Combinatorial Chemistry

DOI: 10.1002/anie.201310864

Multi-Objective Molecular De Novo Design by Adaptive Fragment Prioritization**

*Michael Reutlinger, Tiago Rodrigues, Petra Schneider, and Gisbert Schneider**

Abstract: *We present the development and application of a computational molecular de novo design method for obtaining bioactive compounds with desired on- and off-target binding. The approach translates the nature-inspired concept of ant colony optimization to combinatorial building block selection. By relying on publicly available structure–activity data, we developed a predictive quantitative polypharmacology model for 640 human drug targets. By taking reductive amination as an example of a privileged reaction, we obtained novel subtype-selective and multitarget-modulating dopamine D_4 antagonists, as well as ligands selective for the sigma-1 receptor with accurately predicted affinities. The nanomolar potencies of the hits obtained, their high ligand efficiencies, and an overall success rate of 90% demonstrate that this ligand-based computer-aided molecular design method may guide target-focused combinatorial chemistry.*

Traditional combinatorial chemistry aims at the generation of large diverse compound arrays for bioactivity screening.[1] It has been realized that multiple "adaptive" synthesis-and-test cycles using smaller, focused compound libraries might be better suited, faster, and more economical to find lead-like bioactive compounds.[2,3] Computational molecular design methods offer the additional advantage of generating bioactive compounds while considering multiple objectives in parallel,[4] and combinatorial libraries with desired properties can be obtained by relying on chemistry-oriented computational molecular design.[5,6] Though potentially appealing, these methods have rarely been prospectively applied. Here, we present the comprehensive application of a computational concept for designing combinatorial libraries that exhibit an accurately predicted bioactivity profile. We show that the molecular ant algorithm (MAntA)[7] effectively transfers a nature-inspired optimization principle to chemistry-driven molecular design.

For a proof-of-concept we focused on the reductive amination reaction as a scheme for combinatorial synthesis. By automated structure optimization, MAntA generated small compound libraries with lead-like qualities, high hit rates, and nanomolar activities. It implements a new design strategy that is applicable to all kinds of chemistry-driven computational methods,[8] and neither requires prior knowledge about the bioactivity of scaffold classes nor is limited to privileged scaffolds. In a retrospective study, ant colony optimization turned out to perform better or on par with other optimization methods.[7b] Here, we pioneer the concept of polypharmacology-based molecular de novo design using combinatorial chemistry. We demonstrate that both target-selective, and multitarget-modulating members of large combinatorial compound libraries are rapidly identified without the need for full library enumeration and synthesis.

The molecular design method requires 1) a scheme for compound synthesis, 2) a method for predicting the affinity of the virtual products, and 3) a technique for optimizing the building blocks. For our concept study, we chose the reductive amination reaction and aldehydes/ketones and amines as building blocks. We applied MAntA to the products of single-step reductive amination starting from commercially available building blocks. The reaction products have a high likelihood of possessing desirable druglike features, as visualized in Figure 1, which presents a map of the known bioactivity space. Virtual reaction products (green dots) cluster in a densely populated area, and the reductive amination may be regarded as a preferred reaction for drug discovery.

For affinity prediction we trained individual Gaussian process (GP) regression models[9] for 640 human targets annotated in ChEMBL (v14),[10] based on 279866 compounds with 569725 measured bioactivities. Molecules were represented by topological pharmacophore ("CATS2")[11] and substructure (circular Morgan fingerprints)[12] descriptors. The choice of GP regression was motivated by extensive comparison to other modeling techniques using the same training data, where the GP approach performed best (Tables S2 and S3 in the Supporting Information). In addition, GP models compute a data-density-dependent confidence estimate, which we combined with the quantitative bioactivity prediction (p*Affinity*) to obtain a single robust prediction score for each compound.

Equipped with this quantitative affinity prediction model, MAntA performs an adaptive search for optimal combinations of building blocks for the given reaction scheme (Figure 2). The search space consists of all possible substrates

[*] M. Reutlinger, Dr. T. Rodrigues, Dr. P. Schneider, Prof. Dr. G. Schneider
Eidgenössische Technische Hochschule (ETH)
Departement Chemie und Angewandte Biowissenschaften
Vladimir-Prelog-Weg 4, 8093 Zürich (Switzerland)
E-mail: gisbert.schneider@pharma.ethz.ch
Dr. P. Schneider, Prof. Dr. G. Schneider
inSili.com GmbH
Segantinisteig 3, 8049 Zürich (Switzerland)

[**] We thank Dr. Michael Bieler for computing QED values. This research was financially supported by a grant from the OPO Foundation, Zürich.

Supporting information for this article (details on the computational methods, synthesis, analytics, and binding assays) is available on the WWW under http://dx.doi.org/10.1002/anie.201310864.

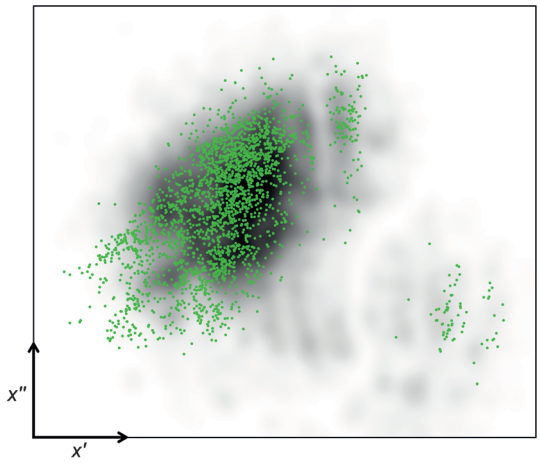


Figure 1. The distribution of 5000 virtual products (green dots) generated by reductive amination in a druglike chemical space. The two-dimensional landscape was calculated from the density of 10000 druglike molecules sampled from the ChEMBL database. The intensity of the gray color indicates the density of known bioactive substances (white: sparsely populated; black: highest local density). The compounds were represented by topological pharmacophores ("CATS2" descriptor)[11] and projected to the plane (x',x'') by stochastic neighbor embedding (SNE), which led to a local-neighborhood preserving map of chemical space. The axes represent nonlinear combinations of the original molecular descriptors. The image was generated with LiSARD.[20]

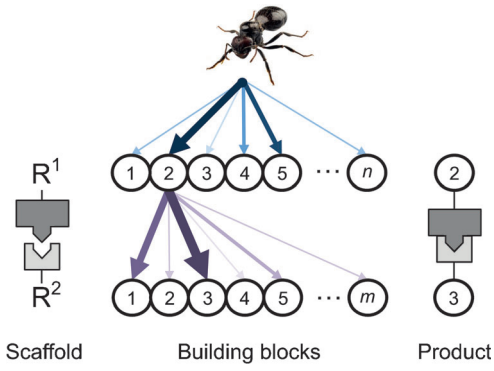


Figure 2. Selection of molecular building blocks by combinatorial ant colony optimization (MAntA). The arrows represent artificial ant paths for this two-component combinatorial library. The widths of the paths correspond to pseudo-probabilities ("pheromone concentrations") that influence the choice made by the ants and thereby determine the actual product spectrum. The pheromone concentrations are adaptive and subject to "evaporation".

labeled with pseudo-probabilities ("pheromones"), according to their contributions to the computed predictive score. Individual ants traverse the search space following pheromone trails and assemble virtual products. These are scored and the pseudo-probabilities on their respective molecular building blocks are adjusted accordingly. Over simulation time, high-scoring combinations of building blocks emerge from the ant colony's optimal path-finding capability.

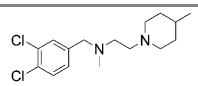
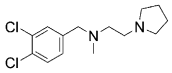
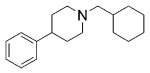
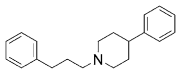
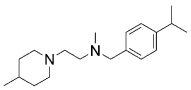
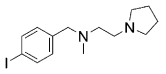
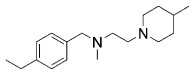
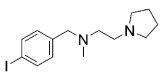
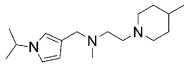
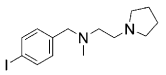
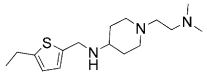
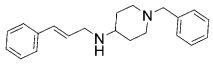
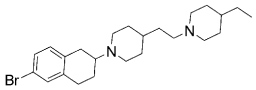
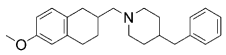
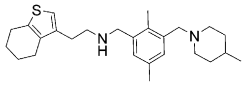
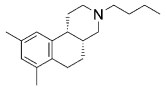
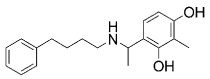
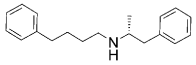
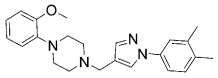
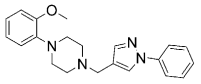
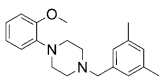
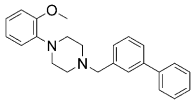
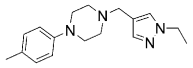
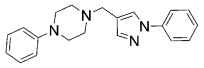
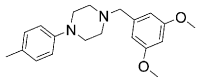
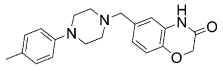
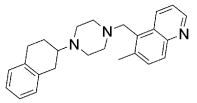
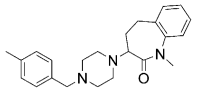
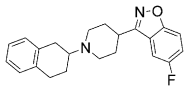
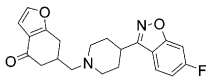
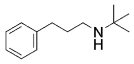
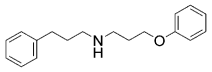
We employed MAntA for the multi-objective design of novel ligands for high-profile macromolecular drug targets that are involved in neuropsychiatric disorders—sigma-1 and dopamine D_4 receptors. The choice of D_4 receptor was also made to allow for a direct comparison to a recent publication by Hopkins and co-workers.[5] In our study, the task was to select a small number of preferred products from a total of approximately 20 million. First, we discarded all designed molecules with undesired structural motifs,[13] and poor predicted absorption, distribution, metabolism, and excretion properties ("negative design").[14] From the 3529 remaining molecules we selected candidates to match different criteria ("positive design"):

1) Potent and selective (sigma-1) or multitarget-modulating (dopamine D_4) ligands
2) Target-subtype-selective ligands
3) Exploratory molecules, lying outside the training domain as expressed by Morgan fingerprint Tanimoto similarities < 0.20
4) Inactive compounds that are nearest neighbors to known high-affinity ligands in ChEMBL bioactivity space.

For the sigma-1 receptor, we selected compounds **1**–**3** according to the high-affinity criterion. Molecules **4** and **5** were designed as receptor-selective ligands (Figure 3a). In fact, the MAntA designs were experimentally validated for their specific goals with accurately predicted pK_i values (Table 1). Compounds **1**–**3** exhibited K_i values of 1.1–2.2 nM, and designs **4** and **5** displayed more than 2500-fold selectivity for the sigma-1 receptor over the δ, κ, and μ opioid receptors. It should be noted that **1**–**4** are equipotent to their nearest-neighbor counterparts from ChEMBL, despite being structurally dissimilar (Tanimoto similarity ≈ 0.45), thereby endorsing the exploratory potential of MAntA. Furthermore, the low molecular weights of **1**–**5**, coupled to low nanomolar K_i values, endow these compounds with high ligand efficiency. Additionally, we synthesized and tested compounds **6**–**8** as scaffold hops from known ChEMBL chemical space (structural Tanimoto similarity to nearest neighbors ≈ 0.20), without critical loss of affinity (sigma-1: $K_i = 10$–210 nM, $\Delta pK_i \approx 0.5$; Table 1, Figure 3b) with the exception of compound **8**. Furthermore, compounds **1**–**8** contain scaffolds that were not present in the training data used for model building (Table S4). Apparently, the low structural resemblance to known small molecules did not considerably affect the algorithm's performance. Finally, compound **9** was designed and validated as a low-affinity sigma-1 ligand ($K_i > 2500$ nM) despite having a highly potent nearest neighbor ($K_i \approx 6$ nM, Tanimoto similarity $= 0.45$, Table 1), pinpointing the adaptive design capabilities of MAntA that go beyond structural similarity analysis. Altogether, the experimental results are in agreement with the landscape projection of the preferred sigma-1 activity islands (Figure 3c; individual target landscapes are shown in Figure S9). Furthermore, as an off-target for the synthesized compounds, MAntA predicted moderate histamine H_3 receptor affinities, which were partly confirmed experimentally.

Next, we designed antagonists for the dopamine D_4 receptor. D_4 receptors are especially implicated in attention-deficit hyperactivity disorder, mood disorders, and Parkinson's disease, among other neuropsychiatric illnesses.[15] From the top 1600 prioritized small molecules with predicted $pK_i >$

Table 1: The designed molecules and their nearest neighbors from the ChEMBL training data with the predicted and experimentally determined binding affinities of compounds **1**–**16**.

MAntA designs					Nearest neighbors (training data)			
ID	Structure	Predicted p*Affinity*	pK_i	LE[a]	Structure	ChEMBL ID[a]	Structural similarity[c]	pK_i
1		9.7	9.0	0.63		143089	0.70	9.4
2		9.4	8.9	0.65		112124	0.44	9.0
3		9.3	8.7	0.58		154397	0.44	8.7
4		9.0	8.8	0.61		154397	0.47	8.7
5		8.1	7.9	0.55		154397	0.34	8.7
6		7.9	7.2	0.50		111909	0.21	7.8
7		8.1	8.1	0.42		179530	0.21	7.7
8		8.0	6.7	0.32		544748	0.20	8.7[d]
9		4.1	n.d.	n.d.		20976	0.45	8.2
10		9.4	8.7	0.44		379602	0.70	9.6
11		9.0	8.3	0.50		285577	0.57	8.1
12		7.9	8.0	0.53		210405	0.41	9.0
13		7.8	7.9	0.46		345552	0.43	8.4
14		7.9	6.6	0.33		305061	0.23	5.4
15		7.3	7.6	0.41		143027	0.27	7.7
16		2.5	n.d.	n.d.		129931	0.40	8.5

[a] Ligand efficiency (LE = $-1.4 \times pK_i$/number of heavy atoms). [b] ChEMBL IDs are given without the "CHEMBL" prefix. [c] Tanimoto similarity index (Morgan fingerprints with radius = 3). [d] pIC_{50} value. n.d.: not determined.

7 for the D_4 receptor, we selected compounds **10** and **11** as high-affinity ligands. Although **10** ($K_i = 2.0$ nM) features an already known scaffold,[16] **11** represents a notably different entity that is more ligand efficient than its ChEMBL nearest neighbor (Tanimoto similarity ≈ 0.6, Table 1). While the selected ligands were primarily designed as high-affinity D_4 receptor antagonists, a polypharmacology profile was not precluded. Accordingly, a promiscuous binding profile to the dopamine D_{1-5} and 5-HT_{1A} receptors was predicted. Subsequent binding tests confirmed the multitarget-modulating profiles of **10** and **11** in agreement with the MAntA-predicted bioactivity spectra and landscape projections (Table 1, Figure 3b, preferred design zones Figure 3d; individual target landscapes are shown in Figure S10). Conversely, compounds

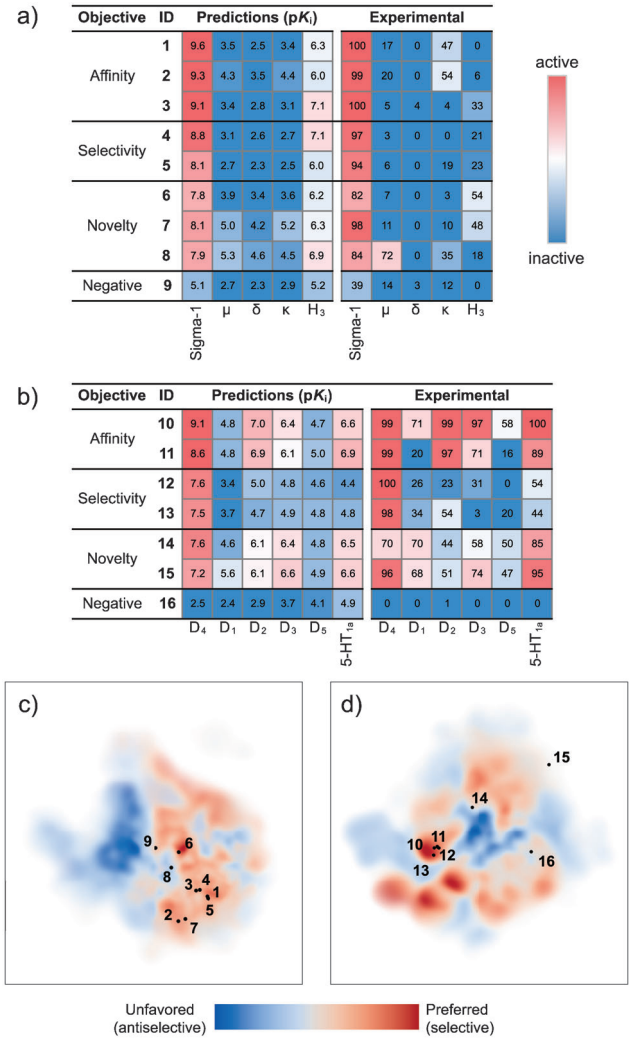


Figure 3. Comparison of the Gaussian process prediction for sigma-1 (a) and dopamine D_4 (b) receptors, together with the observed experimental results. Predictions are variance-corrected p*Affinity* values. Experimental data are expressed as % inhibition (competitive radioligand binding assays). Colors are linearly interpolated from the predicted p*Affinity* intervals [4, 9] (a) and [4, 8] (b), and % inhibition interval [20, 100] for the experimental data. The lower panels presents LiSARD[20] multitarget selectivity landscapes for the sigma-1 receptor (c) and the dopamine D_4 receptor (d) as the respective on-target.

12 and **13** were designed to bind selectively to the D_4 receptor. Selective D_4 antagonists are equally relevant in clinics, as they can prevent stress-induced cognitive dysfunction without extrapyramidal motor symptoms or neuroendocrine side effects.[17] Weak binding affinities of **12** and **13** were predicted for the off-targets in the assay panel. Structural simplicity and low nanomolar affinities ($K_i = 10$–12 nM) distinguish these compounds. Their selectivity for the D_4 receptor is particularly significant, given the high structural similarity to promiscuous molecules **10** and **11**. Of note, 1,4-disubstituted aromatic piperazines have previously been recognized as predominant in promiscuous biogenic amine G-protein-coupled receptor (GPCR) ligands.[18] The opposing target-engagement profiles for the arylpiperazines **10** and **11** as well as **12** and **13** confirm effective building-block selections. The polypharmacology profile of **14** and **15**, which extend the known chemical diversity of D_4 receptor antagonists, is also in agreement with the pK_i predictions. Remarkably, **14** is one log unit more potent against the D_4 receptor than the closest related reference antagonist, which together with the screening results of the designed inactive **16**, demonstrates the successful application of MAntA to dopamine receptors. Evidently, considerably extended experimental GPCR panel activities will be required for the further hit-to-lead progression of the MAntA designs.

With regard to the polygenic nature of most major diseases of the central nervous system and the individual variability of their genetic basis, new drugs with selected polypharmacological activities are desirable.[19] The results of this study suggest a feasible solution for the combinatorial design of new chemical entities with affinity profiles and properties that exceed the average druglike-ness for approved drugs (quantitative estimate of druglike-ness, QED $= 0.72 \pm 0.10$ vs. 0.49).[3] The automated molecular design method should be broadly applicable to other classes of drug targets and chemical reactions, provided reliable structure–activity data are available for constructing affinity prediction models. The actual computational design process is fast (within minutes on a desktop computer), so that focused combinatorial library design and synthesis can be realized within a day of work. A particular advantage of MAntA compared to many other approaches, for example, the meticulous work of Besnard et al. on adaptive drug design,[5] lies in the simultaneous generation of both potent structural analogues and innovative scaffold-hops from known reference compounds. Together with rapid computation, low-cost synthesis, and readily accessible chemical structures, the concept of adaptive building block and fragment prioritization might become widely applicable.

Received: December 15, 2013
Published online: March 12, 2014

Keywords: computer-assisted drug design · GPCR · machine learning · polypharmacology · reductive amination

[1] A. Long, *Curr. Protoc. Pharmacol.* **2012**, *9*, 9–16.
[2] G. Schneider, M. Hartenfeller, M. Reutlinger, Y. Tanrikulu, E. Proschak, P. Schneider, *Trends Biotechnol.* **2009**, *27*, 18–26.
[3] a) M. M. Hann, T. I. Oprea, *Curr. Opin. Chem. Biol.* **2004**, *8*, 255–263; b) G. R. Bickerton, G. V. Paolini, J. Besnard, S. Muresan, A. L. Hopkins, *Nat. Chem.* **2012**, *4*, 90–98.
[4] a) V. J. Gillet, *Methods Mol. Biol.* **2004**, *275*, 335–354; b) C. A. Nicolaou, N. Brown, C. S. Pattichis, *Curr. Opin. Drug Discovery Dev.* **2007**, *10*, 316–324; c) G. Fang, M. Xue, M. Su, D. Hu, X. Li, B. Xiong, L. Ma, T. Meng, Y. Chen, J. Li, J. Li, J. Shen, *Bioorg. Med. Chem. Lett.* **2012**, *22*, 4540–4555.
[5] J. Besnard, G. F. Ruda, V. Setola, K. Abecassis, R. M. Rodriguiz, X. P. Huang, S. Norval, M. F. Sassano, A. I. Shin, L. A. Webster, F. R. Simeons, L. Stojanovski, A. Prat, N. G. Seidah, D. B. Constam, G. R. Bickerton, K. D. Read, W. C. Wetsel, I. H. Gilbert, B. L. Roth, A. L. Hopkins, *Nature* **2012**, *492*, 215–220.
[6] M. Reutlinger, T. Rodrigues, P. Schneider, G. Schneider, *Angew. Chem.* **2014**, *126*, 593–596; *Angew. Chem. Int. Ed.* **2014**, *53*, 582–585.

[7] a) J. A. Hiss, M. Hartenfeller, G. Schneider, *Curr. Pharm. Des.* **2010**, *16*, 1656 – 1665; b) J. A. Hiss, M. Reutlinger, C. P. Koch, A. M. Perna, P. Schneider, T. Rodrigues, S. Haller, G. Folkers, L. Weber, R. B. Baleeiro, P. Walden, P. Wrede, G. Schneider, *Future Med. Chem.* **2014**, *6*, 267 – 280.
[8] a) M. Hartenfeller, H. Zettl, M. Walter, M. Rupp, F. Reisen, E. Proschak, S. Weggen, H. Stark, G. Schneider, *PLoS Comput. Biol.* **2012**, *8*, e1002380; b) M. Hartenfeller, G. Schneider, *WIREs Comput. Mol. Sci.* **2011**, *1*, 742 – 759; c) *De Novo Molecular Design* (Ed.: G. Schneider), Wiley-VCH, Weinheim, **2013**.
[9] C. E. Rasmussen, C. K. I. Williams, *Gaussian Processes for Machine Learning*, MIT, Cambridge, **2006**.
[10] A. Gaulton, L. J. Bellis, A. P. Bento, J. Chambers, M. Davies, A. Hersey, Y. Light, S. McGlinchey, D. Michalovich, B. Al-Lazikani, J. P. Overington, *Nucleic Acids Res.* **2012**, *40*, D1100 – D1107.
[11] M. Reutlinger, C. P. Koch, D. Reker, N. Todoroff, P. Schneider, T. Rodrigues, G. Schneider, *Mol. Inf.* **2013**, *32*, 133 – 138.
[12] D. Rogers, M. Hahn, *J. Chem. Inf. Model.* **2010**, *50*, 742 – 754.
[13] S. Saubern, R. Guha, J. B. Baell, *Mol. Inf.* **2011**, *30*, 847 – 850.
[14] D. Lagorce, O. Sperandio, H. Galons, M. A. Miteva, B. O. Villoutreix, *BMC Bioinf.* **2008**, *9*, 396.
[15] F. I. Tarazi, R. J. Baldessarini, *Mol. Psychiatry* **1999**, *4*, 529 – 538.
[16] S. Löber, H. Hübner, P. Gmeiner, *Bioorg. Med. Chem. Lett.* **2006**, *16*, 2955 – 2959.
[17] A. F. Arnsten, B. Murphy, K. Merchant, *Neuropsychopharmacology* **2000**, *23*, 405 – 410.
[18] S. Löber, H. Hübner, N. Tschammer, P. Gmeiner, *Trends Pharmacol. Sci.* **2011**, *32*, 148 – 157.
[19] P. Giusti-Rodríguez, P. F. Sullivan, *J. Clin. Invest.* **2013**, *123*, 4557 – 4563.
[20] M. Reutlinger, W. Guba, R. E. Martin, A. I. Alanine, T. Hoffmann, A. Klenner, J. A. Hiss, P. Schneider, G. Schneider, *Angew. Chem.* **2011**, *123*, 11837 – 11840; *Angew. Chem. Int. Ed.* **2011**, *50*, 11633 – 11636.

