Highly Sensitive in Vitro Selections for DNA-Linked Synthetic Small Molecules with Protein Binding Affinity and Specificity

Jeffrey B. Doyon, Thomas M. Snyder, and David R. Liu*

Department of Chemistry and Chemical Biology, Harvard University, Cambridge, Massachusetts 02138

Received May 11, 2003; E-mail: drliu@fas.harvard.edu

The discovery of synthetic molecules with desired properties is a central challenge of chemistry. Researchers typically evaluate synthetic molecules using a screen, a process by which compounds are individually assayed. In contrast, researchers (and nature) have discovered functional biological molecules using *selections*,[1] processes that physically separate molecules with favorable properties from inactive molecules. Selections offer much higher potential throughput than screens because all molecules are processed simultaneously, typically do not require sophisticated equipment, and unlike screens can be iterated to multiply the net enrichment of desired molecules. In addition, the outcomes of laboratory and biological selections are typically linked to amplifiable nucleic acids, enabling selections to offer far greater sensitivities than screens.

Despite these attractions, selections for synthetic molecules are largely unexplored because of the challenging requirements associated with their implementation. These requirements include (i) solution-phase libraries, and (ii) a means of identifying each possible species surviving the selection. In addition, each molecule entering a selection is ideally associated with an amplifiable information carrier that uniquely identifies that molecule. Biological systems can satisfy these criteria when selecting proteins or nucleic acids,[1] but are not compatible with most synthetic structures. Solution-phase libraries analyzed using peptide nucleic acid tags[2a] or mass spectrometry[2b,c] can meet the first two requirements, but do not enable signal amplification, a crucial aspect of biological selections that confers their unmatched sensitivity. Solid-phase libraries cosynthesized with nucleic acid tags[2d] offer an amplifiable signal but cannot be subjected to selections because the libraries are not in solution.

The covalent linkage of DNA oligonucleotides to corresponding synthetic molecules, either as a consequence of DNA-templated organic synthesis[3] or as a result of conjugating DNA to synthetic molecules, in theory enables synthetic molecules to satisfy the above requirements. Here, we report general in vitro selections for DNA-linked synthetic small molecules with protein binding affinity and specificity.

We prepared a variety of small molecules conjugated to DNA oligonucleotides such that each small molecule is linked to a unique DNA sequence. Small molecules were chosen either for their known affinities to one of six proteins or as nonbinding negative controls (Figure 1). Solutions containing mixtures of DNA-linked protein ligands and DNA-linked negative controls simulated DNA-templated small molecule libraries containing small fractions of library members with protein binding activities. Selections for protein affinity were performed by incubating these mixtures for 1−2 h with target proteins covalently linked to beads (see the Supporting Information). Nonbinders were removed by washing, and molecules that remained bound to beads were added to PCR reactions to amplify the oligonucleotides surviving selection. Sequences encoding protein ligands or nonligands were distinguished using restriction endonucleases that only cleave the ligand-encoding sequences. The efficiency of each selection was assessed by the degree to which
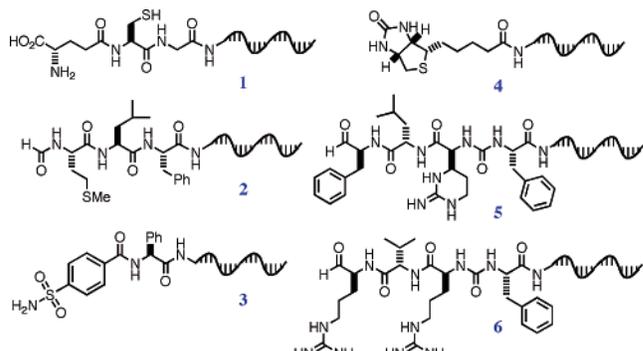

DNA-linked protein ligand	negative control	target protein	predicted activity	enrichment factor	sensitivity (mol)
1	2	glutathione S-transferase	$K_d = 10\ \mu M^{4a}$	2,500	10^{-20}
3	2 or 4	carbonic anhydrase	$K_d = 0.9\ nM^{4b}$	330	10^{-20}
4	2	streptavidin	$K_d = 40\ fM^{4c}$	4,400	10^{-18}
5	4	papain	$IC_{50} = 14\ \mu M^{4d}$	64	10^{-16}
5	4	chymotrypsin	$IC_{50} = 290\ nM^{4d}$	76	10^{-16}
6	4	papain	$IC_{50} = 270\ nM^{4d}$	98	10^{-18}
6	4	trypsin	$K_d = 100\ nM^{4d}$	125	10^{-17}

Figure 1. Enrichment factors for a single affinity selection containing a 1:1000 ratio of protein-binding:nonbinding molecules are shown in column 5. Minimum quantities of protein-binding molecules to achieve an enrichment factor of at least 50-fold are listed in column 6.

DNA-linked protein ligands were enriched relative to DNA-linked nonbinders (the "enrichment factor").

Glutathione amide binds to glutathione S-transferase (GST) with modest affinity ($K_d = \sim 10\ \mu M^{4a}$) and therefore represents a stringent test of protein binding selections. $10^3 - 10^7$ DNA-linked glutathione amide molecules (**1**) were combined with a 100- to 10^6-fold molar excess of a DNA-linked negative control (**2**), and the resulting mixtures were selected for binding to GST-linked agarose beads. The selection enriched as few as 10^4 copies of the DNA-linked glutathione by 100- to $>10^4$-fold (Figure 2, top and the Supporting Information). This 10^{-20} mol sensitivity is at least 10^8-fold greater than that of state-of-the-art general screening methods[5a] for protein binding small molecules. Although the concentration of DNA-linked molecules during selections was lower than micromolar, the selections were successful because the effective concentration of immobilized GST exceeded $\sim 10\ \mu M$, enabling a significant fraction of **1** to remain bound to GST. These results demonstrate that selections for modest protein affinities are possible in this format.

To evaluate the generality of this approach, we performed similar selections for binding to streptavidin, carbonic anhydrase, papain, trypsin, and chymotrypsin in addition to GST (Figure 1). Collectively, these six diverse proteins bind the ligands in Figure 1 with predicted affinities that span more than 8 orders of magnitude.[4]
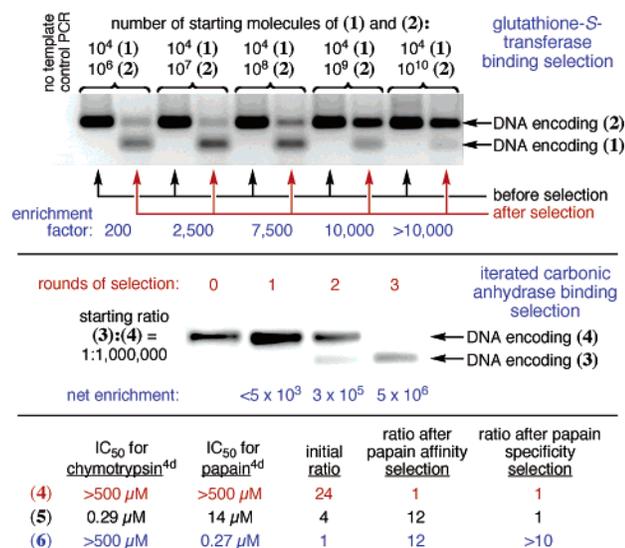
Figure 2. (Top) **1** and **2** were combined in the quantities listed and subjected to a single selection for GST binding. DNA surviving selection was amplified by PCR and digested with *Hin*dIII to determine the enrichment factors for **1**. (Middle) Three successive selections for carbonic anhydrase binding (without PCR amplification between selections) were performed on an initial 1:1 000 000 mixture of **3**:**4**. DNA surviving each selection was amplified by PCR and digested with *Hin*dIII to reveal the ratio of **3**:**4** after each iterated selection. (Bottom) A 24:4:1 mixture of **4**:**5**:**6** was subjected to two iterated selections for papain affinity, resulting in the enrichment of both nonspecific and specific protease inhibitors **5** and **6**, respectively. Repeating the selection in the presence of excess free chymotrypsin resulted in the exclusive enrichment of the papain-specific binder (**6**) (see Supporting Information).

In each case, affinity selection enriched as little as $10^{-16}-10^{-20}$ mol of the known DNA-linked ligand by at least 50-fold over the nonbinder (Figure 1), indicating that DNA conjugation does not impair the ability of these ligands to bind their protein targets and suggesting that these selections may be applicable to a variety of unrelated proteins.

In theory, selections can be iterated to multiply the net enrichment of desired molecules. To test this possibility with DNA-linked synthetic molecules, we subjected a 1:1000 mixture of DNA-linked phenyl sulfonamide (**3**):DNA-linked fMet-Leu-Phe (**2**) to a selection for binding carbonic anhydrase. Molecules surviving the first selection were eluted and directly subjected to a second selection (without PCR between selections) using fresh immobilized carbonic anhydrase. The first selection yielded a 1:3 ratio of **3**:**2**, representing a 330-fold enrichment for the DNA-linked phenyl sulfonamide. The second selection further enriched **3** by at least 30-fold, such that the ratio of **3**:**2** following two iterated selections exceeded 10:1 ($>10^4$-fold net enrichment, see Supporting Information). Similarly, three iterated selections were used to enrich a $1:10^6$ starting ratio of **3**:DNA-linked biotin (**4**) by a factor of 5×10^6 into a solution containing predominantly **3** (Figure 2 middle). These findings indicate that enormous net enrichments for DNA-linked synthetic molecules can be achieved through iterated selection and suggest that desired molecules as rare as 1 part in 10^6 within DNA-templated synthetic libraries may be isolated in this manner.

In addition to affinity, binding specificity is an important property of synthetic molecules. Selections for specificity in principle can be performed in a single experiment by selecting simultaneously for target binding as well as for the inability to bind one or more nontargets. To validate selections for specificity among DNA-linked synthetic molecules, we combined into one solution DNA-linked biotin (**4**), DNA-linked chymostatin (**5**), and DNA-linked antipain (**6**) in a 24:4:1 ratio, respectively (Figure 2, bottom). This mixture

simulates a library containing predominantly nonbinding molecules (**4** does not bind chymotrypsin or papain) with a minor fraction of nonspecific binders (**5** binds both proteases) and an even smaller fraction of a target-specific binder (**6** binds papain only). When this mixture is subjected to two iterated selections for binding to papain, both **5** and **6** are enriched as expected. When the above mixture is instead selected for binding to papain in the presence of excess free chymotrypsin, however, only the papain-specific ligand (**6**) is enriched (Figure 2, bottom and Supporting Information). The ability of these selections to separate target-specific and nonspecific synthetic molecules suggests their use to discover molecules that exclusively bind one member of a family of related proteins.

In summary, we have described general in vitro selections for DNA-linked synthetic small molecules with protein binding affinity and specificity. Spatially separated screens[5] are more flexible in the properties they can evaluate, are not sensitive to the possibility of interference from a macromolecular tag, and can examine direct effects on living cells. On the other hand, the selections described above require only readily accessible equipment, can be applied to a variety of proteins unrelated in function, yield high degrees of enrichment for active molecules, offer sensitivities far greater than those of small molecule screening methods, and can be iterated to multiply their effectiveness. The application of methods developed here to DNA-templated libraries may play an important role in the discovery of synthetic molecules with desired properties using powerful selection and amplification strategies previously available only to biological molecules.

Note Added after ASAP: In version published 9/13/2003, **5** was incorrect in Figure 1 and Supporting Information. Final version published 9/16/2003 and in print is correct.

Supporting Information Available: Experimental details, oligonucleotide sequences and structures, and additional data (PDF). This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) (a) Wilson, D. S.; Szostak, J. W. *Annu. Rev. Biochem.* **1999**, *68*, 611–647. (b) Taylor, S. V.; Kast, P.; Hilvert, D. *Angew. Chem., Int. Ed.* **2001**, *40*, 3310–3335. (c) Lin, H.; Cornish, V. W. *Angew. Chem., Int. Ed.* **2002**, *41*, 4402–4425. (d) Bull, J. J.; Wichman, H. A. *Annu. Rev. Ecol. Syst.* **2001**, *32*, 183–217.
(2) (a) Winssinger, N.; Harris, J. L.; Backes, B. J.; Schultz, P. G. *Angew. Chem., Int. Ed.* **2001**, *40*, 3152–3155. (b) Falb, D.; Jindal, S. *Curr. Opin. Drug Discovery Dev.* **2002**, *5*, 532–539. (c) Verdugo, D. E.; Cancilla, M. T.; Ge, X.; Gray, N. S.; Chang, Y. T.; Schultz, P. G.; Negishi, M.; Leary, J. A.; Bertozzi, C. R. *J. Med. Chem.* **2001**, *44*, 2683–2686. (d) Needels, M. C.; Jones, D. G.; Tate, E. H.; Heinkel, G. L.; Kochersperger, L. M.; Dower, W. J.; Barrett, R. W.; Gallop, M. A. *Pro. Natl. Acad. Sci. U.S.A.* **1993**, *90*, 10700–10704.
(3) (a) Gartner, Z. J.; Liu, D. R. *J. Am. Chem. Soc.* **2001**, *123*, 6961–6963. (b) Gartner, Z. J.; Kanan, M. W.; Liu, D. R. *Angew. Chem., Int. Ed.* **2002**, *41*, 1796–1800. (c) Gartner, Z. J.; Kanan, M. W.; Liu, D. R. *J. Am. Chem. Soc.* **2002**, *124*, 10304–10306. (d) Calderone, C. T.; Puckett, J. W.; Gartner, Z. J.; Liu, D. R. *Angew. Chem., Int. Ed.* **2002**, *41*, 4104–4108. (e) Gartner, Z. J.; Grubina, R.; Calderone, C. T.; Liu, D. R. *Angew. Chem., Int. Ed.* **2003**, *42*, 1370–1375 and references cited therein. (f) Li, X.; Liu, D. R. *J. Am. Chem. Soc.* **2003**, *125*, in press.
(4) (a) D'Silva, C. *Biochem. J.* **1990**, *271*, 161–165. (b) Jain, A.; Whitesides, G. M.; Alexander, R. S.; Christianson, D. W. *J. Med. Chem.* **1994**, *37*, 2100–2105. (c) Green, N. M. *Methods Enzymol.* **1990**, *184*, 51–67. (d) Otto, H. H.; Schirmeister, T. *Chem. Rev.* **1997**, *97*, 133–172.
(5) (a) Clemons, P. A.; Koehler, A. N.; Wagner, B. K.; Sprigings, T. G.; Spring, D. R.; King, R. W.; Schreiber, S. L.; Foley, M. A. *Chem. Biol.* **2001**, *8*, 1183–1195. (b) Mayer, T. U.; Kapoor, T. M.; Haggarty, S. J.; King, R. W.; Schreiber, S. L.; Mitchison, T. J. *Science* **1999**, *286*, 971–974. (c) Korbel, G. A.; Lalic, G.; Shair, M. D. *J. Am. Chem. Soc.* **2001**, *123*, 361–362. (d) Torrance, C. J.; Agrawal, V.; Vogelstein, B.; Kinzler, K. W. *Nat. Biotechnol.* **2001**, *19*, 940–945.

JA036065U