doi:10.2533/chimia.2019.990 Chimia 73 (2019) 990–996

Exploring Machine Learning Tools for the Prediction of the Stability of New Togni-type Reagents

Shungo Koichi*[a] and Hans P. Lüthi*[b]

Abstract: In the context of the prediction of the (in-)stability of chemical compounds using machine learning tools, we are often confronted with a basic issue: Whereas much information is available on stable (existing) compounds, little is known about compounds that might well exist, but that have not yet been successfully synthesized, or compounds that are inherently unstable (kinetically and thermodynamically). In the search for Togni-type reagents, many of them kinetically instable, the stability of the prospects can be assessed based on the transition state for the conversion to their non-hypervalent inactive isomer. In earlier work, we determined the barriers of conversion for over one-hundred reagents, still not enough information to train a tool such as a vector support machine. Here, instead, we focus on the early intermediate structures expressed along the isomerization pathway, *i.e.* transition state searches are replaced by finding (local) minima. Based on an array of 382 Togni-type reagents whose behaviour was known in advance, we show that it is possible to have the machine predict the intermediate form expressed. The approach introduced here can be used to make predictions on the stability and possibly also the reactivity of Togni-type reagents in general.

Keywords: Machine learning · Togni-type reagents

Hans P. Lüthi obtained his MSc and PhD degrees in chemistry from ETH Zürich and the University of Zurich. After post-doctoral studies at the IBM Research Center in San Jose, California, and a visiting professorship at Minnesota Supercomputer Institute, he returned to ETH Zurich in 1987, where he was involved in the build-up of the Swiss National Supercomputing Center. Later, he joined the ETH Department of Chemistry and Applied Biosciences as an adjunct professor (Privatdozent). He is currently the treasurer of the Swiss Chemical Society (SCS) and director of the SCS Foundation.

Shungo Koichi obtained his PhD in Information Science and Technology from the University of Tokyo. Then, he moved to Nagoya, Japan in 2008 to join Department of Information Systems and Mathematical Sciences, Nanzan University as an assistant professor. He is currently an associate professor at Nanzan University. His research interests include the application of information science to chemical data. He visited ETH Zürich and the University of Zurich during his sabbatical leave from Nanzan University in 2016/17.

1. Introduction

In organic synthesis, hypervalent iodine reagents ('iodanes'), such as Togni's reagent (see structure a in Fig. 1), have become very established for the transfer of electrophilic substituents to arenes or other nucleophiles.[1] In the case of Togni's reagent, the substituent transferred is a trifluoromethyl group (CF_3), but many other reagents of this type have been reported.[2] Based on the choice of electrophilic substituent and modification of the benzoiodoxole scaffold, many more reagents can be designed. Given the large number of candidate reagents, hundreds if not thousands, the prediction of their stability and reactivity becomes a potentially rewarding venture, also for computation.[3]

The Togni-type reagents all contain a five-membered heterocycle ring, which is part of the benziodoxole scaffold, and which carries the iodine atom and the second ligand (oxygen, in case of the original reagent; see structure a in Fig. 1). The electrophilic substituent E, the iodine atom I, and the second ligand L (also referred to as leaving-group) express the 3-center bond typical for iodanes (denoted E-I-L bond in this article). The 3-center bond was shown to be responsible for much of the reactivity of the iodanes.[4] For an extensive discussion of the structure and bonding in iodanes see ref. [5].

With the advances of machine learning, data-driven modelling has become an option to expedite the search of promising derivatives from a lead such as the Togni reagent. However, it often turns out that the data-situation is an obstacle: Whereas much information is available on existing compounds, there is very little information on compounds that are not stable, even if it is for a good reason, or that have not been synthesized successfully. This also applies to the Togni reagent and its derivatives.

In an attempt to improve the data situation, we computationally determined the stability of an array of well over one-hundred Togni-type reagents. For many of the reagents, the hypervalent iodine form is only kinetically stable,[6] *i.e.* protected by a barrier from conversion to a thermodynamically more stable, but inactive non-hypervalent isomeric form (structure b in Fig. 1). In this conversion, the electrophilic substituent E is transferred towards the ligand L locked into the 5-membered-ring heterocycle, to form a new E-L bond. The intramolecular transfer of E towards L oc-

*Correspondence: Prof. S. Koichi[a], E-mail: shungo@nanzan-u.ac.jp, Dr. H. P. Lüthi[b], E-mail: luethi@ethz.ch, [a]Department of Systems and Mathematical Science, Nanzan University, Nagoya, Japan, [b]Department of Chemistry and Applied Biosciences, ETH Zurich, Zurich, Switzerland.

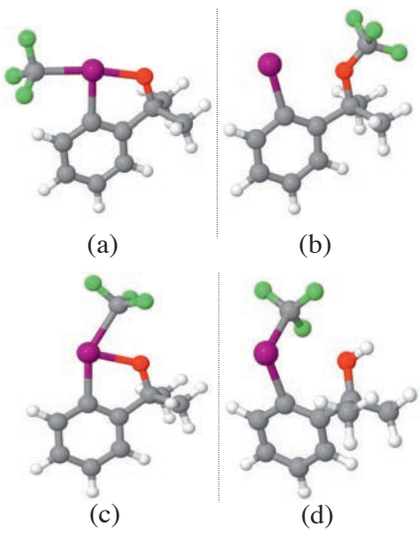

Fig. 1. (a) Hypervalent iodine iodane form of Togni's reagent expressing an E-I-L 3-center bond (equilibrium structure). (b) Open (acyclic) non-hypervalent isomer. This form is not active as a reagent for trifluoromethylation. (c) Alternative iodane structure expressing an E-I-C 3-center bond; an intermediate on the in-plane isomerization path. (d) Iodonium form of protonated Togni's reagent, an intermediate on the out-of-plane isomerization path. The reagent comes in two forms, one with a dimethyl-substituted carbon in the 5-membered-ring heterocycle as shown here (Togni I, 'alcohol reagent'), and one with a keto group instead (Togni II; 'acid reagent').

curs either in-plane, or, alternatively, above the molecular plane. In the first case, an alternative hypervalent structure is expressed, with a 3-center bond that involves the carbon atom of the phenyl ring (denoted E-I-C, or alternative hypervalent bond; structure c in Fig. 1). In the out-of-plane case, so far observed for protonated species only, we find an iodonium-like structure (structure d in Fig. 1), where the E-I-L angle is close to 90 degrees, the I-L bond becomes very long, and the iodine atom usually carries a significantly enhanced positive charge. Typically, the alternative E-I-C iodane and iodonium-like structures are minima on the potential energy surface and can be considered as intermediates on these two isomerization reaction pathways. However, there also were bi-and even tri-molecular transition states found for the isomerization reaction.[7]

As part of this study,[7] we explored the relative stability of the two isomers for 628 different reagents; for 121 of these, we were able to find the transition state geometry and to determine the barrier to isomerization. This allowed us to classify this subset of reagents according to their thermodynamic and kinetic stabilities. The array of compounds explored contained twenty different electrophilic substituents E, six different ligands L, plus modifications of the benziodoxole scaffold (see Fig. 3 in ref. [7]).

Still, the list of unstable compounds found, *i.e.* thermodynamically unstable compounds with a small barrier towards isomerization, was too short to train a support vector machine for the prediction of the stability of these reagents. Another, computationally less demanding approach to relate structure with stability – and possibly reactivity – needs to be found.

We therefore shifted our attention to the early stage of the isomerization reaction. The type of intermediate structure taken before reaching the transition state is indicative of the isomerization pathway taken. The availability of an iodonium-like intermediate, if low in energy, might indicate high mobility of the electrophilic substituent in the out-of-plane direction (and vice versa). The mobility of the electrophilic substituent may not only relate to the stability of the reagent, but also be decisive for its reactivity with an incoming nucleophile ('nucleophilic attack'). As a matter of fact, earlier *ab initio* molecular dynamics studies[8] showed that in trifluoromethylation reactions with Togni's reagent the transfer of the CF_3 group to the nucleophile can not be attributed to a single clearly defined reaction mechanism. Instead, the reactions might proceed through several concomitant mechanisms of similar probability. The study also showed that the out-of-plane mechanism *via* an iodonium-like intermediate thus may offer an alternative which is lower in energy than the in-plane path, and which may be available only to specific reagents. The enhanced mobility of the CF_3 group might be part of the explanation why Bronsted activation of Togni's reagent leads to higher reactivity towards nucleophiles.

Seeking to categorize an array of Togni-type reagents with respect to the intermediate structures expressed (alternative iodane- or iodonium-like) on the reaction pathway to isomerization may be a valid alternative to predict the stability of these compounds. In particular, we wish to correctly identify reagents expressing an iodonium-like intermediate as these may also show enhanced reactivity for the transfer of the electrophilic substituent to a nucleophile. This is a computationally much less demanding, but hopefully still useful approach. Knowing the correct answer in advance for the entire array of compounds studied will obviously lead to finding new, promising derivatives. This study explores the capabilities of different machine learning tools using a chemical model that gives easier access to information also on unstable or yet unobserved compounds.

2. Generating Data for the Description of the Reagent Stability and Reactivity

In order to profile these two distinct isomerization reaction paths for an entire array of compounds, we performed potential energy surface scans varying their E-I-L angles (θ and ϕ in Fig. 2) in the in-and out-of-plane direction between zero and 120°. The energies of these structures were computed by single point calculation (no geometry relaxation) using density functional theory (BP86 functional[9,10]) with aug-cc-pVDZ[11,12] and aug-cc-pVDZ-PP (iodine)[13] basis sets.

From Figs 3 and 4, we see that for the in-plane conversion the barrier for the neutral species (marked red) is generally lower, whereas for the protonated species (blue) the out-of-plane distortion results in lower-energy profiles. For Togni's reagent, this difference is quite dramatic (compare energy-profile marked in cyan in these two figures). The availability of a lower energy out-of-plane pathway is a strong indication that the protonated reagent may transfer its CF_3 group along this path (for isomerization or in the reaction with a nucleophile). For the transfer of the CF_3 group for the neutral form of Togni's reagent, no such conclusion can be drawn on the basis of these crude (unrelaxed) potential energy surface scans.

In order to investigate this further, we determined the energy of the intermediate structures along these two coordinates by means of full geometry relaxation for a subset of 382 of the 628 compounds scanned (Figs 3 and 4). Each of the 382 scaffolds can be

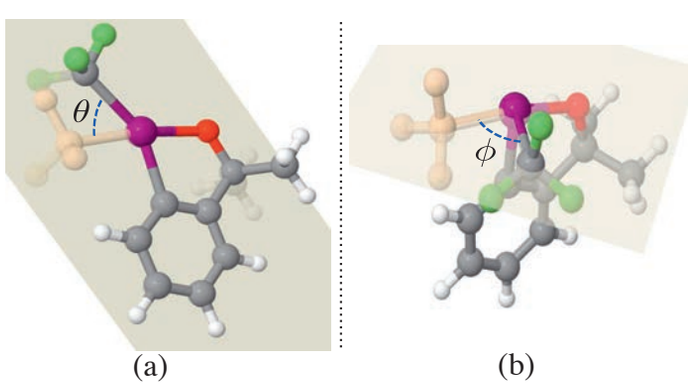

Fig. 2. Scanning the in-plane and out-of-plane isomerization reaction paths from the equilibrium structure towards the alternative iodane and the iodonium intermediate structures (c and d in Fig. 1).

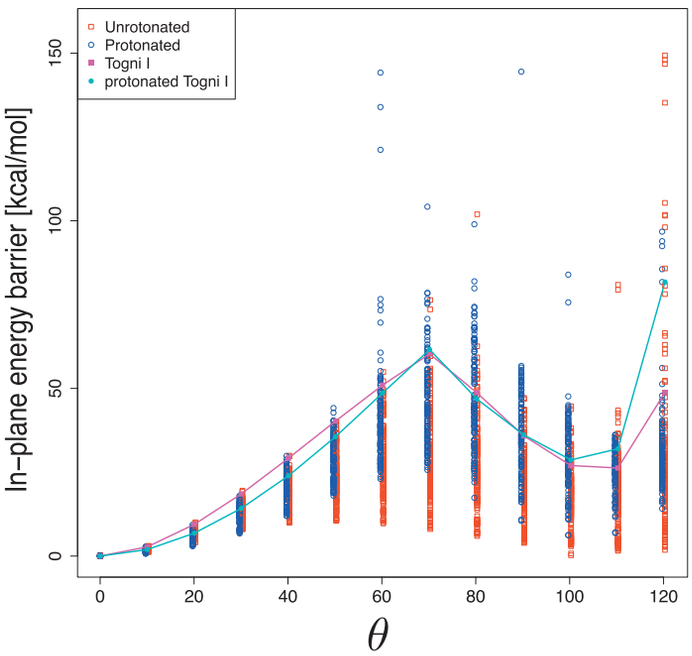

Fig. 3. The energy profiles for an array of 264 neutral as well as 252 protonated Togni-type reagents computed along the in-plane distortion path (angle θ as shown in Fig. 2). The energies were computed for angles θ for values of zero to 120° (scans in steps of 10° without relaxation of the distorted structures). The energies computed for Togni's reagent in its neutral and its protonated form are marked magenta and cyan, respectively. Their intermediate structures are located near 110° and 100°, respectively

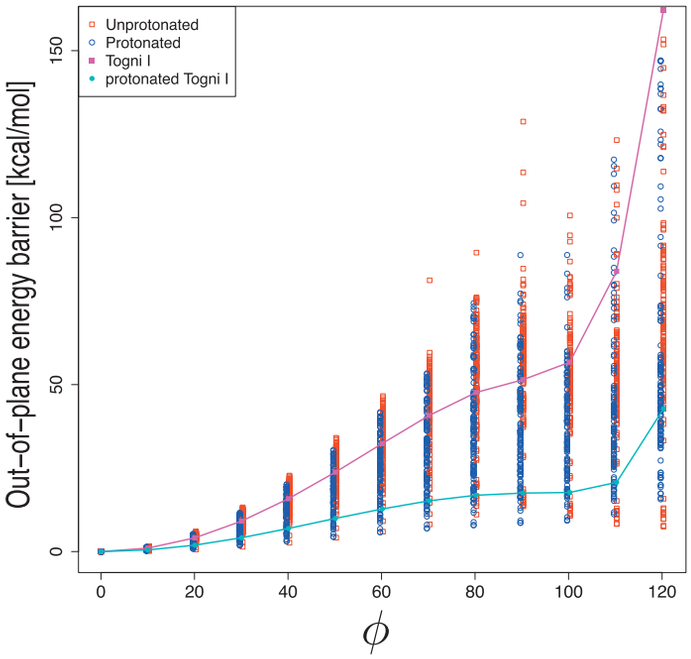

Fig. 4. The out-of-plane energy profiles computed for the same array of compounds in the same way as specified in Fig. 3. For the protonated form of Togni's reagent, a local minimum near an angle ϕ = 100°, corresponding to an iodonium-like structure, is visible.

labelled according to the intermediate structures they express, *i.e. E-I-C iodane* or *iodonium-like*. The computations were performed with Gaussian[14] and a special version of TURBOMOLE[15] generating output in eXtensible Markup Language (XML) for import into a database.[16]

The output includes energy and geometry information, charges and multipoles as well as data items from the natural bond orbital (NBO) analysis,[17] in particular hypervalent contributions and shared electron numbers ('overlap populations') in the 3-center bond.

We found that all neutral (218), plus several protonated reagents (56), express an alternative hypervalent iodane form with a E-I-C 3-center bond as shown in Fig. 1(c). The majority of the protonated reagents (111), however, adopt an iodonium-like form as shown in Fig. 1(d). We did not find any neutral reagent that would express an iodonium-like intermediate, *i.e.* this intermediate is expressed by protonated species only. Also, we did not find any compounds that express both intermediate structures. In general, the iodonium-type intermediates are lower in energy than their iodane-type counterparts (6 versus 3kcal/mol on average). In some cases, we observe that the iodonium-type intermediate is lower in energy than the iodane reagent, *i.e.* the reagent is thermodynamically unstable relative to the out-of-plane intermediate. We have no information about the transition state between the reagent and the intermediate, but we expect the barrier to be much smaller than the one of the isomerization to an non-hypervalent form or of the nucleophilic attack.

To make sure that the iodonium-form of the reagent can serve as a starting point for the reaction of the reagent with an incoming nucleophile (such as acetonitrile), we tried to find the corresponding transition state. The transition state found (see Fig. 5) has a very similar and even somewhat lower energy as the in-plane transition state for the same reaction. These two barriers are also much lower in energy than the one computed for the transfer of the CF_3 group

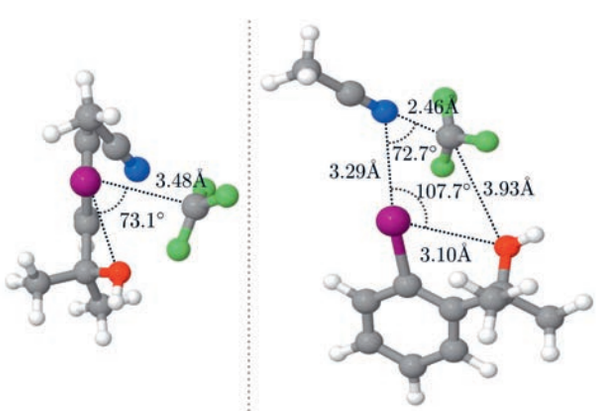

Fig. 5. Transition state geometry (two different views) for the out-of-plane CF_3 group transfer between Togni's reagent in its protonated form and an acetonitrile molecule leading to the formation of an acetonitrilium ion.

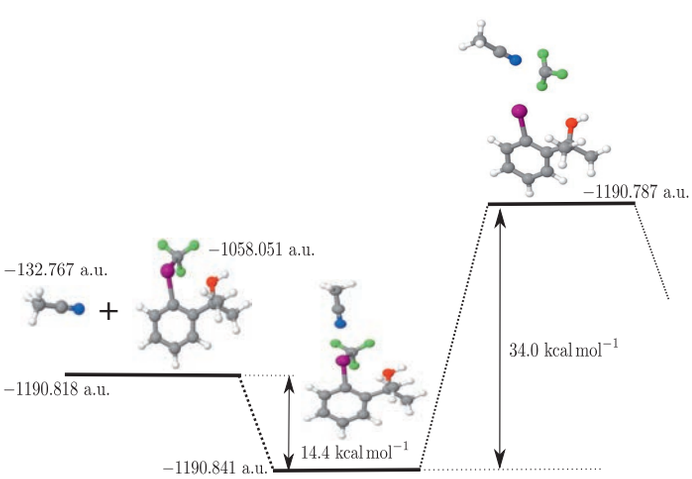

Fig. 6. Energy profile for the out-of-plane transfer of the CF_3 group to acetonitrile using Togni's reagent (protonated form; for the neutral reagent this path is not accessible). The step from the non-interacting reactants in their equilibrium form to the iodoinum-like intermediate and the acetonitrile, requiring a heat of formation of about 6 kcal/mol, is not shown here.

using the neutral reagent. The energy profile for the reaction starting from the iodonium-type form of the reagent is shown in Fig. 6.

The next step will be to find what properties determine the preference for one of the two intermediate structures from the data collected on the respective local minima, and to use these for the categorization of the reagents in view of their preferred intermediate structure.

3. Finding Descriptors and Reagent Categorization

We first applied the random forest method provided by the identically named package[18] in R[19] to find descriptors for the preferences of the 382 compounds in view of their intermediate structure taken. The random forest method, which is a classification method, constructs a number of decision trees based on random sampling of the training data combined with random selection of descriptors. A decision tree is an expression of judgment based on *if-else* decisions as shown in Fig. 7 and 8. The classification obtained by the random forest method is essentially a majority vote based on the decision trees it constructed. By preparing a many-fold of decision trees at random, the trees obtained by applying this approach are known to be robust against irregular data.

Our input contains about fifty numeric data items ('properties') per compound, each of which has its own unit. Hence, it is not trivial to interpret, for example, a sum of those numeric data; however, judgements in a decision tree are individual to each feature. Given the diversity of potential descriptors provided through the input data, we decided to apply the random forest method to find the most important features and to then use these same descriptors with other tools for comparison of the respective categorizations.

Taking the data from our array of 382 compounds as input, using the caret package[20] in R, a program that allows to setup an environment for various machine learning models, we evaluated the results of ten independent classification runs, each based on four-fold cross-validation, *i.e.* based on randomly partitioning the original sample into four equally sized sub-samples, one of which is retained as test set in each of the four validation rounds. The samples are split three to one (training versus test set). The (average) accuracy attained by this procedure was about 96%.

The two most important descriptors found by the random forest method through this procedure are the hypervalent contribution of the iodine atom and the shared electron number between the iodine atom and (the representative atom of) the leaving-group L, which were both obtained from the natural bond orbital analysis. In fact, the two features appear to be closely related, and therefore, to avoid overfitting, we decided to use, next to the shared electron number between I and L, the overlap populations between I and E, as well as the shared charge over the three centers I, E, and L (denoted s(I-L), s(I-E) and s(E-I-L), respectively).

Although we used slightly different and therefore potentially suboptimal features, the random forest method still shows an accuracy of more than 92% with four-fold cross-validation, and, as shown in Table 3, 100% accuracy in the resubstitution evaluation, *i.e.* when resubmitting the training set as a test set.

Using these three features, we also applied a simple decision tree method as provided by the rpart package[21] in R, and then obtained the results shown in Fig. 7 and 8, which depend on the user-defined parameter determining their complexity, *i.e.* the height of the output tree. The result shows that none of the output decision trees took s(E-I-L) as a descriptor, which implies that compared to s(I-L) and s(I-E) this particular descriptor is not important for the categorization. Even a simple decision tree as

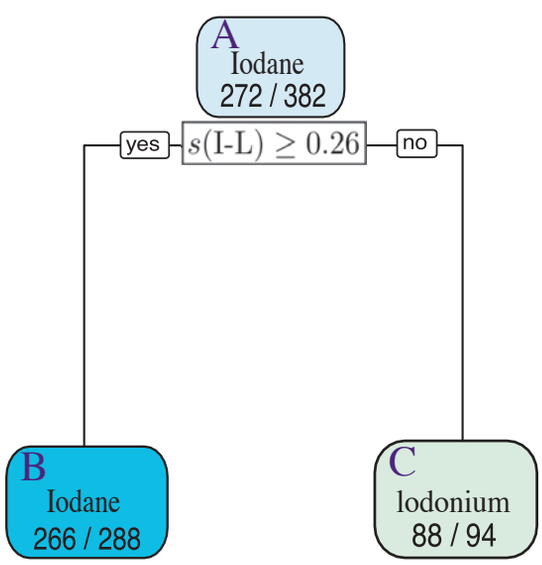

Fig. 7. Evaluation of a simple decision tree for the classification of the array of reagents in view of their predicted intermediate structure (E-I-C iodane versus iodonium-like form). The top node contains 382 compounds, 272 of which are known to take an iodane structure, whereas the remaining 110 compounds will prefer the iodonium form. The test of the array against the s(I-L) population results in a categorization saying that 288 reagents will prefer an iodane, whereas 94 reagents will prefer an iodonim intermediate structure. Comparison to the known correct result, *i.e.* 266 (out of 288) iodane and 88 (out of 94) iodoinum-like structures, shows that the overall correctness of prediction of this simple decision tree is 92.7%.

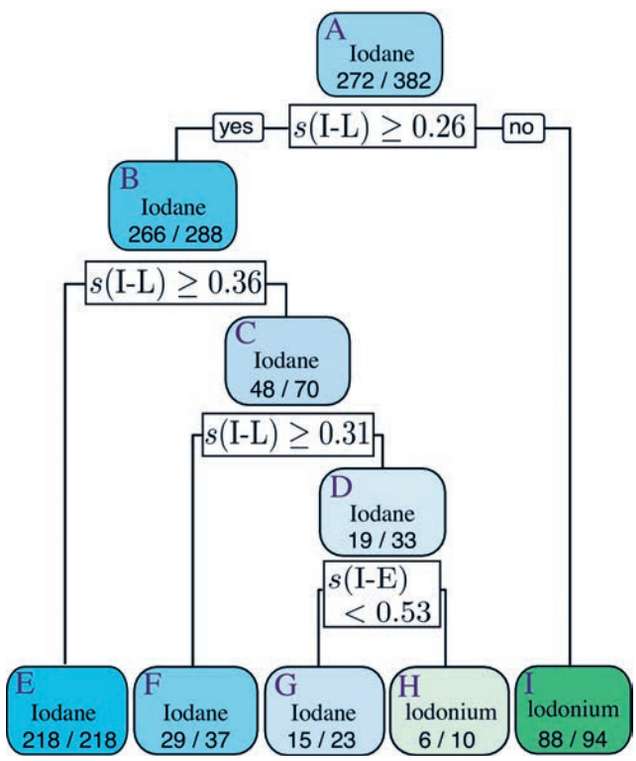

Fig. 8. Using all descriptors, *i.e.* s(I-E), s(I-L) and s(E-I-L), respectively, leads to an improved prediction of the iodonium intermediate structures. After the first test, which is the same as in the simple decision tree, the population in node B is tested against a higher threshold for s(I-L), which leads to a perfect categorization within this subset of the array: all of the 218 compounds that passed the two tests will indeed prefer an iodane intermediate structure. Most of the reagents preferring an iodonium-like intermediate structure are found in nodes H and I. In node H there are 10 compounds for which the descriptors take values of $0.26 \leq s(\text{I-L}) < 0.31$ and $s(\text{I-E}) \geq 0.53$. But only 6 of these compounds will actually take an iodonium-like form as their intermediate structure. Together with the iodane-expressing structures in node I, a total 94 assignments are correct, a slight improvement over the small decision tree (85.5 as opposed to 80.0% of iodonium-expressing reagents predicted correctly) Overall, the complex decision tree shows a correctness of 93.2% (as opposed to 92.7% for the simple decision tree.

shown in Fig. 7 achieves an accuracy of (266 + 88)/382 = 92.7% (resubstitution evaluation), which means that the shared electron number between an iodine atom and the leaving-group mainly determines the preference; namely, if a hypervalent iodine compound has *s*(I-L) ≥ 0.26, it takes (with high probability) an E-I-C iodane form as its intermediate structure; otherwise it is likely to take an iodonium-like form.

The multilayer decision tree shown in Fig. 8 also makes use of the *s*(I-E) feature, leading to a slightly improved prediction of iodonium structures: Now 94 out of 110 structures are correctly predicted (See confusion matrix in Table 2). Still there are 16 cases that are falsely predicted as iodane-structure preferring (boxes F and G in Fig. 8 or Table 2). On the iodane side, we now have 262 correct and 10 false assignments. From a chemical perspective, finding iodonium-structure-expressing reagents may be more rewarding, thus making the multilayer decision tree more attractive to use. Both Tables 1 and 2 show the confusion matrices for the resubstitution evaluation.

Fig. 9 shows a scatter plot of our data whose x-and y-axes are the shared electron number between I and E and between I and L, respectively. One can see that these two features show a weak correlation. In this case, the decision tree method might be inappropriate for the classification, because every border given by the decision tree method is perpendicular to some axis. Therefore, we simply used the linear support vector machine provided by the kernlab package[22] in R with the two features shown in Fig. 9. For the training, the C-parameter set to one, *i.e.* no bias on the penalties for misclassifying training data. (Here we also do not need to take care of their units.) This linear support vector machine, trained using a set of 191 compounds (130 iodane- and 61 iodonium-preferring cases), yielded the green categorization-borderline displayed in Fig. 9. The best predictability (93%) using an SVM was achieved with four-fold cross-validation using the complete training set of 382 compounds. Hence the performance

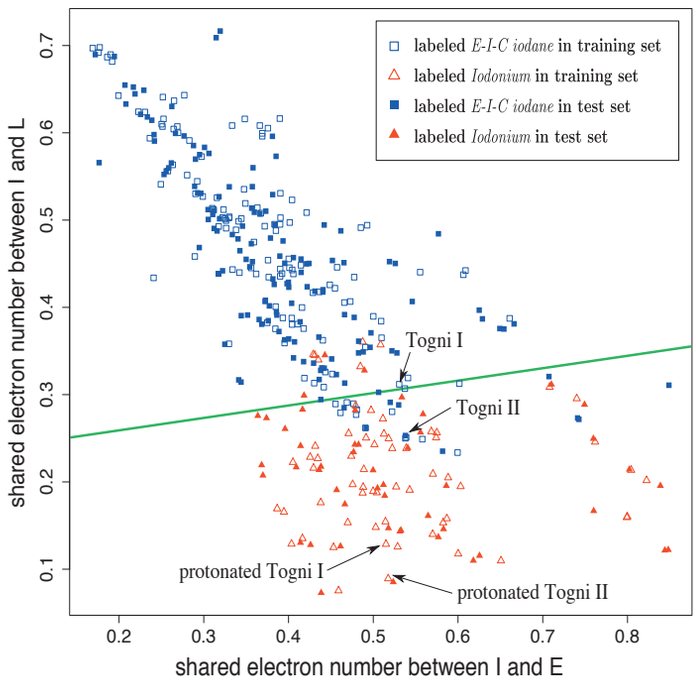

Fig. 9. The borderline (marked green) between reagents predicted to prefer either an iodane- or an iodonium-like structure as intermediates. The borderline is obtained from a linear support vector machine. Open symbols (squares and triangles) represent data items used for training; filled symbols represent members of the test set. The blue and red symbols point at compounds that are known to take iodane and iodonium-like forms, respectively. The axes denote the two descriptors *s*(I-E) and *s*(I-L) in units of number of electrons. On the right-hand-side, below the borderline, we see some of the few examples of failed categorization: The three blue squares represent cases we find in node H of Fig. 8, *i.e.* a node with only 60% accuracy.

Table 1. The confusion matrix for the prediction by the simple decision tree in Fig. 7

		Predicted class	
		iodonium	E-I-C iodane
Actual class	iodonium	88	22
	E-I-C iodane	6	266

Table 2. The confusion matrix for the prediction by the multilayer decision tree in Fig. 8

		Predicted class	
		iodonium	E-I-C iodane
Actual class	iodonium	94	16
	E-I-C iodane	10	262

Table 3. The confusion matrix for the prediction by the random forest method upon resubmission (*i.e.* using the training set as a test set)

		Predicted class	
		iodonium	E-I-C iodane
Actual class	iodonium	110	0
	E-I-C iodane	0	272

Table 4. The confusion matrix for the prediction by the support vector machine

		Predicted class	
		iodonium	E-I-C iodane
Actual class	iodonium	96	14
	E-I-C iodane	14	258

of the support vector machine method is almost the same as that of the random forest and decision tree methods. The confusion matrices are shown in Table 4.

It also shows that all the species preferring an iodonium-type intermediate are protonated compounds. The neutral compounds falling underneath the green line in Fig. 9 are cases of failed categorization. It is interesting to note that the non-protonated Togni reagents I and II, both reactive also in their neutral form, are positioned near the border, whereas the protonated species clearly fall into the iodonium-like intermediate category.

In fact, although unprotonated Togni reagents I and II take an E-I-C iodane form as intermediate structures, closer inspection of the isomerization path reveals that the CF_3 group goes outside the molecular plane to then fall back to the molecular plane to form the intermediate structure. Apparently, the CF_3 group enjoys substantial mobility also in the out-of-plane direction.

Exploring the area of strong preference of iodonium-like intermediates in the output of the SVM (bottom and bottom-right in Fig. 10), we find, next to the two (protonated) Togni compounds, species with the same scaffold, but carrying cyano (CN), alkynyl (CCH) and azide (N_3) E-groups. The substituent to the phenyl group (only NO_2 was present in the array of compounds) appears to play a very minor role only. For all of these compounds, we observe that the energy of iodonium intermediate is distinctly lower than the energy of the alternative iodane intermediate. This is true also for protonated compounds expressing an alternative iodane (rather than an iodonium) intermediate. In some cases, mostly observed in the presence of NH and PH as second ligand, the iodonium intermediate is lower in energy than the reagent in its iodane equilibrium structure.

The availability of a low-energy intermediate may be beneficial for the reactivity of the compound. Unfortunately, we were not able to find a strong correlation between *s*(I-E), *s*(I-L) (or any other feature) and the energy of the intermediate for the protonated compounds. Still, the most reactive species are expected to be those Bronsted-activated compounds that express an iodonium-type intermediate.

Another descriptor for the same categorization can be derived from natural localized molecular orbital (NLMO[23]) analysis of the 3-center bond. Based on the analysis of the contribution of the iodine atom to the bond for 186 reagents, we found that the percentage of *s*-orbital character in the NLMO composing the bond between I and E has to be greater than 2.7% for the compound to express an iodonium-like form. Greater *s*-orbital contribution means less *p*-orbital involvement in this bond, which allows the electrophilic group to rotate around the iodine atom more easily, thus enjoying greater mobility.

Knowing that electronic structure features render good descriptors for the structure and reactivity of iodanes invites to consider quantities derived from the Domain Averaged Fermi-Hole analysis used to show the presence (and non-presence) of 3-center bonds in these compounds.[24] However, the analysis is computationally demanding and difficult to automate as manual interventions are frequently required.

4. Outlook and Discussion

Using a machine learning approach (random forest method), we found that the shared electron number between I and L (*s*(I-L)) is the key feature that determines whether in the isomerization reaction the reagent will take an E-I-C iodane or an iodonium-like form as its intermediate structure (boxes B and C in Fig. 7). The smaller the number of electrons shared between I and L, the more easily the bond is broken and eventually an iodonium-like form will be expressed: The decision tree method indicates that most of these reagents show an *s*(I-L) of 0.26 electrons or less; there is no single reagent expressing an iodonium-like intermediate with

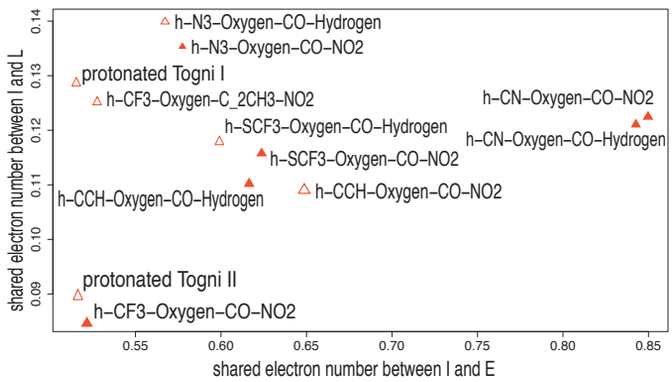

Fig. 10. A close-up of the area containing those reagents, including the two forms of the protonated Togni reagent, the SVM predicts to be most reactive. The tags h-E-L-X-Y used to label the compounds stand for protonated (h; all compounds), electrophilic substituent (E; CF_3, SCF , N , CN, and CCH), leaving group or second ligand (L; all oxygen), group adjacent to L (X; CO, $C(CH_3)_2$) in the five-membered heterocycle, and substituent to the phenyl group (Y; H, NO_2).

s(I-L) >0.36 electrons (box E in Fig. 8). Still, the prediction of iodonium-preferring structures using these two features is slightly more difficult (less than 90% accurate).

For the categorization of reagents showing a shared electron number between 0.26 and 0.36, *s*(I-E), the number of electrons shared between the iodine atom and the electrophilic substituent serves as an additional, secondary feature used in a more complex decision tree method: A high value of *s*(I-E) (greater than 0.53 electrons) increases the likelihood that the compound will express an iodonium-like intermediate.

With these descriptors and the two different decision tree methods we were able to categorize our array of 382 compounds with an overall accuracy of more than 93%. For the prediction of iodane and iodonium intermediate structures we observe a prediction correctness of 96.3 and 85.5%, respectively.

With an SVM trained with a set of 191 compounds using these two descriptors, we achieved a categorization (see Fig. 9) of very similar quality as the one obtained based on the decision tree methods. From the output of the SVM, we can backtrack and find those (protonated) compounds predicted to express an iodonium intermediate. These are expected to be among the most reactive, most likely more reactive (and less stable) than those protonated compounds that express and alternative iodane structure.

The relationship between the shared electron number and the bond strength in hypervalent compounds is not a new discovery, but has been reported already in the 1980s.[25] Here, the novel aspect is that it was the machine – not humans – that selected the shared electron number as descriptor out of a manifold of molecular properties.

We will apply these methods towards the categorization of other Togni-like reagents, *i.e.* reagents where we have no advance knowledge of their reactivity and preferred intermediate structure. Some of these may be very efficient reagents, certainly when Bronsted activated, and may deserve to be explored experimentally. In this same context, we will need to further validate the correlation between the trend of reagents expressing iodonium-like intermediate structure and their enhanced reactivity with nucleophiles.

The model introduced here allowed to bypass the difficulties encountered with the missing data on unstable compounds needed for training. A similar lack of negative information is observed with the data sets for chemical reactions (see article of Nair *et al.*[26] in this issue of CHIMIA). To close this gap, the community will need to establish a mechanism for the publication and collection of negative information.

Acknowledgements

Part of this work was performed when S. K. was on a sabbatical leave from Nanzan University, visiting ETH Zurich and the University of Zurich.

Received: October 31, 2019

[1] J. Charpentier, N. Früh, A. Togni, *Chem. Rev.* **2015**, *115*, 650.
[2] V.V. Zhdankin, 'Hypervalent Chemistry: Preparation, Structure and Synthetic Applications of Polyvalent Iodine Compounds', Wiley, Chichester, 1st edn, **2014**.
[3] H. Jiang, T.-Y. Sun, Y. Chen, X. Zhang, Y.-D. Wu, Y. Xie, H. F. Schaefer, *Chem. Commun.* **2019**, *55*, 5667.
[4] H. Pintode Magalhães, H. P. Lüthi, A. Togni, *J. Org. Chem.* **2014**, *79*, 8374.
[5] B. Olofson, Z. Rappoport, I. Marek, 'Chemistry of Functional Groups (Patai Series): The Chemistry of Hypervalent Halogen Compounds', Wiley, Chichester, 134th edn, **2018**.
[6] X. Wang, H. Geng, Y. Xie, Y.-D. Wu, X. Zhang, H. F. Schaefer, *Chem. Commun.* **2016**, *52*, 5371.
[7] S. Koichi, B. Leuthold, H. P. Lüthi, *CPPC* **2017**, *19*, 32179.
[8] O. Sala, H. P. Lüthi, A. Togni, M. Iannuzzi, J. Hutter, *J. Comp. Chem.* **2015**, *36*, 785.
[9] A. D. Becke, *Phys. Rev. A* **1988**, *38*, 3098.
[10] J. P. Perdew, *Phys. Rev. B* **1986**, *33*, 8822.
[11] T. H. Dunning, Jr., *J. Chem. Phys.* **1989**, *90*, 1007.
[12] D. Woon, T. H. Dunning, Jr., *J. Chem. Phys.* **1993**, *98*, 1358.
[13] K. Peterson, D. Figgen, E. Goll, H. Stoll, M. Dolg, *J. Chem. Phys.* **2003**, *119*, 11113.
[14] M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G. A. Petersson, H. Nakatsuji, M. Caricato, X. Li, H. P. Hratchian, A. F. Izmaylov, J. Bloino, G. Zheng, J. L. Sonnenberg, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, J. A. Montgomery, Jr., J. E. Peralta, F. Ogliaro, M. Bearpark, J. J. Heyd, E. Brothers, K. N. Kudin, V. N. Staroverov, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, N. Rega, J. M. Millam, M. Klene, J. E. Knox, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, R. L. Martin, K. Morokuma, V. G. Zakrzewski, G. A. Voth, P. Salvador, J. J. Dannenberg, S. Dapprich, A. D. Daniels, O. Farkas, J. B. Foresman, J. V. Ortiz, J. Cioslowski, D. J. Fox, Gaussian09, Revision D.03, **2009**, Gaussian Inc. Wallingford CT.
[15] A. Gls, M. P. Brändle, W. M. Klopper, H. P. Lüthi, *Mol. Phys.* **2012**, *110*, 2523.
[16] H. P. Lüthi, S. Heinen, G. Schneider, A. Gls, M. Brändle, R. King, E. Pyzer-Knapp, F. Alharbi, S. Kais, *J. Comp. Science* **2016**, *15*, 65.
[17] E. D. Glendening, C. R. Landis, F. Weinhold, *J. Comp. Chem.* **2013**, *34*, 1429.
[18] A. Liaw, M. Wiener, *RNews* **2002**, *2*, 18.
[19] R Core Team, 'R:A Language and Environment for Statistical Computing', R Foundation for Statistical Computing, Vienna, Austria, **2017**.
[20] M. K. C. from Jed Wing, S. Weston, A. Williams, C. Keefer, A. Engelhardt, T. Cooper, Z. Mayer, B. Kenkel, the RCoreTeam, M. Benesty, R. Lescarbeau, A. Ziem, L. Scrucca, Y. Tang, C. Candan, T. Hunt, 'caret: Classification and Regression Training', **2017**.
[21] T. Therneau, B. Atkinson, B. Ripley, 'rpart: Recursive Partitioning and Regression Trees', **2017**.
[22] A. Karatzoglou, A. Smola, K. Hornik, A. Zeileis, *J. Stat. Software* **2004**, *11*, 1.
[23] A. E. Reed, F. Weinhold, *J. Chem. Phys.* **1985**, *83*, 1736.
[24] H. Pinto de Magalhães, H. P. Lüthi, P. Bultinck, *PhysChemChemPhys* **2016**, *18*, 846.
[25] C. Ehrhardt, R. Ahlrichs, *Theor. Chim. Acta* **1985**, *68*, 231.
[26] V. H Nair, P. Schwaller, T. Laino, *Chimia* **2019**, *73*, 997.