ChemComm

Interrogating Target-Specificity by Parallel Screening of a DNA-Encoded Chemical Library against Closely Related Proteins

Journal:	*ChemComm*
Manuscript ID:	CC-COM-02-2015-001230.R1
Article Type:	Communication
Date Submitted by the Author:	03-Mar-2015
Complete List of Authors:	Franzini, Raphael; ETH Zürich, Department of Chemistry and Applied Biosciences Nauer, Angela; ETZ Zürich, Department of Chemistry and Applied Biosciences Scheuermann, Joerg; ETH Zurich, DCHAB Neri, Dario; ETH Zurich, Institute of Pharmaceutical Sciences

Chemical Communications

RSCPublishing

COMMUNICATION

Interrogating Target-Specificity by Parallel Screening of a DNA-Encoded Chemical Library against Closely Related Proteins†

Raphael M. Franzini,[a] Angela Nauer,[a] Jörg Scheuermann,[a] Dario Neri[a,*]

Cite this: DOI: 10.1039/x0xx00000x

Received ooth January 2012,
Accepted ooth January 2012

DOI: 10.1039/x0xx00000x

www.rsc.org/

Parallel affinity screening of a DNA-encoded chemical library against rat, bovine and human serum albumin allowed the identification of small-molecule ligands with distinctive binding specificities to the individual proteins.

The achievement of an adequate target-specificity often represents a main challenge in small-molecule drug development. Screening compound libraries, the most widely used method to identify early drug hits, provides little information on target-specificity. This parameter is generally assessed and optimized during hit-to-lead development. However, iterative cycles of small-molecule synthesis, *in vitro* evaluation and identification of structural features responsible for target-specificity are expensive and time-consuming. In principle, it would be possible to obtain valuable information regarding protein-specificity directly by repeating library screening on a series of related target proteins. Unfortunately, the high costs associated with conventional screening methods usually preclude this approach.

Screening DNA-encoded chemical libraries (DECLs) is increasingly being used as a method for early hit discovery.[1] DECLs are collections of small-molecules encoded by DNA sequence tags, serving as amplifiable identification barcodes. Several DECL designs and synthetic strategies have been described.[2] Hit identification involves affinity screening against the immobilized protein of interest, followed by sequencing of the DNA barcodes. The technology has been demonstrated to be effective for pharmaceutical hit identification.[3]

The screening of DECLs is remarkably rapid and inexpensive, opening the possibility to screen multiple targets in parallel and to gain information on target-specificity at the hit discovery stage.[4] Here, we tested this concept by performing a comparative analysis of DECL screening experiments with serum albumins from three different species (rat serum albumin, RSA; bovine serum albumin, BSA; human serum albumin, HSA). Using this procedure, we identified ligands with a binding preference to either protein as well as ligands with affinity to all three targets. We screened DAL-100K (Fig. S1 in ESI†), a recently described DECL containing 103,200 branched diamide compounds, for these targets.[5] Affinity screening of DAL-100K against immobilized RSA and BSA resulted in highly-enriched normalized sequence counts (NSC) for multiple compounds in the library indicative of affinity binding to these proteins similar to previously reported results for HSA[5] (Fig. 1).

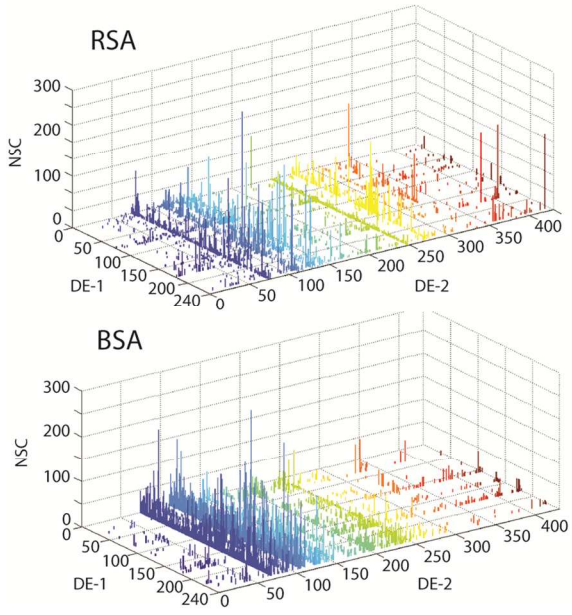


Fig. 1. DAL-100K screening results for RSA and BSA. The compounds in the library are distributed in the x,y-plane and the z-axis provides the normalized sequence counts (NSC) for each compound. Cutoff NSC > 5; color used to enhance visibility. Results for HSA[4] are shown in Fig. S2 in ESI†. DE: diversity element.

The screening results for the three albumins exhibited noticeable differences. Displaying the NSC-values of hit compounds in a 3-D scatter plot revealed protein-selective enrichment for numerous hit compounds (Fig. 2a). The differences in selectivity were even more pronounced when plotting the protein selectivity (PS) and the additive sequence counts (ASC) for the different hits (Fig. 2b; for definitions of PS and ASC see formulas E1-E3 in ESI†). Compounds predicted to be selective are clustered on the top of this graph (PS > 0.8), while binders predicted to have the highest protein affinities are found at the right of this graph. Based on the screening results, we selected four sets of ten representative compounds each (structures shown in Tables S1-S4 in ESI†); three sets of predicted preferential RSA, BSA and HSA ligands (high PS values) and one set of structures with sequence enrichment for all serum albumins (low PS values). Analysis was complicated by the fact that the enrichment levels varied for the three proteins; many compounds were highly enriched for BSA, whereas for HSA the number of compounds with high NSC-values was lower but even compounds with NSC-values of 5-10 have been found to be high-affinity binders.[5]

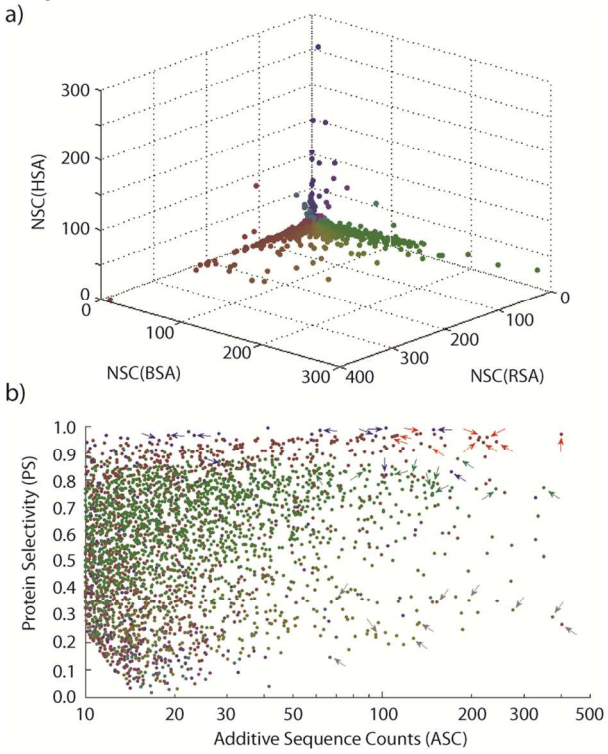


Fig. 2. Comparison of selection results for RSA, BSA, and HSA. a) 3-D scatter plot of NSC-values for library screening against RSA (x-axis), BSA (y-axis) and HSA (z-axis) for each compound. b) Scatter plot of protein selectivity (PS) versus additive sequence counts (ASC) for compounds in the library (a cloase-up with labelling of hit compounds is provided in Figure S4 in ESI†). Arrows indicated compounds selected for further analysis (Fig 3). Only compounds with NSC ≥ 5 are displayed. The hue of the marker indicates the albumin-preference of the corresponding compounds (RSA: red, BSA: green, HSA: blue). Formulas (E1-E3) used for calculating specificity and hue are provided in the ESI†.

We synthesized fluorescein-conjugates (Fig. S3 in ESI†) based on the 40 selected structures and determined the dissociation constants (K_d) by fluorescence polarization (Fig. 3). Several small-molecule conjugates with nanomolar K_d values were identified for each protein (the fluorophore and C_{12}-linker substantially contribute to the affinity of these conjugates).[5] Overall, the measured binding preference agreed well with the predictions based on the DAL-100K screening results. All predicted RSA and HSA ligands bound with the highest affinity to the anticipated protein (synthesis of three structures selected for HSA was unsuccessful and one compound showed no protein binding in the tested concentration range; Table S-3 in ESI†). Seven of the anticipated BSA-ligands bound this protein with the lowest K_d values from the three serum albumins. The NSC-values of most BSA-hits were also elevated for the other two serum albumins (PS values < 0.9) consistent with the challenge of identifying BSA-specific binders from screening DAL-100K. Most universal albumin binders showed affinity binding to all of the three proteins with few exceptions. Although some measured specificities deviated from the predictions based on sequence counts, it was thus generally possible to predict binding preference correctly. Different interactions of the proteins with fluorescein and the linker as well as differences between ligand binding of biotinylated proteins on solid support and the unmodified proteins in solution may account for these discrepancies.

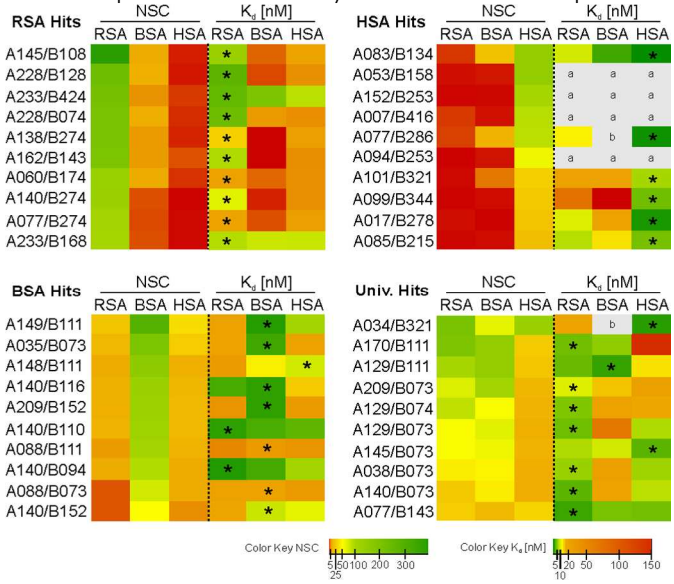


Fig. 3. Heat map of screening results (NSC values, three left columns) and measured affinities (K_d values, three right columns) for selected compounds corresponding to the four classes. Asterisks indicate the protein with the highest affinity to the compound. A table containing all NSC and K_d values as well as the chemical structures is provided in the ESI†. a: hit compound could not be synthesized or no binding was observed; b: broad transition unsuitable for sigmoidal curve fitting.

For each serum albumin, we were able to identify a small-molecule compound that bound with K_d values < 10 nM and exhibited ≥10-fold specificity over the other two (Fig. 4). A228/B128 bound RSA with a K_d of 4.5 nM (K_d(BSA) = 121 nM; K_d(HSA) = 45 nM), A209/B152 to BSA with a K_d of 2.3 nM (K_d(RSA) = 40 nM; K_d(HSA) = 33 nM), and A099/B344 to HSA with a K_d of 5.4 nM (K_d(RSA) = 77 nM; K_d(BSA) = 300 nM). Screening only a single protein would not have revealed these structures. For instance, the selective HSA binder A099/B344 is only moderately enriched in DAL-100K selection and it is unlikely that it would have been chosen solely based on the HSA selection results. Indeed, the K_d(HSA) value of A099/B344 is >10-fold above that of the best HSA-binder (A083/B134, K_d = 0.4 nM). In addition, this compound

is structurally distinct from most other HSA-hits and its discovery illustrates the potential of comparative analysis of parallel screening results for the isolation of target-selective compounds.

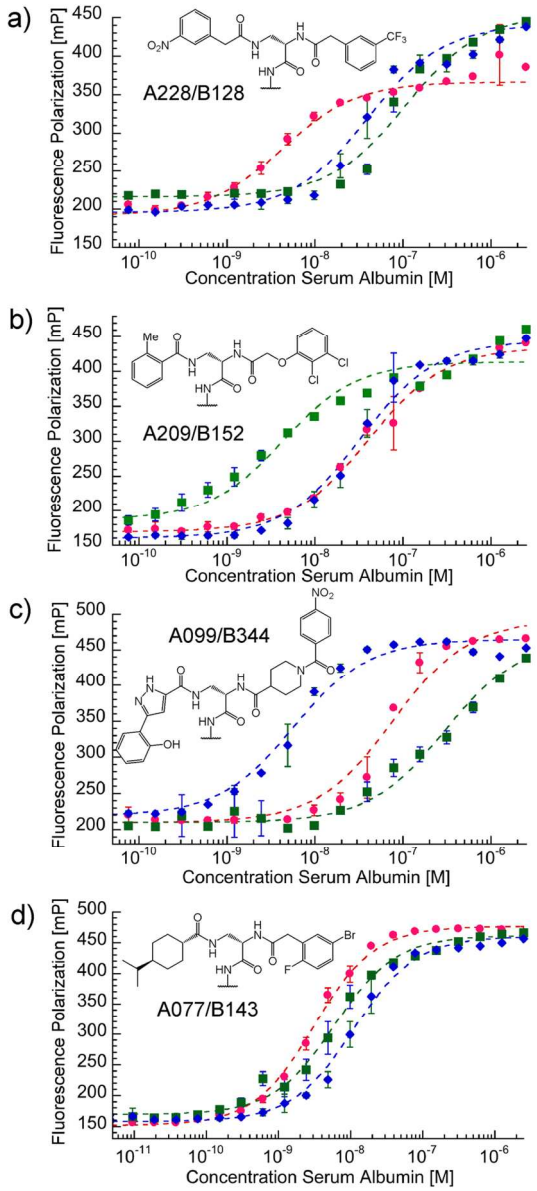


Fig. 4. Dependence of fluorescence polarization on the concentration of serum albumins for selected compounds (red: RSA; green: BSA; blue: HSA). a) Selected RSA-specific ligand A228/B128. b) Selected BSA-specific ligand A209/B152. c) Selected HSA-specific ligand A099/B344. Selected general serum albumin binderA077/B143.

In conclusion, the data in this report illustrate that parallel DECL screening of related proteins can provide valuable information concerning the target-specificity of identified hit compounds. The ease and speed of DECL screening allows to routinely incorporating multi-target screening procedures at early stages of hit discovery in drug development.

Financial support from the ETH Zürich, the Swiss National Science Foundation and Philochem AG are gratefully acknowledged. RMF acknowledges a postdoctoral research fellowship (VPFW-ETH fellowship) by ETHZ and Marie Curie Actions. We thank Mauro Zimmerman for assistance with LC-MS analysis of the conjugates and the Functional Genomics Center Zurich for high-throughput sequencing.

Notes and references

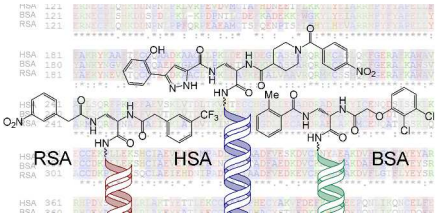
[a] Institute of Pharmaceutical Sciences, ETH Zürich, Vladimir-Prelog-Weg 3, 8093 Zürich (Switzerland). *E-mail: dario.neri@pharma.ethz.ch.*

† Electronic Supplementary Information (ESI) available: Detailed experimental procedures, supporting figures and tables. See DOI: 10.1039/c000000x/

1. (a) K. K. Sadhu and N. Winssinger, *Chimia*, 2013, **67**, 905-909; (b) R. M. Franzini, D. Neri and J. Scheuermann, *Acc. Chem. Res.*, 2014, **47**, 1247-1255.
2. (a) Z. J. Gartner, B. N. Tse, R. Grubina, J. B. Doyon, T. M. Snyder and D. R. Liu, *Science*, 2004, **305**, 1601-1605; (b) S. Melkko, J. Scheuermann, C. E. Dumelin and D. Neri, *Nat. Biotechnol.*, 2004, **22**, 568-574; (c) K. I. Sprinz, D. M. Tagore and A. D. Hamilton, *Bioorg. Med. Chem. Lett.*, 2005, **15**, 3908-3911; (d) F. Debaene, J. A. Da Silva, Z. Pianowski, F. J. Duran and N. Winssinger, *Tetrahedron*, 2007, **63**, 6577-6568; (e) S. J. Wrenn, R. M. Weisinger, D. R. Halpin and P. B. Harbury, *J. Am. Chem. Soc.*, 2007, **129**, 13137-13143; (f) L. Mannocci, Y. Zhang, J. Scheuermann, M. Leimbacher, G. De Bellis, E. Rizzi, C. Dumelin, S. Melkko and D. Neri, *Proc. Natl. Acad. Sci. U. S. A.*, 2008, **105**, 17670-17675; (g) M. H. Hansen, P. Blakskjaer, L. K. Petersen, T. H. Hansen, J. W. Hojfeldt, K. V. Gothelf and N. J. Hansen, *J. Am. Chem. Soc.*, 2009, **131**, 1322-1327; (h) N. Svensen, J. J. Diaz-Mochon and M. Bradley, *Chem. Biol.*, 2011, **18**, 1284-1289; (i) Y. Li, P. Zhao, M. Zhang, X. Zhao and X. Li, *J. Am. Chem. Soc.*, 2013, **135**, 17727-17730; (j) J. P. Daguer, M. Ciobanu, S. Alvarez, S. Barluenga and N. Winssinger, *Chem. Sci.,* 2011, **2**, 625-632; (k) C. Cao, P. Zhao, Z. Li, Z. Chen, Y. Huang, Y. Bai and X. Li, *Chem. Commun.*, 2014, **50**, 10997-10999; (l) M. Wichert, N. Krall, W. Decurtins, R. M. Franzini, F. Pretto, P. Schneider, D. Neri and J. Scheuermann, *Nat. Chem.,* 2015, **7**, 241-249.
3. (a) M. A. Clark, *et al., Nat. Chem. Biol.*, 2009, **5**, 647-654; (b) R. E. Kleiner, C. E. Dumelin, G. C. Tiu, K. Sakurai and D. R. Liu, *J. Am. Chem. Soc.*, 2010, **132**, 11779-11791; (c) S. Melkko, L. Mannocci, C. E. Dumelin, A. Villa, R. Sommavilla, Y. Zhang, M. G. Grutter, N. Keller, L. Jermutus, R. H. Jackson, J. Scheuermann and D. Neri, *ChemMedChem*, 2010, **5**, 584-590; (d) N. Svensen, J. J. Diaz-Mochon, K. Dhaliwal, S. Planonth, M. Dewar, J. D. Armstrong and M. Bradley, *Angew. Chem. Int. Ed.*, 2011, **50**, 6133-6136; (e) H. Deng, *et al., J. Med. Chem.*, 2012, **55**, 7061-7079; (f) M. Leimbacher, Y. Zhang, L. Mannocci, M. Stravs, T. Geppert, J. Scheuermann, G. Schneider and D. Neri, *Chem. Eur. J.*, 2012, **18**, 7729-7737; (g) J. S. Disch, *et al., J. Med. Chem.*, 2013, **56**, 3666-3679; (h) L. Encinas, *et al., J. Med. Chem.*, 2014, 57, 1276-1288; (i) C. S. Kollmann, *et al., Bioorg. Med. Chem.*, 2014, **22**, 2353-2365; (j) J.-P. Daguer, C. Zambaldo, M. Ciobanu, P. Morieux, S. Barluenga, and N. Winssinger, *Chem. Sci.,* 2015, **6**, 739-744.

4. H. D. Urbina, F. Debaene, B. Jost, C. Bole-Feysot, D. E. Mason, P. Kuzmic, J. L. Harris and N. Winssinger, *ChemBioChem,* **2006**, *7*, 1790-1797.
5. R. M. Franzini, T. Ekblad, N. Zhong, M. Wichert, W. Decurtins, A. Nauer, M. Zimmermann, F. Samain, J. Scheuermann, P. J. Brown, J. Hall, S. Gräslund, H. Schüler, and D. Neri, *Angew. Chem. Int. Ed.*, manuscript in press, *DOI: 10.1002/anie.201410736R1*.

Table of contents entry:



TOC text:

Screening DNA-encoded chemical libraries against closely related proteins enables to predict target specificity as illustrated for serum albumins from different species.