ACS central science

http://pubs.acs.org/journal/acscii

Research Article

Toward in Silico Modeling of Dynamic Combinatorial Libraries

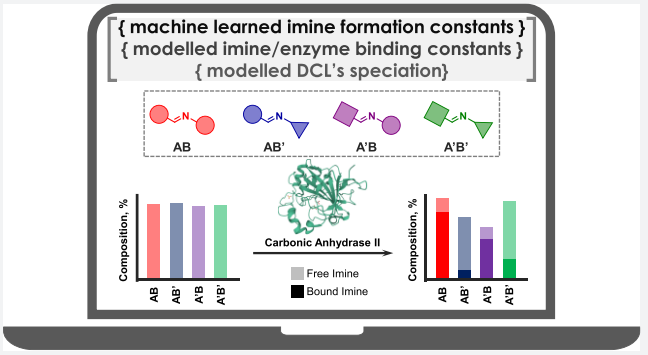
Iuri Casciuc, Artem Osypenko, Bohdan Kozibroda, Dragos Horvath, Gilles Marcou, Fanny Bonachera, Alexandre Varnek,* and Jean-Marie Lehn*

Cite This: *ACS Cent. Sci.* 2022, 8, 804–813

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: Dynamic combinatorial libraries (DCLs) display adaptive behavior, enabled by the reversible generation of their molecular constituents from building blocks, in response to external effectors, e.g., protein receptors. So far, chemoinformatics has not yet been used for the design of DCLs—which comprise a radically different set of challenges compared to classical library design. Here, we propose a chemoinformatic model for theoretically assessing the composition of DCLs in the presence and the absence of an effector. An imine-based DCL in interaction with the effector human carbonic anhydrase II (CA II) served as a case study. Support vector regression models for the imine formation constants and imine-CA II binding were derived from, respectively, a set of 276 imines synthesized and experimentally studied in this work and 4350 inhibitors of CA II from ChEMBL. These models predict constants for all DCL constituents, to feed software assessing equilibrium concentrations. They are publicly available on the dedicated website. Models rationally selected two amines and two aldehydes predicted to yield stable imines with high affinity for CA II and provided a virtual illustration on how effector affinity regulates DCL members.



INTRODUCTION

Dynamic combinatorial chemistry (DCC) implements the generation of sets of dynamic molecular (or supramolecular) entities by the recombination of building blocks linked by covalent (or non-covalent) bonds formed in a variety of reversible chemical reactions.[1−8] The central feature of such dynamic combinatorial libraries (DCLs) is their operation under thermodynamic control, in comparison with "classical" combinatorial libraries, which may be considered as static in view of the high kinetic stability of the covalent bonds that build up their members. The members of the DCL (called *constituents*) are in equilibrium with one another through constant exchange of building blocks (called *components*) via reversible covalent (or noncovalent) reactions. As a consequence, such a DCL can *adapt* to the action of physical stimuli or chemical entities (called *effectors*), resulting in *amplification of the fittest* constituent(s),[7,9−16] for that specific physical or chemical agent, through *selection* and *exchange of components*. The agent can be of variable nature—a physical stimulus, like a change of temperature,[17] or a chemical effector like a metal ion,[18] a protein/enzyme,[19,20] or properties of the medium (solvent, pH, viscosity).[21,22]

Along with their numerous applications,[1−8] DCLs are of particular interest for drug discovery,[23−30] where they have been used to identify binders/inhibitors to proteins/enzymes,[23−25] nucleic acids,[26−29] and even living cells.[30] Addition of a biological target (e.g., an enzyme) to a DCL of potential inhibitors has been shown to drive the *selection* of the most potent binder/inhibitor in the DCL, causing its *amplification* with respect to the distribution in the absence of the target protein.[19,20,23−25] Hence, the DCLs may be implemented for lead generation in drug discovery. Enabling the protein to actively enhance the formation of its preferred ligand(s), from the *pool of virtual binders*, in a sort of "*The Lock generates its Key*" process, provides an approach that can be beneficial over the high-throughput screening (HTS)[31] of individual compounds of classical "*static*" combinatorial libraries[23−30] obtained by mixing sets of reagents of the same category—typically, *n* nucleophilic species N_1, N_2, ...N_n and *m* electrophilic reagents E_1, E_2, ...E_m. The key benefit expected from DCLs is maintenance of the simple "mixture" strategy but for a set of equilibrating constituents while improving the chances that strong affinity products will emerge—because they are dynamically selected and amplified. The final DCL consists of the equilibrium population of the constituents representing all of the possible combinations generated by the reversible connection of the components. The addition of an effector will modify the

Received: January 18, 2022
Published: May 27, 2022

distribution of constituents depending on their affinity for the target entity, amounting to *an adaptation of the DCL to the effector*.[7,32−34] Note that a procedure of dynamic deconvolution may be applied to complex DCLs.[23,35,36]

Chemoinformatics is a key player in HTS library design. On one hand, it may help to "focus" on compounds most likely to bind the screened target (thereby eliminating testing of species predicted to be inactive).[37] On the other hand, it is also widely used to design generic "diverse" libraries,[38] to be used in HTS against targets with no ligand structure−affinity information on which to base a focusing strategy. Diversifying a library means maximizing chemical space coverage and ensuring that included compounds are not redundant. So far, however, chemoinformatics has, to our knowledge, not yet been invoked for the design of DCLs, given that such design comprises a radically different set of challenges. Note that unsupervised machine learning (including PCA, LDA, and cluster analysis) has been used for statistical DCL data analysis.[39−41]

First, as in classical drug design, chemoinformatics may help to select appropriate building blocks that are highly likely to lead to products that fulfill the (steric, electronic, pharmacophoric) constraints required for activity. In this context, there is no need to precisely identify which of the possible products will be most active because the DCL strategy *per se* provides a powerful search mechanism for the latter. This is fortunate because typically chemoinformatics approaches are not accurate enough (except for, perhaps, costly free energy perturbation simulations)[42] to explain subtle activity differences between strongly related members of a combinatorial library. They are, however, well suited to quickly discard building block combinations that are almost certainly unlikely to lead to active products. In principle, a DCL should not be prebiased but based on sets of the highest molecular component diversity without any preconceived ideas. However, the DCL may be simplified by not including building blocks that are with a high probability either not expected to engender actives or predicted to be highly active when combined with at least some of the other partners. (Bio)activity prediction models—either based on machine-learned quantitative structure−activity relationships (QSAR) or on ligand-site interaction models (pharmacophores, docking)—are thus important for DCL design, as they are for classical library design.

Second, for the application of chemoinformatics to DCLs, the partner building blocks should be selected so as to present comparable reactivity and to form products of comparable thermodynamic stability. The effector may displace the equilibrium concentration in favor of its preferred binders unless other concurrent reactions lead to some extremely stable adducts. *Thermodynamic stability of DCL constituents and target/library constituents association* are properties that *can be machine learned* on the basis of experimentally studied cases.[43] From such data, quantitative structure−property relationships (QSPR) models can be used to predict, on one hand, product stability as a function of its structure, and, on the other hand, the affinity of DCL constituents for the effector. The stability problem is, however, complicated by the impact of the solvents, as that used for DCL experiments (usually water for biological targets) may differ from that for which the QSPR model has been calibrated. Extrapolating measured equilibrium constants to a solvent different from that in which the measurement was performed (chloroform to water, for example) can be estimated on the basis of partition coefficients (log P) between the two concerned immiscible solvents. Once all the equilibrium constants are presumed known, the equilibrium concentrations—and their effector-induced shifts—can be calculated by a speciation algorithm[44,45] so that the DCL behavior can be simulated in silico.

Chemical diversity considerations are particularly important. Selected building blocks should be chemically as diverse as permitted by the above constraints (matching activity requirements and ensuring a balanced distribution of relative product stability). If there are building blocks based on distinct chemotypes predicted to be compatible with the constraints above, then they should be selected—instead of limiting the DCL to a redundant collection of building block homologues.

This work tentatively explores all the three key points above, in order to (i) provide a concrete and technically detailed illustration for an in silico DCL design strategy, (ii) prepare required data—both from public databases and in-house experimental measures—and (iii) finally build the models in view of a future DCL design campaign, followed by an experimental assessment. Building on seminal work in this area,[19] human carbonic anhydrase II (CA II)[46] was chosen as an effector to model the adaptive behavior of the imine-based DCL.

RATIONALE AND WORKFLOW OF SPECIATION MODELING OF DCLs

The three steps of the modeling workflow are shown in Figure 1 and in Figure S1 (see Supporting Information). They comprised the following operations.

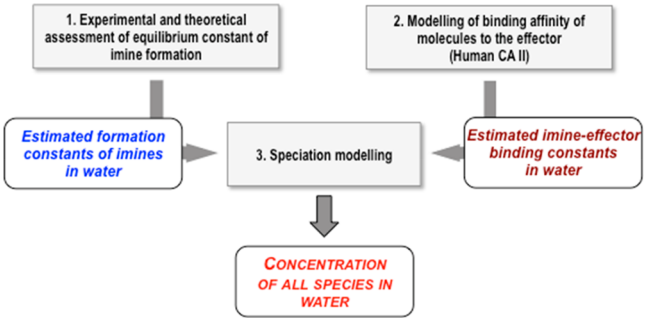


Figure 1. Main steps and outputs of speciation modeling workflow.

Part 1. Experimental and Theoretical Assessment of Equilibrium Constants for Imine Formation.

(1) Preparation of the training data set.
   (a) Preselection of the aldehydes and amines based on their "popularity" estimated by the number of references in a scientific database (primary data set: 400 aldehydes and 300 amines);
   (b) Selection of small diverse (nonredundant) pools of amines and aldehydes.
(2) Experimental determination of the formation constants of 276 imines in deuterated chloroform ($CDCl_3$) from the selected training data set.
(3) Building of a predictive machine-learning model for the logarithm of imine formation constant (log K_C) in chloroform as a function of the structure.
(4) Since DCL−effector protein interaction is occurring in water, imine stability in water needs to be assessed using the predicted stability in $CDCl_3$. This was achieved with the help of the predictive model for the chloroform−water partition coefficient (log $P_{C/w}$) prepared in this work.

https://doi.org/10.1021/acscentsci.2c00048
ACS Cent. Sci. 2022, 8, 804−813

Part 2. Preparation of the Model for the Affinity of Organic Molecules to the Effector. The model for the logarithm of the dissociation constant (pK_i) of organic molecules from human CA II was prepared using experimental data extracted from the ChEMBL database.[47−49]

Part 3. Speciation Modeling of DCL in the Presence of the Enzyme As Effector. The dynamic behavior of a simple DCL (2 amines × 2 aldehydes) was simulated by applying speciation software to compute equilibrium concentrations of free and protein-bound imines, given their estimated stability in water and affinities to the protein effector.

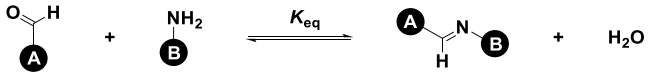
Pending experimental validation, this work focused on the key steps of the envisaged strategy, with evaluation of the strengths and potential pitfalls of the models.

RESULTS AND DISCUSSION

Part 1. Experimental and Theoretical Assessment of Equilibrium Constant of Imine Formation. In the field of dynamic combinatorial chemistry, imines,[50−55] (Scheme 1) formed by reversible amine-aldehyde condensation, represent a class of compounds of particular significance for several reasons:

(i) they display high diversity in terms of structures and physicochemical properties;
(ii) their building blocks, aldehydes and amines, are readily synthetically or commercially available;
(iii) in most cases, they offer a convenient range of exchange kinetics at room temperature in various media, including neat conditions, organic solvents (such as chloroform, toluene, or DMSO) and water.

Scheme 1. Generalized Reaction Scheme of Imine Formation from an Aldehyde and an Amine; and the Corresponding Expression for Its Equilibrium Constant (Equation 1)



$$K_{eq} = \frac{[\text{imine}][H_2O]}{[\text{amine}][\text{aldehyde}]} \quad (1)$$

Although imine formation and component exchange in aqueous medium are of special interest, the presence of several possible intermediates and various side reactions such as amine protonation as well as the formation of aldehyde-hydrate and hemiaminals are serious challenges for experimental investigation, as we have shown elsewhere.[55]

Thus, it was decided to use deuterated chloroform as a medium for the reaction. Chloroform is the most widely used NMR solvent, and imine formation in chloroform usually leads to negligible amounts of side-products (such as aldehyde hydrates, hemiaminals, and aminals).

Selection of the Experimental Data Set. First, amine and aldehyde building blocks were taken as the top 400 most cited aromatic aldehydes and top 300 primary amines according to SciFinder, using the following protocol: (a) the compounds were sorted by the frequency of their use; (b) only molecules with a molecular weight ≤ 400 Da were selected; (c) compounds with only one aldehyde/primary amine group were chosen; multifunctional compounds would produce much more complex dynamic sets and represent a further step of investigation; (d) preselected sets were manually checked, for compatibility reasons (duplicates, functional group incompatibility, aggregate state incompatibility, solubility, availability, etc.). Note that aliphatic aldehydes were excluded from the study because experimental tests revealed various side reactions, making the analysis challenging.[51,54]

This procedure resulted in a set of 120 000 possible imines (400 × 300) serving as a reference pool out of which a small combinatorial sublibrary of 360 imines was selected for experimental assessment of their thermodynamic stability. This core was defined as the combination of maximum diversity reagents: the MaxMin[56] algorithm was applied separately to the amine and aldehyde sets in ISIDA fragment descriptor space (see Table S1 in Supporting Information for details), picking subsets of 24 aromatic aldehydes and 15 primary amines, respectively (Figure 2).

These reagent subsets should in principle span as broad as possible reactivity ranges, in order to yield an informative pool of imines of significantly different stability, from which machine learning would easily identify structural features enhancing and respectively decreasing stability. Intuitively, it is therefore legitimate to ask whether this diversity selection should not been rather conducted in a quantum-chemical descriptor space, as the latter is perceived as most directly related to reactivity issues. However, there are several strong arguments in favor of the herein adopted strategy:

(1) ISIDA fragment descriptors are excellent descriptors of reactivity—as will be proven further on, when discussing their propensity to fit to experimentally measured equilibrium constants. This simply means that key quantum-chemical descriptors (HOMO or LUMO energies, for example) are effectively covariant with the presence of specific (electron-withdrawing or -donating) fragments captured by the ISIDA fingerprint.
(2) Quantum-chemical descriptors alone fail to account for sterical hindrance, which is better rendered by fragment counts—albeit in an implicit way. Also, they are geometry-dependent—HOMO/LUMO energy differences in response of a conformational change may be actually larger than differences between analogous molecules in comparable geometries.
(3) In view of that mentioned above, it is no surprise that machine-learning models of the imine stability based on 30 quantum-chemical descriptors issued from DFT calculations (see their list in Section 3.3 in Supporting Information) are not better than the much easier to use and much faster fragment descriptor counterparts (see Table S2 in Supporting Information). Moreover, ISIDA-descriptor-driven diversity is perfectly suited to select amines and aldehydes that are also "diverse" in terms of quantum-chemical terms, as it is shown on the example of HOMO/LUMO energies distribution in Table S3 in Supporting Information.
(4) Finally, yet importantly, time-consuming DFT calculations can hardly be recommended to calculate descriptors for large combinatorial libraries of 120 000 virtual imines. On the other hand, the generation of ISIDA descriptors is very fast, which makes them particularly attractive when working with big chemical data.

Out of the above-mentioned 360 pairwise combinations, 276 imines were synthesized, and the equilibrium constants for their formation were measured using ^{1}H NMR spectroscopy.

https://doi.org/10.1021/acscentsci.2c00048
ACS Cent. Sci. 2022, 8, 804−813

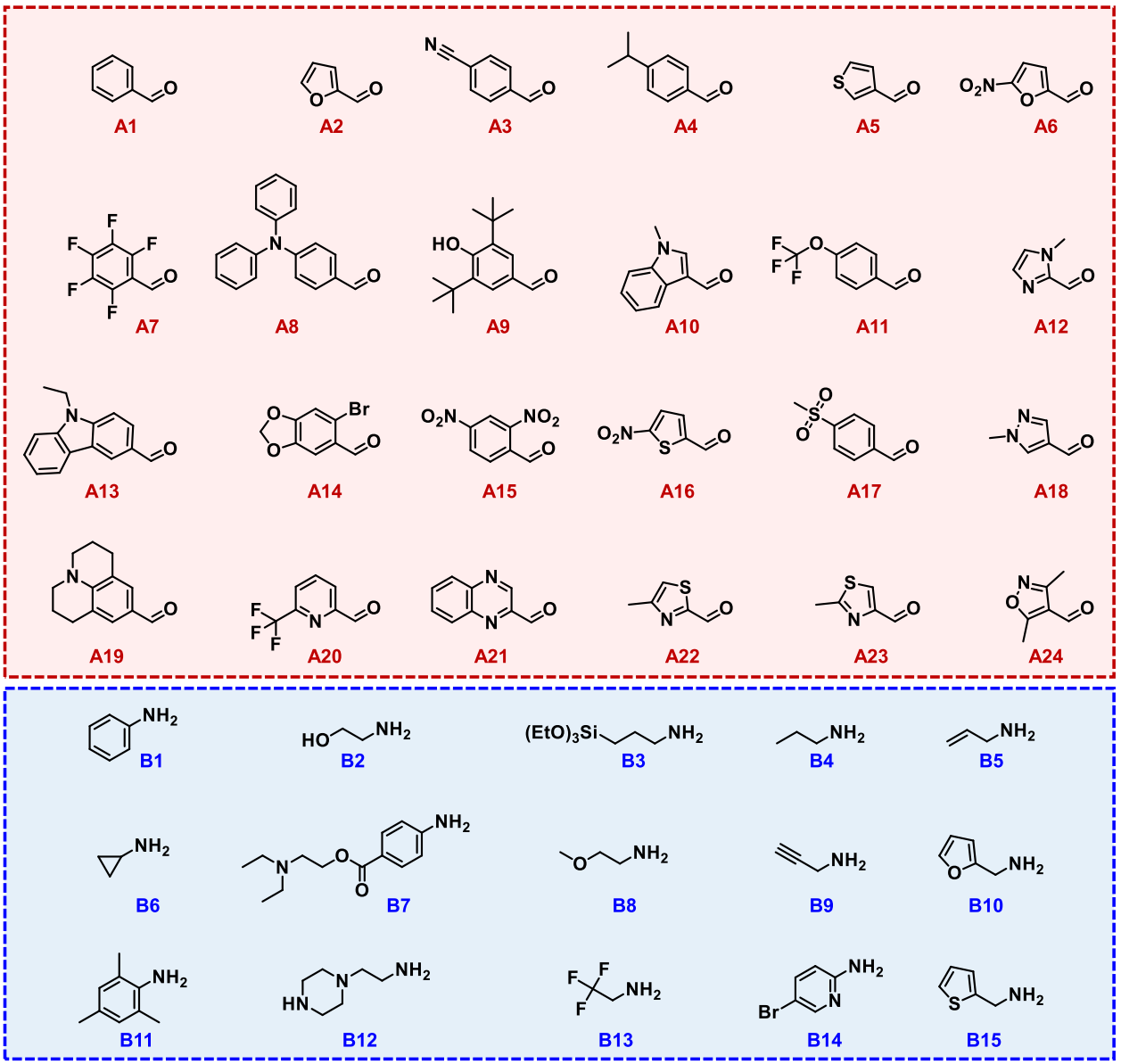


Figure 2. Chemical structures of selected aldehydes and amines for experimental determination of imine formation reaction constants.

From the structural point of view, the set of aldehydes is quite diverse (Figure 2 (top)): half of the molecules are heterocyclic aldehydes, e.g., containing furan (**A2** and **A6**), thiophene (**A5**, **A16**), and thiazole (**A22** and **A23**) cores. The set incorporates aldehydes presenting either electron-donating groups (e.g., **A4**, **A8**, **A9**, **A13**, **A19**), or electron-withdrawing groups (e.g., **A3**, **A6**, **A7**, **A11**, **A15**).

The set of amines, on the other hand, predominantly consists of various aliphatic amines (11 out of 15 molecules), three anilines (**B1**, **B7**, and **B11**), and one heterocyclic amine (**B14**); see Figure 2 (bottom).

Measurement of Stability Constants (log K_C). Stock solutions of all the aldehydes and amines were prepared in deuterated chloroform. Prior to use, $CDCl_3$ was filtered through basic alumina to remove the possible traces of acid; then, it was saturated with water to ensure a constant water content of 73.8 mM,[57] and hexamethyldisiloxane (HMDSO) was added as an internal standard. Imines were prepared directly in NMR tubes by mixing the stock solutions of aldehydes and amines to reach a concentration of 20 mM. To speed up the reaction, 2 mol % of trifluoroacetic acid (TFA) was added to each tube, and the reactions were equilibrated for 24 h at room temperature. Notice that kinetics of equilibration for several checked samples was well below 1 h.

Thus, from a virtual pool of 120 000 imines, 276 were synthesized. The reaction constant for each was calculated from direct measurement of the concentrations of the imine and of the residual aldehyde and amine by integrating their corresponding NMR signals relative to an internal standard (see the "Experimental measurements" section in Supporting Information). In most cases, the integrals could be measured so as to provide stability constant (K_C) values with a reasonable precision of 0.15 log K units (Figure 3, green bars), but where reaction was limited or strongly favored or where signal overlap occurred, errors were large and the K_C values can only be described as "estimated" (Figure 3, orange bars).

As expected, the imines having high log K_C, **A6B6** (5.40), **A7B6** (5.48), **A21B9** (5.54), etc., are formed by highly nucleophilic amines and highly electrophilic aldehydes. Note that most of the imines with log K_C > 5 contain the

https://doi.org/10.1021/acscentsci.2c00048
ACS Cent. Sci. 2022, 8, 804−813

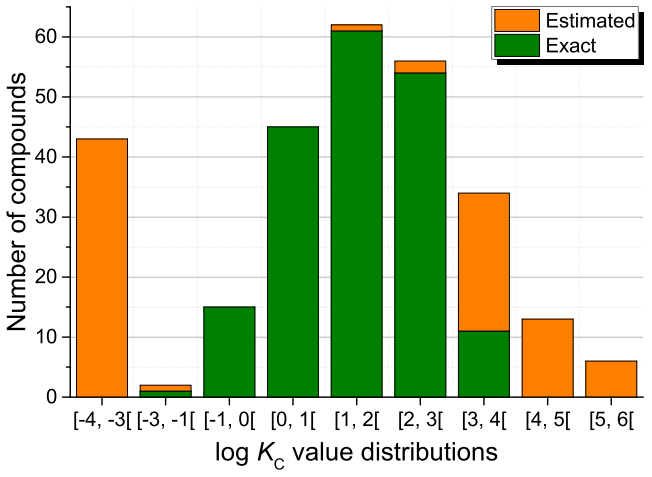


Figure 3. log K_C values distribution. Data are annotated as "exact" and "estimated", respectively. "Estimated" labels were assigned in cases featuring (i) too low concentrations of reactants/products or (ii) overlapping signals, leading to difficulties in quantitative identification of compounds.

cyclopropylamine fragment (**B6**). The imines with very low log K_C are formed from electron-poor amines and electron-rich aldehydes. For instance, electron-deficient amines such as **B7** and **B14** are poorly reactive in reactions with most aldehydes (log $K_C < -3$). Some amines (e.g., aniline **B11**) lead to imines with a broad range of stability from −3.25 (**A19B11**) to 1.91 (**A15B11**). In this case, steric effects apparently play a significant role in modulating the stability of constituents.

Predictive Model of Imine Stability in Chloroform. The data obtained were used to calibrate and validate the model. Seven support vector regression (SVR)[58] individual models, each built on a particular type of ISIDA descriptor[59,60] (see Supporting Information), contributed to consensus calculations. Their predictive performance was assessed in five-fold cross-validation. Finally, experimental versus predicted (cross-validated) log K_C values were compared (Figure 4a). For most molecules, the predicted log K_C values were close to those determined experimentally (root-mean-squared error (RMSE) is 0.62 log K units, see Figure 4), whereas most erroneous predictions were found for the compounds labeled as "estimated".

Aside from its predictive utility, another important criterion characterizing the obtained model is chemical space coverage, identified as the applicability domain (AD) of the model. The role of the AD is to define the boundaries in the chemical space within which a model can be used and provide reliable and accurate predictions. According to Vapnik,[61] statistical models

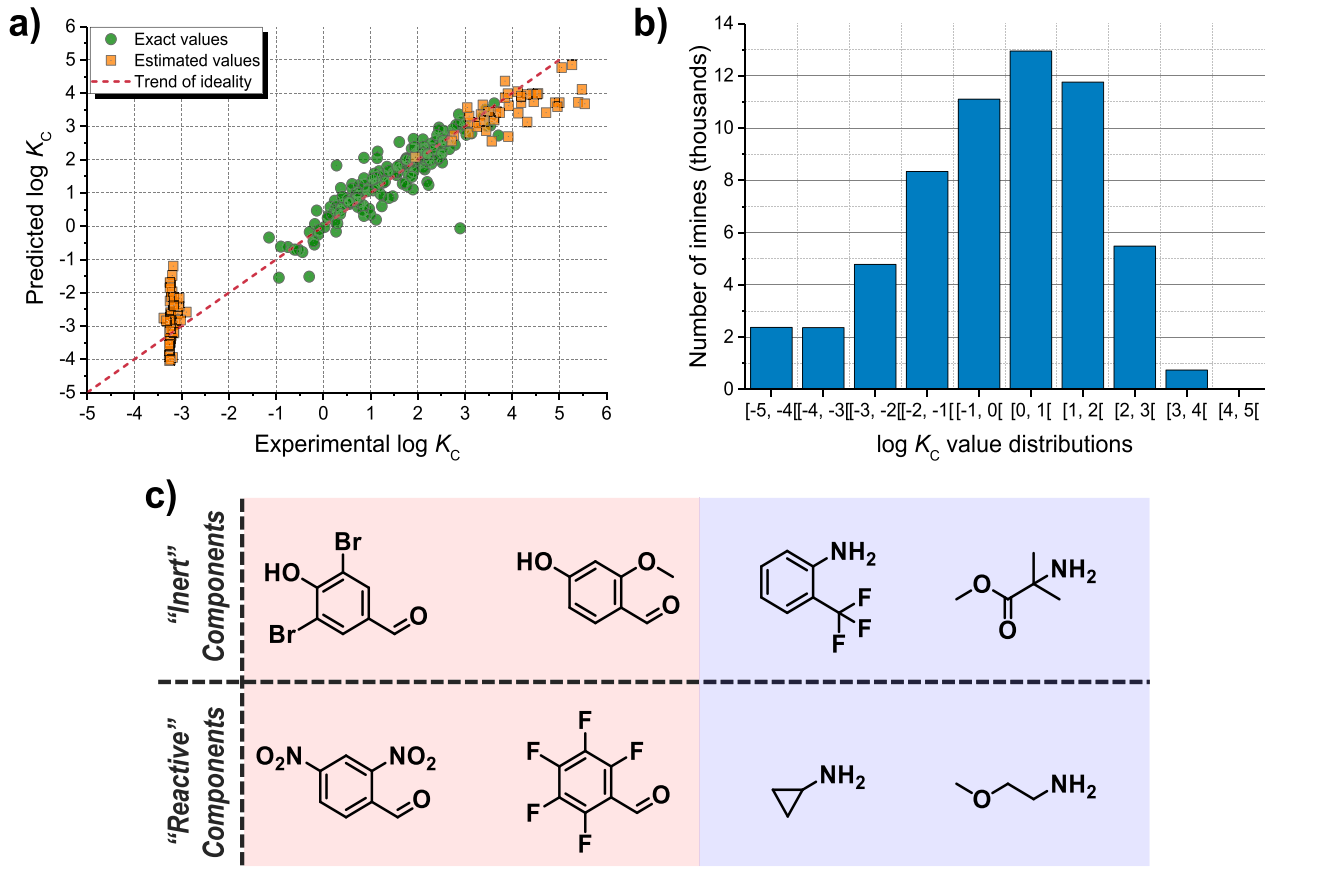


Figure 4. (a) Experimental vs predicted (cross-validated) log K_C values plot of the consensus SVR model with $Q^2 = 0.93$ and RMSE = 0.62 log K units (see details in Supporting Information). The dotted line corresponds to ideal predictions. (b) Distribution of predicted values of log K_C of imine formation in chloroform. (c) (Top) Examples of "inert" aldehydes (left) and "inert" amines (right). Their interactions with any other aldehyde and amine, respectively, lead in approximately 60% of cases to negative predicted log K. (Bottom) Examples of "reactive" aldehydes (left) and "reactive" amines (right). Their interactions with other aldehydes and amines, respectively, lead in more than 60% of the cases to log $K_C > 1$.

https://doi.org/10.1021/acscentsci.2c00048
ACS Cent. Sci. 2022, 8, 804−813

are directly applicable to any test instance drawn from the statistical distribution describing a training set; i.e., loosely speaking, the training and test molecules should not be too different. Here, we used the *fragment control* approach[59] to identify the AD. If a test compound contains an ISIDA fragment absent in the training set structures, it is considered to be out of the AD and, therefore, should be discarded. In this context, the SVR consensus model trained on log K_C of 276 imines should provide reliable predictions for almost 50% of the considered imines (59 935 out of 120 000). For approximately half of the imines within the AD (Figure 4b), the log K_C has been predicted as ≤0, for around 30 000 imines the predicted log K_C values were in the range between 0 and 3, and for 768 imines the predicted values of log K_C were >3 (Figure 4b). Thus, the latter group can be considered as suitable candidates for a DCL.

For these 59 935 imines, the chemotypes of their source reactants were analyzed. Some aldehydes and amines have been identified as "inert": with >60% of the coupling partners, their products have negative log K_C values. By contrast, "reactive" compounds have been identified as those with log K_C values >1 in approximately 60% of the reactions involving them. As expected, inert/reactive amines have, respectively, electron-acceptor/electron-donor substituents, which reduce/increase the reagent's basicity (Figure 4c). Conversely, inert/reactive aldehydes carry, respectively, electron-donor/electron-acceptor substituents.

Estimation of Imine Formation Constants in Water. Predicting the speciation of dynamic imine networks in the presence of biological molecules as effectors requires the prediction of the equilibrium constant of imine formation in water (log K_W) instead of the chloroform (log K_C), considered so far. The conversion of the constant in chloroform to that in water can be related to differences in solvation of the involved species, which are nothing but expressions of water−chloroform partition coefficients:

$$\log K_W = \log K_C - \log P_{C/W}(\text{imine}) + \log P_{C/W}(\text{aldehyde}) + \log P_{C/W}(\text{amine}) - \log P_{C/W}(H_2O) \quad (2)$$

The detailed derivation of eq 2 is given in Supporting Information, eqs S1−S3.

However, the required log $P_{C/W}$ values have not been experimentally assessed for all the DCL reagents and even less so for the large pool of possible products. Therefore, a computational predictive log $P_{C/W}$ model was successfully developed (see Supporting Information) on the basis of a training set containing 50 compounds from the ChEMBL database[48,49] with experimentally measured chloroform/water partition coefficients log $P_{C/W}$. However, because of the relatively small size of the training set (50 fragment-like molecules), the applicability domain of the model is very restricted. Thus, reliable predictions have been obtained only for 64 imines constituted from 14 amines and 22 aldehydes, with structures given in the DCL_data.zip file in Supporting Information. Application of eq 2 to the set of these 64 imines shows that formation constants in water are always larger than that in chloroform, i.e., log K_W > log K_C. However, this notwithstanding, the corresponding imine concentrations will be lower in water (which shifts the equilibrium toward the reagents—amine and aldehyde). As water concentrations are constant both in aqueous (55.56 M) and chloroform environments (saturation concentration of 73.8 mM) and hence do not need to be monitored in the subsequent speciation calculations, it makes sense to introduce "effective" stability constants instead of the thermodynamic values employed so far:

$$K^{\text{eff}} = \frac{K}{[H_2O]} \quad (3)$$

where $[H_2O]$ stands for the above-mentioned water concentrations in the respective phases.

For 64 imines within a chloroform/water partition coefficient AD, a simple relation between effective constants of imines formation in water and in chloroform has been observed (Figure S4):

$$\log K_W^{\text{eff}} = \log K_C^{\text{eff}} - 1 \quad (4)$$

Thus, effective stability constants reflect the intuitive expectation of a net decrease of effective stability paralleling the net decrease of imine concentrations in water. It is also a useful shortcut for the speciation simulations. A linear dependence of unit slope, involving a simple constant offset is also expected as far as the intervening players displaying "ideal" solvation behavior in both chloroform and water so that their respective log P values may be considered as additive in terms of functional group contributions. If so, the only net difference is expected to stem from the replacement of the oxygen of the aldehyde carbonyl by the nitrogen of the amine which loses most of its basicity when converted to the =N− of the imine. Contributions of conserved functional groups on the aldehyde or the amine to the chemical potential of solvation will be roughly the same, irrespective of whether they are carried by the reagents of the product and hence cancel out according to eq 2—hence the constant offset practically observed in Figure S4. Of course, this simple assumption is no longer valid if functional groups would mutually interact in the product and/or reagents. A state-of-the-art chemoinformatics model of log P might indeed be trained to capture such effects—but, unfortunately, not in this case, given the sparseness of measured chloroform/water partition coefficient values log $P_{C/W}$. Thus, we assume in the following that eq 4 offers so far the best available estimation of imine stability in water and can be applied to all 59 935 virtual imines found within AD for the log K_C model.

Part 2. Modeling of Binding Affinity to Human CA II. The ChEMBL database was used as a source for experimental ligand binding affinity data (cited as the negative logarithm of the dissociation or "instability" constant, pK_i). The training data for the modeling contained 4350 unique inhibitors of human CA II with experimentally measured pK_i varying from 0 to 11 (Figure S5 in Supporting Information). This set included 41 imines, most of which had a pK_i in the range between 6 and 9. The developed consensus SVR model (refer to Supporting Information for details) of R^2 = 0.96 and RMSE = 0.27 log K_i units was used to predict pK_i for the set of 59 935 imines within the applicability domain of the model for imine equilibrium constants. For these molecules, the predicted pK_i values vary from 4 to 8, and the distribution function has a maximum at pK_i = 5−6 (Figure S6 in Supporting Information).

Part 3. Speciation Modeling of the DCL. To illustrate the operation of the speciation workflow, we decided to select the simplest DCL consisting of two aldehydes, two amines, and the related four imines. Ideal imines selected for the DCL should fulfill the following requirements: (i) their formation constants should be similar and high enough in order to provide comparable and rather high concentrations, and (ii) one of

https://doi.org/10.1021/acscentsci.2c00048
ACS Cent. Sci. 2022, 8, 804−813

the imines should have a much larger affinity for the effector than the other DCL members and their binding blocks, in order to reveal the effector-induced dynamic enhancement.

After obtaining the 59 935 effector affinity constants for imines within the model AD, this set was filtered by the requirement of a predicted log $K_W^{eff} > 3$ (stable in water). It was achieved by applying the predictive model of thermodynamic stability in chloroform, converting the result to the "effective" constant in chloroform (eq 3) and then eventually to the effective constant in water (eq 4). A total of 3615 imines passed this test. This pool is a collection of individual products, not a combinatorial library. "Singletons" were removed from this collection, in the sense that an aldehyde (or amine) **A** was kept if and only if there was at least another reagent **A′** of the same class, as well as two partners **B** and **B′** such that all combinations (**AB**, **AB′**, **A′B**, **A′B′**) were among the 3615 selected. This led to a restrained subset of 3091 of the above 3615 imines, forming a sparse matrix of 278 aldehydes × 89 amines as a mosaic of several complete combinatorial sublibraries (Figure S8).

One of these 2 × 2 sublibraries (see Scheme 2) was chosen to illustrate the speciation analysis, the final step of the present workflow. First, effector affinities for CA II (pK_i) were also estimated for the amine and aldehyde reagents, as these might also interact with the protein (see Figure 5a and Table S5 in Supporting Information). These values were used as an input to the ChemEqui speciation software[62] in order to calculate the species concentrations in the absence and in the presence of the CA II protein receptor (Figure 5b and Table S6 in Supporting Information).

As shown in Figure 5a, **A′B′** has the largest predicted pK_i value, although it does not stand out in comparison with the others in this respect. As expected, in the absence of the effector, the concentration of all imines is larger than that of their building blocks, and **A′B′** is the dominant product. In the presence of the CA II enzyme, the interplay between the different ligand-enzyme stabilities results in significant changes of the constituent distribution of the DCL. The imine **A′B′**, which has the highest binding affinity for the effector CA II (pK_i = 6.70), becomes involved in a shift of the global equilibrium toward this ligand−enzyme complex. Consequently, the concentrations of its free building blocks in solution decrease, the increase of concentration —"amplification"—of the dynamically selected **A′B′** leading to a decrease or "down-regulation" of the poorly bound **AB′** and **A′B** (Figure 5c). To sum up, the addition of the human CA II to the solution increased the overall concentration of **AB** by 12% and A′B′ by 27% with respect to their concentrations in the absence of the effector associated with a decrease of the concentrations of **AB′** and **A′B** by 26% and 28%, respectively.

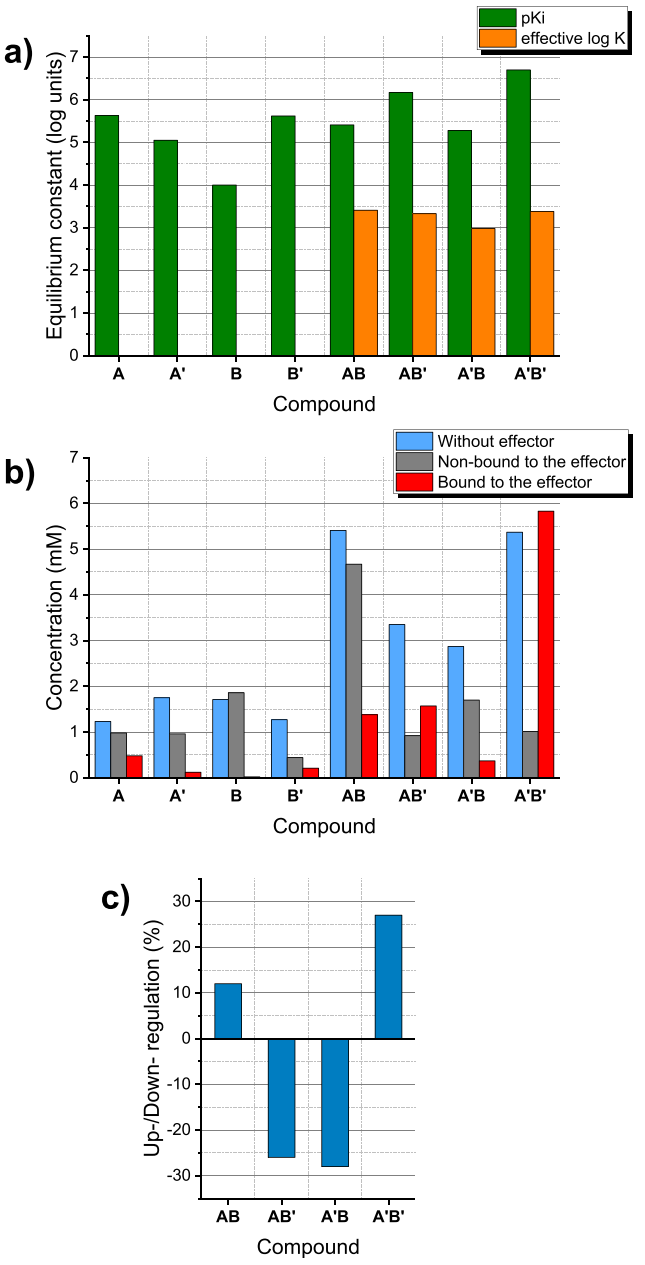


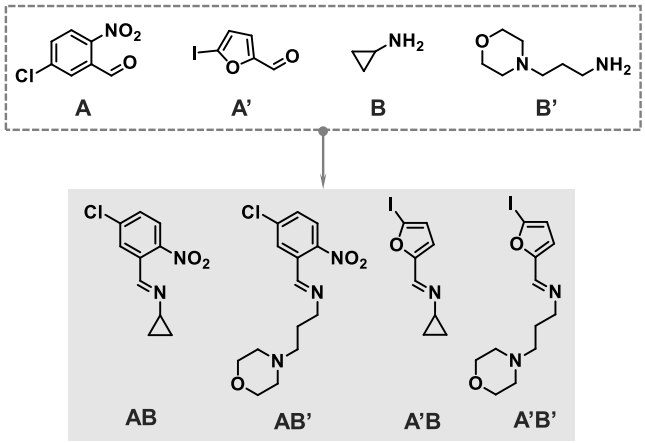
Figure 5. Calculated thermodynamic and speciation parameters for aldehydes (**A**, **A′**), amines (**B**, **B′**), and corresponding imines (**AB**, **AB′**, **A′B**, **A′B′**). (a) Predicted log K_W^{eff} (in orange) and pK_i values (in green). (b) The concentrations of the species in the absence (blue) and in the presence of human CA II protein (gray for uncomplexed, and red for complexed species). (c) Effect of dynamic amplification (up-regulation) and down-regulation (%).

DISCUSSION

In the present study, the stability constants for imine formation, log *K*, and the affinity constants toward carbonic anhydrase CA II, pK_i (predicted), of almost 60 000 imines were determined. With help from the speciation tool, a focused array of *n* aldehydes × *m* amines could be picked such as to ensure that (a) there are putative strong CA binders among the *n* × *m* imines, and (b) these putative binders are not penalized by an intrinsic instability that might jeopardize their "selection" by the protein site. Of this pool of imines, the results show that there was no "minimalistic" DCL obtained from a pairwise reaction of two aldehydes and two amines, which would result in the exclusive complexation of the human CA II enzyme with only one imine. This is not surprising, as it echoes an already known feature of combinatorial libraries—the high degree of relatedness of its members: near neighbors sharing a parent may also share comparable activity levels for the target. The positive aspect of this result is that the discovery of a series of active analogues may help the subsequent hit-to-lead optimization efforts. However, it is clear that the DCL investigated so far is incomplete, failing to include important structural features, notably in this case a (phenyl)sulfonamide group because (i) of its low solubility in chloroform, and (ii) it would overwhelmingly bias the DCL, as it is expected to interact very strongly with the Zn(II) cation in the

https://doi.org/10.1021/acscentsci.2c00048
ACS Cent. Sci. 2022, 8, 804−813

Scheme 2. Aldehydes (A, A′), Amines (B, B′), and Corresponding Imines (AB, AB′, A′B, A′B′) Selected for the Speciation Experiments



active site of the enzyme, thus overshadowing any other constituents. Investigation of other building blocks is required. This is nevertheless not a liability at this stage because any "primary hits" detected by a DCL would not have direct applications in drug discovery. DCLs are key tools to probe the protein binding patterns and provide structure−affinity information for refinement of affinity prediction models (machine-learned, pharmacophore-based, or docking-based).

The availability of experimental data on imine formation in a given solvent and on effector-imine affinity is crucial for machine-learning models. Models trained on small training sets have restricted applicability domains, which may significantly reduce the number of the considered DCL candidates.

Clearly, chemoinformatics will not predict a sole winner, given the inherent inaccuracies of the underlying models. Even the most accurate affinity prediction tool—computer-intensive free energy perturbation calculations—would fall short of this goal. Actually, even if all stability and affinity constants of individual DCL members were experimentally measured, the intrinsic experimental error of measurements (typically on the order of 0.5 log units) would still introduce significant uncertainty in the output of predicted speciation. Also note that predicted protein−ligand interactions are only prone to happen at the "envisaged" binding site for which the affinity model was tuned (as far as training data are binding site specific, as is the case of the classical K_i determinations from dose−response reference ligand displacement curves). Should some ligands bind at different protein sites—possibly modulating the protein activity—they would be selected by the DCL but not recognized as privileged ligands according to the predictive models. Such binding may give rise to secondary site bioactivity, for instance, by operation of an allosteric effect. This eventuality is an especially attractive feature of the DCL approach, more than direct binding to the "main" receptor site as highlighted by crystallographic data. It amounts to exploration of potential (virtual) sites versus design for a known site—but cannot benefit from chemoinformatics support, which is conditioned by prior knowledge. From the drug discovery point of view, it would suggest new regions for exploration of structure/activity relationships. In practice, the application of a DCL is a task of identification of the best/optimal binder(s), and it implicitly is much facilitated by an *a priori* knowledge of the protein structure and hence the knowledge about the binding site(s).[63]

CONCLUSIONS

The present study shows that detailed *in silico* predictions of the behavior of DCLs is technically feasible, pending experimental validation to prove that such insights gained from simulations may indeed help to rationally design DCLs maximizing the expectation to discover useful new protein inhibitors, metal ion chelators, and synthetic receptors. So far, training data quantity and quality are not sufficient to build ideally predictive models, with extrapolation capacities such as to render predicted equilibrium constant values accurate enough to support a prediction of equilibrium concentrations in such a complex system as a DCL. However, this ultimate goal is not the actual objective of chemoinformatics, which has proven of great utility in spite of the inaccuracy of its predictions. Total "*computational deconvolution*" of a DCL is hardly an achievable goal. Fortunately, this is not needed because the DCL is *per se* an outstanding search tool for the optimal binder, allowing for simultaneous "testing" of large numbers of competing structures.[23,35,36] The approach may, however, be sufficiently accurate to ensure that a computer-designed DCL stands enhanced chances of success compared to some random mixture of reagent pools. Discovery is not expected to come from one initial "perfect" prediction but from cycles of prediction—experimentation—model reassessment and refinement, taking into account the latest experimental results. The present work outlines the technical feasibility of the computational part, leaving the experimental validation challenge open for future work.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acscentsci.2c00048.

Details of the diverse library selection, statistical parameters of individual SVR models predicting logarithms of stability constants of imine formation and CAII-ligand complexation, information about the CAII binding site, experimentally measured formation constants for 276 imines in chloroform, and a user guide for the predictor tool implementing developed models for log K (in chloroform) and pK_i (PDF)

DCL_data.zip archive that includes five EXCEL files containing information about the raw data used for machine learning (ZIP)

AUTHOR INFORMATION

Corresponding Authors

Alexandre Varnek − *Laboratoire de Chémoinformatique UMR 7140 CNRS, 67081 Strasbourg, France;* orcid.org/0000-0003-1886-925X; Email: varnek@unistra.fr

Jean-Marie Lehn − *Laboratoire de Chimie Supramoléculaire, Institut de Science et d'Ingénierie Supramoléculaires (ISIS), Université de Strasbourg, 67000 Strasbourg, France;* orcid.org/0000-0001-8981-4593; Email: lehn@unistra.fr

Authors

Iuri Casciuc − *Laboratoire de Chémoinformatique UMR 7140 CNRS, 67081 Strasbourg, France; Laboratoire de Chimie Supramoléculaire, Institut de Science et d'Ingénierie Supramoléculaires (ISIS), Université de Strasbourg, 67000 Strasbourg, France*

https://doi.org/10.1021/acscentsci.2c00048
ACS Cent. Sci. 2022, 8, 804−813

Artem Osypenko − *Laboratoire de Chimie Supramoléculaire, Institut de Science et d'Ingénierie Supramoléculaires (ISIS), Université de Strasbourg, 67000 Strasbourg, France;* orcid.org/0000-0001-8251-7457

Bohdan Kozibroda − *Laboratoire de Chimie Supramoléculaire, Institut de Science et d'Ingénierie Supramoléculaires (ISIS), Université de Strasbourg, 67000 Strasbourg, France; Institute of High Technologies, Taras Shevchenko National University of Kyiv, 03022 Kyiv, Ukraine;* orcid.org/0000-0002-6654-1454

Dragos Horvath − *Laboratoire de Chémoinformatique UMR 7140 CNRS, 67081 Strasbourg, France;* orcid.org/0000-0003-0173-5714

Gilles Marcou − *Laboratoire de Chémoinformatique UMR 7140 CNRS, 67081 Strasbourg, France;* orcid.org/0000-0003-1676-6708

Fanny Bonachera − *Laboratoire de Chémoinformatique UMR 7140 CNRS, 67081 Strasbourg, France*

Complete contact information is available at: https://pubs.acs.org/10.1021/acscentsci.2c00048

Notes

The authors declare no competing financial interest.

Model predicting imines formation constants in chloroform and binding affinity to CA II enzyme are available for the users at http://infochim.u-strasbg.fr/cgi-bin/predictor_mixtures.cgi (log K_C) and http://infochim.u-strasbg.fr/webserv/VSEngine.html (pK_i). Detailed instructions for models application are given in Supporting Information.

■ ACKNOWLEDGMENTS

The authors thank the ERC (Advanced Research Grant SUPRADAPT 290585) and the University of Strasbourg Institute for Advanced Study (USIAS) for financial support. I.C. thanks Région Grand Est for a Ph.D. fellowship. We thank Dr. Vitaly Solov'ev for providing us with the ChemEqui program for speciation assessment and Prof. Jack Harrowfield for valuable comments and discussions.

■ REFERENCES

(1) Dynamic Combinatorial Chemistry. In *Drug Discovery, Bioorganic Chemistry, and Materials Science*; Miller, B. L., Ed.; John Wiley & Sons, Inc.: Hoboken, NJ, 2009.

(2) *Dynamic Combinatorial Chemistry*; Reek, J. N. H., Otto, S., Eds.; Wiley-VCH: Weinheim, Germany, 2010.

(3) *Constitutional Dynamic Chemistry*; Barboiu, M., Ed.; Topics in Current Chemistry; Springer: Berlin, 2012; Vol. *322*.

(4) *Dynamic Covalent Chemistry: Principles, Reactions, and Applications*; Zhang, W., Jin, Y., Eds.; John Wiley & Sons, Ltd.: Chichester, UK, 2017.

(5) Rowan, S. J.; Cantrill, S. J.; Cousins, G. R. L.; Sanders, J. K. M.; Stoddart, J. F. Dynamic Covalent Chemistry. *Angew. Chem., Int. Ed.* **2002**, *41*, 898−952.

(6) Corbett, P. T.; Leclaire, J.; Vial, L.; West, K. R.; Wietor, J.-L.; Sanders, J. K. M. M.; Otto, S. Dynamic Combinatorial Chemistry. *Chem. Rev.* **2006**, *106*, 3652−3711.

(7) Lehn, J.-M. Dynamic Combinatorial Chemistry and Virtual Combinatorial Libraries. *Chem. - Eur. J.* **1999**, *5*, 2455−2463.

(8) Lehn, J.-M. From Supramolecular Chemistry towards Constitutional Dynamic Chemistry and Adaptive Chemistry. *Chem. Soc. Rev.* **2007**, *36*, 151−160.

(9) Amplification of the fittest constituent can be complicated by the sensitivity of the selection process to the influence of several parameters, namely, the (i) size of the DCL; (ii) concentration of the constituents; (iii) relative values of binding constants of the constituents with the effector/receptor; (iv) stoichiometry, symmetry, and isomerism of the components (1:1, dimer, trimer, etc.). Greater detail on these issues is available elsewhere.[10−16]

(10) Severin, K. The Advantage of Being Virtual—Target-Induced Adaptation and Selection in Dynamic Combinatorial Libraries. *Chem. - Eur. J.* **2004**, *10*, 2565−2580.

(11) Saur, I.; Severin, K. Selection Experiments with Dynamic Combinatorial Libraries: The Importance of the Target Concentration. *Chem. Commun.* **2005**, *11*, 1471−1473.

(12) Grote, Z.; Scopelliti, R.; Severin, K. Adaptive Behavior of Dynamic Combinatorial Libraries Generated by Assembly of Different Building Blocks. *Angew. Chem., Int. Ed.* **2003**, *42*, 3821−3825.

(13) Ludlow, R. F.; Otto, S. The Impact of the Size of Dynamic Combinatorial Libraries on the Detectability of Molecular Recognition Induced Amplification. *J. Am. Chem. Soc.* **2010**, *132*, 5984−5986.

(14) Corbett, P. T.; Otto, S.; Sanders, J. K. M. Correlation between Host−Guest Binding and Host Amplification in Simulated Dynamic Combinatorial Libraries. *Chem. - Eur. J.* **2004**, *10*, 3139−3143.

(15) Corbett, P. T.; Sanders, J. K. M.; Otto, S. Competition between Receptors in Dynamic Combinatorial Libraries: Amplification of the Fittest? *J. Am. Chem. Soc.* **2005**, *127*, 9390−9392.

(16) Corbett, P. T.; Otto, S.; Sanders, J. K. M. What Are the Limits to the Size of Effective Dynamic Combinatorial Libraries? *Org. Lett.* **2004**, *6*, 1825−1827.

(17) Giuseppone, N.; Lehn, J.-M. M. Protonic and Temperature Modulation of Constituent Expression by Component Selection in a Dynamic Combinatorial Library of Imines. *Chem. - Eur. J.* **2006**, *12*, 1715−1722.

(18) Gu, R.; Lehn, J. Metal Ion-Driven Constitutional Adaptation in Dynamic Covalent C = C/C = N Organo-Metathesis. *Chem. - Asian J.* **2021**, *16*, 44−48.

(19) Huc, I.; Lehn, J.-M. J.-M. Virtual Combinatorial Libraries: Dynamic Generation of Molecular and Supramolecular Diversity by Self-Assembly. *Proc. Natl. Acad. Sci. U. S. A.* **1997**, *94*, 2106−2110.

(20) See also a recent review on the use of DCLs targeting Carbonic Anhydrases: Su, D.; Zhang, Y.; Ulrich, S.; Barboiu, M. Constitutional Dynamic Inhibition/Activation of Carbonic Anhydrases. *ChemPlusChem.* **2021**, *86*, 1500−1510.

(21) Hafezi, N.; Lehn, J.-M. Adaptation of Dynamic Covalent Systems of Imine Constituents to Medium Change by Component Redistribution under Reversible Phase Separation. *J. Am. Chem. Soc.* **2012**, *134*, 12861−12868.

(22) Gu, R.; Lehn, J.-M. Constitutional Dynamic Selection at Low Reynolds Number in a Triple Dynamic System: Covalent Dynamic Adaptation Driven by Double Supramolecular Self-Assembly. *J. Am. Chem. Soc.* **2021**, *143*, 14136−14146.

(23) Ramström, O.; Lehn, J.-M. Drug Discovery by Dynamic Combinatorial Libraries. *Nat. Rev. Drug Discovery* **2002**, *1*, 26−36.

(24) Mondal, M.; Hirsch, A. K. H. Dynamic Combinatorial Chemistry: A Tool to Facilitate the Identification of Inhibitors for Protein Targets. *Chem. Soc. Rev.* **2015**, *44*, 2455−2488.

(25) Hartman, A. M.; Gierse, R. M.; Hirsch, A. K. H. Protein-Templated Dynamic Combinatorial Chemistry: Brief Overview and Experimental Protocol. *Eur. J. Org. Chem.* **2019**, *2019* (22), 3581−3590.

(26) Bugaut, A.; Jantos, K.; Wietor, J.-L.; Rodriguez, R.; Sanders, J. K. M.; Balasubramanian, S. Exploring the Differential Recognition of DNA G-Quadruplex Targets by Small Molecules Using Dynamic Combinatorial Chemistry. *Angew. Chem., Int. Ed.* **2008**, *47*, 2677−2680.

(27) Whitney, A. M.; Ladame, S.; Balasubramanian, S. Templated Ligand Assembly by Using G-Quadruplex DNA and Dynamic Covalent Chemistry. *Angew. Chem., Int. Ed.* **2004**, *43*, 1143−1146.

(28) Ladame, S.; Whitney, A. M.; Balasubramanian, S. Targeting Nucleic Acid Secondary Structures with Polyamides Using an Optimized Dynamic Combinatorial Approach. *Angew. Chem., Int. Ed.* **2005**, *44*, 5736−5739.

(29) McAnany, J. D.; Miller, B. L. Dynamic Combinatorial Chemistry as a Rapid Method for Discovering Sequence-Selective RNA-Binding

Compounds. *Methods Enzymol.* **2019**, *623*, 67−84, DOI: 10.1016/bs.mie.2019.05.012.
(30) Carbajo, D.; Pérez, Y.; Bujons, J.; Alfonso, I. Live-Cell-Templated Dynamic Combinatorial Chemistry. *Angew. Chem., Int. Ed.* **2020**, *59*, 17202−17206.
(31) Michael, S.; Auld, D.; Klumpp, C.; Jadhav, A.; Zheng, W.; Thorne, N.; Austin, C. P.; Inglese, J.; Simeonov, A. A Robotic Platform for Quantitative High-Throughput Screening. *Assay Drug Dev. Technol.* **2008**, *6*, 637−657.
(32) Lehn, J.-M. Constitutional Dynamic Chemistry: Bridge from Supramolecular Chemistry to Adaptive Chemistry. *Topics in Current Chemistry* **2012**, *322*, 1−32.
(33) Lehn, J.-M. Perspectives in Chemistry-Steps towards Complex Matter. *Angew. Chem., Int. Ed.* **2013**, *52*, 2836−2850.
(34) Ayme, J.-F.; Lehn, J.-M. From Coordination Chemistry to Adaptive Chemistry. In *Adv. Inorg. Chem.*; Academic Press, 2018; Vol. *71*, pp 3−78.
(35) Bunyapaiboonsri, T.; Ramström, O.; Lohmann, S.; Lehn, J.-M.; Peng, L.; Goeldner, M. Dynamic Deconvolution of a Pre-Equilibrated Dynamic Combinatorial Library of Acetylcholinesterase Inhibitors. *ChemBioChem.* **2001**, *2*, 438−444.
(36) Ramström, O.; Lehn, J.-M. M. In Situ Generation and Screening of a Dynamic Combinatorial Carbohydrate Library against Concanavalin A. *ChemBioChem.* **2000**, *1*, 41−48.
(37) Schneider, G.; Nettekoven, M. Ligand-Based Combinatorial Design of Selective Purinergic Receptor (A2A) Antagonists Using Self-Organizing Maps. *J. Comb. Chem.* **2003**, *5*, 233−237.
(38) Selzer, P.; Ertl, P. Applications of Self-Organizing Neural Networks in Virtual Screening and Diversity Selection. *J. Chem. Inf. Model.* **2006**, *46*, 2319−2323.
(39) For general information on the use of unsupervised machine-learning methods applied to dynamic combinatorial chemistry see for instance Misuraca, M. C.; Moulin, E.; Ruff, Y.; Giuseppone, N. Experimental and Theoretical Methods for the Analyses of Dynamic Combinatorial Libraries. *New J. Chem.* **2014**, *38*, 3336−3349 For instance, PCA, LDA, and cluster analysis of DCLs in sensing applications has been actively used by the group of Severin[40] and recently by our group[41].
(40) Severin, K. J. Pattern-Based Sensing with Simple Metal−dye Complexes. *Curr. Opin. Chem. Biol.* **2010**, *14*, 737−742.
(41) Osypenko, A.; Dhers, S. S.; Lehn, J.-M. Pattern Generation and Information Transfer through a Liquid/Liquid Interface in 3D Constitutional Dynamic Networks of Imine Ligands in Response to Metal Cation Effectors. *J. Am. Chem. Soc.* **2019**, *141*, 12724−12737.
(42) Jorgensen, W. L.; Thomas, L. L. Perspective on Free-Energy Perturbation Calculations for Chemical Equilibria. *J. Chem. Theory Comput.* **2008**, *4*, 869−876.
(43) Gimadiev, T. R.; Madzhidov, T. I.; Nugmanov, R. I.; Baskin, I. I.; Antipin, I. S.; Varnek, A. Assessment of Tautomer Distribution Using the Condensed Reaction Graph Approach. *J. Comput. Aided. Mol. Des.* **2018**, *32*, 401−414.
(44) Gans, P.; Sabatini, A.; Vacca, A. Investigation of Equilibria in Solution. Determination of Equilibrium Constants with the HYPERQUAD Suite of Programs. *Talanta* **1996**, *43*, 1739−1753.
(45) Solov'ev, V. P.; Tsivadze, A. Y. Supramolecular Complexes: Determination of Stability Constants on the Basis of Various Experimental Methods. *Prot. Met. Phys. Chem. Surfaces* **2015**, *51*, 1−35.
(46) Carbonic anhydrases are metalloenzymes that catalyze the reversible hydration of carbon dioxide to a bicarbonate ion and a proton in two steps: Supuran, C. T. Carbonic Anhydrases: Novel Therapeutic Applications for Inhibitors and Activators. *Nat. Rev. Drug Discovery* **2008**, *7*, 168−181.
(47) Gaulton, A.; Bellis, L. J.; Bento, A. P.; Chambers, J.; Davies, M.; Hersey, A.; Light, Y.; McGlinchey, S.; Michalovich, D.; Al-Lazikani, B.; et al. ChEMBL: A Large-Scale Bioactivity Database for Drug Discovery. *Nucleic Acids Res.* **2012**, *40*, D1100−D1107.
(48) Gaulton, A.; Hersey, A.; Nowotka, M.ł; Bento, A. P.; Chambers, J.; Mendez, D.; Mutowo, P.; Atkinson, F.; Bellis, L. J.; Cibrian-Uhalte, E.; Davies, M.; Dedman, N.; Karlsson, A.; Magarinos, M. P.; Overington, J. P.; Papadatos, G.; Smit, I.; Leach, A. R. The {ChEMBL} Database in 2017. *Nucleic Acids Res.* **2017**, *45*, D945−D954.
(49) Mendez, D.; Gaulton, A.; Bento, A. P.; Chambers, J.; De Veij, M.; Félix, E.; Magariños, M. P.; Mosquera, J. F.; Mutowo, P.; Nowotka, M.; et al. ChEMBL: Towards Direct Deposition of Bioassay Data. *Nucleic Acids Res.* **2019**, *47*, D930−D940.
(50) Belowich, M. E.; Stoddart, J. F. Dynamic Imine Chemistry. *Chem. Soc. Rev.* **2012**, *41*, 2003−2024.
(51) Sprung, M. A. A Summary of the Reactions of Aldehydes with Amines. *Chem. Rev.* **1940**, *26*, 297−338.
(52) Layer, R. W. The Chemistry of Imines. *Chem. Rev.* **1963**, *63*, 489−510.
(53) *Carbon-Nitrogen Double Bonds (1970)*; Patai, S., Ed.; John Wiley & Sons, Ltd.: Chichester, UK, 1970.
(54) Clegg, J. K.; Harrowfield, J.; Kim, Y.; Lee, Y. H.; Lehn, J.-M.; Lim, W. T.; Thuéry, P. Chelation-Controlled Molecular Morphology: Aminal to Imine Rearrangements. *Dalt. Trans.* **2012**, *41*, 4335−4357.
(55) Godoy-Alcántar, C.; Yatsimirsky, A. K.; Lehn, J.-M. M. Structure-Stability Correlations for Imine Formation in Aqueous Solution. *J. Phys. Org. Chem.* **2005**, *18*, 979−985.
(56) Holliday, J. D.; Willett, P. Definitions of "Dissimilarity" for Dissimilarity-Based Compound Selection. *J. Biomol. Screen.* **1996**, *1*, 145−151.
(57) Masterton, W. L.; Gendrano, M. C. Henry's Law Studies of Solutions of Water in Organic Solvents. *J. Phys. Chem.* **1966**, *70*, 2895−2898.
(58) Drucker, H.; Burges, C. J. C.; Kaufman, L.; Smola, A.; Vapnik, V. Support Vector Regression Machines. In *Advances in Neural Information Processing Systems*; Mozer, M. C., Jordan, M., Petsche, T., Eds.; MIT Press, 1997; Vol. *9*, pp 155−161.
(59) Varnek, A.; Fourches, D.; Horvath, D.; Klimchuk, O.; Gaudin, C.; Vayer, P.; Solov'ev, V.; Hoonakker, F.; Tetko, I.; Marcou, G. ISIDA - Platform for Virtual Screening Based on Fragment and Pharmacophoric Descriptors. *Curr. Comput. Aided-Drug Des.* **2008**, *4*, 191−198.
(60) Ruggiu, F.; Marcou, G.; Solov'ev, V.; Horvath, D.; Varnek, A. *ISIDA Fragmentor 2015 User Manual*, 2015.
(61) Vapnik, V. N. *Statistical Learning Theory*; Wiley, 1998.
(62) Solov'ev, V. P.; Tsivadze, A. Y. Supramolecular Complexes: Determination of Stability Constants on the Basis of Various Experimental Methods. *Prot. Met. Phys. Chem. Surfaces* **2015**, *51*, 1−35.
(63) Ringe, D.; Mattos, C. Analysis of the Binding Surfaces of Proteins. *Med. Res. Rev.* **1999**, *19*, 321−331.