Affinity Measurements Hot Paper

International Edition: DOI: 10.1002/anie.201600077
German Edition: DOI: 10.1002/ange.201600077

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

*Farzad Jalali-Yazdi, Lan Huong Lai, Terry T. Takahashi, and Richard W. Roberts**

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 those obtained with other methods of affinity measurement.*

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 current challenge is to rank the molecules based on their desirable properties, in particular in terms of their target affinity.[5]

Initially, we hypothesized that the affinity of a ligand would directly correlate with 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 observed that the rank of a sequence does not accurately predict its 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 versus ultrahigh-affinity sequences can thus depend on other factors during selection, such as PCR bias, transcription bias, ligation bias, translation efficiency, and the efficiency of the fusion formation. This means that whereas after the convergence of a pool of ligands, the highest represented sequences exhibit functionality, their rank order is poorly correlated to their affinity.

[*] Dr. F. Jalali-Yazdi, Prof. R. W. Roberts
Department of Chemical Engineering and Materials Science
University of Southern California
Los Angeles, CA, 90089 (USA)
E-mail: richrob@usc.edu
L. Huong Lai, Prof. T. T. Takahashi
Department of Chemistry
University of Southern California
Los Angeles, CA, 90089 (USA)
Prof. R. W. Roberts
Department of Molecular Computational Biology
USC Norris Comprehensive Cancer Center
University of Southern California
Los Angeles, CA, 90089 (USA)
Supporting information and the ORCID identification number(s) for the author(s) of this article can be found under http://dx.doi.org/10.1002/anie.201600077.

Owing 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 by using radioactivity,[5a] SPR or fluorescence microarrays,[6] and ELISA assays.[5b] However, all of these methods require individually expressed and purified ligands, greatly reducing their throughput. Measuring the K_d value for thousands of potential ligands simultaneously has not yet been achieved.

Herein, 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 the applicability of our method, we chose two enriched pools from our selection against B-cell lymphoma extra-large protein (Bcl-x_L).[11] 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.

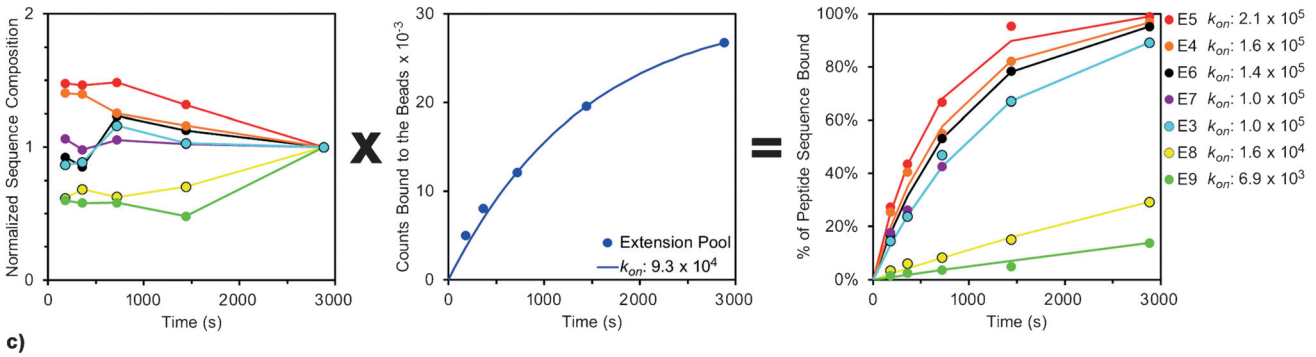
To obtain on-rates by high-throughput sequencing kinetics (HTSK), a library of mRNA–peptide fusions was first mixed with Bcl-x_L immobilized on magnetic beads. A portion of the beads was removed at various time points, washed, PCR amplified, and sent for next-generation sequencing. A sample calculation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

We separately measured the total amount of peptide bound to the beads at each time point using the radiolabeled samples (Figure 1b, middle panel). The amount of radioactivity at each time point represents the sum of all of the peptides bound to the beads at that point. For ligand pools of small diversity, where the composition of a pool can vary significantly during the association and dissociation phases, it will be prudent to normalize the radiolabeled signal to the

a)

	Association					Dissociation					
	Time Point 1	Time Point 2	Time Point 3	Time Point 4	Time Point 5	Time Point 1	Time Point 2	Time Point 3	Time Point 4	Time Point 5	Time Point 6
Time (s)	180	360	720	1440	2880	1980	9000	30540	80340	171600	428940
E5's Frequency in Pool	115	102	115	115	73	101	103	120	114	109	89
Total number of Sequences in Pool	2851754	2542137	2838312	3204532	2671767	3159350	3330810	3331534	3019767	3056863	2903743
E5's Fractional Composition (PPM)	40	40	41	36	27	32	31	36	38	36	31
Pool's Radiolabeled Counts on Beads	5017	8031	12142	19567	26750	24175	23006	22634	17437	14599	9962
Amount of E5 on Beads (PPM x Counts)	202316	322234	491958	702195	730883	772841	711424	815264	658 269	520563	305336

b)



c)

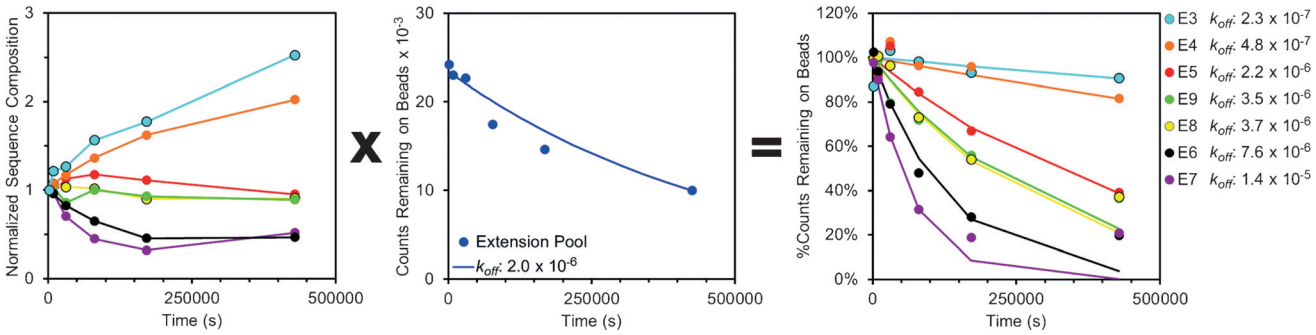


Figure 1. Obtaining kinetic rates for ligands by HTSK. a) Transforming high-throughput sequencing data for peptide E5 into usable form. The pool of mRNA–peptide fusion molecules was incubated with Bcl-x_L (immobilized on beads). At each time point, a fraction of the beads was washed and subjected to next-generation sequencing (association phase). The behavior of the pool was characterized by using radiolabeled ligands. To determine the dissociation rate, ligand-bound beads were placed in a solution with excess Bcl-x_L to prevent dissociated ligands from re-binding to the beads. Time points were obtained similar to the association phase (where $t = 0$ is the start of the dissociation phase). Dividing the frequency of E5 by the total number of sequences in each pool yields the fractional composition at each time point. By multiplying the composition fraction of E5 obtained by the radiolabeling approach, we obtain the ligand's contribution to the total radiolabeled binding at each time point. b) Determining 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 binding of a pool was determined at each time point (middle). By multiplying the composition fraction of each ligand 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) Determining 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).

average number of methionine residues per peptide in each pool to obtain a more accurate measure of the total number of peptides bound to the beads. This normalization will only be needed when radiolabeling is used to measure the total amount of bound peptide at each time point and when the pool diversity is low. The radiolabeling approach is not the only method of measuring the total amount of peptide bound to the 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 the fractional composition of each ligand by the total amount of peptide bound to the beads. This results in a measure for the binding of 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 the Supporting Information, Figure S1). As 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 bound ligand is not a function of the ligand concentration. The contribution of the dissociation rate to the binding equation has been removed because on the

small time scale of this experiment (ca. 45 min) and given the low off-rate of the sequences tested ($2\times10^{-6}\ s^{-1}$ on average), the contribution of the dissociation rate is minimal. These assumptions enable the independent calculation of the on- and off-rates (see Figure S1).

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 the beads was removed at various time points, washed, PCR amplified, and sent for next-generation sequencing (Figure 1c, left panel). By multiplying the fractional composition of each sequence by the total number of 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. We were also able to obtain the on- and off-rates of the pool 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 the radiolabeling approach (<30% deviation; see Table S2).

Figure 2a shows the K_d values obtained for the 50 highest-frequency ligands in each tested pool. As expected, on average, the ligands in the doped pool exhibit a higher affinity than the ligands in the extension pool. It is also clear that the frequency rank poorly correlates with sequence affinity. To show the reproducibility of the obtained kinetic constants, we compared the values obtained 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.

To confirm 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 k_{off} values versus the off-rates obtained with radiolabeled peptide. The two sets of values correlate very well; however, there is a consistent bias between the two methods. The measured bias is approximately sevenfold for the fastest off-rate clone, and less than twofold for the lowest off-rate clones. This bias is relatively small in comparison to those obtained for other established methods for affinity measurements, which frequently vary by as much as a factor of 60.[6b,8] One factor that contributes to this difference could be the context of binding. The HTSK results are obtained for mRNA–DNA–peptide fusion molecules whereas the k_{off} values obtained by the radiolabeling approach 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] The E1 peptide has a K_d value of 39 ± 6 pM by ELISA and 23 ± 2 pM by HTSK, and the D1 peptide has a K_d value of 9 ± 2 pM by ELISA and 15 pM by HTSK (see Table S3).

Using HTSK, we identified peptide D79 (frequency rank 79 in the doped selection pool) with a k_{off} value of 5.9×10^{-7}, which is over three times lower than that of the previously identified lowest off-rate peptide ligand (D1) or the biotin–streptavidin interaction (Figure 2d). We also identified peptide E1452 (frequency rank 1452 in the extension selection pool) with a k_{off} value of 8.5×10^{-7}, which is over twofold

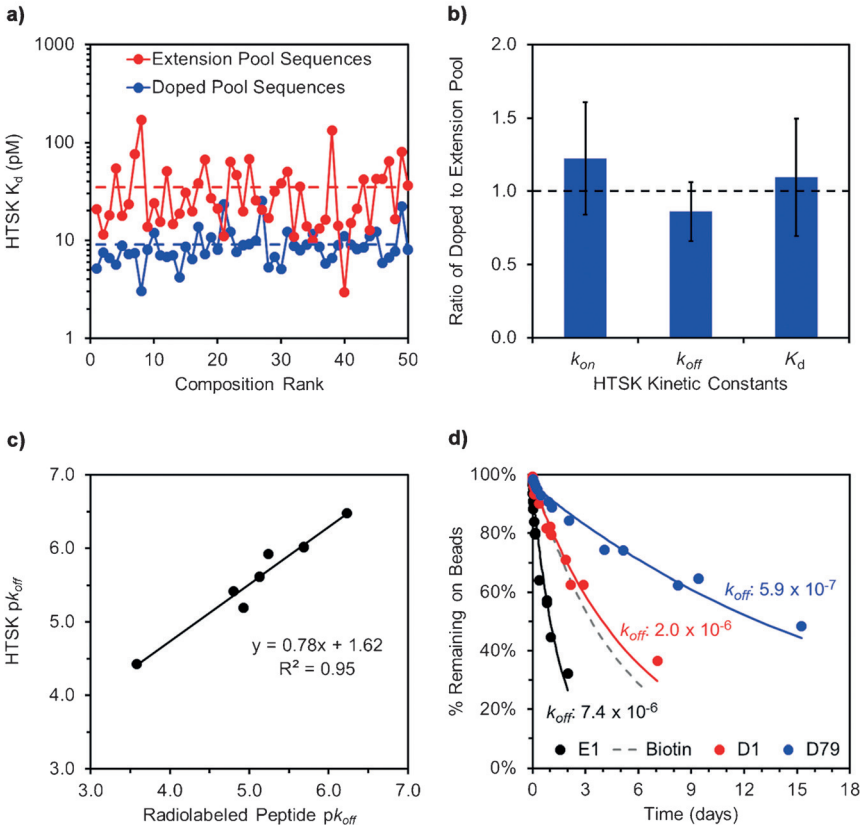


Figure 2. The HTSK results are reproducible and accurate. a) The obtained K_d values for the top 50 clones in the extension and doped pools. Whereas on average, the extension pool (red line) consists of lower-affinity binders than the doped pool (blue line), some sequences in the extension pool showed higher affinity than the doped-pool average. b) The obtained HTSK values are reproducible. Forty 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} values obtained by HTSK correlate well with the values obtained using radiolabeled peptides. There is a consistent bias in the measured off-rate values for the two methods. d) The off-rates for the radiolabeled peptide of the previously identified sequences E1 and D1, and the HTSK-identified sequence D79. The off-rate for sequence D79 is over three times lower than the off-rate of D1, the previously identified highest-affinity binder. The lowest reported value for the off-rate of biotin and streptavidin in the literature (2.4×10^{-6})[7] is shown as a reference.

lower than that of D1 (see Figure S2). Indeed, for this modest chain length (21 amino acids), using HTSK, we have identified thousands of sequences with K_d values of 10 pM or better by HTSK (ca. 2600 in the supporting dataset). The presence of rare sequences that display higher affinities than the most frequent sequences suggests the need for testing less abundant sequences for functionality. Whereas 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.

Although we used mRNA display for determining the 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 discovered thus far.

Acknowledgements

This work was supported by the NIH (R01AI085583 to R.W.R. and R01CA170820 to R.W.R. and T.T.T.) and the Ming Hsieh Institute for Research on Engineering-Medicine for Cancer (R.W.R.). We 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 the high-throughput DNA sequencing results.

Keywords: affinity measurements · dissociation constants · high-throughput screenings · kinetics · protein–protein interactions

How to cite: *Angew. Chem. Int. Ed.* **2016**, *55*, 4007–4010
Angew. Chem. **2016**, *128*, 4075–4078

[1] J. McCafferty, A. D. Griffiths, G. Winter, D. J. Chiswell, *Nature* **1990**, *348*, 552–554.
[2] L. C. Mattheakis, R. R. Bhatt, W. J. Dower, *Proc. Natl. Acad. Sci. USA* **1994**, *91*, 9022–9026.
[3] R. W. Roberts, J. W. Szostak, *Proc. Natl. Acad. Sci. USA* **1997**, *94*, 12297–12302.
[4] C. A. Olson, J. Nie, J. Diep, I. Al-Shyoukh, T. T. Takahashi, L. Q. Al-Mawsawi, J. M. Bolin, A. L. Elwell, S. Swanson, R. Stewart, J. A. Thomson, H. T. Soh, R. W. Roberts, R. Sun, *Angew. Chem. Int. Ed.* **2012**, *51*, 12449–12453; *Angew. Chem.* **2012**, *124*, 12617–12621.
[5] a) A. C. Larsen, A. Gillig, P. Shah, S. P. Sau, K. E. Fenton, J. C. Chaput, *Anal. Chem.* **2014**, *86*, 7219–7223; b) F. Jalali-Yazdi, J. M. Corbin, T. T. Takahashi, R. W. Roberts, *Anal. Chem.* **2014**, *86*, 4715–4722.
[6] a) J. P. Landry, Y. Fei, X. Zhu, *Assay Drug Dev. Technol.* **2012**, *10*, 250–259; b) D. Wassaf, G. Kuang, K. Kopacz, Q. L. Wu, Q. Nguyen, M. Toews, J. Cosic, J. Jacques, S. Wiltshire, J. Lambert, C. C. Pazmany, S. Hogan, R. C. Ladner, A. E. Nixon, D. J. Sexton, *Anal. Biochem.* **2006**, *351*, 241–253; c) K. Usui-Aoki, K. Shimada, M. Nagano, M. Kawai, H. Koga, *Proteomics* **2005**, *5*, 2396–2401.
[7] U. Piran, W. J. Riordan, *J. Immunol. Methods* **1990**, *133*, 141–143.
[8] a) P. Estep, F. Reid, C. Nauman, Y. Liu, T. Sun, J. Sun, Y. Xu, *MABS* **2013**, *5*, 270–278; b) J. P. Landry, Y. Ke, G. L. Yu, X. D. Zhu, *J. Immunol. Methods* **2015**, *417*, 86–96.
[9] F. Jalali-Yazdi, T. T. Takahashi, R. W. Roberts, *Anal. Chem.* **2015**, *87*, 11755–11762.
[10] a) B. C. Cunningham, J. A. Wells, *Science* **1989**, *244*, 1081–1085; b) C. A. Olson, N. C. Wu, R. Sun, *Current Biol.* **2014**, *24*, 2643–2651.
[11] T. T. Takahashi, R. W. Roberts, unpublished results.

Received: January 4, 2016
Revised: February 2, 2016
Published online: February 23, 2016

