

Encoded Self-Assembling Chemical Libraries

Samu Melkko^a, Jens Sobek^b, Greta Guarda^a, Jörg Scheuermann^a, Christoph E. Dumelin^a, and Dario Neri^{*a}

Abstract: The display of chemical moieties at the extremity of synthetic oligonucleotides allows the self-assembly of large chemical libraries in which every chemical moiety is associated to a unique DNA sequence, serving as an 'identification bar code'. In this article, we describe the principles, the potential, and the challenges of this novel technology, which promises to facilitate the isolation of high-affinity binders to protein targets of biological and pharmaceutical interest.

Keywords: Affinity maturation · Combinatorial chemistry · DNA-templated chemistry · Encoded self-assembling chemical libraries · Lead identification

1. Introduction

The identification of molecules that bind specifically to target proteins of interest remains a formidable challenge. Technologies which facilitate the isolation of binding molecules will have profound implications for research in chemistry, biology and pharmaceutical sciences (most drug development programs rely on the ability to isolate small organic binding compounds to a given protein). With the aging population and with an increased understanding of the mechanisms of disease at a molecular level, biomedical scientists are facing the demand for more and better drugs. Additionally, the elucidation of the biological function of proteins will, in many cases, require the access to specific ligands (an ap-

proach often termed as 'chemical genetics' [1]). Techniques for the general, fast, and inexpensive isolation of small organic binding molecules are lacking at present. So far, antibodies are the only class of molecules that can be rapidly generated against a variety of different targets [2], but the inherent inability of antibodies to penetrate cells is a major limitation for many applications.

In the last decade, work with very large antibody phage display libraries (>10⁹ members) has shown that the probability of isolating high-affinity binders to target antigens increases with library size [3]. This technology relies on the ability to generate large combinatorial libraries of antibodies, together with an efficient selection methodology (phage biopanning) which allows the facile identification and amplification of binding antibodies [4]. In principle, it should be possible to use similar strategies for the isolation of small organic binders to target proteins, provided that large libraries of compounds can be synthesized, and that methods are available for the identification of binding molecules.

2. ESACHEL

We have developed a methodology, termed encoded self-assembling chemical libraries (ESACHEL) [5], which (i) allows the construction of very large libraries of chemical compounds, and (ii) provides a method for the isolation and identification of specific binding molecules. The self-assembly of chemical moieties, which

represent putative binding fragments and which are covalently coupled to molecular species capable of dimer formation, represents an avenue for the facile generation of large chemical libraries, starting from smaller sub-libraries. For example, the self-assembly (heterodimerization) of two libraries containing 1000 members would yield 1'000'000 different combinations, *i.e.* 1'000'000 different chemical entities (Fig. 1A). Notably, the hetero-trimerization of libraries containing 1000 members would yield a library containing 1'000'000'000 different chemical entities (Fig. 1B). Obviously, similar procedures can be used for the affinity maturation of known lead binding compounds, which can be paired with libraries of binding molecules for the identification of synergistic pharmacophores (Fig. 1C). DNA derivatives are ideal candidates for the construction of ESACHEL libraries since suitable oligonucleotides can mediate the self-assembly of the libraries, providing them with a unique identification 'bar code'. ESACHEL libraries can be screened using panning procedures which are conceptually similar to phage display methodologies [2]. The identification of the library compounds that are isolated by the panning procedures can be achieved by the amplification of the DNA coding tags, followed by hybridization on DNA microarrays (Fig. 1D) or DNA sequencing.

For the construction of ESACHEL libraries, oligonucleotides with appropriate linker moieties have to be conjugated to individual organic molecules (representing putative binding fragments), and the reac-

*Correspondence: Prof. Dr. D. Neri^a

Tel.: +41 44 633 74 28

Fax: +41 44 633 13 58

E-Mail: neri@pharma.ethz.ch

^aSwiss Federal Institute of Technology Zurich

Institute of Pharmaceutical Sciences

Department of Chemistry and Applied Biosciences

ETH Hönggerberg

Wolfgang-Pauli-Str. 10

CH-8093 Zurich

^bFunctional Genomics Center Zurich

Winterthurerstrasse 190

CH-8057 Zurich

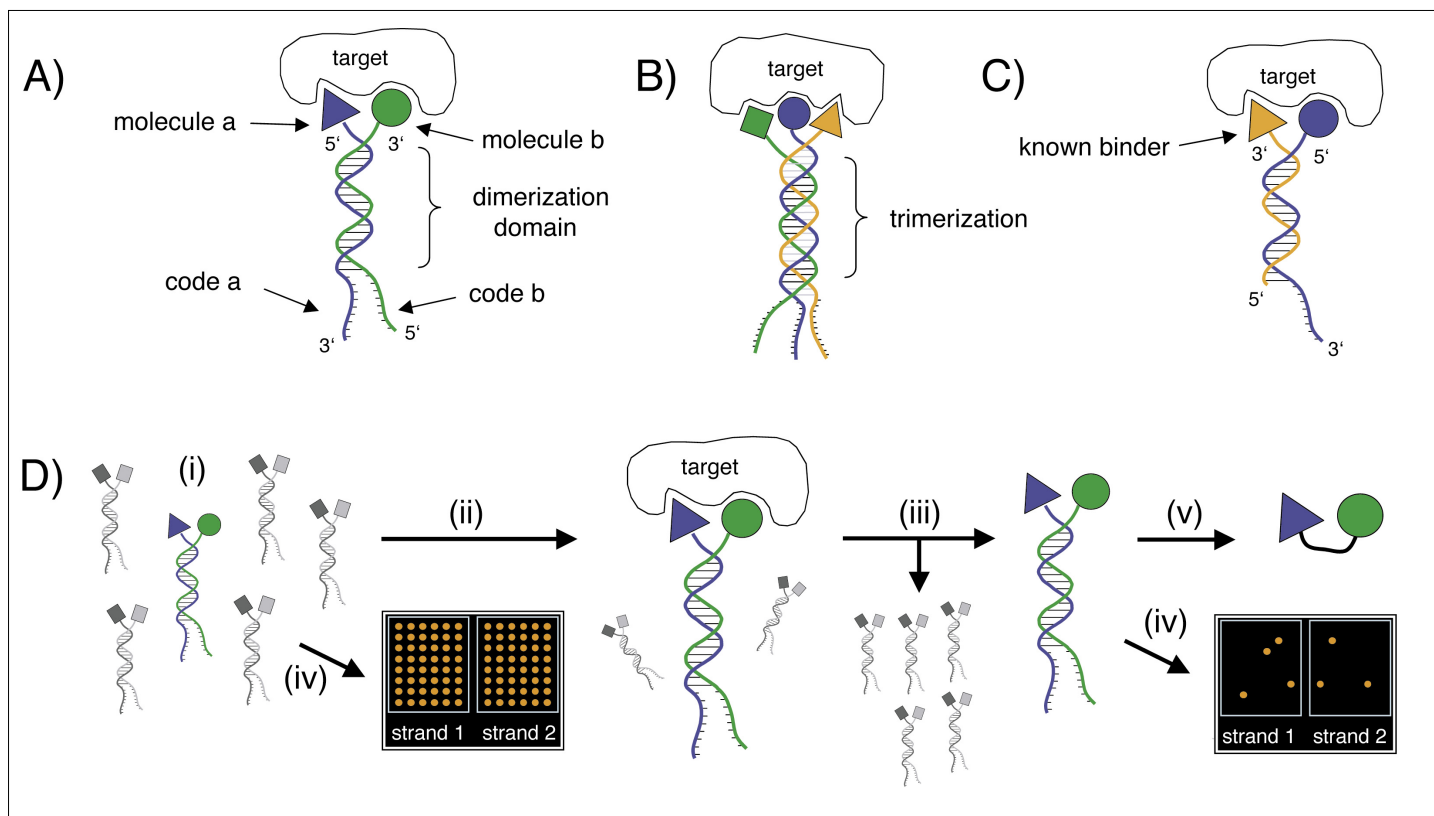


Fig. 1. Schematic representation of encoded self-assembling chemical library (ESACHEL) technology. Library compounds consist of chemical entities (pharmacophores), which are covalently coupled to molecular species (typically, oligonucleotides) capable of complex formation and a DNA coding tag, unambiguously identifying the putative binding molecule. Large libraries can be obtained by the self-assembly of smaller libraries of pharmacophores to form heterodimeric (A) or heterotrimeric (B) library compounds. (C) Furthermore, a known binder to a target protein (a lead compound) may be improved (affinity matured) by conjugating the known binder to a shorter oligonucleotide only containing a hybridization domain, pairing it with a small ESACHEL library and selecting pharmacophores which may interact with adjacent binding sites of the target protein. (D) Illustration of the process of isolating specific binders to target proteins by ESACHEL technology. Large ESACHEL libraries (i) are incubated with a target protein of choice (ii) and captured on a solid support. While non-binders are washed away (iii), library compounds capable of retaining binding to the target protein are specifically enriched in this process. The identity of the pharmacophores responsible for the binding can be detected by reading the DNA codes assigned to library compounds, for example using oligonucleotide microarray technology (iv). The comparison of microarray signals of the library before selection with microarray signals after selection reveals the compounds which are preferentially enriched in the affinity selection procedure. (v) For most practical applications, it will be convenient to conjugate the pharmacophores identified in the ESACHEL selections by means of a linker molecule, thus yielding a bidentate binder. The chemical nature of the linker will influence the properties (binding affinity, selectivity, solubility etc.) of the final product.

tion product has to be isolated and characterized. Oligonucleotides carrying primary amino groups or thiol groups at the 5' or 3' extremity can be purchased from commercial suppliers. These chemical groups enable a variety of chemical reactions for the covalent modification of oligonucleotides [6], including the conjugation of organic molecules carrying activated acid groups, such as reactive esters or acyl chlorides or sulphonyl chloride groups for the primary amines, and several nucleophilic substitutions and Michael addition reactions for the thiols. Alternatively, modified phosphoramidites can be incorporated in solid-phase synthesis protocols [7], to obtain oligonucleotides that display small organic molecules.

Until now, for the construction of ESACHEL libraries we have mainly used the *in situ* activation of carboxylic acids or isothiocyanate derivatives for the conjugation of chemical moieties to oligonucleotides carrying primary amino groups at

their 3' or 5' extremity. After coupling, the individual reaction products were purified using HPLC, and the identity of the conjugate was confirmed using mass spectrometry (Fig. 2).

ESACHEL compounds, which are preferentially enriched in affinity selections with the target protein of interest, can be identified ('decoded') by virtue of the unique DNA code associated to each chemical moiety. We have developed two different methods to experimentally read the code of ESACHEL fragments, based on either sequencing of the DNA tags or on microarray technology.

ESACHEL decoding by DNA sequencing bears some analogies to Serial Analysis of Gene Expression protocols [8]. For this method, the DNA fragments should contain suitable restriction sites flanking the identification code. After ESACHEL selections, preferentially enriched oligonucleotide derivatives are PCR-amplified, followed by digestion with restriction enzymes and

ligation to form concatamers. These DNA fragments, which contain a series of coding fragments, are ligated into sequencing vectors, which can then be transfected into bacteria and sequenced, thus providing information about the identity and the relative frequency of DNA codes of ESACHEL compounds before and after selection.

When using microarray technology for ESACHEL decoding, a modified PCR reaction amplifies the DNA tags to be analyzed (Fig. 1D). One of the oligonucleotide primers is labeled with a fluorophore, and is used in the PCR at a higher concentration than the second oligonucleotide ('asymmetric PCR'), resulting in the preferential production of a fluorescently labeled, single-stranded PCR product, which contains the DNA bar-code sequence of the original ESACHEL compound(s). This sample can then be hybridized to microarrays, spotted at distinct positions with suitable complementary oligonucleotides, carrying different codes corresponding to the different

ESACHEL bar codes. The procedure results in the detection of fluorescent spots on the microarrays, where the preferentially enriched ESACHEL compounds give rise to the strongest fluorescence signal.

Fig. 3 presents the decoding results of model ESACHEL selections using the microarray and the sequencing method. Fig. 3A illustrates how a 10-fold or 100-fold enrichment of a certain ESACHEL oligonucleotide derivative in a library of ESACHEL compounds can be detected by microarray-based decoding. In the experiment, a small library of 138 compounds, covalently attached at the extremity of 138 different oligonucleotides and paired with a single oligonucleotide derivative (Fig. 1C), was spiked with a 10-fold or 100-fold excess of the oligonucleotide derivative **76**. After asymmetric PCR amplification with a fluorescence-labeled oligonucleotide, Fig. 3A

shows that the relative intensity of the microarray spots corresponding to compound **76** get progressively stronger (compared to the spots of the other library compounds) in the microarrays corresponding to a 10-fold and 100-fold enrichment. A quantification of these experiments, based on the integration of the fluorescence data and on the normalization relative to the signals of the original library, can be found in Fig. 3B and 3C. Fig. 3D shows the results of the microarray-based decoding of a model ESACHEL selection experiment, in which a biotinylated oligonucleotide is preferentially enriched over a library of 137 compounds after capture on streptavidin resin. A comparison of the decoding of the library before and after selection immediately reveals the preferential capture of the biotinylated oligonucleotide; the visual impression is confirmed by a quantitation of the ratios of the

fluorescence signals of the spots (Fig. 3E). The same selection experiment, decoded using the concatamer formation and DNA sequencing methodology described above, confirms the preferential enrichment of the biotinylated oligonucleotide (Fig. 3F). The fact that 11/13 sequence tags after selection correspond to the biotinylated oligonucleotide, while the relative abundance of this compound before selection was 1/138, allows the computation of an enrichment factor >700 for this selection.

3. Final Remarks

A number of technologies have recently been developed which aim at the identification of high-affinity protein binders by the tethering of chemical moieties that interact with adjacent epitopes on the protein of interest. Such fragment-based screening

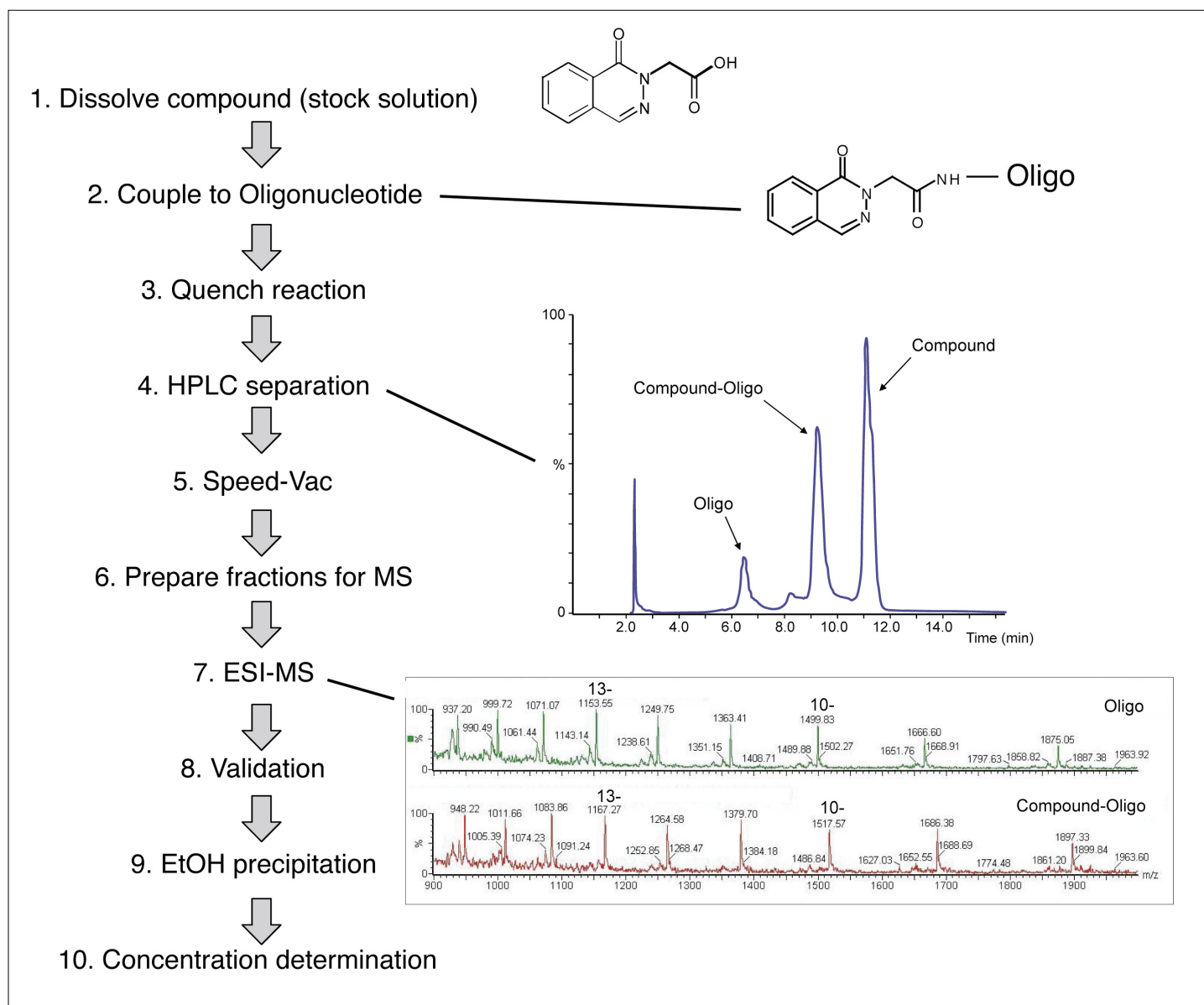


Fig. 2. Illustration of the standardized preparation of a large sub-library of ESACHEL compounds. The preparation includes the individual HPLC purification of each oligonucleotide-pharmacophore conjugate, and mass determination by ESI-MS for structural confirmation.

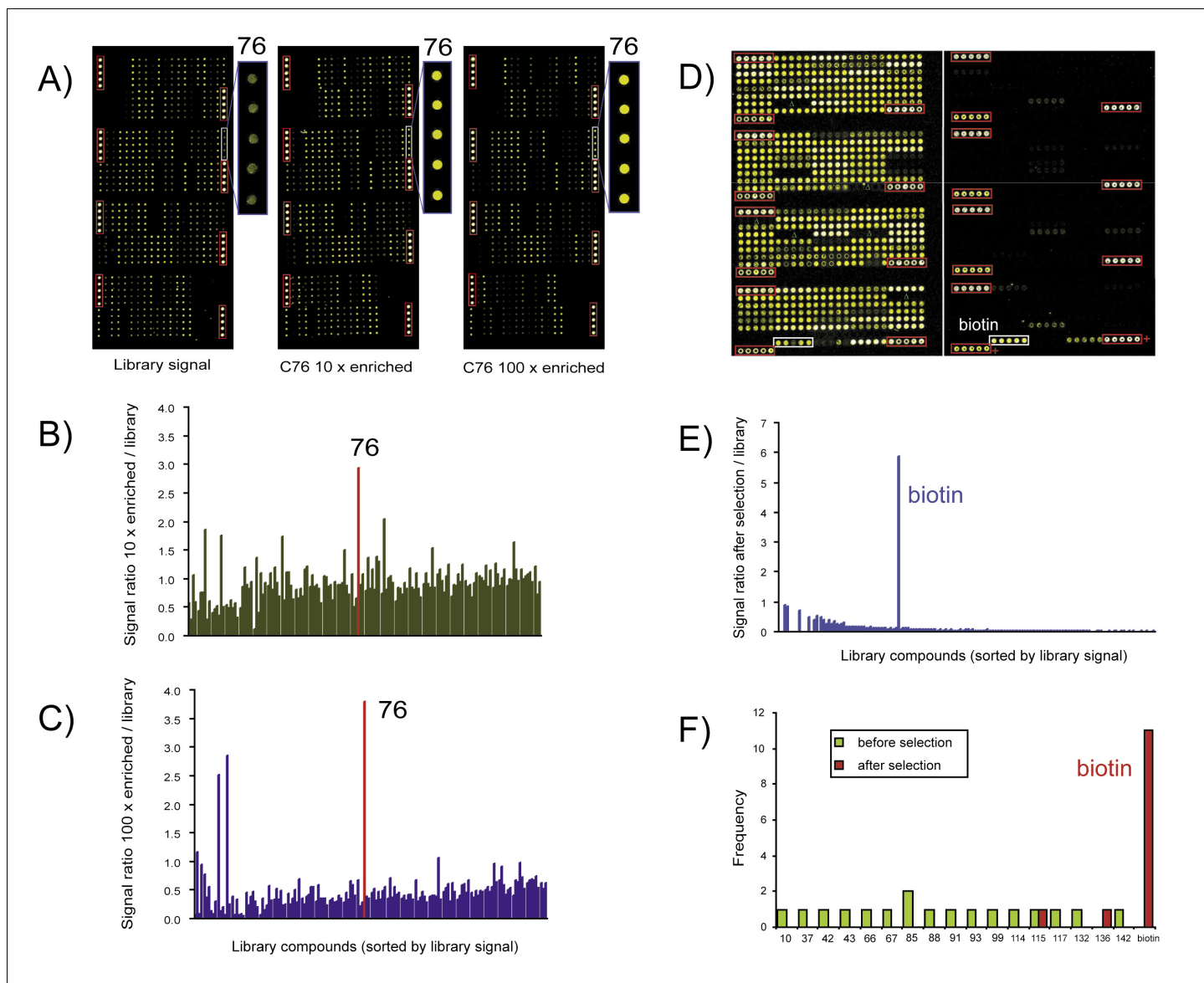


Fig. 3. (A) The microarray images show a 138-compound ESACHEL library with compound **76** present at the same concentration (left panel), at a 10-fold higher concentration (middle panel) and at a 100-fold higher concentration (right panel) relative to the other library compounds. Spots were arrayed in quintuplicate for each library member. Fluorescence-labeled positive control oligonucleotides (red frames) confirm the homogeneity of spotting. The spots corresponding to compound **76** are highlighted in a white frame and are shown as a close-up. (B) Plot of the ratios of average microarray signal intensities for each compound, with compound **76** in a 10-fold higher concentration relative to the library signals. A ten-fold enrichment of a library compound can be detected by microarray technology. (C) A 100-fold enrichment of an individual library compound is detected by microarray technology with a further increased signal-to-background ratio. The compounds are ordered on the x-axis according to their signal intensities in the original library, thus values on the far-left of the x-axis derive from divisions of small numbers and hence should be treated with caution. (D) Decoding of a model ESACHEL selection against streptavidin. The microarray images show the signal intensities before selection (library signal) and after panning the 138-compound ESACHEL library against streptavidin-sepharose. A selective intensity increase for the microarray spots corresponding to biotin can be observed. (E) Plot of the ratios of average microarray signal intensities for individual compounds, after and before selection. The selective enrichment of biotinylated ESACHEL compound is reflected in the higher value of the ratio, compared to the other members of the library. (F) Decoding of ESACHEL selections on streptavidin by concatamer formation and DNA sequencing. Before selection, no preferential enrichment of ESACHEL compounds was observed. After selection, 11 of the 13 codes sequenced correspond to the biotinylated oligonucleotide.

approaches include SAR by NMR [9][10], SAR by MS [11], combinatorial ligand-guided target assembly [12], and tethering [13], and have already yielded high-affinity protein binders, some of which are depicted in Fig. 4. In a similar fashion, small ESACHEL libraries have already allowed the affinity maturation of lead compounds with millimolar to micromolar affinity to standard target antigens, such as albumin and carbonic anhydrase. Moreover, the direct

selection of a library of >500 compounds chemically coupled to individual single-stranded oligonucleotides has allowed the identification of a cyclic urea which bears a striking resemblance to certain structural features of biotin and which displays a nanomolar dissociation constant for streptavidin [14] (Fig. 4).

Large ESACHEL sub-libraries containing >700 compounds are now available in our laboratory. The next challeng-

es include the application of ESACHEL technology for the affinity-maturation of lead compounds against target proteins of pharmaceutical relevance (according to the scheme depicted in Fig. 1C), as well as the implementation of ESACHEL selections with very large libraries, according to the scheme of Fig. 1A. The choice of suitable organic molecules which serve as scaffolds for the pharmacophores identified using ESACHEL selections (Fig. 1D) will greatly




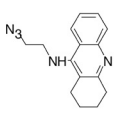
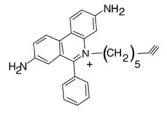
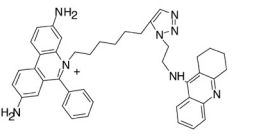
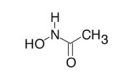
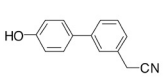
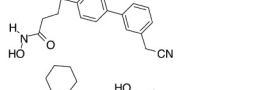
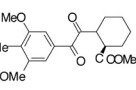
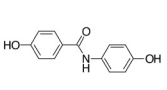
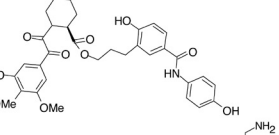
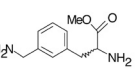
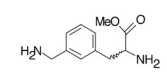
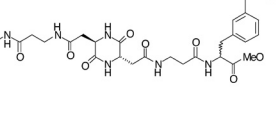
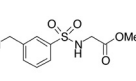
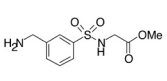
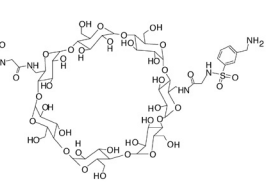
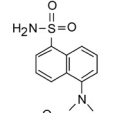
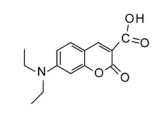
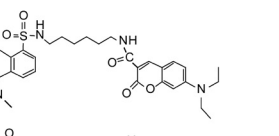
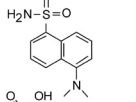
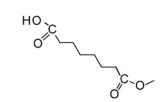
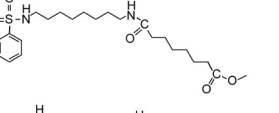
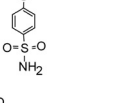
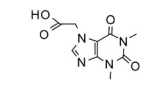
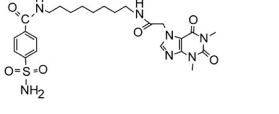
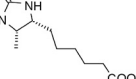
	K_d		K_d		K_d
Selected examples from literature					
	10 μ M		10 μ M		8 fM (A)
	17 mM		20 μ M		15 nM (B)
	2 μ M		100 μ M		49 nM (C)
	14 μ M		14 μ M		10 nM (D)
	17 μ M		17 μ M		0.6 nM (E)
ESACHEL - derived ligands					
	146 μ M		41 μ M		4 μ M (F)
	146 μ M		n.d.		4 μ M (G)
	1 μ M		n.d.		25 nM (H)
	2 nM (I)				

Fig. 4. The conjugation of pharmacophores that interact with adjacent binding sites of a target protein can yield ligands with markedly increased binding affinities. Examples from the literature include (A) acetylcholinesterase [15], (B) the matrix metalloprotease stromelysin [9], (C) the FK506 binding protein [10], and (D, E) beta-tryptase [16][17] as target proteins. Using an 138-compound ESACHEL library, we have affinity matured the human serum albumin binder dansylamide (F, G), and a sulfonamide derivative that inhibits carbonic anhydrase (H) in the implementation of ESACHEL technology depicted in Fig. 1C [5]. In a single strand selection, using an ESACHEL sub-library containing >500 compounds, we detected desthiobiotin as a streptavidin binder with affinity in the nanomolar range (I) [14].

determine the performance of the resulting bidentate ligands. Geometrical features (e.g. distance and exit vectors), flexibility, lipophilicity and other chemical parameters can influence the magnitude of the chelate effect and, thus, the affinity gain for bidentate ligands.

Finally, it is worth mentioning that ESACHEL technology may be compatible with the covalent linkage of the two pharmacophores at the extremity of the oligonucleotides after the self-assembly of the sub-libraries, thus creating more compact and rigid chemical structures, which may

be more suitable for certain pharmaceutical applications.

Acknowledgements

The authors acknowledge a financial contribution from the ETH Zürich (CLE' INIT), the Bundesamt für Bildung und Wissenschaft (BBW) for the EU Project STROMA and from Philogen. C.E.D. is the recipient of a bursary from the Roche Research Foundation. The experimental support of Ms. Madalina Jaggi for the quality controls of the library is gratefully acknowledged.

- [1] R.L. Strausberg, S.L. Schreiber, *Science* **2003**, *300*, 294.
- [2] G. Winter, A.D. Griffiths, R.E. Hawkins, H.R. Hoogenboom, *Annu. Rev. Immunol.* **1994**, *12*, 433.
- [3] A.D. Griffiths, S.C. Williams, O. Hartley, I.M. Tomlinson, P. Waterhouse, W.L. Crosby, R.E. Kontermann, P.T. Jones, N.M. Low, T.J. Allison, *Embo. J.* **1994**, *13*, 3245.
- [4] J. McCafferty, A.D. Griffiths, G. Winter, D.J. Chiswell, *Nature* **1990**, *348*, 552.
- [5] S. Melkko, J. Scheuermann, C.E. Dumelin, D. Neri, *Nat. Biotechnol.* **2004**, *22*, 568.
- [6] G.T. Hermanson, 'Bioconjugate Techniques', Academic Press, **1996**.
- [7] S. Josephson, E. Lagerholm, G. Palm, *Acta Chem. Scand. B* **1984**, *38*, 539.
- [8] V.E. Velculescu, L. Zhang, B. Vogelstein, K.W. Kinzler, *Science* **1995**, *270*, 484.
- [9] P.J. Hajduk, R.P. Meadows, S.W. Fesik, *Science* **1997**, *278*, 497.
- [10] S.B. Shuker, P.J. Hajduk, R.P. Meadows, S.W. Fesik, *Science* **1996**, *274*, 1531.
- [11] E.E. Swayze, E.A. Jefferson, K.A. Sannes-Lowery, L.B. Blyn, L.M. Risen, S. Arakawa, S.A. Osgood, S.A. Hofstadler, R.H. Griffey, *J. Med. Chem.* **2002**, *45*, 3816.
- [12] D.J. Maly, I.C. Choong, J.A. Ellman, *Proc. Natl. Acad. Sci. USA* **2000**, *97*, 2419.
- [13] D.A. Erlanson, J.W. Lam, C. Wiesmann, T.N. Luong, R.L. Simmons, W.L. DeLano, I.C. Choong, M.T. Burdett, W.M. Flanagan, D. Lee, E.M. Gordon, T. O'Brien, *Nat. Biotechnol.* **2003**, *21*, 308.
- [14] C.E. Dumelin, J. Scheuermann, S. Melkko, D. Neri, submitted.
- [15] W.G. Lewis, L.G. Green, F. Grynszpan, Z. Radic, P.R. Carlier, P. Taylor, M.G. Finn, K.B. Sharpless, *Angew. Chem. Int. Ed.* **2002**, *41*, 1053.
- [16] N. Schaschke, G. Matschiner, F. Zettl, U. Marquardt, A. Bergner, W. Bode, C.P. Sommerhoff, L. Moroder, *Chem. Biol.* **2001**, *8*, 313.
- [17] N. Schaschke, A. Dominik, G. Matschiner, C.P. Sommerhoff, *Bioorg. Med. Chem. Lett.* **2002**, *12*, 985.