Published in final edited form as:

Angew Chem Int Ed Engl. 2016 March 14; 55(12): 4007–4010. doi:10.1002/anie.201600077.

High Throughput Measurement of Binding Kinetics via mRNA Display and Next-Generation Sequencing

Dr. Farzad Jalali-Yazdi,

Department of Chemical Engineering and Materials Science

Lan Huong Lai,

Department of Chemistry

Prof. Terry T. Takahashi, and

Department of Chemistry

Prof. Richard W. Roberts

Department of Chemical Engineering and Materials Science. Department of Chemistry.

Department of Molecular Computational Biology, USC Norris Comprehensive Cancer Center, University of Southern California, Los Angeles, Ca, 90089 (USA)

Richard W. Roberts: richrob@usc.edu

Abstract

There is great demand for high-throughput methods to characterize ligand affinity. By combining mRNA display with next-generation sequencing, we determined the kinetic on- and off-rates for over twenty thousand ligands, without the need for synthesis or purification of individual members. Our results are reproducible and as accurate as other methods of affinity measurement.

Graphical Abstract

We were able to obtain the kinetic on- and off-rates of over twenty thousand individual ligands for their target protein, without the need to synthesize each individual ligand separately. We accomplished this by combining mRNA display and high throughput DNA sequencing

Keywords

High-throughput screening; Protein-protein interactions; Dissociation Constant (K_d); Affinity Measurement; Kinetic on- and off-rates

Correspondence to: Richard W. Roberts, richrob@usc.edu.

Supporting information for this article is given via a link at the end of the document.

Various *in vitro* selection techniques (e.g., phage display,^[1] ribosome display,^[2] and mRNA display^[3]) facilitate the generation of polypeptide ligands against targets of interest. Recent advances combining *in vitro* selection with high-throughput sequencing have greatly accelerated the process of generating large lists of potential ligands.^[4] The challenge, increasingly, is to rank the molecules based on desirable properties, chiefly, their affinity for their targets.^[5]

Initially, we hypothesized that the affinity of a ligand would directly correlate to its frequency rank-order in an affinity-enriched pool of ligands. Although we have shown with mRNA display that higher ranked sequences do exhibit functionality,^[4] we observe that a sequence's rank does not accurately predict binding affinity. One source of this variation is that all sequences above the threshold of each enrichment step will show near quantitative pull-down. The enrichment efficiency of high affinity vs. ultrahigh affinity sequences thus can depend on other factors during selection such as PCR bias, transcription bias, ligation bias, translation efficiency and the efficiency of fusion formation. This means that while after a pool's convergence, the highest represented sequences exhibit functionality, their rank order is poorly correlated to their affinity.

Due to this effect, there is a great need for methods that evaluate the affinity of a ligand for its target in a high throughput manner. Advances in the field have increased the throughput of K_d measurements using radioactivity,^[5a] SPR or fluorescent microarrays,^[6] and ELISA assays.^[5b] However, all these methods require individually expressed and purified ligands, greatly reducing their throughput. Measuring the K_d for thousands of potential ligands simultaneously has not yet been achieved.

In this work, we combined high throughput DNA sequencing with mRNA display to obtain kinetic on- and off-rates, and consequently K_d values, for tens of thousands of ligands simultaneously. To demonstrate our method, we chose two enriched pools from our selection against B-cell lymphoma extra-large protein (Bcl-x_L) (Takahashi and Roberts, manuscript in preparation). The first pool is the final enriched pool from a selection resulting in 21 amino acid long peptide ligands against Bcl-x_L (extension selection). The second pool is the final enriched pool of a doped (biased) selection based on one of the top sequences (E1) from the extension selection to further optimize binding. mRNA from both pools were ligated to a 3' DNA linker attached to puromycin, *in vitro* translated, purified and reverse transcribed.^[3] A small fraction of each pool was also translated using radiolabeled methionine to track pool binding.

To obtain on-rates by High Throughput Sequencing Kinetics (HTSK), a library of mRNA-peptide fusions was first mixed with Bcl-x_L immobilized on magnetic beads. A portion of the beads was removed at various time points, washed, PCR amplified, and sent for next-generation sequencing. A sample calculation is shown for peptide sequence E5 in Figure 1a. High throughput sequencing of each time point allowed identification of all the ligands bound to the beads at that point as well as each sequence's frequency. We were able to calculate each sequence's fractional composition by dividing the sequence frequency by the total number of sequences at each time point (Figure 1b, left panel). Sequences with fast on-rates bind to the target quickly; therefore, they have a high fractional composition at early

time points. As time passes, more slow on-rate ligands bind to the beads, reducing the composition fraction of the fast on-rate ligands.

Separately, using the radiolabeled samples, we measured the total amount of peptide bound to the beads at each time point (Figure 1b, middle panel). The amount of radioactivity at each time point represents the sum of all the peptides bound to the beads at that point. In cases with ligand pools of small diversity, where the composition of pool can vary significantly during the association and dissociation phases, it will be prudent to normalize the radiolabeled signal to the average number of methionine residues per peptide in each pool to obtain a more accurate measure of total peptides bound to beads. This normalization will only be needed where radiolabeled binding is being used to measure the total peptide bound at each time point and where diversity is low. The radiolabeled binding is not the only method of measuring the total amount of peptide bound to beads. Similar calculations can be performed using immunosorbent or fluorescence assays, or simply by quantitating the amount of DNA/RNA bound to the beads. To obtain the kinetic on-rates for each ligand, we simply multiplied each ligand's fractional composition by the total amount of peptide bound to beads. This results in a measure of binding for each sequence as a function of time (Figure 1b, right panel).

Using this analysis, and knowing the concentration of immobilized Bcl-x_L, we obtained the kinetic on-rate for each sequence by fitting the binding data to a simple kinetic on-rate equation (see Figure S1 in the Supporting Information). Since the concentration of Bcl-x_L (7 nM) was much higher than the concentration of the mRNA-peptide fusion molecules (<1 nM), the fraction of ligand bound is not a function of ligand concentration. The contribution of the dissociation-rate to the binding equation has been removed because in the small time scale of this experiment (~45 minutes) and given the slow off-rate of the sequences tested ($2 \times 10^{-6} \text{ s}^{-1}$ on average), the contribution of the dissociation rate is minimal. This allows for independent calculation of on- and off-rates (see Figure S1 in the Supporting Information).

To obtain the HTSK off-rates, we followed a similar approach. After the kinetic on-rate experiment, the remaining beads were washed and excess Bcl-x_L was added in solution to prevent rebinding of dissociated ligands to the beads under pseudo-first order binding conditions. A small fraction of beads was removed at various time points, washed, PCR amplified, and sent for next-generation sequencing (Figure 1c, left panel). By multiplying each sequence's fractional composition by the total radiolabeled peptides still bound at each time point (Figure 1c, middle panel), we were able to obtain the amount of each peptide still bound as a function of time. A simple exponential decay fit was then used to calculate the kinetic off-rate (Figure 1c, right panel). The data for generating Figure 1c is presented in Table S1 in the Supporting Information. We were also able to obtain the pool's on- and off-rates by quantitating the amount of DNA on the beads at each time point. The kinetic rates obtained by DNA quantitation matched the rates obtained by radiolabeled binding (<30% deviation, Table S2 in the Supporting Information).

Figure 2a shows the K_d obtained for the 50 highest frequency ligands in each tested pool. As expected, the ligands in the doped pool exhibit a higher affinity on average than the ligands in the extension pool. It is also clear that frequency rank poorly correlates to sequence

affinity. To show the reproducibility of the obtained kinetic constants, we compared the obtained values for the 40 ligands that appeared in both the extension and doped pools (Figure 2b). The results show that the HTSK values are remarkably reproducible and highly precise.

In order to check the validity of the obtained results, we tested the off-rate of several ligands using *in vitro* translated radiolabeled peptides. The peptide ligands were made using a C-terminal HA tag, and affinity purified. The off-rate of the radiolabeled peptides was then measured. Figure 2c shows the HTSK vs. radiolabeled peptide off-rates. The HTSK off-rates correlate very well to the radiolabeled peptide off-rates, however, there is a consistent bias between the two methods. The measured bias is ~7-fold for the fastest off-rate clone, and less than 2-fold for the slowest off-rate clones. This bias is relatively small in comparison to biases measured between other established methods for affinity measurement, which frequently vary by as much as 60-fold.^[6b, 8] One contributing factor to this difference could be the context of binding. The HTSK results are obtained for mRNA-DNA-peptide fusion molecules whereas the radiolabeled k_{off} values are for the peptide with a short C-terminal HA tag. To further demonstrate the accuracy of this assay, we compared the HTSK K_d values for two of the peptides with previously published results.^[9] E1 peptide has a K_d value of 39 ± 6 pM by ELISA, and 23 ± 2 pM by HTSK, and D1 peptide has a K_d value of 9 ± 2 pM by ELISA and 15 pM by HTSK (see Table S3 in the Supporting Information).

Using HTSK, we identified peptide D79 (frequency rank of 79 in the doped selection pool) with a k_{off} value of 5.9×10^{-7} , over three times slower than the previously identified slowest off-rate peptide ligand (D1) or the biotin-streptavidin interaction (Figure 2d). We also identified peptide E1452 (frequency rank of 1452 from the extension selection pool) with the k_{off} value of 8.5×10^{-7} , over two fold slower than D1 (see Figure S2 in the Supporting Information). Indeed, in this modest chain length (21 amino acids long), using HTSK, we have identified thousands of sequences with 10 pM K_d or better (~2,600 in Supporting Dataset). The presence of rare sequences with higher affinities than the most frequent sequences suggests the need for testing lower abundant sequences for functionality. While this is not practical when individual sequences must be synthesized and tested, our HTSK method provides a viable approach to testing thousands of sequences simultaneously.

While we used mRNA display for determining HTSK of an *in vitro* library, the HTSK approach is directly transferable to aptamer selection techniques or any monomeric genotype-phenotype linked display system (e.g., ribosome display). The results from such a high-throughput analysis could be used not only to find the highest affinity binders, but also to obtain structural information from the mutational analysis of a protein.^[10] Here, we have shown our HTSK method to be reproducible and accurate, and have identified the highest affinity peptide-protein interaction yet discovered.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This work was supported by NIH grants R01AI085583 (R.W.R.) and R01CA170820 (R.W.R. and T.T.T.) and the Ming Hsieh Institute for Research on Engineering-Medicine for Cancer (RWR). The authors made use of the USC Nanobiophysics core and the USC Genome & Cytometry Core as part of this work. We thank Mehmet Cetin for providing the initial version of the python code used in the analysis of high throughput DNA sequencing results.

References

1. McCafferty J, Griffiths AD, Winter G, Chiswell DJ. Nature. 1990; 348:552–554. [PubMed: 2247164]
2. Mattheakis LC, Bhatt RR, Dower WJ. Proceedings of the National Academy of Sciences of the United States of America. 1994; 91:9022–9026. [PubMed: 7522328]
3. Roberts RW, Szostak JW. Proceedings of the National Academy of Sciences of the United States of America. 1997; 94:12297–12302. [PubMed: 9356443]
4. Olson CA, Nie J, Diep J, Al-Shyoukh I, Takahashi TT, Al-Mawsawi LQ, Bolin JM, Elwell AL, Swanson S, Stewart R, Thomson JA, Soh HT, Roberts RW, Sun R. Angew Chem Int Ed Engl. 2012; 51:12449–12453. [PubMed: 23125174]
5. a Larsen AC, Gillig A, Shah P, Sau SP, Fenton KE, Chaput JC. Analytical chemistry. 2014; 86:7219–7223. [PubMed: 24970615] b Jalali-Yazdi F, Corbin JM, Takahashi TT, Roberts RW. Analytical chemistry. 2014; 86:4715–4722. [PubMed: 24749546]
6. a Landry JP, Fei Y, Zhu X. Assay and drug development technologies. 2012; 10:250–259. [PubMed: 22192305] b Wassaf D, Kuang G, Kopacz K, Wu QL, Nguyen Q, Toews M, Cosic J, Jacques J, Wiltshire S, Lambert J, Pazmany CC, Hogan S, Ladner RC, Nixon AE, Sexton DJ. Analytical biochemistry. 2006; 351:241–253. [PubMed: 16510109] c Usui-Aoki K, Shimada K, Nagano M, Kawai M, Koga H. Proteomics. 2005; 5:2396–2401. [PubMed: 15880825]
7. Piran U, Riordan WJ. Journal of immunological methods. 1990; 133:141–143. [PubMed: 2212686]
8. a Estep P, Reid F, Nauman C, Liu Y, Sun T, Sun J, Xu Y. mAbs. 2013; 5:270–278. [PubMed: 23575269] b Landry JP, Ke Y, Yu GL, Zhu XD. Journal of immunological methods. 2015; 417:86–96. [PubMed: 25536073]
9. Jalali-Yazdi F, Takahashi TT, Roberts RW. Analytical chemistry. 2015; 87:11755–11762. [PubMed: 26485531]
10. a Cunningham BC, Wells JA. Science. 1989; 244:1081–1085. [PubMed: 2471267] b Olson CA, Wu NC, Sun R. Current biology : CB. 2014; 24:2643–2651. [PubMed: 25455030]

Figure 1.

Obtaining kinetic rates for ligands using HTSK. **a)** Transforming high throughput sequencing data for peptide E5 into usable form. The pool of mRNA-peptide fusion molecules was incubated with Bcl-xL (immobilized on beads). At each time points, a fraction of beads were washed and sequenced via next-generation sequencing (association phase). The behavior of the pool was characterized by radiolabeled binding. To characterize the dissociation rate, ligand-bound beads were placed in a solution with excess Bcl-xL to prevent dissociated ligands from re-binding to the beads. Time points were obtained similar to the association phase (where time = 0 is the start of dissociation phase). Dividing E5's frequency by the total number of sequences in each pool yields the fractional composition at each time point. By multiplying E5's composition fraction by the radiolabeled binding, we obtain the ligand's contribution to the total radiolabeled binding at each time point. **b)** Obtaining the kinetic on-rate. The fraction of each ligand at each time point was calculated from the sequencing data and normalized with respect to the final data point (left). Separately, the pool's binding was determined at each time point (middle). By multiplying each ligand's composition fraction by the radiolabeled binding at each data point, we obtained the ligand's contribution to the radiolabeled binding, and subsequently, the on-rate

(right). **c)** Obtaining the kinetic off-rate. The fraction of each ligand at each time point was calculated from the sequencing data and normalized with respect to the first data point (left). The counts remaining on the beads at each time point were measured using the radiolabeled sample (middle). By multiplying each ligand's composition fraction by the radiolabeled binding at each data point, we obtained the ligand's contribution to the radiolabeled binding and the off-rate (right).

Figure 2.

The HTSK results are reproducible and accurate. **a)** The obtained K_d for the top 50 clones in the extension and doped pools. While the extension pool on average (dashed red line) is comprised of lower affinity binder than the doped pool (dashed blue lines), some sequences in the extension pool show higher affinity than the doped pool average. **b)** The obtained HTSK values are reproducible. 40 sequences appeared in both the extension and the doped pools. Comparing the kinetic constants for these sequences shows that the results are reproducible. **c)** The k_{off} value obtained by HTSK correlate well to the values obtained using radiolabeled peptides. There is a consistent bias in the measured off-rate values for the two methods of measurements. **d)** The radiolabeled peptide off-rate for the previously identified sequences E1 and D1, and the HTSK identified sequence D79. The off-rate for sequence D79 is over 3 times slower than the off-rate of D1, the previously identified highest affinity binder. The slowest reported value for the off-rate of biotin and streptavidin in the literature (2.4×10^{-6})^[7] is shown as a reference.