# 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 KD 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 enrichment in certain ligands is primarily driven by 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.

Unfortunately, currently existing experimental methods do not allow one to closely monitor selection from the libraries with diversities 109 to 1012 unique ligands. 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.

Thermodynamic ligand-target complex formation-dissociation equilibrium for a given ligand (n) is described by equations:

LT ⇌ Ln + T

kD,n = [Ln]・[T]/[LnT]

[LnT] = [Ln]・[T]/kD,n

The probability of a target protein being bound (PT,b) is described by equation:

Pb,T= [LT]/([T] +[LT]) = ([L]/kD)/(1 + [L]/kD)

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

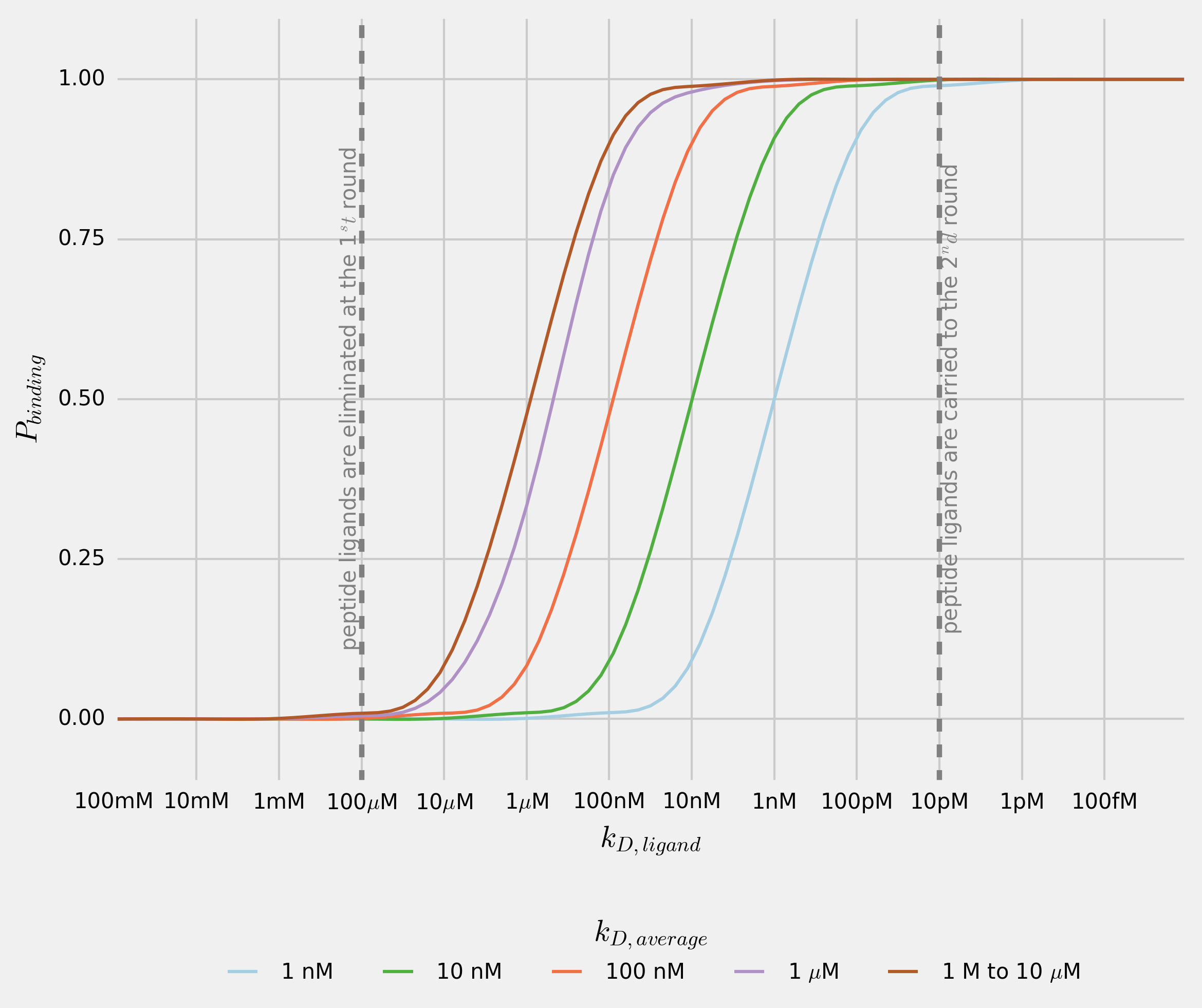
Thus, the amount of the unbound target protein ([T]) can be expressed as follows:

[T] = (1 - P(b,T))・T

The probability of a given ligand being bound (Pn,b) to a target protein is described by equation:

Pb,n = [LnT]/([Ln] +[LnT]) = ([T]/KD,n)/(1 + [T]/KD,n).

Figure\_\_ shows the probability of a ligand being bound to a target protein depending on a given ligand KD, and the average KD 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.



Figure\_\_\_. Probability of a ligand (n) being bound to a target protein dependence on a ligand-target complex dissociation constant (KD,n). Every line shows the probability of a ligand (n) being bound to a target protein for total-ligand-population average KD in the range 1 nM to 10 µM.

The concentration of a target protein in during incubation with cDNA-mRNA-peptide fusion libraty 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 KD values are found after selection, total library average KD can be safely assumed to be at 10 µM or higher. It can be seen that a ligand which complex with the target protein KD 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 present 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 enrichment 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. While it is not possible to completely eliminate the effect of PCR bias, introduces 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.

Previously suggested introduction of a ‘warhead’ can be defined more specifically: to ensure the recovery of the ligand after the first cycle of selection, the ‘warhead’ has to ensure KD at 1 nM or lower. 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 KD value, and the more likely it is that he peptides bound will have lower KD values.

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.

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

While it may be impossible to eliminate the undesirable interference between a ligand and a target protein or to 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, thus eliminating the PCR bias. It was shown that some peptides with higher affinity towards target protein yet amplifying at a lower rate, can be easily disregarded. It is hoped that the analysis suggested in this paper may uncover such ligands, thus increasing the rate of successful ligand discoveries.