ChemPubSoc Europe

DOI: 10.1002/chem.201605499

CHEMISTRY A European Journal Communication

Retrosynthetic Design

Neural-Symbolic Machine Learning for Retrosynthesis and Reaction Prediction

Marwin H. S. Segler[a] and Mark P. Waller*[a, b]

Abstract: Reaction prediction and retrosynthesis are the cornerstones of organic chemistry. Rule-based expert systems have been the most widespread approach to computationally solve these two related challenges to date. However, reaction rules often fail because they ignore the molecular context, which leads to reactivity conflicts. Herein, we report that deep neural networks can learn to resolve reactivity conflicts and to prioritize the most suitable transformation rules. We show that by training our model on 3.5 million reactions taken from the collective published knowledge of the entire discipline of chemistry, our model exhibits a top10-accuracy of 95% in retrosynthesis and 97% for reaction prediction on a validation set of almost 1 million reactions.

The deliberate and controlled synthesis of molecules is one of the central pillars of well-being in modern society, as it provides a wide range of chemicals from medicines to materials. To rationally synthesize new molecules, two intimately related problems, reaction prediction and retrosynthesis, have to be solved. In reaction prediction, the task is to infer how a set of molecules (starting materials) will react and what the product will be. In retrosynthesis, the aim is to propose how to make a target molecule, by deducing reactions in which the target molecule is the product. Computer programs that could assist in these tasks are desirable for practical synthesis, but even more so for automated synthesis and virtual space exploration in de novo molecular design.[1] However, current approaches are perceived to be limited.[2]

The standard methodology for retrosynthesis and reaction prediction are rule-based expert systems.[2,3] Transformation rules, some of which are named after their inventor(s) in the chemical curriculum, encode the functional groups that are involved in a reaction. The rules are applied to the reactants to obtain the product in reaction prediction, or in reverse, to the product, for retrosynthesis. This is essentially a symbolic pattern recognition process. The great advantage of rules is that they are straightforward to interpret. However, the rule-based approach has several drawbacks: First, rule-based expert systems cannot predict anything outside of their knowledge, which renders them unable to discover novel chemistry.[4] Second, the rules have to be compiled and curated. They can either be laboriously encoded by human experts, or extracted algorithmically from data. This procedure can be difficult because mechanistic understanding is usually needed to encode which neighbouring functional groups influence the outcome of a reaction. However, a rule is generally not enough to reliably predict reactions, because the context of the complete molecule is ignored (Figure 1). The rest of the molecule might contain functional groups that could interfere or compete with the intended reaction. Then, even though the rule formally matches, the reaction would not yield the desired product. Therefore, reaction rules have to be annotated with additional information about tolerated functional groups and selectivity to filter and prioritize the rules when a match has been found. This problem remains an unsolved challenge. Recently, Grzybowski and co-workers have argued, that it can only be tackled by tens of thousands of complex rules encoded by human experts, a process which takes years even for large teams.[2] In contrast, Kayala and Baldi concluded that the need for selectivity and tolerance rules renders the construction of expert systems impossible at scale.[5] An additional limitation of rules is that they lack an inherent ranking mechanism. However, given that complex targets can formally be made in hundreds of different ways, prioritizing rules is crucial. Similar to a search engine, the user legitimately expects to see the desired results ranked highly, possibly within the first 10 items shown. Importantly, rule-based expert systems for retrosynthesis have never been rigorously evaluated with large hold out test sets. Only anecdotal evidence based on a few examples has been provided for these systems so far.

Given the inherent challenges of rule-based systems, it is appealing to have the machine learn to predict reactions and perform retrosynthesis from experimental data. Earlier systems were shown to have excellent performance, but were focused on the prediction of restricted sets of special reaction classes.[3e,5,6] In 2005, Aires de Sousa and co-workers proposed to combine machine learning with the rule-based approach.[7] A model was trained on the reactant or product molecules to predict which is the most probable rule, which can then subse-

[a] M. H. S. Segler, Prof. M. P. Waller
Organisch-Chemisches Institut and Center for Multiscale Theory and Computation
Westfälische Wilhelms-Universität Münster
Corrensstr. 40, 48149 Münster (Germany)
E-mail: m.waller@uni-muenster.de

[b] Prof. M. P. Waller
Department of Physics and International Center for Quantum and Molecular Structures
Shanghai University, Shangda Road 99
200444 Shanghai (China)

Supporting information for this article can be found under http://dx.doi.org/10.1002/chem.201605499.

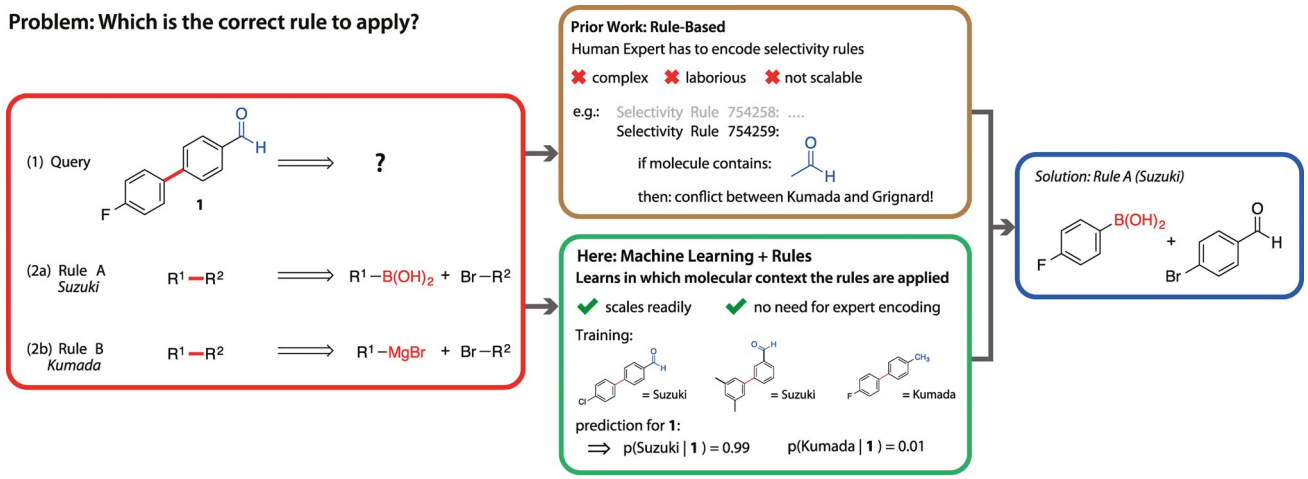

Figure 1. The challenge in retrosynthesis and reaction prediction is to select the correct rule among possibly tens or hundreds of matching rules. In this example, both a Suzuki and a Kumada coupling (among others) formally match the biaryl moiety. However, the aldehyde in the molecular context would be in conflict with a Grignard reagent. Therefore, the Kumada coupling cannot be used to make target **1**. This information has to be encoded by hand in expert systems. Our system simply learns it from data.

quently applied to the molecules. They first encoded molecules by unsupervised pre-training of self-organizing maps, a type of neural network,[8] which then serves as an input to a random forest classifier.[7] They conducted a study with a dataset of 278 reactions from the literature and 8 reaction classes. During the preparation of this manuscript, we became aware that Aspuru-Guzik and co-workers recently reported a neural network approach for reaction prediction, in which reactant and reagent molecules were encoded as neural[9] or extended connectivity fingerprints, which are concatenated.[10] They trained their models to predict 16 reaction types on an artificial dataset of 3200 reactions.

Herein, we propose a novel neural-symbolic model, which can be used for both reaction prediction and retrosynthesis. Our model uses neural networks in order to predict the most likely transformation rules to be applied to the input molecule(s) (see Figures 1 and 2).

In short, the computer has to learn which named reaction was used to make a molecule (or under which rule the starting materials reacted). We trained neural networks by showing them millions of examples of known reactions and the corresponding correct reaction rule. The goal is to learn patterns in the molecules' functional groups that allow the machine to generalize to molecules it has not seen before. In this problem the model learns to predict the probability over all rules.[11] We hypothesize that the advantage of combining machine learning with symbolic rules is that we retain the familiar concept of rules, whereas the model learns to prioritize the rules and to estimate selectivity and compatibility from the provided training data, which are successfully performed experiments. We test our hypothesis in the first large-scale systematic investigation of machine learning and rule-based systems.

Several metrics are reported in Table 1 and Table 2 to evaluate the models. Accuracy shows how many reactions are correctly predicted when the rule with the highest predicted probability is evaluated. In top-*n* accuracy, we examine if the correct reaction rule is among the *n* highest ranked rules, similar to being on the first page of the results of a search engine. This means we allow the algorithm to propose more than one applicable rule, which is reasonable, since the same molecules can react differently, or several routes are possible to synthesize a target molecule in retrosynthesis. Furthermore, we

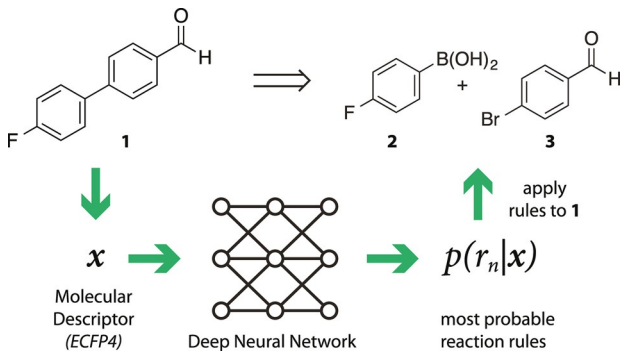

Figure 2. Overview of our neural-symbolic ansatz.

Table 1. Results for the study on 103 hand coded rules.

Task/Model	Acc	Top **3**-Acc	MRR	W. Prec.
Reaction prediction				
random	0.03	0.12	0.04	0.03
expert system	0.07	0.33	0.12	0.46
logistic regression	0.86	0.97	0.91	0.86
highway network	0.92	0.99	0.96	0.92
FC512 ELU	0.92	0.99	0.96	0.92
Retrosynthesis				
random	0.03	0.12	0.04	0.03
expert system	0.05	0.30	0.06	0.11
logistic regression	0.64	0.95	0.77	0.62
highway network	0.77	0.98	0.86	0.77
FC512 ELU	0.78	0.98	0.87	0.78

Table 2. Results for the study on 8720 automatically extracted rules.

Task/model	Acc	Top **10**-Acc.	MRR	W. Prec.
Reaction prediction				
random	0.00	0.00	0.00	0.00
expert system	0.02	0.18	0.02	0.06
logistic regression	0.41	0.65	0.49	0.31
highway network	0.78	0.98	0.86	0.77
FC512 ELU	0.77	0.97	0.85	0.76
Retrosynthesis				
random	0.00	0.00	0.00	0.00
expert system	0.02	0.19	0.01	0.06
logistic regression	0.31	0.59	0.41	0.23
highway network	0.64	0.95	0.75	0.63
FC512 ELU	0.62	0.94	0.74	0.62

report mean reciprocal rank (MRR), which is a metric from information retrieval.[11] It measures the ability of a system to highly rank the correct result. A value of 1.0 indicates perfect ranking, a value of 0.5 could arise if all correct results are ranked second. Weighted precision (w. Prec.) measures the ability of the system to not create false positives.[11] For each task, we compare our best neural-symbolic models, a neural network with one hidden layer (FC512 ELU)[12] and a deep highway network,[13] to a purely rule-based expert system operating with the same rule set. Further comparisons are made to logistic regression (which is a simpler linear machine learning model) and to a random stratified baseline, which just picks rules at random based on their distribution, without performing any rule matching. Unfortunately, no hand-annotated expert systems are free or open source, and the annotations themselves are not published in the public domain making a direct comparison unfeasible. Fortunately, our model performs so well, that a head to head comparison is not required because any marginal gains made by a hand-annotated system are greatly outweighed by the cost associated with curated rules. Therefore, the machine-learning algorithm obviates the need for tedious expert encoding of information about the applicability, the context, and the priority for individual rules by annotations.

Hand-coded reactions

In this experiment, the models had to predict the correct rule amongst 103 hand-coded rules for retrosynthesis and reaction prediction, see Table 1. For reaction prediction, the accuracy of the rule-based system is 0.07, whereas our model reaches an accuracy of 0.92. This implies that the rule-based system gets 7 out of 100 reactions right, whereas our system correctly predicts 92 out of 100. In retrosynthesis, the expert system yields an accuracy of 0.05 and an MRR of 0.01. Our single-layer neural network reaches an accuracy of 0.78 and an MRR of 0.87, which implies that the model is very well able to rank the true reactions highly. The neural-symbolic models thus outperform both the simpler logistic regression, and the rule-based expert system by a large margin.

Encouraged by these results we sought out a much tougher challenge for the machine-learning algorithm. We decided to see if the performance would be maintained when we use a much larger rule set, which would cover a large percentage of all reactions ever published in the chemical literature.

Automatically extracted rules

Table 2 shows the results for the prediction with a large set of 8720 automatically extracted rules. This makes the model fully data-driven. In the forward reaction prediction task, the expert system matches a mean of 44.5 rules per query. This clearly demonstrates that prioritizing of the rules is essential for reliable predictions, and is reflected in the poor accuracy (0.02) of the expert system. The accuracy of the best neural-symbolic model is 0.78. This indicates that the model is 39 fold better than the purely rule based system, and if an expert user of the model is content with inspecting the highest 10 proposed reactions the prediction accuracy reaches an almost perfect score of 0.98. Logistic regression is again much weaker than the neural network models, which demonstrates the need for a nonlinear model.

In the retrosynthesis task on 8720 rules, random guessing is 0.00 in all metrics. The expert system only obtains an accuracy of 0.02, a top10-accuracy of 0.19 and a MRR of 0.01. The best neural-symbolic model reaches an accuracy of 0.64 and a top10-accuracy of 0.95. This means the model is capable of making very good predictions, if it is allowed to propose just 10 different routes. The mean reciprocal rank is 0.75, which indicates good ability to prioritize the reactions. Logistic regression again performs worse than the neural networks.

There are several observations to be discussed. First, the neural-symbolic models outperform expert systems in all experiments because the rule-based system matches tens or hundreds of rules. Most of these rules are irrelevant or even incorrect because of reactivity conflicts. In contrast, the neural-symbolic models have learnt to focus on the relevant rules, indicated by the high MRR and top-*n* accuracy scores. We conclude that after seeing millions of examples, the neural network has not only learned which functional groups are involved in a rule, but also which molecular contexts are tolerated. If, for example, a Kumada coupling never occurs in the presence of an aldehyde, but a Suzuki coupling does, the system will learn to assign a higher probability to the Suzuki coupling if presented with aldehyde-bearing coupling partners. Second, rule-based systems without additional information about the molecular context perform only slightly better than random, even if the rule set is small. Reaction-driven de novo molecular design approaches would therefore benefit from neural-symbolic models. Thirdly, the overall metrics for retrosynthesis are lower than in the reaction prediction task. This is expected, because in retrosynthesis the system has less information available (just the product), whereas in forward reaction prediction the reactive functional groups are still present in the reactants.

Finally, the difference in performance between the hand coded and the algorithmically extracted rules can be attributed

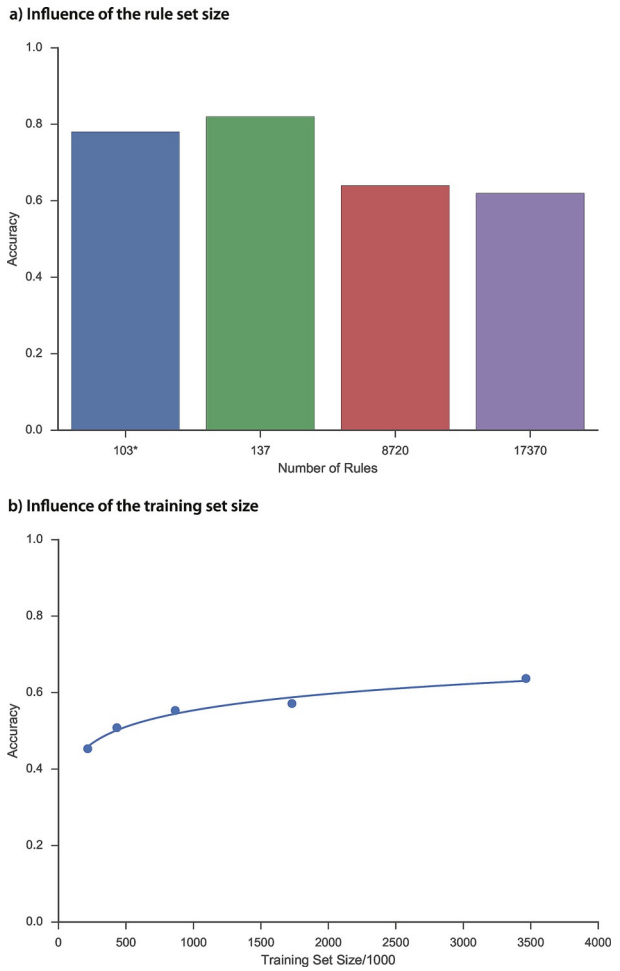

Figure 3. a) Performance improves with smaller rule sets. Interestingly, the hand-coded rules [103*(hc)] are outperformed by the algorithmically extracted rules when the rule set size is similar. b) Model performance improves when training on larger datasets (retrosynthesis, 8270 rules).

to the size of the rule sets. A model with 137 algorithmically extracted rules performed surprisingly better than the system with 103 hand-coded reactions (see Figure 3a and Supporting Information). Furthermore, as expected in a machine-learning experiment, the performance improves with training set size (Figure 3b). Importantly, reasonable performance is already achieved with datasets in the size of typical in-house electronic laboratory notebooks (ELN).

Timing

Training our largest neural networks takes 6 h using an nvidia Tesla K80 GPU (graphical processing unit). To calculate retrosynthetic analyses for 1000 random drug-like molecules from the CHEMBL database, our model needs 25 s on a 2013 MacBook Pro. In contrast, the rule-based expert system takes 62 min and 24 s. The neural symbolic is thus 150 times faster. This can be understood because the rate-determining step for both approaches is the rule matching (or, more formally: subgraph isomorphism). The rule based system tries to match all 8720 rules against the input molecule. The neural symbolic system looks at the query molecule(s) once to calculate the fingerprint, then runs the neural network to determine the 20 most relevant rules, which are subsequently applied. The neural-symbolic model thus has to perform less work in the rate-determining step, which outweighs the additional overhead of the neural network.

To provide an illustrative example, Figure 4 shows the retrosynthetic analysis of a molecule from the recent medicinal chemistry literature. The retrosynthetic target was not included in our training data. More examples are deposited in the Supporting Information. Herein, we employed the highway network operating on 8720 rules. For each step, the top prediction was re-entered into the model. Our model predicts that target acid **4** can be made by oxidation of aldehyde **5**, which is suggested to be obtained by formylation of indole **6**. The model predicts that a Suzuki coupling should be employed to make the biaryl moiety of **6**. This was also the route that was chosen by Cameron et al.[14]

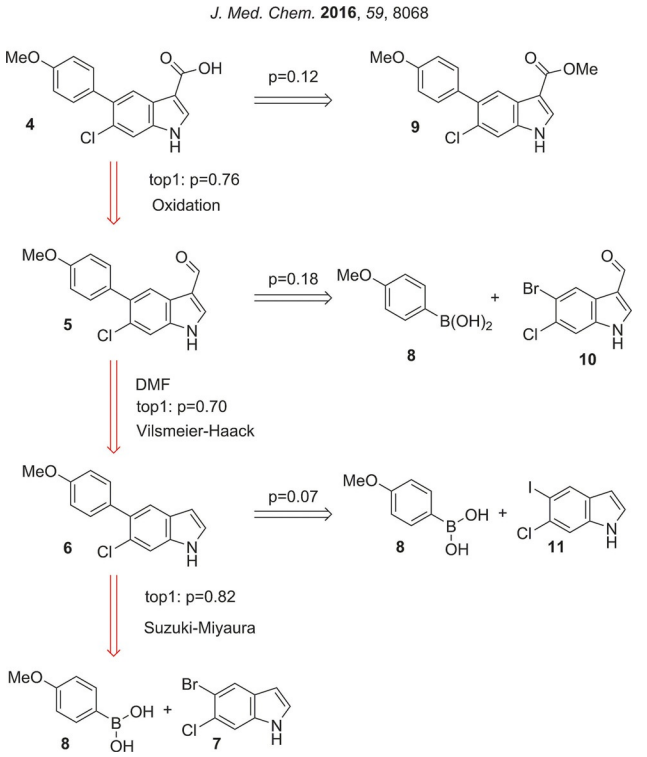

Figure 4. Retrosynthetic analysis of a drug molecule (**4**) based on 8720 rules. The top 1 prediction was recursively fed into the system. The red branch is the retrosynthetic tree with the highest scoring. It matches the reported route.

Limitations

There are a number of shortcomings of our present model that we would like to discuss. Most of the limitations stem from the underlying rules, and not the machine-learning component. The first limitation, which our system shares with other rule-based systems, is that it cannot predict anything outside its rule base. It does not solve the dilemma of rules: Either, one defines rules that are too general, which would

generate a lot of noise or rules that are too specific, which can only predict the substrate used to derive the rule. This is especially problematic for reaction types that only occur a few times. This limitation might be important when trying to plan complex natural product syntheses. However, we think it is more important for computer-aided synthesis systems to assist in the "everyday" syntheses of typical drug-like molecules, than covering more exotic corner cases. Furthermore, we have recently shown how this limitation of rules can be overcome with a model of chemical reasoning based on knowledge graphs.[4] The second limitation is that our system does not take stereochemistry into account. This issue has also not been convincingly addressed in traditional rule-based expert systems yet, and we are curious if it can be solved with a global model without involving quantum chemistry, for example using stereochemically aware descriptors.[21] We defer this difficult challenge to follow-up studies.

In summary, we report on a hybrid neural-symbolic approach for both retrosynthesis and reaction prediction that can be trained with large reaction sets from databases. The experiments show that neural networks can learn to which molecular context particular rules can be applied, and can prioritize the rules for both retrosynthesis and reaction prediction using either hand-coded or automatically extracted rule sets. This is transformational because it allows one to simply train an algorithm on a single computer overnight, instead of having a highly skilled team of chemists working for years. We anticipate that neural-symbolic models will be a key building block in future systems for computer-aided synthesis design, robot synthesis, virtual chemical space exploration, and de novo drug design.

Experimental Section

Data

As the dataset, we used all reactions with up to three reactants that lead to a single reported product from the Reaxys database, published from 1771 until 2015.[15] This represents essentially the complete published knowledge of organic chemistry in this period. The hand-coded and the extracted rules were assigned to every reaction. This left us with 3 million reactions for the hand-coded rules, and 4.9 million reactions for the extracted rules. The data were split randomly into a training set, a development set and a validation set (7:1:2). For model development, models were trained on the training set and tested on the development set. After selection of the two best models based on the development set, final performance is reported on the validation set.

Reaction rules

The reaction rules were obtained in two different ways. First, we entered 103 rules of common reactions by hand (see Supporting Information for a full list). In the second approach, we extracted very general rules algorithmically, following an established, shell-based Scheme, which is also used by state-of-the-art rule-based systems.[3c,d,16] The rules we extracted contained the reaction center, which are the atoms and bonds changed in the reaction. Atoms that are direct neighbors to the reaction center are also included. The algorithm currently ignores stereochemistry. Only rules that occurred at least 50, 100, and 5000 times were used to maintain robustness, leading to 17 370, 8720 and 137 rules, respectively. Rules were assigned with RDKit.[17]

Molecular descriptors

The molecules need to be encoded in a representation that the computer can process. For this purpose, we generate counted Extended-Connectivity Fingerprints (ECFP4)[18] with CDK 1.5.13.[19] These fingerprints are vectors $\boldsymbol{x}_i$ that essentially contain the count of the occurring functional groups, for every reactant molecule if the task is reaction prediction, or the product for retrosynthesis. As reactions can have more than one reactant, the encoding has to be order-invariant for the reaction prediction task. Then, we take the sum of these vectors to obtain a single vector $\boldsymbol{x}$, which serves as our order-invariant descriptor.

Neural networks

We define our problem as a multiclass classification. As our classifier, we evaluated different neural network architectures with one or more fully connected hidden layer(s), and Highway Networks.[13] As the non-linearity, we apply the exponential linear unit developed by Hochreiter and co-workers.[12] The last layer of the neural network is a *softmax*, which gives the probability distribution over the reaction rules.[11] Keras[20] was used as the machine-learning framework. The complete set of the studied hyperparameters and the architecture of the best models are listed in the Supporting Information.

Acknowledgements

We gratefully acknowledge Deutsche Forschungsgemeinschaft (SFB858) for funding. We thank D. Evans (RELX Intellectual Properties) and J. Swienty-Busch (Elsevier Information Systems) for the reaction dataset. M.S. thanks M. Jansen, D. Barton, T. Kogej, C. Tyrchan and A. Finkelmann for fruitful discussions.

Keywords: artificial intelligence · machine learning · retrosynthesis · synthesis design · total synthesis

[1] a) S. V. Ley, D. E. Fitzpatrick, R. Ingham, R. M. Myers, *Angew. Chem. Int. Ed.* **2015**, *54*, 3449 – 3464; *Angew. Chem.* **2015**, *127*, 3514 – 3530; b) S. V. Ley, D. E. Fitzpatrick, R. Ingham, R. M. Myers, *Angew. Chem.* **2015**, *127*, 3514 – 3530; c) F. Chevillard, P. Kolb, *J. Chem. Inf. Model.* **2015**, *55*, 1824 – 1835; d) M. Hartenfeller, M. Eberle, P. Meier, C. Nieto-Oberhuber, K.-H. Altmann, G. Schneider, E. Jacoby, S. Renner, *J. Chem. Inf. Model.* **2011**, *51*, 3093 – 3098; e) M. Hartenfeller, G. Schneider, *WIREs Comput. Mol. Sci.* **2011**, *1*, 742 – 759.

[2] S. Szymkuć, E. P. Gajewska, T. Klucznik, K. Molga, P. Dittwald, M. Startek, M. Bajczyk, B. A. Grzybowski, *Angew. Chem. Int. Ed.* **2016**, *55*, 5904 – 5937; *Angew. Chem.* **2016**, *128*, 6004 – 6040.

[3] a) A. Cook, A. P. Johnson, J. Law, M. Mirzazadeh, O. Ravitz, A. Simon, *WIREs Comput. Mol. Sci.* **2012**, *2*, 79 – 107; b) O. Ravitz, *Drug Discovery Today: Technologies* **2013**, *10*, e443 – e449; c) A. Bøgevig, H.-J. Federsel, F. Huerta, M. G. Hutchings, H. Kraut, T. Langer, P. Löw, C. Oppawsky, T. Rein, H. Saller, *Org. Process Res. Dev.* **2015**, *19*, 357 – 368; d) J. Law, Z. Zsoldos, A. Simon, D. Reid, Y. Liu, S. Y. Khew, A. P. Johnson, S. Major, R. A. Wade, H. Y. Ando, *J. Chem. Inf. Model.* **2009**, *49*, 593 – 602; e) W. A. Warr, *Mol. Inform.* **2014**, *33*, 469 – 476; f) M. H. Todd, *Chem. Soc. Rev.* **2005**, *34*, 247 – 266.

[4] M. Segler, M. P. Waller, *Chem. Eur. J.* **2016**, DOI: 10.1002/chem.201604556.

[5] a) M. A. Kayala, P. Baldi, *J. Chem. Inf. Model.* **2012**, *52*, 2526 – 2540; b) M. A. Kayala, C.-A. Azencott, J. H. Chen, P. Baldi, *J. Chem. Inf. Model.* **2011**, *51*, 2209 – 2222.
[6] a) J. H. Chen, P. Baldi, *J. Chem. Inf. Model.* **2009**, *49*, 2034 – 2043; b) J. Gasteiger, T. Engel, *Chemoinformatics: A Textbook*, Wiley-VCH, Weinheim, **2003**, DOI: 10.1002/3527601643; c) G. Marcou, J. Aires de Sousa, D. A. Latino, A. de Luca, D. Horvath, V. Rietsch, A. Varnek, *J. Chem. Inf. Model.* **2015**, *55*, 239 – 250.
[7] Q.-Y. Zhang, J. Aires-de-Sousa, *J. Chem. Inf. Model.* **2005**, *45*, 1775 – 1783.
[8] J. Zupan, J. Gasteiger, *Neural Networks in Chemistry and Drug Design*, Wiley, **1999**, ISBN: 978-3-527-29779-5.
[9] D. K. Duvenaud, D. Maclaurin, J. Iparraguirre, R. Bombarell, T. Hirzel, A. Aspuru-Guzik, R. P. Adams, in *Advances in Neural Information Processing Systems* (Eds.: C. Cortes, N. D. Lawrence, D. D. Lee, M. Sugiyama, R. Garnett), MIT Press, **2015**, pp. 2224 – 2232.
[10] J. N. Wei, D. Duvenaud, A. n. Aspuru-Guzik, *ACS Cent. Sci.* **2016**, *2*, 725 – 732.
[11] K. P. Murphy, *Machine Learning: A Probabilistic Perspective*, MIT Press **2012**, ISBN: 9780262018029.
[12] D.-A. Clevert, T. Unterthiner, S. Hochreiter, arXiv preprint arXiv:1511.07289, **2015**.
[13] R. K. Srivastava, K. Greff, J. Schmidhuber, in *Advances in Neural Information Processing Systems 28: Annual Conference on Neural Information Processing Systems*, Montreal, **2015**, pp. 2377 – 2385.
[14] K. O. Cameron, D. W. Kung, A. S. Kalgutkar, R. G. Kurumbail, R. Miller, C. T. Salatto, J. Ward, J. M. Withka, S. K. Bhattacharya, M. Boehm, K. A. Borzilleri, J. A. Brown, M. Calabrese, N. L. Caspers, E. Cokorinos, E. L. Conn, M. S. Dowling, D. J. Edmonds, H. Eng, D. P. Fernando, R. Frisbie, D. Hepworth, J. Landro, Y. Mao, F. Rajamohan, A. R. Reyes, C. R. Rose, T. Ryder, A. Shavnya, A. C. Smith, M. Tu, A. C. Wolford, J. Xiao, *J. Med. Chem.* **2016**, *59*, 8068 – 8081.
[15] http://www.reaxys.com, Reaxys is a registered trademark of RELX Intellectual Properties SA used under license.
[16] C. D. Christ, M. Zentgraf, J. M. Kriegl, *J. Chem. Inf. Model.* **2012**, *52*, 1745 – 1756.
[17] G. Landrum, http://www.rdkit.org, **2016** (accessed 18 November 2016).
[18] a) T. Unterthiner, A. Mayr, G. Klambauer, M. Steijaert, J. K. Wegner, H. Ceulemans, S. Hochreiter, in *Deep Learning and Representation Learning Workshop*, NIPS 2014, Montreal, Canada, Dec. 2014; b) D. Rogers, M. Hahn, *J. Chem. Inf. Model.* **2010**, *50*, 742 – 754; c) T. Unterthiner, A. Mayr, G. Klambauer, S. Hochreiter, arXiv preprint arXiv:1503.01445, **2015**.
[19] a) C. Steinbeck, Y. Han, S. Kuhn, O. Horlacher, E. Luttmann, E. Willighagen, *J. Chem. Inf. Comput. Sci.* **2003**, *43*, 493 – 500; b) C. Steinbeck, C. Hoppe, S. Kuhn, M. Floris, R. Guha, E. L. Willighagen, *Current Pharmaceutical Design* **2006**, *12*, 2111 – 2120.
[20] F. Chollet, Github, https://github.com/fchollet/keras, **2015**, (accessed 18 November 2016).
[21] P. Carbonell, L. Carlsson, J.-L. Faulon, *J. Chem. Inf. Model.* **2013**, *53*, 887 – 897.

Manuscript received: November 8, 2016
Accepted Article published: January 30, 2017
Final Article published: February 22, 2017