 to 100 copies of a given ligand are expected in the mixture, such ligand is likely to be eliminated completely from the selection. Ligands with KD lower than 100 μM are likely to be recovered after the first round of selection and amplified, from which point their fraction depends on the amplification efficiency, rather than their affinity towards the target protein.

Based on the results of this study and suggested thermodynamic model of selection, several suggestions for the improvement of selection process can be made.

- PCR AMPLIFICATION REGULATION

While it is not possible to completely eliminate the effect of PCR bias, introduced at the library amplification stage, it may be possible to minimise it using high fidelity and high GC DNA polymerases such as Phusion, Q5 and AccuPrime. To reduce by-

product formation during the PCR amplification, the protocol may benefit from using emulsion PCR, or fewer selection cycles.

To minimise the undesirable interference between cDNA-mRNA part of the ligand and target protein, short random strands of DNA can be introduced into the blocking solution.

- THERMODYNAMIC REGULATION

Previously introduced 'warhead'[Morimoto et al.] can be defined more specifically: to ensure the recovery of the ligand after the first cycle of selection, the 'warhead' has to ensure K_D at 1 nM or lower.

It has been previously suggested that in the first round less 'stringent' conditions are applied to the selection (*e.g.* fewer washes, and lower temperature) [Pande et al.]. Temperature regulation of selection affects complex formation between the ligands and the target proteins in predictable manner: the higher is the temperature at which the peptide-ligands library is incubated with a target protein the lower is the average K_D value, and the more likely it is that he peptides bound will have lower K_D values. Selection at 37°C (as opposed to commonly applied 4°C) was previously used to regulate the selection pressure at later cycles of selection[Hipolito et al.]. Thermodynamic model proposed in this work suggests, that selection may benefit from being carried out at higher temperature from the first cycle.

If the total-ligand to target ratio is reduced to 1:10000 it is possible to ensure that ligands form complexes with a target protein at K_D higher than 10 nM are eliminated from selection, thus ensuring the recovery of peptide ligands with K_D below 10 nM.

Due to the PCR bias, peptide sequences present in the original library and amplifying at a lower rate yet having high affinity towards target protein, are often disregarded and taken out of consideration. Instead, selection results are heavily biased by the rapidly amplifying sequences and their mutants. While it may be impossible to completely eliminate the amplification bias or alter the amplification efficiencies of ligands, the alternative approach to the selection-results analyses offered in this work allows one to identify the peptide sequences present in the original library, yet amplifying at a lower rate due to the PCR bias. It is hoped that the analysis suggested in this paper may uncover such ligands, thus increasing the rate of successful ligand discoveries.

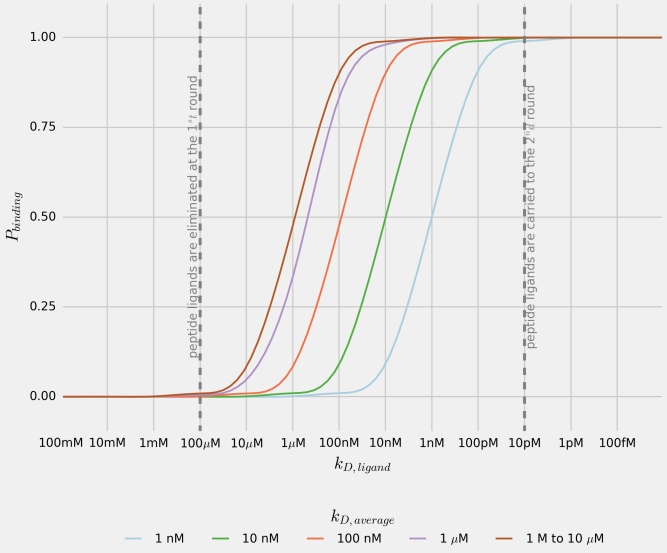

FIGURE 8. MRNA MECHANISM

Probability of a peptide ligand (n) being bound to a target protein dependence on a ligand-target complex dissociation constant ($K_{D,n}$).
Every line shows the probability of a peptide-ligand (n) to be bound to a target-protein for different ligand-population mean K_D in the range 1 nM to 10 μM.

SUPPLEMENTARY DATA

Table 1. Peptide Fractions by Cycle

Peptide Sequence	Count (#) [Fraction(%)]						Rank	# cDNA Mutants	K_D (nM)
	C1	C2	C3	C4	C5	C6			
vwdprtfylsri	0 [0.0%]	0 [0.0%]	2 [0.0%]	137 [3.2%]	7731 [51.9%]	4292 [76.6%]	1	82	3
WDANTIfIKRV.	0 [0.0%]	0 [0.0%]	0 [0.0%]	5 [0.1%]	201 [1.4%]	176 [3.1%]	2	13	4
WNPRTIfIKRA.	0 [0.0%]	0 [0.0%]	0 [0.0%]	4 [0.1%]	540 [3.6%]	132 [2.4%]	3	13	>1000
vwdprtfylsrT	0 [0.0%]	0 [0.0%]	0 [0.0%]	5 [0.1%]	285 [1.9%]	127 [2.3%]	4	18	3
IwdTGtfylsrT	0 [0.0%]	0 [0.0%]	0 [0.0%]	10 [0.2%]	354 [2.4%]	63 [1.1%]	5	11	7
WwNTrSfylsri	0 [0.0%]	0 [0.0%]	0 [0.0%]	22 [0.5%]	1163 [7.8%]	53 [0.9%]	6	12	12
Fwdprtfylsri	0 [0.0%]	0 [0.0%]	0 [0.0%]	2 [0.0%]	99 [0.7%]	38 [0.7%]	7	8	4
vwdpStfylsri	0 [0.0%]	0 [0.0%]	0 [0.0%]	0 [0.0%]	41 [0.3%]	35 [0.6%]	8	4	3
KwdTrtfylsrY	0 [0.0%]	0 [0.0%]	0 [0.0%]	9 [0.2%]	348 [2.3%]	31 [0.6%]	9	6	5
KwdTrtfylsri	0 [0.0%]	0 [0.0%]	0 [0.0%]	2 [0.0%]	97 [0.7%]	29 [0.5%]	10	8	6
Iwdprtfylsri	0 [0.0%]	0 [0.0%]	0 [0.0%]	0 [0.0%]	26 [0.2%]	26 [0.5%]	11	6	1
IwdTGtfylsri	0 [0.0%]	0 [0.0%]	0 [0.0%]	2 [0.0%]	60 [0.4%]	22 [0.4%]	12	5	>1000
vwdprtfylsrM	0 [0.0%]	0 [0.0%]	0 [0.0%]	2 [0.0%]	66 [0.4%]	22 [0.4%]	12	2	4
Awdprtfylsri	0 [0.0%]	0 [0.0%]	0 [0.0%]	4 [0.1%]	117 [0.8%]	18 [0.3%]	13	7	6
vwdSrtfylsri	0 [0.0%]	0 [0.0%]	0 [0.0%]	0 [0.0%]	28 [0.2%]	18 [0.3%]	13	5	3
vwdpGtfylsri	0 [0.0%]	0 [0.0%]	0 [0.0%]	2 [0.0%]	32 [0.2%]	18 [0.3%]	13	8	<1
vwdprtfyMsri	0 [0.0%]	0 [0.0%]	0 [0.0%]	1 [0.0%]	58 [0.4%]	16 [0.3%]	14	5	1
vwdprtfylsrS	0 [0.0%]	0 [0.0%]	0 [0.0%]	2 [0.0%]	74 [0.5%]	16 [0.3%]	14	7	6
vwdprtfylsrV	0 [0.0%]	0 [0.0%]	0 [0.0%]	0 [0.0%]	32 [0.2%]	16 [0.3%]	14	6	3
WNPRTIfIKRV.	0 [0.0%]	0 [0.0%]	0 [0.0%]	0 [0.0%]	19 [0.1%]	14 [0.2%]	15	3	>1000
vRdprtfylsri	0 [0.0%]	0 [0.0%]	1 [0.0%]	1 [0.0%]	26 [0.2%]	14 [0.2%]	15	7	>1000
vwdpKtfylsri	0 [0.0%]	0 [0.0%]	0 [0.0%]	1 [0.0%]	32 [0.2%]	13 [0.2%]	16	4	14
vwdprtfylsrN	0 [0.0%]	0 [0.0%]	0 [0.0%]	0 [0.0%]	68 [0.5%]	13 [0.2%]	16	5	>1000
FRFpFYIQRR..	0 [0.0%]	0 [0.0%]	0 [0.0%]	3 [0.1%]	79 [0.5%]	11 [0.2%]	17	3	>1000
Total Count (#)	**11122**	**12026**	**15578**	**4279**	**14882**	**5601**			

Enrichment of the 24 most frequent peptides found after the final (6^{th}) cycle of selection.
The peptide sequences are formatted so that amino acids in each position, if different from *the top peptide* (vwdprtfylsri), are capitalised; if the peptide sequence is shorter, the missing amino acid is indicated by the '.' at the end of the sequence.

Experimental

PHD2 immobilisation using magnetic beads

For the immobilisation of the target protein PHD2-bio, streptavidin magnetic Dynabeads M 280 Streptavidin (Life Technologies) were used. For 1 cycle of selection 100 µg (10 µL of beads slurry) was incubated with approximately 40 pmol PHD2-bio at slow rotation at 4°C for 30 min; with approximately 25 pmol PHD2 being immobilised on the beads. For the immobilisation of the target protein PHD2-bio, streptavidin magnetic Dynabeads M 280 Streptavidin (Life Technologies) were used.

PHD2-Activity Confirmation

Activity assay was initiated by mixing 50 µL (10 µM immobilised PHD2, and 600 µM 2OG in 100 mM NaCl and 50 mM Tris-Cl at pH 7.5) and 50 µL (50 µM CODD, 1 mM L-Asc, and 100 µM Fe(ii) and in 100 mM NaCl and 50 mM Tris-Cl at pH 7.5).

The reaction was incubated for 1 h at 37°C. The beads were removed from the solution. Activity of PHD2 was confirmed by observing a +16 Da mass shift on MALDI spectrum.

Selection

Selection against PHD2 from the mRNA-cyclic-peptides library was carried using a protocol previously described in details [Hayashi et al.]. The original cDNA library was transcribed into mRNA and ligated to puromycin using T4 ligase. Translation was performed using Met (−) FIT system and ClAc-D-Tyr-tRNA$^{fMet}_{CAU}$, used to initiate the peptide chain and to enable post-translational cyclisation with Cys residue in a peptide, thus enabling an intramolecular thioester.

To produce c-DNA-mRNA-cyclic-peptides library, the mRNA part of thus generated fusion was reverse transcribed using MMLV RNase H (−) (Promega). 100 mM Tris-HCl (pH 7.5) supplied with 200 mM NaCl and 0.1% tween was used as 2 × selection buffer (100 mM NaCl, 50 mM Tris-HCl at pH 7.5, and 0.1% Tween-20). 0.2% BSA in 2 × selection buffer was used as 2 × blocking solution.

To block active streptavidin, 1 µL 500 µM biotin was added to the beads and they were incubated for another 30 min. For a dummy incubation two 5 µL of bead slurry aliquots were used: one was incubated with 1 µL 500 µM biotin in 100 µL selection buffer for 30 min, another was incubated in 100 µL selection buffer for 30 min.

After incubation both aliquots were washed 3 times using 1 × selection buffer and combined together. Prior to the incubation with the immobilised PHD2, c-DNA-mRNA-cyclic-peptides library was diluted with equal volume of 2 × blocking solution.

SPR Analysis

K_D of the cyclic-peptides complexes with PHD2 were estimated using a Biacore T200 machine (GE Healthcare) equipped with Sensor Chip SA. PHD2-bio immobilisation on the chip Biotin CAPture Kit was performed following the standard protocol. The running buffer was 50 mM Tris-HCl (pH 7.5) supplied with 100 mM NaCl and 0.05% tween, and 0.1% DMSO. Complex formation was tested by injecting varying concentrations (4 nM to 1 024 nM) at a flow rate of 30 µl min−1 and measured by single-cycle kinetics method. All data were fitted to the standard 1:1 binding model.

Peptide Synthesis

Peptide synthesis was performed using Biotage® Syro Wave™. For a single peptide synthesis 48 mg NovaPEG Rink Amide resin was used. Protected amino acids and solvents were measured according to the standard protocol. After final de-protection step of Fmoc-synthesis 0.2 M ClAc-NHS (MW = 174 g • mol-1) solution in DMF, was added. The procedure was carried out twice for 30 min using 600 µL for per peptide per repeat. Resin was washed 5 times with DMF and 5 times with CH2Cl2; afterwards it was dried under vacuum for 30 min. For de-protection TFA containing TIS, EDT and H2O 2.5% (v/v) each was used. This cocktail was added to each sample to cover the resin (~2 mL per sample) and mixed gently. Samples were incubated at RT with constant mixing for 1.5 hours. Supernatant was collected and the resin was washed twice using 1 mL TFA. The washes were combined with collected supernatant. The samples were concentrated for 30-60 min using centrifugal evaporator at 40°C. Peptides were precipitated by addition of of 10 mL ice-cold Et2O and following manual centrifugation. Peptides have been washed 3 times using 5 mL ice-cold Et2O and dried briefly (~5 min) in the centrifugal evaporator at 20°C. Peptides were reconstituted in 500 µL DMSO, followed by addition of 5 mL MeCN. For cyclisation, the peptide solution was alkalised using 20 µL TEA;

and the mixture has been incubated at 20ºC for 1 hour at constant mixing. To quench the reaction the mixture has been quenched using 50 µL FA. Cyclisation was confirmed using MALDI-MS and α-CHCA matrix. HPLC purification Peptides were purified using Aeris PEPTIDE column, 250 x 21.2 mm, C18 5 µm, 100 Å, which was operated at 20 mL•min-1. For separation, linear gradient

10 mM ammonium formate and 30 mM formic acid in 5% MeCN to 10mM ammonium formate and 10 mM formic acid 95% MeCN over 40 min was used. 5 mL sample was loaded on the column. Peptide fractions were dried from MeCN for 60-90 min using centrifugal evaporator at 40ºC and lyophylised using freeze-drying. The resulting peptides were reconstituted in DMSO. The purity was confirmed using MALDI-MS and α-CHCA matrix.