Encoded Libraries Hot Paper

International Edition: DOI: 10.1002/anie.201410736
German Edition: DOI: 10.1002/ange.201410736

Identification of Structure–Activity Relationships from Screening a Structurally Compact DNA-Encoded Chemical Library**

*Raphael M. Franzini, Torun Ekblad, Nan Zhong, Moreno Wichert, Willy Decurtins, Angela Nauer, Mauro Zimmermann, Florent Samain, Jörg Scheuermann, Peter J. Brown, Jonathan Hall, Susanne Gräslund, Herwig Schüler, and Dario Neri**

Abstract: *Methods for the rapid and inexpensive discovery of hit compounds are essential for pharmaceutical research and DNA-encoded chemical libraries represent promising tools for this purpose. We here report on the design and synthesis of DAL-100K, a DNA-encoded chemical library containing 103 200 structurally compact compounds. Affinity screening experiments and DNA-sequencing analysis provided ligands with nanomolar affinities to several proteins, including prostate-specific membrane antigen and tankyrase 1. Correlations of sequence counts with binding affinities and potencies of enzyme inhibition were observed and enabled the identification of structural features critical for activity. These results indicate that libraries of this type represent a useful source of small-molecule binders for target proteins of pharmaceutical interest and information on structural features important for binding.*

DNA-encoded chemical libraries (DECLs) are promising tools for pharmaceutical hit discovery.[1] DECLs are compound collections in which each molecule is uniquely encoded by an appended DNA sequence, serving as amplifiable identification barcodes. Affinity screening of DECLs against proteins of interest enriches target-binding molecules, which are identified by high-throughput sequencing of the corresponding DNA tags.[2]

DECL technology allows for the rapid screening of small-molecule libraries of unprecedented size in a one-pot protocol, requiring only microgram quantities of the target proteins. Several groups have described DECLs of various designs and modes of preparation[3] and selections performed with DECLs have yielded ligands for diverse proteins of biological and pharmacological interest.[3e,4]

Here we describe the design and synthesis of DAL-100K, a branched diamide DECL and its application to the discovery of protein ligands. DAL-100K comprises predominantly drug-like compounds in contrast to libraries assembled from 3 or more sets of building blocks, for which encoded compounds rarely adhere to Lipinski's rule of 5 given in Ref. [5]. Furthermore, the library format of DAL-100K allows for thorough quality control, which in combination with the high structural diversity enabled the recognition of structure–activity relationships (SAR) directly from screening data. Patterns of structural similarities among hit compounds have been previously reported.[3m,4d,i] Evaluations of possible SARs and quantitative analyses were however missing.

DAL-100K contains 103 200 compounds assembled from 240×430 unique, structurally diverse carboxylic acids attached through amide bonds to two diversity elements (DEs) on a DNA-terminal diamine (Figure 1). Carboxylic acids are readily available and allow for selecting highly diverse structures; furthermore, amide bond formation is a robust reaction for DNA-conjugation.[6]

DAL-100K was prepared by a split-and-pool strategy using consecutive synthesis and encoding steps (Figures S1 and S2 in the Supporting Information). The building blocks at DE-1 were attached on solid support to ensure high coupling yields. Thorough quality control (i.e. purification after the first coupling step and extensive reactivity tests for the second coupling step) was implemented during library synthesis. The selected carboxylic acids mostly complied with the "Rule of 3"[7] ($\text{ClogP} \leq 3$, $\text{MW} \leq 300$ Da) to favor drug-likeness (average molecular weight, MW, of 413 Da and ClogP of 1.6; Figure S3 and Table S3).

[*] Dr. R. M. Franzini, M. Wichert, W. Decurtins, A. Nauer, M. Zimmermann, Dr. J. Scheuermann, Prof. Dr. J. Hall, Prof. Dr. D. Neri
Institute of Pharmaceutical Sciences, ETH Zürich
8093 Zürich (Switzerland)
E-mail: dario.neri@pharma.ethz.ch
Dr. N. Zhong, Dr. P. J. Brown, Dr. S. Gräslund
Structural Genomics Consortium (SGC), University of Toronto
Toronto, M5G 1L7 (Canada)
Dr. T. Ekblad, Prof. Dr. H. Schüler
Department of Medical Biochemistry and Biophysics
Karolinska Institutet, 17177 Stockholm (Sweden)
Dr. F. Samain
Philochem AG, 8112 Otelfingen (Switzerland)

[**] We thank ETH Zürich, Philochem and Krebsliga Schweiz (grant number KFS-2839-08-2011) for financial support. R.M.F. acknowledges a postdoctoral research fellowship (VPFW-ETH fellowship) by ETHZ and Marie Curie Actions. H.S. acknowledges support from the IngaBritt and Arne Lundbergs Research Foundation, the Swedish Foundation for Strategic Research, and the SGC. The SGC, a registered charity (number 1097737) receives funds from AbbVie, Bayer Pharma AG, Boehringer Ingelheim, the Canada Foundation for Innovation, Genome Canada, GlaxoSmithKline, the Innovative Medicines Initiative, Janssen, Lilly Canada, Merck & Co., the Novartis Research Foundation, the Ontario Ministry of Economic Development and Innovation, Pfizer, the São Paulo Research Foundation, Takeda, and the Wellcome Trust. Instant JChem (ChemAxon) was used for structure and data management (http://www.chemaxon.com). We acknowledge Dr. Max Pillong (ETHZ) for assistance with the in silico analysis of the encoded structures.

Supporting information for this article is available on the WWW under http://dx.doi.org/10.1002/anie.201410736.

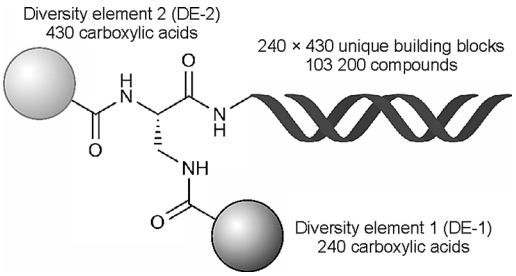

Figure 1. Structure of the DAL-100K DNA-encoded chemical library.

Affinity screening of DAL-100K against several targets provided elevated normalized sequence counts (NSC, Formula S1) compared to a survey of DNA codes prior to ligand selection indicating the recognition of a specific protein–ligand interaction (Figure 2). In these graphs, enriched compounds on a line have a conserved building block at one of the DEs indicating that this fragment is mainly responsible for the binding affinity. In contrast, both building blocks are predicted to contribute to high-affinity binding in the case of isolated peaks. Screening experiments for two model targets (streptavidin and carbonic anhydrase IX, CAIX) led to the enrichment of compounds with structural features anticipated for binders of these proteins. For streptavidin, a series of compounds containing B087 (nomenclature: building blocks at DE-1 and DE-2 are specified by A and B, respectively, followed by the number of the building block) was apparent in the sequence enrichment plot corresponding to desthiobiotin, a known high-affinity ligand of streptavidin (Figure 2b). Four distinct series of enriched molecules were observed in the screening results for CAIX (Figure 2c) all containing aromatic sulfonamides (A227, A240, B138, and B157), which are frequently present in CAIX binders.[8] These results provide support for successful DECL preparation and an efficient screening protocol.

During this study, we further identified de novo ligands for three proteins of pharmaceutical interest (Figure 2): prostate-specific membrane antigen (PSMA), human serum albumin (HSA), and tankyrase 1 (TNKS1).

Distinct enrichment fingerprints were obtained when screening PSMA (Figure 2d), a marker of prostate cancer.[9]

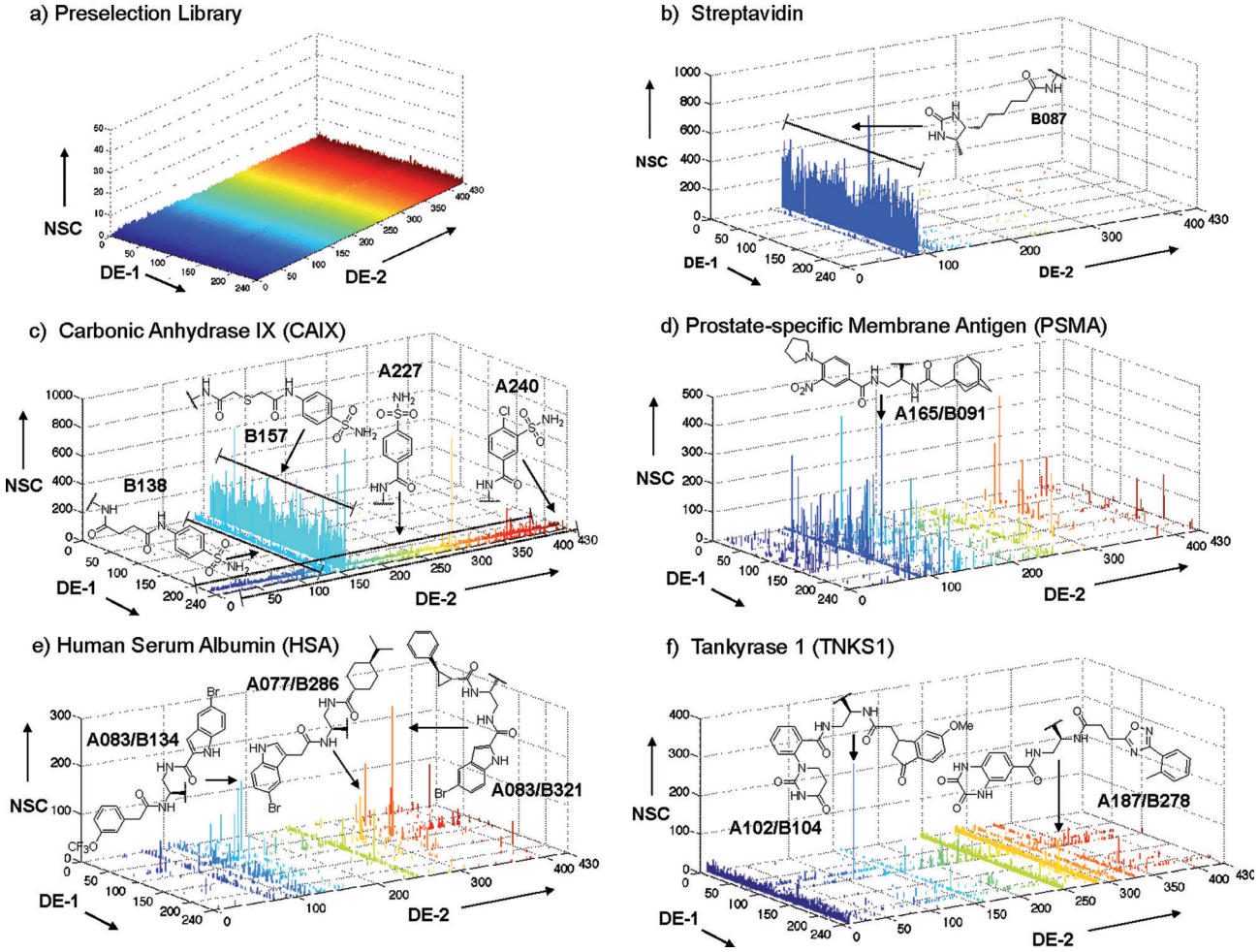

Figure 2. High-throughput DNA-sequencing results of affinity screening experiments of DAL-100K with several proteins including a) a preselection library (original library prior to affinity screening), b) streptavidin, c) CAIX, d) PSMA, e) HSA, and f) TNKS1. Each peak represents one library compound identified by the building blocks specified in the *x*/*y* plane; the *z* axis indicates the normalized sequence counts (NSC, Formula S1). Compounds with NSC values smaller than 5 are omitted to improve visibility.

A fluorescein derivative (Chart S1) of the most-highly enriched compound A165/B091 bound to PSMA with a dissociation constant (K_d) of 0.83 μM as determined by fluorescence polarization (FP; Figure S5). Interestingly, this molecule contains no obvious Zn^{2+}-binding functionality in contrast to reported small-molecule PSMA ligands.

The sequencing fingerprint for HSA (Figure 2e) revealed numerous isolated peaks of elevated NSC values suggesting synergistic binding interactions of the corresponding compounds involving both building blocks. Several hit compounds contained a 5-bromoindole moiety (A083, B286) and two molecules (A083/B321, A077/B321) were further evaluated. Fluorescein derivatives of A077/B286 and A083/B321 (Figure 2e and Chart S1) bound HSA with subnanomolar dissociation constants [K_d (A083/B321) = 0.9 nM; K_d (A077/B286) = 0.7 nM] as determined by FP (Table 1 and Figure S6). Replacement of either of the building blocks by an acetamide group (AcNH) drastically decreased the affinity of the ligands to HSA. Particularly, the 5-bromoindole moieties were critical for HSA binding and their substitution raised the K_d values by more than 2000-fold [K_d (AcNH/B321) = 3.1 μM, K_d (A077/AcNH) = 1.5 μM]. The HSA affinity of these derivatives originates in part from the fluorophore and the linker and replacing the lipophilic (LPL) by a hydrophilic (HPL) linker decreased the affinity substantially (Table 1, Chart S1). Nevertheless, the hit compounds enhanced the affinity of these conjugates by more than 10000-fold. To the best of our knowledge, these compounds are the highest-affinity small-molecule binders of HSA reported and bear potential as tags for the half-life extension of peptide drugs and imaging agents.[10]

Table 1: Dissociation constants (K_d) of derivatives of A083/B321 and A077/B286 and HSA.[a]

DE-1	DE-2	Linker	K_d [nM][b]
A083	B321	LPL	0.94 ± 0.10
A077	B286	LPL	0.70 ± 0.04
A083	AcNH	LPL	130 ± 40
AcNH	B321	LPL	3100 ± 200
A077	AcNH	LPL	1500 ± 200
AcNH	B286	LPL	930 ± 170
A083	B321	HPL	1000 ± 200
A077	B286	HPL	670 ± 80
AcNH	AcNH	LPL	> 10000

[a] Structures of building blocks and linkers are shown in Figure 2e and Chart S1, respectively. [b] K_d values were determined by fluorescence polarization.

A more detailed analysis of the HSA hit compounds revealed patterns of structural similarities. For instance, several indoles, benzofurans, and benzothiophenes were enriched in combination with A077 or B321 (Table 2). In addition to A077/B286 and A083/B321 such compounds included A017/B321, A085/B321, and A077/B107. These structures all had the carboxylic acid modification at C-2 or C-3 and an electron-withdrawing and hydrophobic substituent at C-5 of the ring system, whereas related heterocycles lacking these structural features generally had NSC values smaller than 5 (Table 2, Table S4).

Table 2: Structural features identified among HSA hit compounds.

	Building block	K_d [nM][a]	NSC
	in combination with B321		
A083	5-bromoindole-2-carboxylic acid	0.94 ± 0.10	270
A085	5-chlorobenzofuran-2-carboxylic acid	4.0 ± 0.5	74
A017	5-trifluoromethoxyindole-2-carboxylic acid	2.5 ± 0.5	38
A053	5-chloro-7-methoxybenzofuran-2-carboxylic acid	6.7 ± 0.8	8
A020	5-methoxyindole-2-carboxylic acid	10.5 ± 1.0	6
A024	indole-2-carboxylic acid	9.6 ± 1.0	5
A042	benzofuran-5-carboxylic acid	41 ± 13	2
A018	indole-5-carboxylic acid	130 ± 30	2
	in combination with A077		
B286	5-bromoindole-3-acetic acid	0.70 ± 0.04	90
B107	5-chlorobenziothiophene-3-acetic acid	1.4 ± 0.1	14
B182	indole-3-acetic acid	20 ± 5	1
B168	benzothiophene-3-acetic acid	9.3 ± 1.0	1

[a] K_d values were determined by fluorescence polarization.

The structural features linked to elevated NSC values were experimentally confirmed to be important for high-affinity binding (Table 2, Table S4). Compounds containing the listed features had dissociation constants smaller than 9 nM [K_d (A017/B321) = 2.5 nM; K_d (A085/B321) = 4.0 nM; K_d (A077/B107) = 1.4 nM] whereas related heterocycles without the corresponding C-5 substituent had K_d values greater than 9 nM [K_d (A020/B321) = 10.5 nM; K_d (A024/B321) = 9.6 nM; K_d (A077/B182) = 20 nM; K_d (A077/B168) = 9.3 nM] and K_d values for compounds with the carboxylic acid modification at positions other than C-2 or C-3 were greater than 40 nM [K_d (A042/B321) = 41 nM; K_d (A018/B321) = 130 nM]. Additional patterns of structural similarities were observed for other families of enriched compounds including several phenylacetic and benzoic acids modified with hydrophobic substituents and the measured affinities largely agreed with the observed sequence counts (Table S4).

Furthermore, we observed a clear connection between the NSC values and the experimental affinities (Figure 3). Most compounds with low sequence counts (empirical threshold NSC ca. 5) had K_d values greater than 10 nM. Compounds with NSC values greater than 5 had K_d values below 10 nM and the affinities largely correlated with the sequence counts.

The screening experiment for TNKS1 resulted in strong enrichment of several compounds in the sequencing fingerprint (Figure 2f). TNKS1 is a poly(ADP-ribose) polymerase involved in telomerase regulation and the Wnt signalling pathway; TNKS1 has been proposed as a potential drug

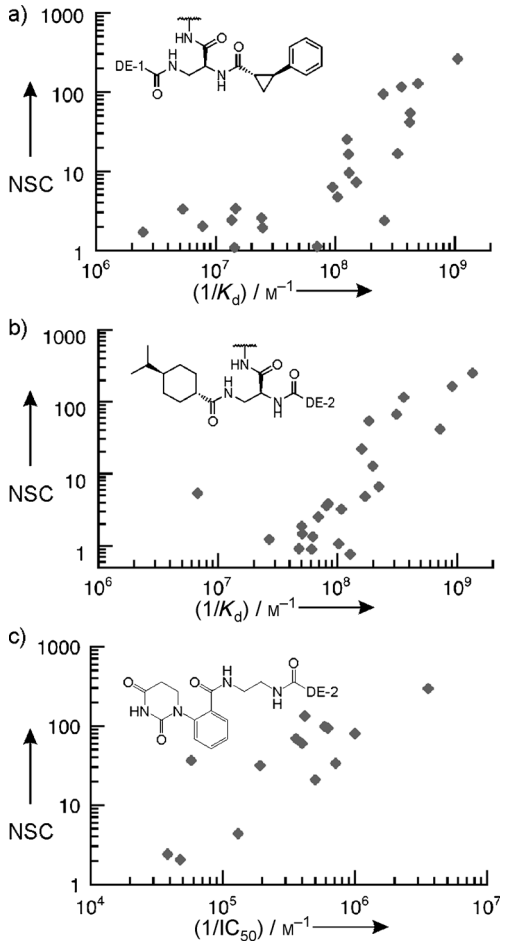

Figure 3. Dependence of binding affinity and inhibition and sequence counts. Series of HSA-binders with a conserved a) B321 and b) A077 building block. c) Inhibitors of TNKS1 with a conserved A102 building block. For structures see Tables S4 and S7.

target.[11] Three small-molecule derivatives of the hit compounds A102/B104 and A187/B278 were synthesized (en: ethylenediamine, pip: piperazine, dap: L-2,3-diaminoproanamide; Table S6) and tested as inhibitors in an in vitro PARylation assay (Table 3 and Figure S7).[12] All A102/B104 derivatives were potent TNKS1 inhibitors with half maximal inhibitory concentration (IC_{50}) values of 290–370 nM. A187-Dap-B278 inhibited TNKS1 with an IC_{50} value of 3.9 μM, whereas A187-Pip-B278 and A187-En-B278 were less potent.

A series of further A102-containing hit compounds was synthesized and the IC_{50} values of TNKS1 inhibition were measured (Table S7).[12] Overall, a clear trend between the NSC values of these compounds and the measured potencies ($1/IC_{50}$) was observed (Figure 3c). The most highly enriched compound A102/B104 was the most potent inhibitor (IC_{50} = 290 nM) and other A102-containing hit compounds generally had IC_{50} values in the range of 1–7 μM, which is substantially better than structurally related compounds that were not enriched (IC_{50} = 9–20 μM). Some of these fragments (e.g. 4-oxo-4-phenylbutanoic acids and 3-propionic acids of 2-phenylazoles) were reminiscent of previously described TNKS1 inhibitors.[11]

Table 3: In vitro inhibition of TNKS1 activity by selected hit compounds.[a]

IC_{50}	En [μM]	Pip [μM]	Dap [μM]
A102/B104	0.29	0.29	0.37
A187/B278	40	33	3.9

[a] IC_{50} values were determined by an enzymatic PARylation assay (pIC_{50} values and standard errors are listed in Table S6).

In summary, we designed and synthesized DAL-100K, a DECL containing more than 100000 compact compounds composed of structurally diverse carboxylic acid building blocks. DAL-100K was successfully applied to the identification of small-molecule binders for several targets of pharmaceutical interest. For instance, nanomolar inhibitors for TNKS1 were identified directly from the screening results. Correlations between NSC values and the experimental affinity constants were observed for HSA and TNKS1. The possibility to derive SARs directly from a single screening experiment is likely to be important for medicinal chemistry optimization steps and for computer-assisted docking procedures.[4f]

Keywords: combinatorial chemistry · drug discovery · encoded libraries · high-throughput screening · structure–activity relationships

How to cite: *Angew. Chem. Int. Ed.* **2015**, *54*, 3927–3931
Angew. Chem. **2015**, *127*, 3999–4003

[1] a) R. M. Franzini, D. Neri, J. Scheuermann, *Acc. Chem. Res.* **2014**, *47*, 1247–1255; b) N. Winssinger, *Chimia* **2013**, *67*, 340–348.

[2] a) L. Mannocci, Y. Zhang, J. Scheuermann, M. Leimbacher, G. De Bellis, E. Rizzi, C. Dumelin, S. Melkko, D. Neri, *Proc. Natl. Acad. Sci. USA* **2008**, *105*, 17670–17675; b) P. Zhao, Z. Chen, Y. Li, D. Sun, Y. Gao, Y. Huang, X. Li, *Angew. Chem. Int. Ed.* **2014**, *53*, 10056–10059; *Angew. Chem.* **2014**, *126*, 10220–10223.

[3] a) S. Brenner, R. A. Lerner, *Proc. Natl. Acad. Sci. USA* **1992**, *89*, 5381–5383; b) Z. J. Gartner, B. N. Tse, R. Grubina, J. B. Doyon, T. M. Snyder, D. R. Liu, *Science* **2004**, *305*, 1601–1605; c) S. Melkko, J. Scheuermann, C. E. Dumelin, D. Neri, *Nat. Biotechnol.* **2004**, *22*, 568–574; d) F. Debaene, J. A. Da Silva, Z. Pianowski, F. J. Duran, N. Winssinger, *Tetrahedron* **2007**, *63*, 6577–6586; e) S. J. Wrenn, R. M. Weisinger, D. R. Halpin, P. B. Harbury, *J. Am. Chem. Soc.* **2007**, *129*, 13137–13143; f) M. A. Clark, et al., *Nat. Chem. Biol.* **2009**, *5*, 647–654; g) M. H. Hansen, P. Blakskjaer, L. K. Petersen, T. H. Hansen, J. W. Hojfeldt, K. V. Gothelf, N. J. Hansen, *J. Am. Chem. Soc.* **2009**, *131*, 1322–1327; h) F. Buller, Y. Zhang, J. Scheuermann, J. Schafer, P. Buhlmann, D. Neri, *Chem. Biol.* **2009**, *16*, 1075–1086; i) N. Svensen, J. J. Diaz-Mochon, K. Dhaliwal, S. Planonth, M. Dewar, J. D. Armstrong, M. Bradley, *Angew. Chem. Int. Ed.* **2011**, *50*, 6133–6136; *Angew. Chem.* **2011**, *123*, 6257–6260; j) N. Svensen, J. J. Diaz-Mochon, M. Bradley, *Chem. Commun.* **2011**, *47*, 7638–7640; k) C. Cao, P. Zhao, Z. Li, Z. Chen, Y. Huang, Y. Bai, X. Li, *Chem. Commun.* **2014**, *50*, 10997–10999; l) A. Novoa, T. Machida, S. Barluenga, A. Imberty, N. Winssinger, *ChemBioChem* **2014**, *15*, 2058–2065; m) J.-P. Daguer, C. Zambaldo, M. Ciobanu, P. Morieux, S. Barluenga, N. Winssinger, *Chem. Sci.* **2015**, *6*, 739–744.

[4] a) R. E. Kleiner, C. E. Dumelin, G. C. Tiu, K. Sakurai, D. R. Liu, *J. Am. Chem. Soc.* **2010**, *132*, 11779–11791; b) 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, D. Neri, *ChemMedChem* **2010**, *5*, 584 – 590; c) V. Baeriswyl, H. Rapley, L. Pollaro, C. Stace, D. Teufel, E. Walker, S. Chen, G. Winter, J. Tite, C. Heinis, *ChemMedChem* **2012**, *7*, 1173 – 1176; d) H. Deng, et al., *J. Med. Chem.* **2012**, *55*, 7061 – 7079; e) M. Leimbacher, Y. Zhang, L. Mannocci, M. Stravs, T. Geppert, J. Scheuermann, G. Schneider, D. Neri, *Chem. Eur. J.* **2012**, *18*, 7729 – 7737; f) J. S. Disch, et al., *J. Med. Chem.* **2013**, *56*, 3666 – 3679; g) P. L. Podolin, et al., *Prostaglandins Other Lipid Mediat.* **2013**, *104 – 105*, 25 – 31; h) C. S. Kollmann, et al., *Bioorg. Med. Chem.* **2014**, *22*, 2353; i) L. Encinas, et al., *J. Med. Chem.* **2014**, *57*, 1276 – 1288.

[5] C. A. Lipinski, F. Lombardo, B. W. Dominy, P. J. Feeney, *Adv. Drug Delivery Rev.* **2001**, *46*, 3 – 26.

[6] R. M. Franzini, F. Samain, M. Abd Elrahman, G. Mikutis, A. Nauer, M. Zimmermann, J. Scheuermann, J. Hall, D. Neri, *Bioconjugate Chem.* **2014**, *25*, 1453 – 1461.

[7] M. Congreve, R. Carr, C. Murray, H. Jhoti, *Drug Discovery Today* **2003**, *8*, 876 – 877.

[8] C. T. Supuran, *Nat. Rev. Drug Discovery* **2008**, *7*, 168 – 181.

[9] A. Ghosh, W. D. Heston, *J. Cell. Biochem.* **2004**, *91*, 528 – 539.

[10] a) C. E. Dumelin et al., *Angew. Chem. Int. Ed.* **2008**, *47*, 3196 – 3201; *Angew. Chem.* **2008**, *120*, 3240 – 3245; b) L. H. Pollaro, C. Heinis, *MedChemComm* **2010**, *1*, 319 – 324.

[11] J. L. Riffell, C. J. Lord, A. Ashworth, *Nat. Rev. Drug Discov.* **2012**, *11*, 923 – 936.

[12] E. Wahlberg et al., *Nat. Biotechnol.* **2012**, *30*, 283 – 288.

Received: November 4, 2014
Published online: February 3, 2015

