A generally applicable quantitative reactivity model for nucleophilic aromatic substitution built from simple descriptors

Jingru Lu, Irina Paci* and David C. Leitch*

Department of Chemistry, University of Victoria, 3800 Finnerty Rd. Victoria BC, CANADA, V8P 5C2.

*ipaci@uvic.ca; dcleitch@uvic.ca.

ABSTRACT. We report a multivariate linear regression model able to make accurate predictions for the rate and regioselectivity of nucleophilic aromatic substitution (S_NAr) reactions based on the electrophile structure. This model uses a diverse training/test set from experimentally-determined relative S_NAr rates between benzyl alcohol and 74 unique electrophiles, including heterocycles with multiple substitution patterns. There is a robust linear relationship between the experimental S_NAr free energies of activation and three molecular descriptors that can be obtained computationally: the *LUMO* energy of the electrophile; the average molecular electrostatic potential (*ESP*) at the carbon undergoing substitution; and the sum of average *ESP* values for the *ortho* and *para* atoms relative to the reactive center. Despite using only simple descriptors calculated from ground state wavefunctions, this model demonstrates excellent correlation with previously measured S_NAr reaction rates, and is able to accurately predict site selectivity for multihalogenated substrates: 91% prediction accuracy across 82 individual examples. The excellent agreement between predicted and experimental outcomes makes this easy-to-implement reactivity model a potentially powerful tool for synthetic planning.

Introduction

Making reliable predictions about the reactivity of organic molecules under specific conditions is the cornerstone of organic synthesis.[1] Every organic chemist learns to qualitatively predict and/or rationalize reactivity based on the properties of functional groups and substituents, and to use these predictions in designing effective syntheses.[2,3] Quantitative predictions of reactivity and selectivity are generally more challenging to achieve, and rely on sufficient experimental data to build structure-reactivity correlations, extensive theoretical calculations, or a combination of the two.[4–9] Recent advances in this area combine techniques such as high-throughput experimentation, descriptor generation, multivariate statistical analysis, and machine learning to generate robust quantitative structure-reactivity relationships (QSRR) and/or quantitative structure-selectivity relationships (QSSR) for specific reactions.[10–22] However, many significant challenges remain, including reliable data collection for a large enough set of chemical space, broad applicability of the resulting models beyond the specific training/test sets examined, and deployment in complex molecule synthesis planning and design.[23–25]

One class of organic reactions for which accurate predictive models would be invaluable is nucleophilic aromatic substitution (S_NAr). S_NAr is one of the most important and well-studied transformations in organic synthesis.[26–29] It is extensively used in total synthesis of natural products,[30–37] medicinal chemistry and agrochemistry,[38–43] and manufacturing of active pharmaceutical and agrochemical ingredients.[44–48] For example, S_NAr reactions are particularly powerful for the synthesis and functionalization of *N*-heterocycles, which are among the most ubiquitous structural components in active pharmaceutical ingredients.[49–51]

Because of its importance in synthesis, designing efficient and highly selective S_NAr reactions involving complex molecules is crucial. Substantial research over the past 100 years has been devoted to understanding the operative reaction mechanisms, whether stepwise or concerted,[26,52–54] and in collecting experimental reactivity and selectivity data for myriad substrate combinations. For example, Hammett[55] and/or Mayr parameters[4] are often used as mechanistic probes and to correlate/predict S_NAr reactivity (Fig. 1A).[56–62]

Theoretical and computational methods have been used to develop predictive models for specific subsets of S_NAr chemistry (Fig. 1B). Early work focused on stability of the σ-complex intermediates using I_π-repulsion theory,[63,64] or frontier molecular orbital considerations[65] to explain and predict regioselectivity.[66] Baker and Muir[67,68] as well as Brinck, Svensson, and co-workers[69–71] have published several works on predicting regioselectivity for S_NAr reactions using DFT-calculated transition state energies and/or stability of the σ-complex intermediates (*SS*).[71]

Quantum chemical transition state calculations are undeniably a powerful tool to explore reaction mechanisms and provide theoretical evidence to support experimental findings; however, the computational cost of performing transition state analyses remains high, and the complexity and nuance of these calculations make them beyond the expertise of many synthetic research groups. More desirable from an end-user perspective are models built from easily obtained molecular descriptors. In addition to established electronic and steric descriptors,[72,55,73] in 2016 Brinck and co-workers introduced the local electron attachment energy (a concept similar to the local electron affinity) as a molecular descriptor for electrophilicity,[74] and have applied it toward reactivity/selectivity predictions for S_NAr reactions.[75] While this descriptor correlates well with sets of experimental rates, and is able to provide qualitative selectivity predictions in multihalogenated systems, there is a need for new and more varied data and descriptor sets as foundations to build broadly applicable models for synthetic planning.

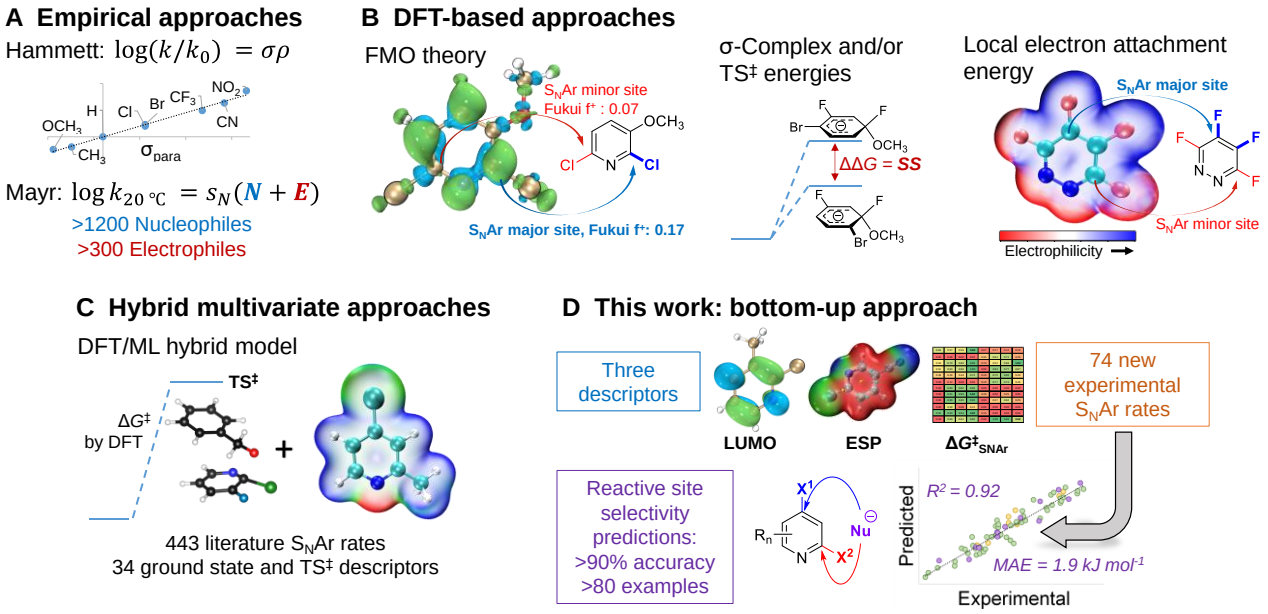

Fig. 1 Approaches to developing quantitative structure-reactivity relationships (QSRR) for S_NAr reactions. A) Empirical parameters derived from experimental data. B) Calculated descriptors from DFT analysis (FMO = frontier molecular orbital theory; TS = transition state). C) Recent hybrid DFT/ML approach. D) Bottom-up approach combining new experimental data with simple calculated descriptors.

Recently, Jorner, Brinck, Norrby, and Buttar reported the use of a hybrid DFT/machine learning (ML) approach to predicting experimental activation energies (Fig. 1C).[21] This study collates more than 440 S_NAr reaction rates from prior studies, and uses 34 ground state and transition state descriptors as the training/test set. Notably, the DFT-calculated transition state energies are a crucial descriptor in the best-performing model. This hybrid approach is demonstrably powerful; however, a reliance on previously reported experimental rates means there are key gaps in the training data, such as an overemphasis on nitroarenes, and relatively few heterocyclic electrophiles. It also still requires transition state calculations for maximum accuracy, especially if relatively few data points are available.

In this work, we consider the following three aspects of a predictive model to have equal importance: 1) the prediction accuracy the model provides, especially for new (external) predictions; 2) the breadth of applicability the model affords across chemical space; and 3) the ease and simplicity of applying the model to new systems. In the previously described examples, reaction rate/selectivity data used to train and validate the QSRR/QSSR models are taken from literature values, skewing the chemical space coverage toward well-studied systems. For our training/test dataset, we measured relative reaction rates for 74 individual electrophiles – including many nitrogen heterocycles relevant to pharmaceutical synthesis – using a competition experiment approach. Having control over the composition of our training set gives us the flexibility to have a varied and balanced distribution of structural features, which is necessary to ensure both accuracy and applicability in making new predictions. To make the model easy to implement, and to reduce the computational cost required, we combined simple and easy-to-obtain ground state molecular descriptors with our own experimentally determined S_NAr rates. From this combination of factors, we have constructed a QSRR model for S_NAr reactions with excellent performance in predicting reactivity trends and site selectivity for many different electrophiles, including for multiple external test sets with significantly different molecular structures (Fig. 1D).

Results and Discussion

Creating the training/test set.

An efficient approach to collect a large and diverse data set of reaction rates is critical to our bottom-up approach. To determine a large number of reaction rates in a timely manner, we followed a workflow of high-throughput competition experimentation shown in Fig. 2. This experimental approach can be summarized in three steps: first, we monitored the reaction progress of three touchstone reactions under *pseudo* first order conditions to determine absolute rate constants and free energies of activation ($\Delta G^{\ddagger}_{SNAr}$) for S_NAr between benzyl alkoxide and 2-chloropyridine, 2-chloro-6-methylpyridine, or 2-chloro-5-methoxypyridine as the electrophile (Fig. 2A). Next, we determined relative rate constants for the electrophile substrate library by a series of 94 individual competition experiments under analogous conditions (Fig. 2B, Table S2). Finally, we calibrated these relative rate constants using the touchstone reactions, giving absolute rate constants and the corresponding $\Delta G^{\ddagger}_{SNAr}$ values for the entire array of S_NAr reactions (Table S3). We used the absolute $\Delta G^{\ddagger}_{SNAr}$ value for the 2-chloropyridine touchstone reaction (88.8 kJ mol^{-1}) as the calibration point, with the other two touchstone reactions (2-chloro-6-methylpyridine, and 2-chloro-5-methoxypyridine) used to confirm the validity of the competition determined $\Delta G^{\ddagger}_{SNAr}$ values. We obtain a percent difference between the competition values and touchstone values of <2% (Fig. S3).

Using this competition approach, we were able to rapidly build a reliable and self-consistent data set from a library of 74 (hetero)aryl halides. This includes electrophiles with many diff-

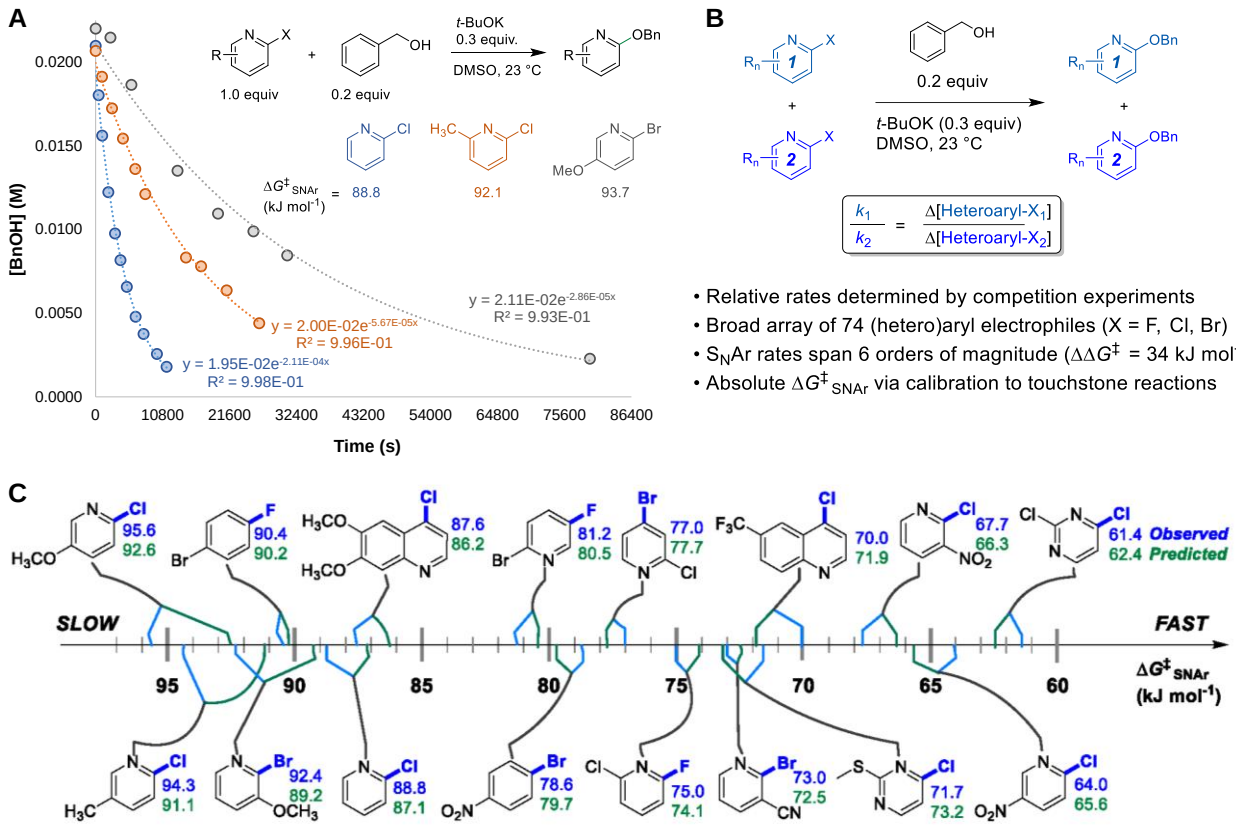

Fig. 2. Experimental approach to collecting free energies of activation for 74 S_NAr reactions; Bn = benzyl. A) Touchstone reaction progress analysis under pseudo first order conditions. B) Competition experiments to establish relative rates across electrophile library. C) Quantitative reactivity scale for representative electrophiles.

erent substitution patterns, and thus a variety of electronic effects. The reactivity of these substrates crosses a broad range, with the reaction rates spanning 6 orders of magnitude; a quantitative reactivity scale for several representative electrophiles is shown in Fig. 2C. As an initial check on the validity of our data set, we assessed the general reactivity trends against the known features of S_NAr reactivity. As expected, electron-deficient arenes react much faster than electron-rich ones; furthermore, the reactivity of the halides leaving groups follows the established trend, with rates decreasing as Ar–F >> Ar–Cl ~ Ar–Br.[76] We also constructed Hammett plots for four sets of 2-X-pyridine substrates (X = Cl, Br), giving linear correlations with rho values of ~4-5 (Fig. S4-S7). Finally, we have prepared and isolated 5 representative S_NAr products (compounds **S1-S5**), and confirmed their structures using NMR spectroscopy and high-resolution mass spectrometry (Fig. S8-S17).

Model generation and performance.

Based on the known aspects of S_NAr reaction mechanisms, and our prior work[77] in applying ground state molecular descriptors[78] to reactivity predictions, we built a quantitative structure-reactivity model for S_NAr electrophiles using only three descriptors. These include a global descriptor in the *LUMO* energy of the electrophile, and two local descriptors based on average molecular electrostatic potentials (*ESP*).[79–82] In addition to the ESP at the carbon undergoing substitution (ESP_1), we also discovered that the sum of *ESP* values for the *ortho* and *para* ring atoms is required for accurate predictions (ESP_2) (Fig. 3A).

By building a multivariate linear correlation between these three ground state descriptors and our experimentally obtained $\Delta G^{\ddagger}_{SNAr}$ values, we have established a unified structure-reactivity model able to accurately predict S_NAr rates for electrophiles with various structural features and leaving groups. There is an excellent linear correlation between the predicted and actual $\Delta G^{\ddagger}_{SNAr}$ values (R^2 = 0.92) and a mean absolute error (MAE) of only 1.8 kJ mol^{-1} (0.43 kcal mol^{-1}) (Fig. 3B). Performing a min/max normalization of the descriptors reveals their percentage contribution to the model, with ESP_1 being most important (50%), followed by ESP_2 (35%), and finally only a modest contribution from the *LUMO* energy (15%). We note including steric-based descriptors was not necessary to obtain good correlations for our data set.

We have assessed the robustness of the model using cross-validation with five different random 60/40 training/test set data splits (Fig. 3C and Fig. S20-23) and one structured split (Fig. S24). All of these regression analyses give essentially identical results, with excellent correlation statistics as indicated by the range of Q^2 values[83] from 0.86 to 0.93, and MAE values from 1.6 to 2.3 kJ mol^{-1} for the test sets. Finally, we also assessed the model performance by analysing the distribution of residuals across the data set, and identifying any possible outliers. As shown in Fig. 3D, the residuals are randomly distributed, almost exclusively in the range -5 to +5 kJ mol^{-1} (i.e. within an order of magnitude of the experimental rate). A box plot reveals only one significant outlier (|residual| > 5 kJ mol^{-1}): 2-(*N*-methylcarboxamide)-4-chloropyridine.

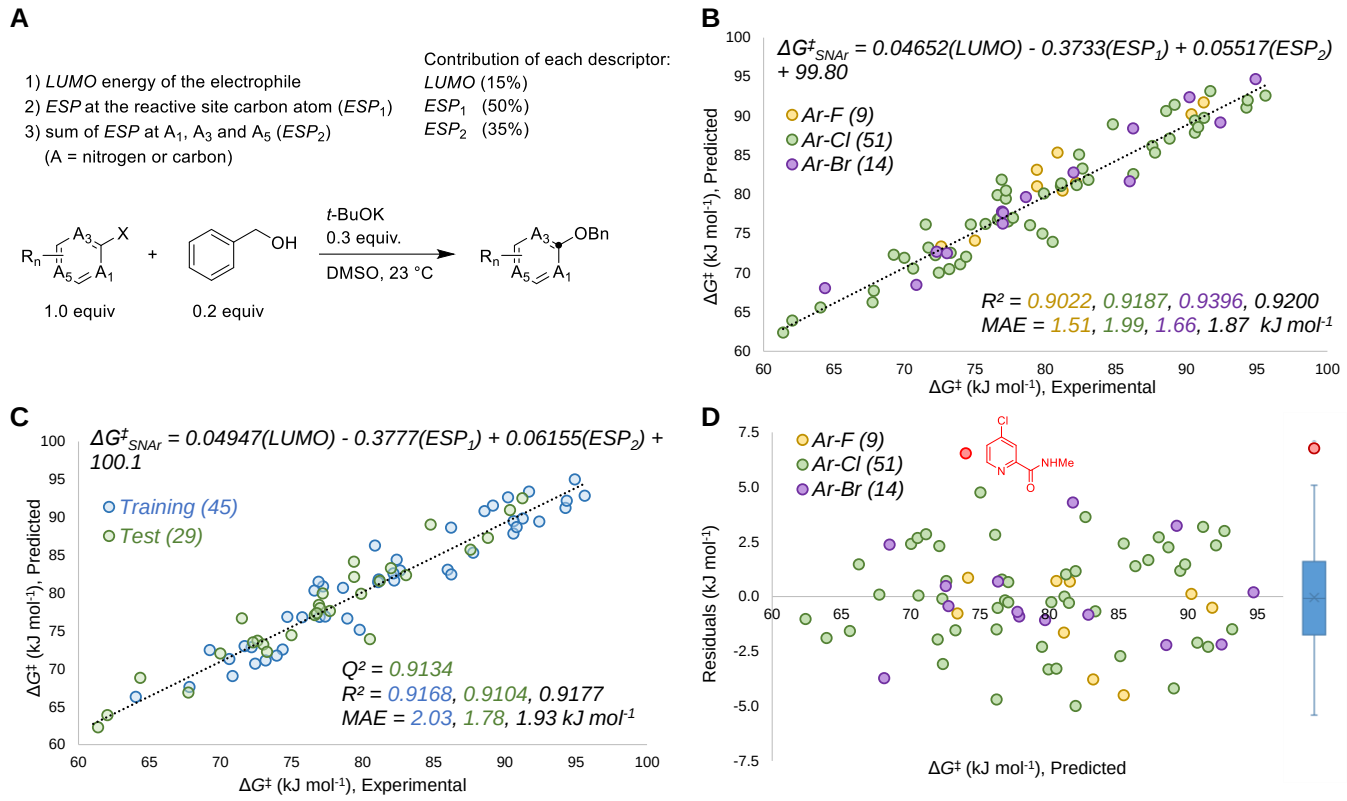

Fig. 3. Quantitative model generation and performance. A) Molecular descriptors used in multivariate regression analysis, with percent contribution determined by min/max normalization; Bn = benzyl. B) All data linear regression analysis for experimental versus predicted $\Delta G^{\ddagger}_{SNAr}$ with accompanying statistics (MAE = mean absolute error); linear correlation uses non-normalized descriptors. C) One of five 60/40 training/test validations, with accompanying statistics. D) Predicted versus residuals plot for the 74 data points, with accompanying box plot (right); one outlier is identified (|R| > 5 kJ mol^{-1}, red point with accompanying structure).

The selection of these specific molecular descriptors was guided by the mechanistic features of nucleophilic aromatic substitution, as well as our previous work on a multivariate model for oxidative addition with (hetero)aryl halides.[77] We also carried out an iterative refinement of the included descriptors based on our experimental observations and model performance (Table S5). The following discussion provides more detail on creation and refinement of the model and its mechanistic basis.

A classic approach to describing nucleophile/electrophile reactivity involves frontier molecular orbital (FMO) theory.[84,85] At a basic level, a lower *LUMO* energy for the electrophile leads to smaller *HOMO-LUMO* gap between nucleophile and electrophile. This results in a lower energy transition state, and therefore a faster reaction. On the other hand, this simple connection between electrophilicity and *LUMO* energy is not necessarily valid for every system: in one recent example, Zipse, Ofial, and Mayr have demonstrated poor correlation between *LUMO* energy and electrophilicity for a series of Michael acceptors.[86] This is attributed to substituent effects that increase π-conjugation (lowering *LUMO* energy), but decrease electrophilicity. Nevertheless, we included *LUMO* energies as a potential molecular descriptor for SNAr reactivity. We obtained these values by calculating each substrate's electron affinity (EA), and applying the DFT-Koopmans's theorem approximation that the *LUMO* energy is the negative of the EA.[87,88] We also investigated an operationally simpler approach to calculating *LUMO* descriptors using Entos Envision,[89] an open online interactive platform for molecular simulation and visualization that performs rapid semi-empirical calculations using GFN1-xTB.[90] We compared the *LUMO* energies computed in Envision with those from EA calculations at the B3LYP level, revealing a strong linear correlation ($R^2 = 0.90$, Fig. S25) and a nearly identical linear regression model (Fig. S26) to that shown in Fig. 3B. While we retained the DFT-derived *LUMO* energies for our subsequent validation and external predictions (*vide infra*), the values obtained from Envision could certainly be a rapid and easy to implement alternative.

To account for substituent effects beyond those on FMO energies, we used average molecular *ESP* at individual aromatic ring atoms as a local descriptor.[79–82] The extent of electron deficiency at the reactive carbon is a key factor in determining SNAr rates, and the corresponding *ESP* is a quantitative descriptor of this molecular feature. Previously, we observed excellent correlation between *ESP*-based descriptors and rates of Ar–X oxidative addition to Pd(0),[77] which shares mechanistic aspects with SNAr reactivity.[91] All *ESP* calculations were performed using the freely available Multiwfn application (version 3.7).[92,93]

We initially constructed a bivariate linear model using just two descriptors: *LUMO* energy and ESP_1 (at the carbon undergoing substitution) (Fig. 4A). This model gives good predictions for halogenated pyridines and quinolines; however, it significantly underestimates the reactivity of halogenated pyrimidines, and overestimates the reactivity of several non-heterocyclic haloarenes. The nature of these outliers led us to consider the electronic structure of the Meisenheimer intermediate and SNAr transition state more generally. During substitution, the excess

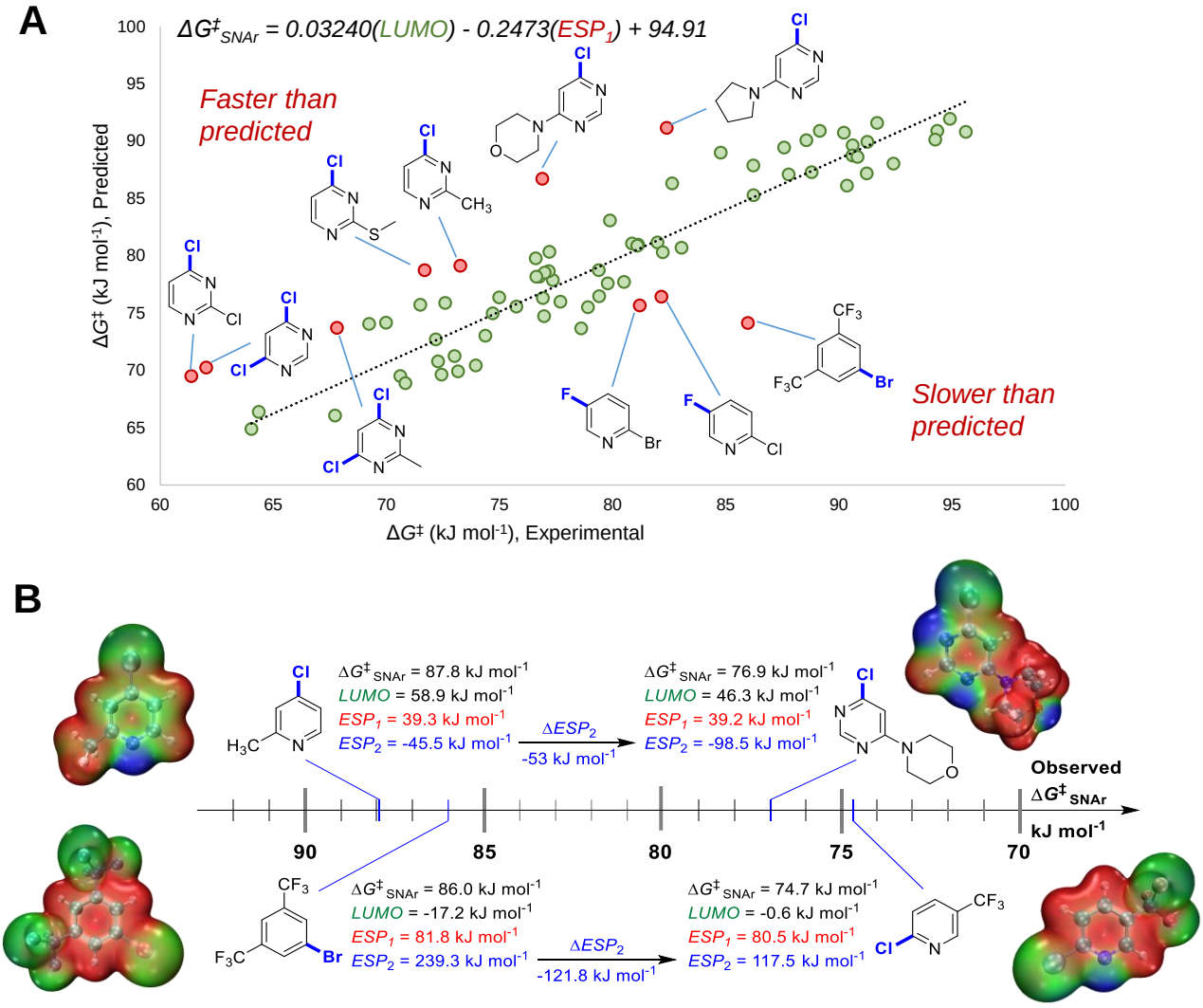

Fig. 4. Importance of ESP_2 descriptor in predicting $\Delta G^{\ddagger}_{SNAr}$ for multiple substrate classes. A) Bivariate model incorporating only *LUMO* and ESP_1 descriptors, with two sets of outliers highlighted in red. B) Comparison of substrate pairs with very similar *LUMO* and ESP_1 values but significantly different $\Delta G^{\ddagger}_{SNAr}$ values, revealing the importance of ESP_2 in differentiating reactivity. *ESP* maps for each substrate structure are shown, with colour gradient indicating local *ESP* (red = maximum positive; green = 0; blue = maximum negative).

negative charge in the intermediate/TS‡ is distributed *via* resonance to the *ortho* and *para* positions relative to the reactive site; the degree to which these atoms can stabilize this negative charge should therefore affect the reaction rate. Thus, we included the ESP_2 descriptor to account for these additional electronic effects, giving the superior model shown in Fig. 3B.

To highlight the importance of ESP_2 in making accurate predictions for multiple electrophile classes, we examined the two largest outliers from the bivariate model on either side of the distribution. We paired these two outliers with halopyridines that have very similar ESP_1 values, but significantly different observed $\Delta G^{\ddagger}_{SNAr}$ (Fig. 4B). In the first case, the faster than predicted outlier 4-chloro-6-morpholinopyridine has very similar *LUMO* energies and nearly identical ESP_1 values to 4-chloro-2-methylpyridine; however, these two electrophiles have a $\Delta\Delta G^{\ddagger}_{SNAr}$ = 10.9 kJ mol^{-1} (~100 fold rate difference at 298 K). These two electrophiles have strikingly different ESP_2 characteristics, with the pyrimidine exhibiting a substantially larger negative value due to the additional nitrogen in the ring. The same situation is observed for the slower than predicted outlier 1-bromo-3,5-bis(trifluoromethyl)benzene and 2-chloro-5-(trifluoromethyl)-pyridine ($\Delta\Delta G^{\ddagger}_{SNAr}$ = 11.3 kJ mol^{-1}): both substrates have nearly identical *LUMO* and ESP_1 descriptor values, but a more than 120 kJ mol^{-1} difference in ESP_2.

Site selectivity in multihalogenated heterocycles.

One of the most powerful applications of quantitative models in synthesis is to predict selectivity for one product over another. Many prior efforts in SNAr reactivity prediction focused on exactly this problem, developing qualitative and quantitative models for site selectivity involving multihalogenated electrophiles.[21,63–71,74,75,94] Within our 74-member substrate training library are several electrophiles with multiple reactive positions. The reactivity of these substrates provides an opportunity to test the model's applicability for quantitative selectivity predictions, despite not being explicitly trained for this purpose. Importantly, the major contributors to the model (ESP_1 and ESP_2) are local descriptors, which is key to enabling differential predictions for each reactive site.[95]

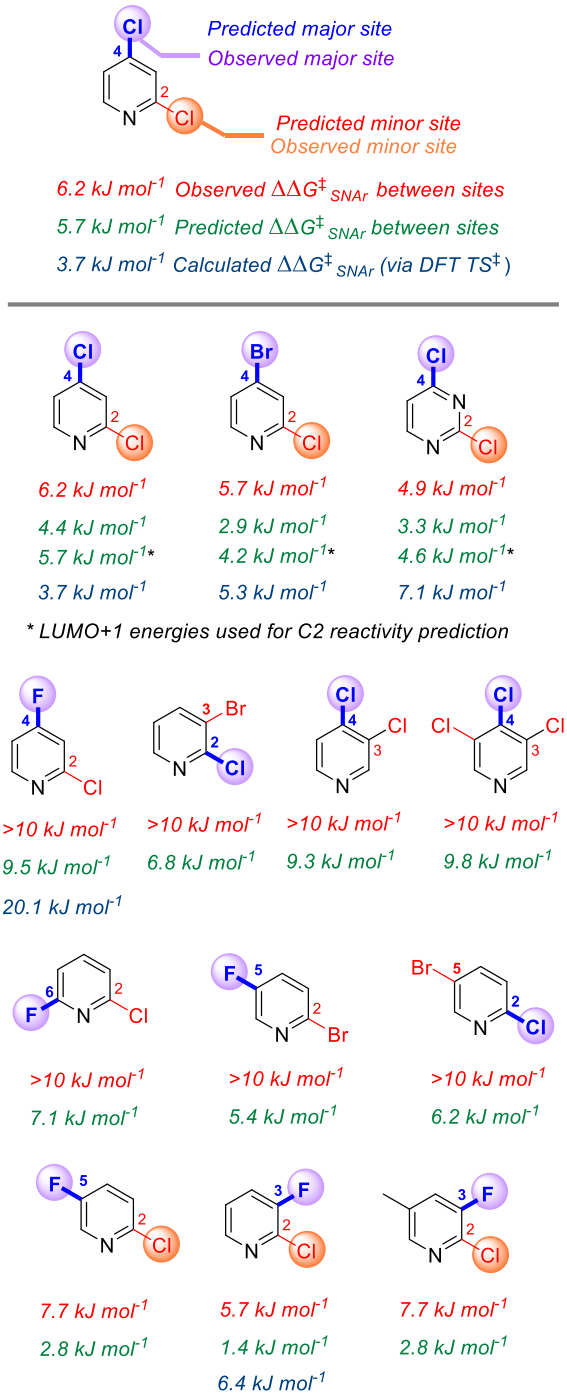

Fig. 5 Site selectivity in multihalogenated heterocycles that are part of the training set.

For the 13 multihalogenated substrates in our library, we determined the experimental site selectivity and compared the resulting $\Delta\Delta G^{\ddagger}_{SNAr}$ to that predicted by our descriptor-based model. We also calculated $\Delta\Delta G^{\ddagger}_{SNAr}$ values for 5 of the substrates from DFT analysis of the corresponding transition states (Fig. 5). In every case, using the three-descriptor model from Fig. 3B to independently predict $\Delta G^{\ddagger}_{SNAr}$ for each site correctly identifies the most reactive position, with reasonable quantitative accuracy that is comparable to that obtained *via* transition state analysis; however, the model-predicted $\Delta\Delta G^{\ddagger}_{SNAr}$ between sites does appear to be systematically low (i.e. selectivity is consistently underestimated).

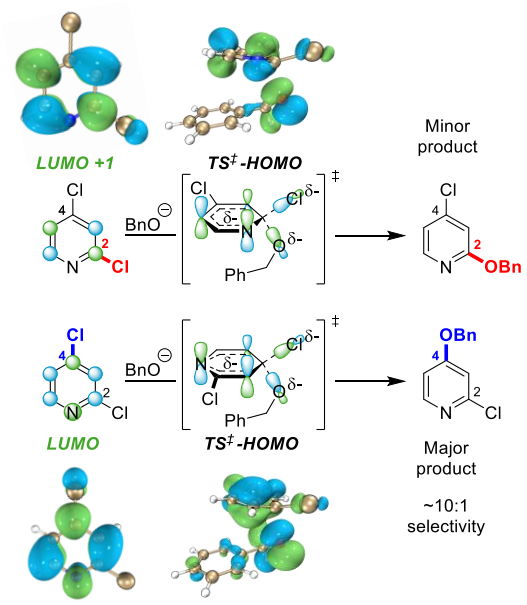

Fig. 6 FMO analysis of S_NAr selectivity with 2,4-dichloropyridine, revealing orbital symmetry effects in the substrate (*LUMO* versus *LUMO*+1) and transition states (*HOMO* contributions from *ortho* and *para* sites); Bn = benzyl.

To identify possible reasons for this systematic underestimation, we considered that our global *LUMO* energy descriptor may not be optimal in these cases, and chose the first three substrates from Fig. 5 for further investigation. To assess the FMOs involved in these specific regioselective S_NAr reactions, we examined the symmetries of the *LUMO* and *LUMO*+1 orbitals of the substrates, and calculated the structures and energies of the S_NAr transition states (Fig. 6 and Fig. S27-S36). In each case, we could not locate a Meisenheimer-type intermediate along the reaction coordinate when using an implicit solvent model (DMSO), but could locate transition states consistent with concerted S_NAr reactions.[29,53,54,96] As shown for 2,4-dichloropyridine in Fig. 6, the relevant electrophile FMO for attack at C4 is the *LUMO*, whereas for attack at C2 it is the *LUMO*+1; this is evident from the *LUMO*/*LUMO*+1 symmetries of the substrate, and the *HOMO* symmetries of two transition states. Taking the calculated *LUMO*/*LUMO*+1 gap into account when applying the model from Fig. 3B for C4 versus C2 predictions of the first three substrates does give increased accuracy, with errors of 0.3-1.5 kJ mol^{-1} for $\Delta\Delta G^{\ddagger}_{SNAr}$.

External case study #1: S_NAr rate correlations.

With our three descriptor model performance validated against internal data, we sought to assess its performance and generality when applied to new predictions beyond the training set. To challenge the scope of applicability to S_NAr reactions with different solvents and/or nucleophile classes, we first examined several correlations between predicted $\Delta G^{\ddagger}_{SNAr}$ values from the model and three sets of experimental $\Delta G^{\ddagger}_{SNAr}$ values from the literature (Fig. 7).[56,97,98] In these experimental data sets, a variety of (hetero)aromatic halides (F, Cl, and Br as leaving groups) are reacted with either alkoxide (Fig. 7A) or amine (Fig. 7B and 7C) nucleophiles. While the absolute $\Delta G^{\ddagger}_{SNAr}$ values from the prediction model are specific to the reaction conditions of the training set, we do obtain good to excellent correlation between the

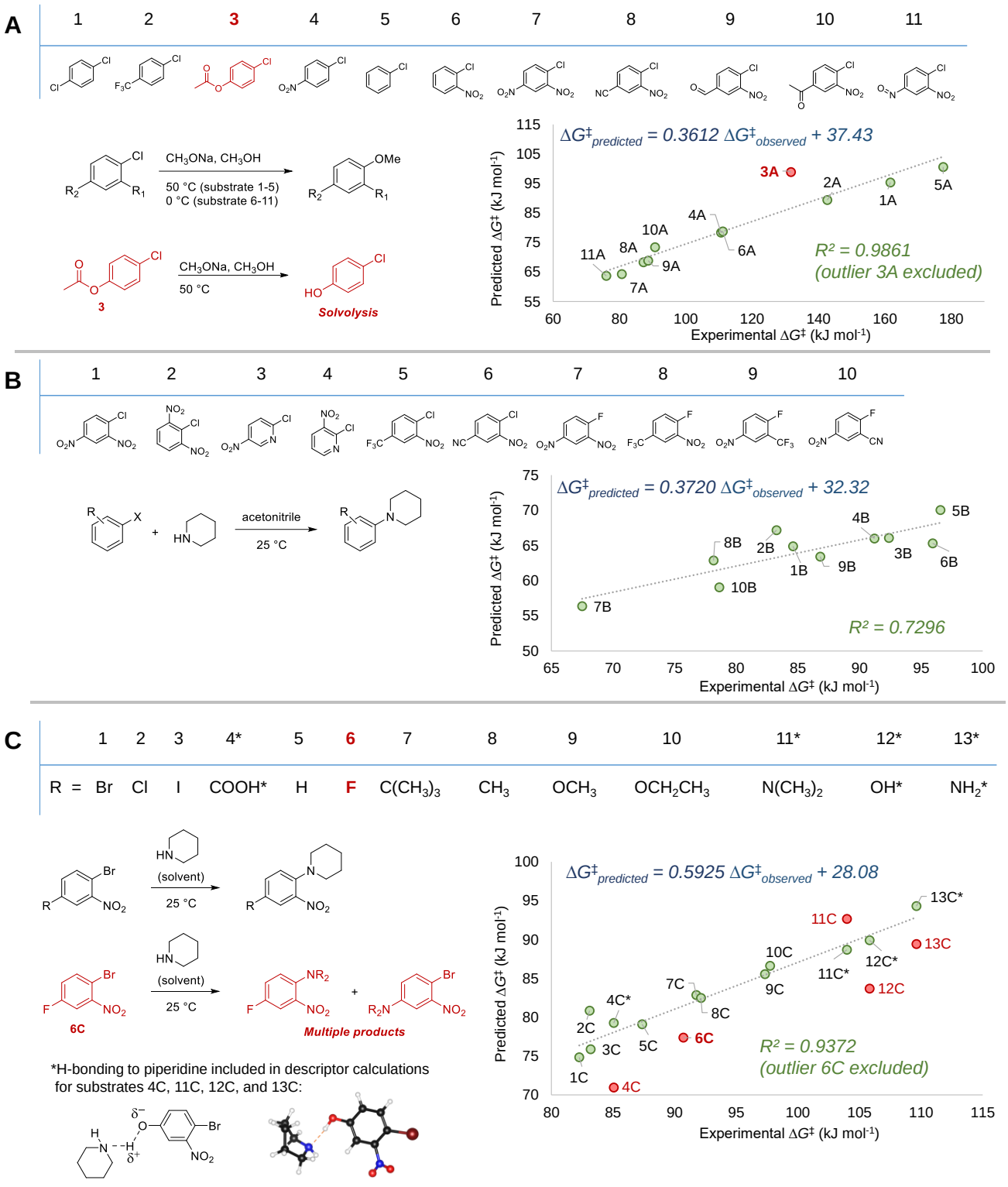

Fig. 7 Model validation through assessing correlations between experimental $\Delta G^{\ddagger}$ values and predicted $\Delta G^{\ddagger}_{SNAr}$ for three external data sets. Outliers are highlighted in red.

predicted $\Delta G^{\ddagger}_{SNAr}$ and experimental $\Delta G^{\ddagger}$ values (R^2 = 0.72-0.99). This is remarkable considering only two of the 34 electrophiles from these data sets are included in our training data (compounds 3B and 4B in Fig. 7B), and these reactions are conducted with different nucleophiles, solvents, and temperatures.

Notably, we are able to account for solvation effects on electrophile reactivity during descriptor generation. In data set C (Fig. 7C), there are several substrates containing acidic or basic functional groups where the initial correlation between experimental and predicted reactivity is poor (Fig. 7C, substrates 4C, 11-13C, red points). Given that these functional groups will hydrogen-bond with the piperidine solvent, significantly altering the electronics of the substrate, we included one explicit solvent molecule and recalculated the *ESP* descriptors for these four

electrophiles.[75] Using these revised *ESP* values, we obtain excellent linear correlation across the entire substrate set.

In addition to the success in applying the *ESP*/*LUMO* model beyond the training set, and in identifying solvation effects on reactivity, we can also identify potential experimental outliers. For example, the data set in Fig. 7A exhibits excellent correlation between experimental and predicted $\Delta G^{\ddagger}_{SNAr}$ with the exception of substrate **3A**. However, that substrate almost certainly will undergo solvolysis of the –OAc group rather than SNAr. Performing an experimental check on the reaction between **3A** and NaOMe/MeOH at 50 °C confirms this, with 4-chlorophenol as the only observed product (Fig. S37). A second example is identified in Fig. 7C, where there is also one significant outlier (**6C**). In this case, **6C** has two potentially reactive positions (Ar–Br and Ar–F). We have experimentally confirmed that reacting **6C** with piperidine leads to a mixture of the two SNAr products, in a 1.5:1 ratio, slightly favouring Ar–Br substitution (Fig. S38).

External case study #2: site selectivity predictions.

To further examine the potential applicability of our *ESP*/*LUMO* model beyond the training set, we assessed 63 external examples of site selectivity in SNAr reactions under a variety of conditions. We first applied predictions to three data sets previously used as a testing ground for site selectivity predictions using other approaches (Fig. 8-10).[69–71,75] These data sets also contain experimentally-determined rates, providing an additional opportunity to test the model's performance.

The first data set involves 7 multiply fluorinated arenes undergoing substitution with ammonia, where 5 substrates have potential for regioisomer formation (Fig. 8).[99] In each case, the predicted major site based on the *ESP*/*LUMO* model matches the experimental site. Furthermore, the predicted $\Delta G^{\ddagger}_{SNAr}$ values correlate well with the experimental ln (*k*) values for these 5 substrates (R^2 = 0.95). Notably, ln (*k*) for substrates **8b** and **8d** do not correlate; this exact situation was noted by Stenlid and Brinck, who also observed these two substrates as significant outliers when correlating ln (*k*) with the local electron attachment energy.[75] While these authors attributed this discrepancy between prediction and experiment to steric effects, there may be a different underlying reason considering the small size of both the nucleophile (ammonia) and the cyano group in **8d**.

The second data set also involves multiply fluorinated arenes, this time undergoing SNAr with the methoxide anion as the nucleophile in methanol solvent (Fig. 9).[100] Across these 10 substrates, 5 have the potential to form regioisomers. In each of these cases, the *ESP*/*LUMO* model correctly predicts the major site of reaction. For substrate **9d**, the predicted second most reactive site is incorrect (C2) based on experimental observation (C3); however, for **9e** the predicted reactivity order from first to third site is correct. While we again observe an underestimation of selectivity based on predicted $\Delta G^{\ddagger}_{SNAr}$ values, we do observe excellent linear correlation with experimental ln (*k*) across the entire substrate set. This is notable in the context of Stenlid and Brinck's prior work with local electron attachment energy, where the experimental ln (*k*) for **9g-j** does not correlate with that descriptor. Here, the *ESP*/*LUMO* model correctly predicts that these four substrates should have similar SNAr rates (within a factor of 10 of each other).

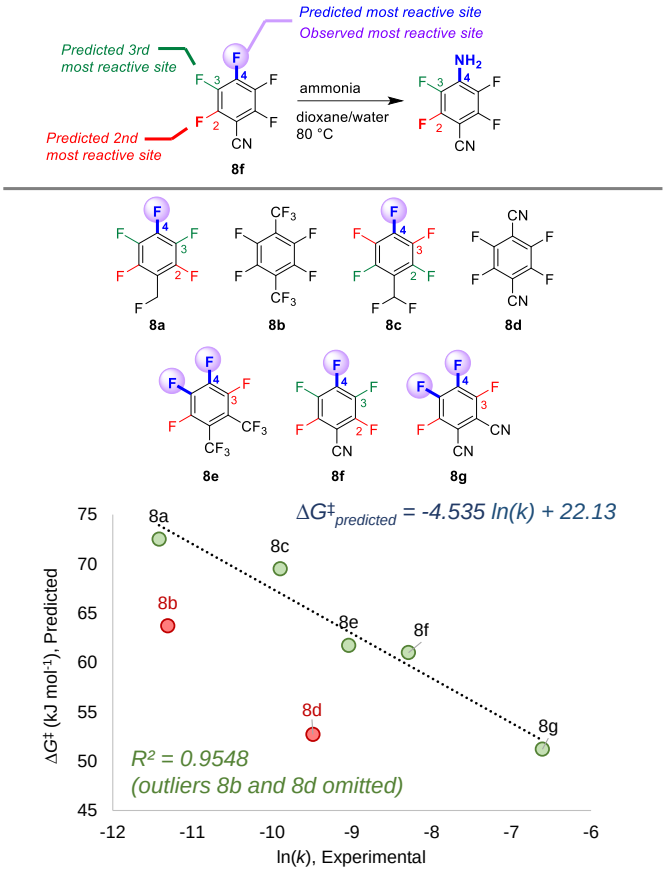

Fig. 8 Site selectivity predictions and rate correlation for SNAr between fluorinated arenes and ammonia. Experimental data from ref 99.

The third data set contains 18 multiply fluorinated nitrogen heterocycles undergoing SNAr with ammonia, with 15 examples where regioisomers can be formed (Fig. 10).[99,101,102] In every case *ESP*/*LUMO* model correctly predicts the major site of reaction, and in all but one case (**10l**) it also predicts the second site of reaction. The quantitative selectivity predictions are also much closer to the experimental values within this data set. We again observe excellent linear correlation between experimental ln (*k*) and predicted $\Delta G^{\ddagger}_{SNAr}$. Note that substrate **10r**, which has a rate "too fast … to measure",[102] is estimated to have an ~10^5-fold larger rate constant than **10d**; this estimated data point is not included in the linear correlation.

Finally, to challenge the qualitative accuracy of the model, we applied it toward a series of more complex SNAr examples with a wider variety of nucleophiles (Fig. 11). Sets A-D were previously collated and categorized by Brinck, Svensson, and coworkers and categorized depending on the nature of the nucleophile/electrophile pairing.[70,102–120] Using only the structure of the electrophile, our *ESP*/*LUMO* model is able to correctly predict the major site of reaction in 26 of the 32 cases. Within sets A and C – (hetero)aryl halides reacting with anionic nucleophiles – the two incorrect predictions are for relatively non-polar fluorinated arenes. For sets B and D, which employ neutral nucleophiles, the incorrect examples all involve secondary amine nucleophiles. In these cases, steric effects appear to play a significant role in overriding the electronic nature of the electrophile; for example, pentachloropyridine reacts preferentially at C4 (as predicted)

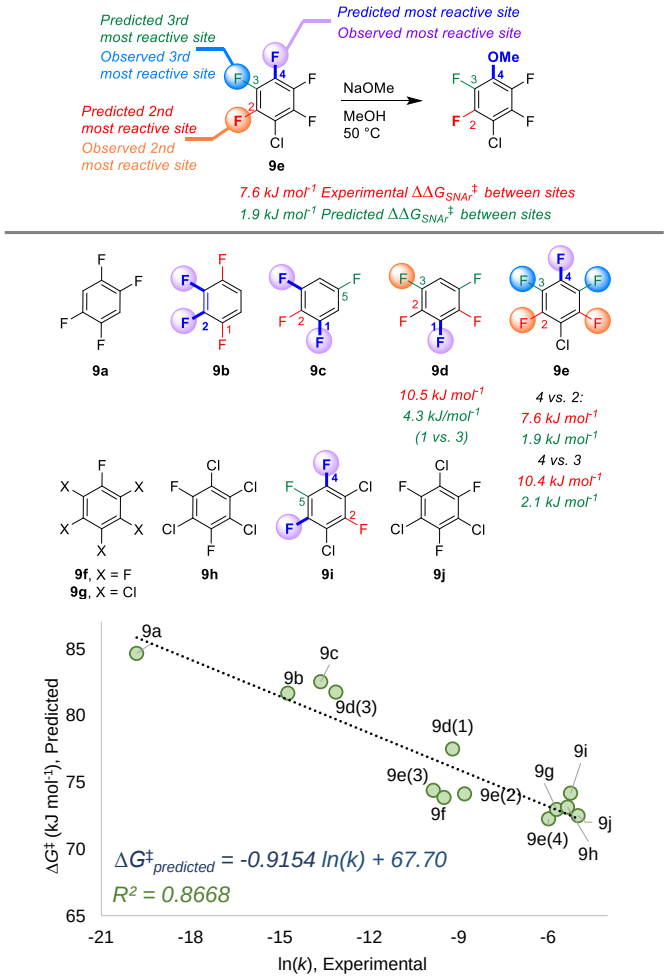

Fig. 9 Site selectivity predictions and rate correlation for S_NAr between fluorinated arenes and methoxide. Experimental data from ref 100.

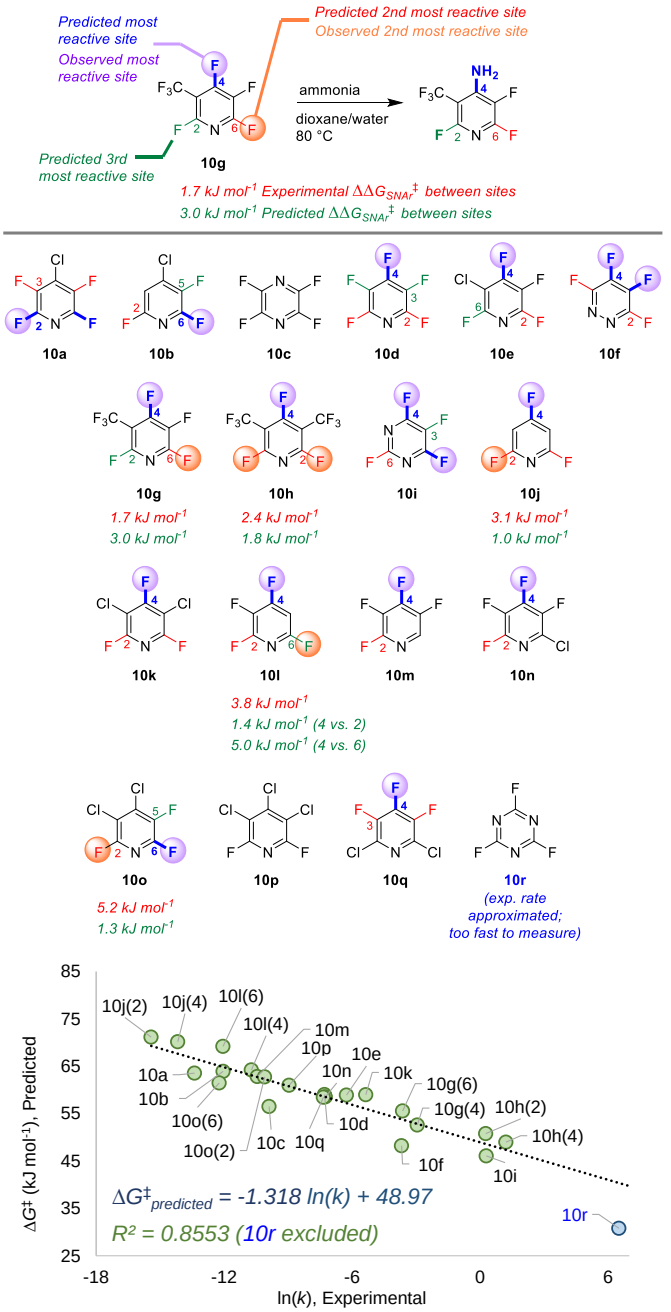

Fig. 10 Site selectivity predictions and rate correlation for S_NAr between fluorinated heterocycles and ammonia. Experimental data from refs 99, 101, and 102.

with alkoxide or ammonia nucleophiles, but switches to C2 selectivity with diethylamine. We also applied predictions to 6 mixed halide electrophiles reacting with a variety of nucleophiles in set E (Fig. 11E), drawn from examples in medicinal/agrochemical discovery.[121–126] The model is able to correctly identify the major site of reactivity for each example, except for a case where the predicted site is at an Ar–F, and the observed reactivity is at a 2-Cl-pyridine site.

External case study #3: complex molecule synthetic planning.

As a test of the *ESP*/*LUMO* model's potential utility in real-world synthetic planning, we sought to validate its predictions against S_NAr reactions used to prepare clinical candidate active pharmaceutical ingredients (APIs). These include recent reports on branebrutinib,[127] an EGFR T790 M inhibitor,[128] a Nav1.7 inhibitor,[129] a tyrosine kinase inhibitor,[130] an SRI/5-HT_{2A} antagonist,[131] an RoRγ inverse agonist,[132] and merestinib[133] (Fig. 12). The first four examples concern site selective S_NAr to generate a variety of targets from structurally complex substrates. In each of these cases, the *ESP*/*LUMO* model is able to predict the correct reactive site. Thus, applying these predictions during synthetic design would help pharmaceutical process chemists to proceed with confidence that selective substitution is feasible. In fact, the chemists at Pfizer used an internal prediction tool (based on Fukui indices) to help guide their synthetic planning toward the EGFR T790 M inhibitor (second example in Fig. 12).[128]

A particularly powerful aspect of *in silico* reactivity predictions is the ability to evaluate multiple options in substrate design before committing experimental resource. We have examined three examples where the substitution pattern of the S_NAr electrophile affects the site selectivity or reactivity. In the first case, synthesis of the target SRI/5-HT_{2A} antagonist requires a site selective S_NAr to install an aryl ether *ortho* to a carbonyl functionality.[131] This was initially performed using an aldehyde moiety; however, the relatively poor site selectivity meant column chromatography was required to purify the intermediate. Further

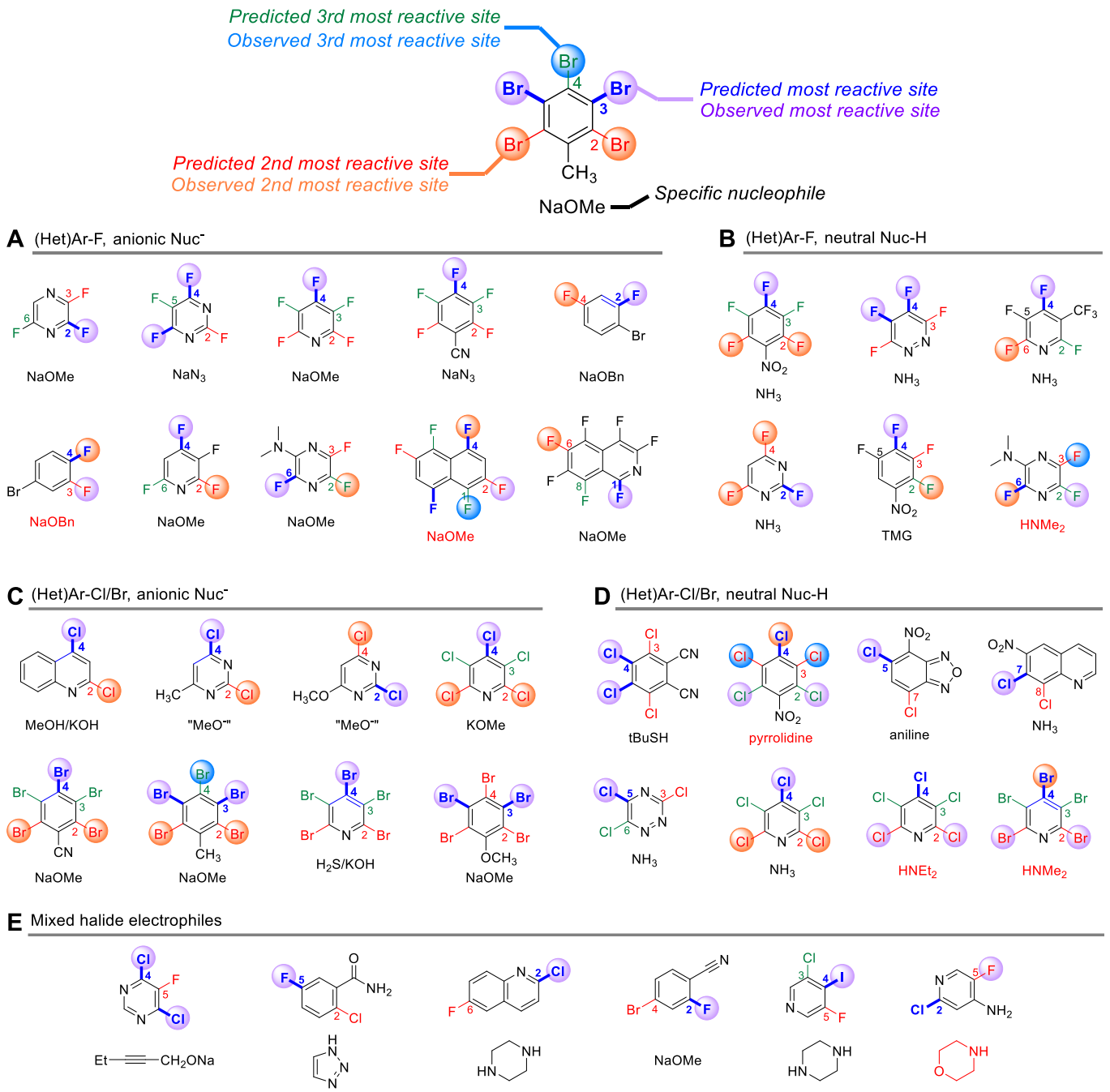

Fig. 11 Qualitative site selectivity predictions for combinations of (hetero)aryl halides with anionic (A and C) and neutral (B and D) nucleophiles, and for mixed halide aromatics (E).

process developments identified an *N*-methylamide as a more selective alternative that retained key functionality for progressing to the target API. This improved selectivity is predicted by the *ESP*/*LUMO* model. A second case involves choice of either an Ar–F or Ar–Cl electrophile for S_NAr with an alkoxide nucleophile.[132] Experimental evaluation of each revealed that both substrates are viable, with the Ar–Cl version requiring slightly higher reaction temperature than the Ar–F analogue. The *ESP*/*LUMO* model predicts that the F for Cl switch would result in a relatively modest reactivity decrease, indicating both should be suitable substrates.

The final example concerns an intramolecular S_NAr to generate an indazole *en route* to merestinib.[133] The final API contains a methoxy group *para* to the indazole nitrogen; however, attempts to perform the intramolecular S_NAr with this strong electron donating group *para* to the substitution site were not successful. Instead, the researchers installed a nitro group to enable the S_NAr to proceed, but which would require multiple functional group interconversions. The substantial difference in reactivity between –OMe and $–NO_2$ derivatives is conceptually obvious (and borne out by the *ESP*/*LUMO* model); however, the orders-of-magnitude difference in predicted rate between the two means that the more desirable –OMe substrate could be ruled out

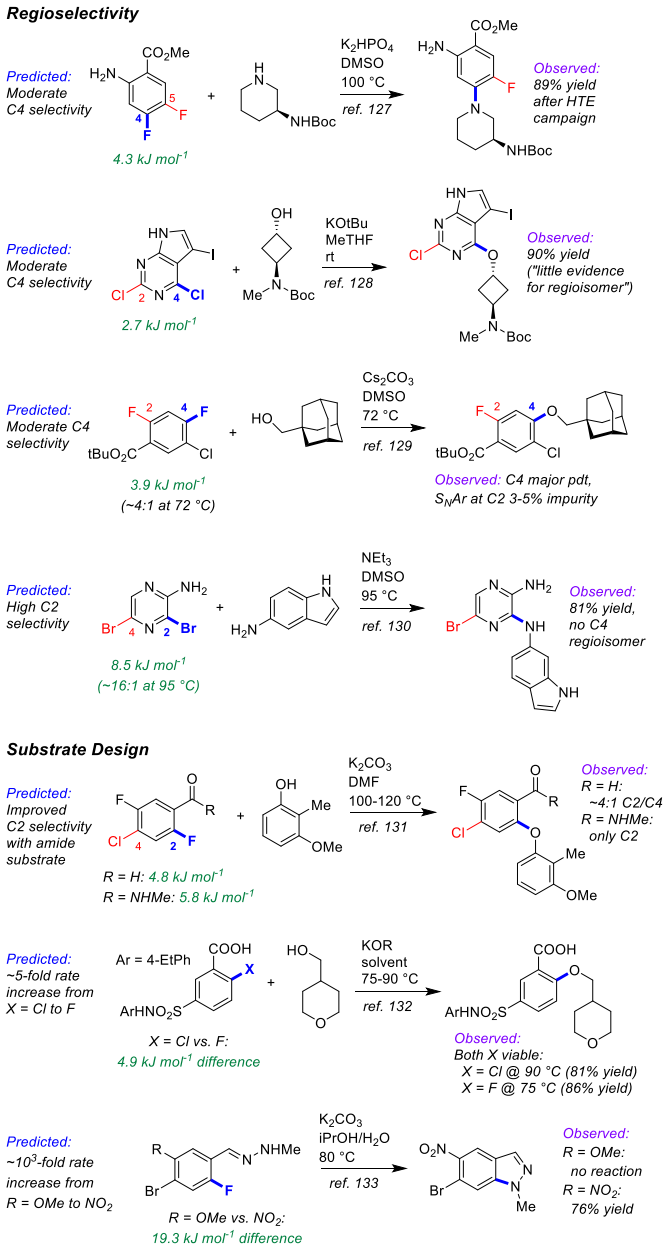

Fig. 12 Example applications of S_NAr predictions to route development for investigational API synthesis, including regioselectivity for specific substrates, and comparison of potential substrate regioselectivity/reactivity.

earlier on in synthetic development. Furthermore, additional hypothetical substrates that retain the required oxygen (such as a sulfonate) could be evaluated using the prediction model (the –OMs derivative has a predicted $\Delta G^{\ddagger}_{SNAr}$ halfway between the $-NO_2$ and –OMe derivatives).

Conclusions

We have demonstrated an effective bottom-up approach to developing a quantitative structure-reactivity model for nucleophilic aromatic substitution reactions. By curating a diverse library of (hetero)aromatic electrophiles, and determining their corresponding S_NAr reaction rates through a series of competition experiments, we rapidly assembled a reliable and diverse data set as an experimental foundation. Pairing this set of reactivity data with simple ground state molecular descriptors – *LUMO* energy and molecular electrostatic potentials – results in a robust multivariate linear correlation between rate and molecular structure.

Importantly, even though the model was trained using only one set of reaction conditions, it is suitable for making correlations and predictions about S_NAr reactivity for a wide variety of nucleophiles, solvents, and temperatures. These include a >90% success rate in predicting the major reaction site for multihalogenated arenes (>80 cases), and examples where substrate design for active pharmaceutical ingredient synthesis can be informed by predicted reactivity. Thus, this simple and easy-to-apply model can generate rapid and accurate predictions with the potential to improve and augment computer-assisted synthesis design, and is complementary to ongoing efforts in machine learning approaches to reactivity prediction. Further work to incorporate nucleophile descriptors, including steric effects, is currently underway in our laboratories.

Author Contributions

J. Lu: Conceptualization, Methodology, Investigation, Validation, Formal analysis, Writing. *I. Paci*: Conceptualization, Methodology, Formal analysis, Supervision, Writing. *D. C. Leitch*: Conceptualization, Methodology, Formal analysis, Supervision, Writing.

Conflicts of interest

There are no conflicts to declare.

Acknowledgements

We acknowledge and respect the Lekwungen peoples on whose traditional territory the University of Victoria (UVic) stands, and the Songhees, Esquimalt and WSÁNEĆ peoples whose historical relationships with the land continue to this day. We also acknowledge funding from the New Frontiers in Research Fund – Exploration (DCL) and NSERC Discovery Grant program (IP and DCL). Supercomputing resources at Westgrid and Compute Canada were integral to this work.

References

(1) Larock, R. C. *Comprehensive Organic Transformations: A Guide to Functional Group Preparations*; Wiley, 2018.

(2) Corey, E. J.; Cheng, X.-M. *The Logic of Chemical Synthesis*; John Wiley & Sons: Nashville, TN, 1995.

(3) *The Art of Writing Reasonable Organic Reaction Mechanisms*; Grossman, R. B., Ed.; Springer: New York, NY, 2003.

(4) Mayr, H.; Patz, M. Scales of Nucleophilicity and Electrophilicity: A System for Ordering Polar Organic and Organometallic Reactions. *Angew. Chem. Int. Ed. Engl.* **1994**, *33*, 938–957.

(5) Mayr, H.; Kempf, B.; Ofial, A. R. π-Nucleophilicity in Carbon−Carbon Bond-Forming Reactions. *Acc. Chem. Res.* **2003**, *36*, 66–77.

(6) Mayr, H.; Ofial, A. R. Kinetics of Electrophile-Nucleophile Combinations: A General Approach to Polar Organic Reactivity. *Pure Appl. Chem.* **2005**, *77*, 1807–1821.

(7) Mayr, H.; Ofial, A. R. Do General Nucleophilicity Scales Exist? *J. Phys. Org. Chem.* **2008**, *21*, 584–595.
(8) Mayr, H.; Ofial, A. R. Philicities, Fugalities, and Equilibrium Constants. *Acc. Chem. Res.* **2016**, *49*, 952–965.
(9) Mayr, H.; Ofial, A. R. Philicity, Fugality, and Equilibrium Constants: When Do Rate-Equilibrium Relationships Break Down? *Pure Appl. Chem.* **2017**, *89*, 729–744.
(10) Sigman, M. S.; Harper, K. C.; Bess, E. N.; Milo, A. The Development of Multidimensional Analysis Tools for Asymmetric Catalysis and Beyond. *Acc. Chem. Res.* **2016**, *49*, 1292–1301.
(11) Niemeyer, Z. L.; Milo, A.; Hickey, D. P.; Sigman, M. S. Parameterization of Phosphine Ligands Reveals Mechanistic Pathways and Predicts Reaction Outcomes. *Nat. Chem.* **2016**, *8*, 610–617.
(12) Wu, K.; Doyle, A. G. Parameterization of Phosphine Ligands Demonstrates Enhancement of Nickel Catalysis via Remote Steric Effects. *Nat. Chem.* **2017**, *9*, 779–784.
(13) Ahneman, D. T.; Estrada, J. G.; Lin, S.; Dreher, S. D.; Doyle, A. G. Predicting Reaction Performance in C–N Cross-Coupling Using Machine Learning. *Science* **2018**, *360*, 186–190.
(14) Maryasin, B.; Marquetand, P.; Maulide, N. Machine Learning for Organic Synthesis: Are Robots Replacing Chemists? *Angew. Chem. Int. Ed.* **2018**, *57*, 6978–6980.
(15) Engkvist, O.; Norrby, P.-O.; Selmi, N.; Lam, Y.; Peng, Z.; Sherer, E. C.; Amberg, W.; Erhard, T.; Smyth, L. A. Computational Prediction of Chemical Reactions: Current Status and Outlook. *Drug Discov. Today* **2018**, *23*, 1203–1218.
(16) Zahrt, A. F.; Henle, J. J.; Rose, B. T.; Wang, Y.; Darrow, W. T.; Denmark, S. E. Prediction of Higher-Selectivity Catalysts by Computer-Driven Workflow and Machine Learning. *Science* **2019**, *363*, eaau5631.
(17) Toyao, T.; Maeno, Z.; Takakusagi, S.; Kamachi, T.; Takigawa, I.; Shimizu, K. Machine Learning for Catalysis Informatics: Recent Applications and Prospects. *ACS Catal.* **2020**, *10*, 2260–2297.
(18) Muratov, E. N.; Bajorath, J.; Sheridan, R. P.; Tetko, I. V.; Filimonov, D.; Poroikov, V.; Oprea, T. I.; Baskin, I. I.; Varnek, A.; Roitberg, A.; Isayev, O.; Curtalolo, S.; Fourches, D.; Cohen, Y.; Aspuru-Guzik, A.; Winkler, D. A.; Agrafiotis, D.; Cherkasov, A.; Tropsha, A. QSAR without Borders. *Chem. Soc. Rev.* **2020**, *49*, 3525–3564.
(19) Mahjour, B.; Shen, Y.; Cernak, T. Ultrahigh-Throughput Experimentation for Information-Rich Chemical Synthesis. *Acc. Chem. Res.* **2021**, *54*, 2337–2346.
(20) Gallegos, L. C.; Luchini, G.; St. John, P. C.; Kim, S.; Paton, R. S. Importance of Engineered and Learned Molecular Representations in Predicting Organic Reactivity, Selectivity, and Chemical Properties. *Acc. Chem. Res.* **2021**, *54*, 827–836.
(21) Jorner, K.; Brinck, T.; Norrby, P.-O.; Buttar, D. Machine Learning Meets Mechanistic Modelling for Accurate Prediction of Experimental Activation Energies. *Chem. Sci.* **2021**, *12*, 1163–1175.
(22) Orlandi, M.; Escudero-Casao, M.; Licini, G. Nucleophilicity Prediction via Multivariate Linear Regression Analysis. *J. Org. Chem.* **2021**, *86*, 3555–3564.
(23) Shen, Y.; Borowski, J. E.; Hardy, M. A.; Sarpong, R.; Doyle, A. G.; Cernak, T. Automation and Computer-Assisted Planning for Chemical Synthesis. *Nat. Rev. Methods Primer* **2021**, *1*, #23.
(24) Betinol, I. O.; Reid, J. P. A Predictive and Mechanistic Statistical Modelling Workflow for Improving Decision Making in Organic Synthesis and Catalysis. *Org. Biomol. Chem.* **2022**. DOI: 10.1039/D2OB00272H.
(25) Beker, W.; Roszak, R.; Wołos, A.; Angello, N. H.; Rathore, V.; Burke, M. D.; Grzybowski, B. A. Machine Learning May Sometimes Simply Capture Literature Popularity Trends: A Case Study of Heterocyclic Suzuki–Miyaura Coupling. *J. Am. Chem. Soc.* **2022**, *144*, 4819–4827.
(26) Meisenheimer, J. Ueber Reactionen Aromatischer Nitrokörper. *Justus Liebigs Ann. Chem.* **1902**, *323*, 205–246.
(27) Bunnett, J. F.; Zahler, R. E. Aromatic Nucleophilic Substitution Reactions. *Chem. Rev.* **1951**, *49*, 273–412.
(28) The S_NAr Reactions: Mechanistic Aspects. In *Modern Nucleophilic Aromatic Substitution*; Wiley-VCH Verlag: Weinheim, Germany, 2013; pp 1–94.
(29) Rohrbach, S.; Smith, A. J.; Pang, J. H.; Poole, D. L.; Tuttle, T.; Chiba, S.; Murphy, J. A. Concerted Nucleophilic Aromatic Substitution Reactions. *Angew. Chem. Int. Ed.* **2019**, *58*, 16368–16388.
(30) Evans, D. A.; Wood, M. R.; Trotter, B. W.; Richardson, T. I.; Barrow, J. C.; Katz, J. L. Total Syntheses of Vancomycin and Eremomycin Aglycons. *Angew. Chem. Int. Ed.* **1998**, *37*, 2700–2704.
(31) Evans, D. A.; Dinsmore, C. J.; Watson, P. S.; Wood, M. R.; Richardson, T. I.; Trotter, B. W.; Katz, J. L. Nonconventional Stereochemical Issues in the Design of the Synthesis of the Vancomycin Antibiotics: Challenges Imposed by Axial and Nonplanar Chiral Elements in the Heptapeptide Aglycons. *Angew. Chem. Int. Ed.* **1998**, *37*, 2704–2708.
(32) Nicolaou, K. C.; Natarajan, S.; Li, H.; Jain, N. F.; Hughes, R.; Solomon, M. E.; Ramanjulu, J. M.; Boddy, C. N. C.; Takayanagi, M. Total Synthesis of Vancomycin Aglycon—Part 1: Synthesis of Amino Acids 4–7 and Construction of the AB-COD Ring Skeleton. *Angew. Chem. Int. Ed.* **1998**, *37*, 2708–2714.
(33) Nicolaou, K. C.; Jain, N. F.; Natarajan, S.; Hughes, R.; Solomon, M. E.; Li, H.; Ramanjulu, J. M.; Takayanagi, M.; Koumbis, A. E.; Bando, T. Total Synthesis of Vancomycin Aglycon—Part 2: Synthesis of Amino Acids 1–3 and Construction of the AB-COD-DOE Ring Skeleton. *Angew. Chem. Int. Ed.* **1998**, *37*, 2714–2716.
(34) Nicolaou, K. C.; Takayanagi, M.; Jain, N. F.; Natarajan, S.; Koumbis, A. E.; Bando, T.; Ramanjulu, J. M. Total Synthesis of Vancomycin Aglycon—Part 3: Final Stages. *Angew. Chem. Int. Ed.* **1998**, *37*, 2717–2719.
(35) Zhang, A. J.; Burgess, K. Total Syntheses of Vancomycin. *Angew. Chem. Int. Ed.* **1999**, *38*, 634–636.
(36) Cheng, L.-J.; Xie, J.-H.; Chen, Y.; Wang, L.-X.; Zhou, Q.-L. Enantioselective Total Synthesis of (−)-Δ8-THC and (−)-Δ9-THC via Catalytic Asymmetric Hydrogenation and SNAr Cyclization. *Org. Lett.* **2013**, *15*, 764–767.
(37) Yamashita, K.; Kume, Y.; Ashibe, S.; Puspita, C. A. D.; Tanigawa, K.; Michihata, N.; Wakamori, S.; Ikeuchi, K.; Yamada, H. Total Synthesis of Mallotusinin. *Chem. Eur. J.* **2020**, *26* (69), 16408–16421.
(38) Bork, J. T.; Lee, J. W.; Chang, Y.-T. The Combinatorial Synthesis of Purine, Pyrimidine and Triazine-Based Libraries. *QSAR Comb. Sci.* **2004**, *23*, 245–260.
(39) Brown, D. G.; Boström, J. Analysis of Past and Present Synthetic Methodologies on Medicinal Chemistry: Where Have All the New Reactions Gone?: Miniperspective. *J. Med. Chem.* **2016**, *59*, 4443–4458.
(40) Preshlock, S.; Tredwell, M.; Gouverneur, V. 18F-Labeling of Arenes and Heteroarenes for Applications in Positron Emission Tomography. *Chem. Rev.* **2016**, *116*, 719–766.
(41) Neumann, C. N.; Ritter, T. Facile C–F Bond Formation through a Concerted Nucleophilic Aromatic Substitution Mediated by the PhenoFluor Reagent. *Acc. Chem. Res.* **2017**, *50*, 2822–2833.
(42) Boström, J.; Brown, D. G.; Young, R. J.; Keserü, G. M. Expanding the Medicinal Chemistry Synthetic Toolbox. *Nat. Rev. Drug Discov.* **2018**, *17*, 709–727.

(43) See, Y. Y.; Morales-Colón, M. T.; Bland, D. C.; Sanford, M. S. Development of SNAr Nucleophilic Fluorination: A Fruitful Academia-Industry Collaboration. *Acc. Chem. Res.* **2020**, *53*, 2372–2383.
(44) Baumann, M.; Baxendale, I. R. An Overview of the Synthetic Routes to the Best Selling Drugs Containing 6-Membered Heterocycles. *Beilstein J. Org. Chem.* **2013**, *9*, 2265–2319.
(45) Flick, A. C.; Leverett, C. A.; Ding, H. X.; McInturff, E.; Fink, S. J.; Helal, C. J.; DeForest, J. C.; Morse, P. D.; Mahapatra, S.; O'Donnell, C. J. Synthetic Approaches to New Drugs Approved during 2018. *J. Med. Chem.* **2020**, *63*, 10652–10704.
(46) Flick, A. C.; Leverett, C. A.; Ding, H. X.; McInturff, E.; Fink, S. J.; Mahapatra, S.; Carney, D. W.; Lindsey, E. A.; DeForest, J. C.; France, S. P.; Berritt, S.; Bigi-Botterill, S. V.; Gibson, T. S.; Liu, Y.; O'Donnell, C. J. Synthetic Approaches to the New Drugs Approved during 2019. *J. Med. Chem.* **2021**, *64*, 3604–3657.
(47) Jeanmart, S.; Edmunds, A. J. F.; Lamberth, C.; Pouliot, M. Synthetic Approaches to the 2010–2014 New Agrochemicals. *Bioorg. Med. Chem.* **2016**, *24*, 317–341.
(48) Jeanmart, S.; Edmunds, A. J. F.; Lamberth, C.; Pouliot, M.; Morris, J. A. Synthetic Approaches to the 2015–2018 New Agrochemicals. *Bioorg. Med. Chem.* **2021**, *39*, 116162.
(49) Vitaku, E.; Smith, D. T.; Njardarson, J. T. Analysis of the Structural Diversity, Substitution Patterns, and Frequency of Nitrogen Heterocycles among U.S. FDA Approved Pharmaceuticals. *J. Med. Chem.* **2014**, *57*, 10257–10274.
(50) Delost, M. D.; Smith, D. T.; Anderson, B. J.; Njardarson, J. T. From Oxiranes to Oligomers: Architectures of U.S. FDA Approved Pharmaceuticals Containing Oxygen Heterocycles. *J. Med. Chem.* **2018**, *61*, 10996–11020.
(51) Das, P.; Delost, M. D.; Qureshi, M. H.; Smith, D. T.; Njardarson, J. T. A Survey of the Structures of US FDA Approved Combination Drugs. *J. Med. Chem.* **2019**, *62*, 4265–4311.
(52) Terrier, F. Rate and Equilibrium Studies in Jackson-Meisenheimer Complexes. *Chem. Rev.* **1982**, *82*, 77–152.
(53) Neumann, C. N.; Hooker, J. M.; Ritter, T. Concerted Nucleophilic Aromatic Substitution with 19F− and 18F−. *Nature* **2016**, *534*, 369–373.
(54) Kwan, E. E.; Zeng, Y.; Besser, H. A.; Jacobsen, E. N. Concerted Nucleophilic Aromatic Substitutions. *Nat. Chem.* **2018**, *10*, 917–923.
(55) Hansch, Corwin.; Leo, A.; Taft, R. W. A Survey of Hammett Substituent Constants and Resonance and Field Parameters. *Chem. Rev.* **1991**, *91*, 165–195.
(56) Miller, J.; Kai-Yan, W. 655. The SN Mechanism in Aromatic Compounds. Part XXIX. Some Para-Substituted Chlorobenzenes. *J. Chem. Soc.* **1963**, 3492–3495.
(57) Fry, S. E.; Pienta, N. J. Effects of Molten Salts on Reactions. Nucleophilic Aromatic Substitution by Halide Ions in Molten Dodecyltributylphosphonium Salts. *J. Am. Chem. Soc.* **1985**, *107*, 6399–6400.
(58) Renfrew, A. H. M.; Taylor, J. A.; Whitmore, J. M. J.; Williams, A. A Single Transition State in Nucleophilic Aromatic Substitution: Reaction of Phenolate Ions with 2-(4-Nitrophenoxy)-4,6-Dimethoxy-1,3,5-Triazine in Aqueous Solution. *J. Chem. Soc. Perkin Trans. 2* **1993**, 1703–1704.
(59) Hunter, A.; Renfrew, M.; Rettura, D.; Taylor, J. A.; Whitmore, J. M. J.; Williams, A. Stepwise versus Concerted Mechanisms at Trigonal Carbon: Transfer of the 1,3,5-Triazinyl Group between Aryl Oxide Ions in Aqueous Solution. *J. Am. Chem. Soc.* **1995**, *117*, 5484–5491.
(60) Um, I.-H.; Im, L.-R.; Kang, J.-S.; Bursey, S. S.; Dust, J. M. Mechanistic Assessment of S_NAr Displacement of Halides from 1-Halo-2,4-Dinitrobenzenes by Selected Primary and Secondary Amines: Brønsted and Mayr Analyses. *J. Org. Chem.* **2012**, *77*, 9738–9746.
(61) ElGuesmi, N.; Berionni, G.; Asghar, B. H. Mechanistic Studies and Quantification of the Electrophilicity of Aromatic Triflones in σ-Complexation and SNAr Reactions. *J. Fluor. Chem.* **2014**, *160*, 41–47.
(62) Mahdhaoui, F.; Zaier, R.; Dhahri, N.; Ayachi, S.; Boubaker, T. SNAr Reactions of Substituted Pyridines with Secondary Amines in Aqueous Solution: Kinetic and Reactivity Indices in Density Functional Theory. *Int. J. Chem. Kinet.* **2019**, *51*, 249–257.
(63) Burdon, J. A Rationalization of Orientation and Reactivity in the Nucleophilic Replacement Reactions of Aromatic Polyhalo-Compounds. *Tetrahedron* **1965**, *21*, 3373–3380.
(64) Burdon, J.; Parsons, I. W. A Recent Attempt to Apply Frontier Orbital Theory to Nucleophilic Aromatic Substitution. *J. Am. Chem. Soc.* **1977**, *99*, 7445–7447.
(65) Epiotis, N. D.; Cherry, W. An MO Study of the Orientational Selectivity in Nucleophilic Substitution Reactions of Polyhalobenzenes. *J. Am. Chem. Soc.* **1976**, *98*, 5432–5435.
(66) Scales, S.; Johnson, S.; Hu, Q.; Do, Q.-Q.; Richardson, P.; Wang, F.; Braganza, J.; Ren, S.; Wan, Y.; Zheng, B.; Faizi, D.; McAlpine, I. Studies on the Regioselective Nucleophilic Aromatic Substitution (S_NAr)Reaction of 2-Substituted 3,5-Dichloropyrazines. *Org. Lett.* **2013**, *15*, 2156–2159.
(67) Muir, M.; Baker, J. A Simple Calculational Model for Predicting the Site for Nucleophilic Substitution in Aromatic Perfluorocarbons. *J. Fluor. Chem.* **2005**, *126*, 727–738.
(68) Baker, J.; Muir, M. The Meisenheimer Model for Predicting the Principal Site for Nucleophilic Substitution in Aromatic Perfluorocarbons — Generalization to Include Ring-Nitrogen Atoms and Non-Fluorine Ring Substituents. *Can. J. Chem.* **2010**, *88*, 588–597.
(69) Liljenberg, M.; Brinck, T.; Herschend, B.; Rein, T.; Rockwell, G.; Svensson, M. A Pragmatic Procedure for Predicting Regioselectivity in Nucleophilic Substitution of Aromatic Fluorides. *Tetrahedron Lett.* **2011**, *52*, 3150–3153.
(70) Liljenberg, M.; Brinck, T.; Herschend, B.; Rein, T.; Tomasi, S.; Svensson, M. Predicting Regioselectivity in Nucleophilic Aromatic Substitution. *J. Org. Chem.* **2012**, *77*, 3262–3269.
(71) Liljenberg, M.; Brinck, T.; Rein, T.; Svensson, M. Utilizing the σ-Complex Stability for Quantifying Reactivity in Nucleophilic Substitution of Aromatic Fluorides. *Beilstein J. Org. Chem.* **2013**, *9*, 791–799.
(72) Hirsch, J. A. Table of Conformational Energies—1967. In *Topics in Stereochemistry Vol. 1*; John Wiley & Sons: Nashville, TN, 1967; pp 199–222.
(73) Clavier, H.; Nolan, S. P. Percent Buried Volume for Phosphine and N-Heterocyclic Carbene Ligands: Steric Properties in Organometallic Chemistry. *Chem. Commun.* **2010**, *46*, 841–861.
(74) Brinck, T.; Carlqvist, P.; Stenlid, J. H. Local Electron Attachment Energy and Its Use for Predicting Nucleophilic Reactions and Halogen Bonding. *J. Phys. Chem. A* **2016**, *120*, 10023–10032.
(75) Stenlid, J. H.; Brinck, T. Nucleophilic Aromatic Substitution Reactions Described by the Local Electron Attachment Energy. *J. Org. Chem.* **2017**, *82*, 3072–3083.
(76) Bartoli, G.; Todesco, P. E. Nucleophilic Substitution. Linear Free Energy Relations between Reactivity and Physical Properties of Leaving Groups and Substrates. *Acc. Chem. Res.* **1977**, *10*, 125–132.
(77) Lu, J.; Donnecke, S.; Paci, I.; Leitch, D. C. A Reactivity Model for Oxidative Addition to Palladium Enables Quantitative Predictions for Catalytic Cross-Coupling Reactions. *Chem. Sci.* **2022**, *13*, 3477–3488.
(78) Geerlings, P.; De Proft, F.; Langenaeker, W. Conceptual Density Functional Theory. *Chem. Rev.* **2003**, *103*, 1793–1874.

(79) Suresh, C. H.; Alexander, P.; Vijayalakshmi, K. P.; Sajith, P. K.; Gadre, S. R. Use of Molecular Electrostatic Potential for Quantitative Assessment of Inductive Effect. *Phys. Chem. Chem. Phys.* **2008**, *10*, 6492–6499.

(80) Sayyed, F. B.; Suresh, C. H. Quantification of Substituent Effects Using Molecular Electrostatic Potentials: Additive Nature and Proximity Effects. *New J. Chem.* **2009**, *33*, 2465–2471.

(81) Remya, G. S.; Suresh, C. H. Quantification and Classification of Substituent Effects in Organic Chemistry: A Theoretical Molecular Electrostatic Potential Study. *Phys. Chem. Chem. Phys.* **2016**, *18*, 20615–20626.

(82) Gadre, S. R.; Suresh, C. H.; Mohan, N. Electrostatic Potential Topology for Probing Molecular Structure, Bonding and Reactivity. *Molecules* **2021**, *26*, 3289.

(83) Consonni, V.; Ballabio, D.; Todeschini, R. Comments on the Definition of the Q2 Parameter for QSAR Validation. *J. Chem. Inf. Model.* **2009**, *49*, 1669–1678.

(84) Fukui, K.; Yonezawa, T.; Nagata, C.; Shingu, H. Molecular Orbital Theory of Orientation in Aromatic, Heteroaromatic, and Other Conjugated Molecules. *J. Chem. Phys.* **1954**, *22*, 1433–1442.

(85) Houk, K. N. Frontier Molecular Orbital Theory of Cycloaddition Reactions. *Acc. Chem. Res.* **1975**, *8*, 361–369.

(86) Allgäuer, D. S.; Jangra, H.; Asahara, H.; Li, Z.; Chen, Q.; Zipse, H.; Ofial, A. R.; Mayr, H. Quantification and Theoretical Analysis of the Electrophilicities of Michael Acceptors. *J. Am. Chem. Soc.* **2017**, *139*, 13318–13329.

(87) Koopmans, T. Über die Zuordnung von Wellenfunktionen und Eigenwerten zu den Einzelnen Elektronen Eines Atoms. *Physica* **1934**, *1*, 104–113.

(88) Perdew, J. P.; Parr, R. G.; Levy, M.; Balduz, J. L. Density-Functional Theory for Fractional Particle Number: Derivative Discontinuities of the Energy. *Phys. Rev. Lett.* **1982**, *49*, 1691–1694.

(89) *Entos Envision*. https://www.entos.ai/envision (accessed 2022-07-07).

(90) Grimme, S.; Bannwarth, C.; Shushkov, P. A Robust and Accurate Tight-Binding Quantum Chemical Method for Structures, Vibrational Frequencies, and Noncovalent Interactions of Large Molecular Systems Parametrized for All *spd*-Block Elements (Z = 1–86). *J. Chem. Theory Comput.* **2017**, *13*, 1989–2009.

(91) Maes, B. U. W.; Verbeeck, S.; Verhelst, T.; Ekomié, A.; von Wolff, N.; Lefèvre, G.; Mitchell, E. A.; Jutand, A. Oxidative Addition of Haloheteroarenes to Palladium(0): Concerted versus S_NAr-Type Mechanism. *Chem. Eur. J.* **2015**, *21*, 7858–7865.

(92) Lu, T.; Chen, F. Multiwfn: A Multifunctional Wavefunction Analyzer. *J. Comput. Chem.* **2012**, *33*, 580–592.

(93) Lu, T.; Chen, F. Quantitative Analysis of Molecular Surface Based on Improved Marching Tetrahedra Algorithm. *J. Mol. Graph. Model.* **2012**, *38*, 314–323.

(94) Wang, B.; Rong, C.; Chattaraj, P. K.; Liu, S. A Comparative Study to Predict Regioselectivity, Electrophilicity and Nucleophilicity with Fukui Function and Hirshfeld Charge. *Theor. Chem. Acc.* **2019**, *138*, 124.

(95) Roy, R. K.; Saha, S. Studies of Regioselectivity of Large Molecular Systems Using DFT Based Reactivity Descriptors. *Annu. Rep. Sect. C Phys. Chem.* **2010**, *106*, 118–162.

(96) Rohrbach, S.; Murphy, J. A.; Tuttle, T. Computational Study on the Boundary Between the Concerted and Stepwise Mechanism of Bimolecular SNAr Reactions. *J. Am. Chem. Soc.* **2020**, *142*, 14871–14876.

(97) Berliner, E.; Monack, L. C. Nucleophilic Displacement in the Benzene Series. *J. Am. Chem. Soc.* **1952**, *74*, 1574–1579.

(98) Crampton, M. R.; Emokpae, T. A.; Isanbor, C. The Effects of Ring Substituents and Leaving Groups on the Kinetics of S_NAr Reactions of 1-Halogeno- and 1-Phenoxy-Nitrobenzenes with Aliphatic Amines in Acetonitrile. *Eur. J. Org. Chem.* **2007**, *2007*, 1378–1383.

(99) Chambers, R. D.; Martin, P. A.; Sandford, G.; Williams, D. L. H. Mechanisms of Reactions of Halogenated Compounds: Part 7. Effects of Fluorine and Other Groups as Substituents on Nucleophilic Aromatic Substitution. *J. Fluor. Chem.* **2008**, *129*, 998–1002.

(100) Bolton, R.; Sandall, J. P. B. Nucleophilic Displacement in Polyhalogenoaromatic Compounds. Part 1. Kinetics of Reaction of Polychlorofluorobenzene Derivatives. *J. Chem. Soc. Perkin Trans. 2* **1976**, 1541–1545.

(101) Chambers, R. D.; Close, D.; Musgrave, W. K. R.; Waterhouse, J. S.; Williams, D. L. H. Mechanisms for Reactions of Halogenated Compounds. Part 2. Orienting Effects of Chlorine Substituents in Nucleophilic Aromatic Substitution. *J. Chem. Soc. Perkin Trans. 2* **1977**, 1774–1778.

(102) Chambers, R. D.; Martin, P. A.; Waterhouse, J. S.; Williams, D. L. H.; Anderson, B. Mechanisms for Reactions of Halogenated Compounds. Part 4. Activating Influences of Ring-Nitrogen and Trifluoromethyl in Nucleophilic Aromatic Substitution. *J. Fluor. Chem.* **1982**, *20*, 507–514.

(103) Chambers, R. D.; Musgrave, W. K. R.; Urben, P. G. Polyhalogenoheterocyclic Compounds. Part XXVI. Nucleophilic Substitution in Trifluoropyrazines. *J. Chem. Soc. Perkin 1* **1974**, 2580–2584.

(104) Banks, R. E.; Prakash, A.; Venayak, N. D. Studies in Azide Chemistry. Part IX. Investigations Involving Fluorinated Azidopyrimidines and 4-Azido-3-Chloro-2,5,6-Trifluoropyridine. *J. Fluor. Chem.* **1980**, *16*, 325–338.

(105) Chambers, R. D.; Seabury, M. J.; Williams, D. L. H.; Hughes, N. Mechanisms for Reactions of Halogenated Compounds. Part 5. Orientating Effects of Fluorine Substituents on Nucleophilic Substitution in Naphthalene and Other Polycyclic Systems. *J. Chem. Soc. Perkin 1* **1988**, 251–254.

(106) Chambers, R. D.; Seabury, M. J.; Williams, D. L. H.; Hughes, N. Mechanisms for Reactions of Halogenated Compounds. Part 6. Investigations into the Activating Effect of Ortho-Fluorine in Nucleophilic Aromatic Substitution. *J. Chem. Soc. Perkin 1* **1988**, 255–257.

(107) Keana, J. F. W.; Cai, S. X. New Reagents for Photoaffinity Labeling: Synthesis and Photolysis of Functionalized Perfluorophenyl Azides. *J. Org. Chem.* **1990**, *55*, 3640–3647.

(108) Dirr, R.; Anthaume, C.; Désaubry, L. Regioselectivity of Fluorine Substitution by Alkoxides on Unsymmetrical Difluoroarenes. *Tetrahedron Lett.* **2008**, *49*, 4588–4590.

(109) Brooke, G. M.; Chambers, R. D.; Drury, C. J.; Bower, M. J. Remarkable Orientational Effects in the Displacement of the Fluorine from Heptafluoro-Isoquinoline and -Quinoline towards Sulfur Nucleophiles. Further Reactions with Oxygen Nucleophiles. *J. Chem. Soc. Perkin 1* **1993**, 2201–2209.

(110) Tanaka, K.; Deguchi, M.; Iwata, S. Ab Initio Study of Nucleophilic Aromatic Substitution of Polyfluorobenzene. *J. Chem. Res. Synop.* **1999**, 528–529.

(111) Schroeder, G.; Eitner, K.; Gierczyk, B.; Rózalski, B.; Brzezinski, B. Reaction of Fluoronitrobenzenes with Tetramethylguanidine in Acetonitrile. *J. Mol. Struct.* **1999**, *478*, 243–253.

(112) Delia, T. J.; Anderson, D. P.; Schomaker, J. M. 2,4,6-Trifluoropyrimidine. Reactions with Nitrogen Nucleophiles. *J. Heterocycl. Chem.* **2004**, *41*, 991–993.

(113) Belli, M. L.; Illuminati, G.; Marino, G. Meta-Substituent Effects on the Kinetics of Methoxydechlorination of Some 2- and 4-Chloroquinolines. *Tetrahedron* **1963**, *19*, 345–355.

(114) Yukawa, M.; Niiya, T.; Goto, Y.; Sakamoto, T.; Yoshizawa, H.; Watanabe, A.; Yamanaka, H. MNDO (Modified Neglect of Diatomic Overlap) Study of the Nucleophilic Substitution Reactions of Chloropyrimidines. *Chem. Pharm. Bull.* **1989**, *37*, 2892–2896.

(115) Flowers, W. T.; Haszeldine, R. N.; Majid, S. A. Synthesis and Reactions of Pentachloropyridine. *Tetrahedron Lett.* **1967**, *8*, 2503–2505.

(116) Collins, I.; Suschitzky, H. Polyhalogeno-Aromatic Compounds. Part VIII. Reactions of Hexabromobenzene with Nucleophiles. *J. Chem. Soc. C Org.* **1969**, 2337–2341.

(117) Volkov, K. A.; Avramenko, G. V.; Negrimovskii, V. M.; Luk'yanets, E. A. Phthalocyanines and Related Compounds: XLIII. Synthesis of Poly[Phenyl(Alkyl)Sulfanyl]-Substituted Phthalonitriles and Some Phthalocyanines Based Thereon. *Russ. J. Gen. Chem.* **2007**, *77*, 1108–1116.

(118) Xinzhuo, Z.; Ibata, T. High Pressure Synthesis of Nitrophenylamines Type Nonlinear Optical Compounds. *Chin. J. Org. Chem.* **2002**, *22*, 778.

(119) Carta, A.; Palomba, M.; Corona, P. Synthesis of Substituted Aminoquinolines as Useful Intermediates for Preparation of Aromatic N-Tricyclic Systems. *Heterocycles* **2006**, *68*, 1715–1722.

(120) Collins, I.; Suschitzky, H. Polyhalogeno-Aromatic Compounds. Part XIV. Nucleophilic Substitution and Peroxy-Acid Oxidation of Pentabromopyridine and Some of Its NN-Dialkylamino- and Bis-(*N*,*N*-Dialkylamino)-Derivatives. *J. Chem. Soc. C: Org.* **1970**, 1523–1530.

(121) Mizuno, H.; Manabe, A. Pyrimidine Compounds and Pests Controlling Composition Containing the Same. WO2004099160A1, November 18, 2004.

(122) Allen, J. M.; Butlin, R. J.; Green, C.; Mccoull, W.; Robb, G. R.; Wood, J. M. Benzothiazoles as Ghrelin Receptor Modulators. WO2009047558A1, April 16, 2009.

(123) Zhou, D.; Stack, G. P.; Lo, J.; Failli, A. A.; Evrard, D. A.; Harrison, B. L.; Hatzenbuhler, N. T.; Tran, M.; Croce, S.; Yi, S.; Golembieski, J.; Hornby, G. A.; Lai, M.; Lin, Q.; Schechter, L. E.; Smith, D. L.; Shilling, A. D.; Huselton, C.; Mitchell, P.; Beyer, C. E.; Andree, T. H. Synthesis, Potency, and in Vivo Evaluation of 2-Piperazin-1-Ylquinoline Analogues as Dual Serotonin Reuptake Inhibitors and Serotonin 5-HT1A Receptor Antagonists. *J. Med. Chem.* **2009**, *52*, 4955–4959.

(124) Weber, O.; Voehringer, V.; Lerchen, H.-G.; Hafner, F.-T.; Keldenich, J.; Schlemmer, K.-H.; Krenz, U.; Riedl, B. Benzofuran and Benzothiophene Derivatives Useful in the Treatment of Cancers of the Central Nervous System. WO2008025509A1, March 6, 2008.

(125) Ahmad, N.; Boyall, D.; Charrier, J.-D.; Davis, C.; Davis, R.; Durrant, S.; Jardi, G. E. I.; Fraysse, D.; Jimenez, J.-M.; Kay, D.; Knegtel, R.; Middleton, D.; O'Donnell, M.; Panesar, M.; Pierard, F.; Pinder, J.; Shaw, D.; Storck, P.-H.; Studley, J.; Twin, H. Compounds Useful as Inhibitors of Atr Kinase. WO2014089379A1, June 12, 2014.

(126) Al-Awar, R.; Isaac, M.; Chau, A. M.; Mamai, A.; Watson, I.; Poda, G.; Subramanian, P.; Wilson, B.; Uehling, D.; Prakesch, M.; Joseph, B.; Morin, J.-A. Inhibitors of the Bcl6 Btb Domain Protein-Protein Interaction and Uses Thereof. WO2019153080A1, August 15, 2019.

(127) Stevens, J. M.; Simmons, E. M.; Tan, Y.; Borovika, A.; Fan, J.; Forest, R. V.; Geng, P.; Guerrero, C. A.; Lou, S.; Skliar, D.; Steinhardt, S. E.; Strotman, N. A. Leveraging High-Throughput Experimentation to Drive Pharmaceutical Route Invention: A Four-Step Commercial Synthesis of Branebrutinib (BMS-986195). *Org. Process Res. Dev.* **2022**, *26*, 1174–1183.

(128) Tao, Y.; Keene, N. F.; Wiglesworth, K. E.; Sitter, B.; McWilliams, J. C. Early Process Development of an Irreversible Epidermal Growth Factor Receptor (EGFR)T790 M Inhibitor. *Org. Process Res. Dev.* **2019**, *23*, 382–388.

(129) Stumpf, A.; Cheng, Z. K.; Beaudry, D.; Angelaud, R.; Gosselin, F. Improved Synthesis of the $Na_V1.7$ Inhibitor GDC-0276 via a Highly Regioselective S_NAr Reaction. *Org. Process Res. Dev.* **2019**, *23*, 1829–1840.

(130) Bremberg, U.; Eriksson-Bajtner, J.; Lehmann, F.; Oltner, V.; Sölver, E.; Wennerberg, J. Development of a Synthesis of Kinase Inhibitor AKN028. *Org. Process Res. Dev.* **2018**, *22*, 1360–1364.

(131) Tao, Y.; Widlicka, D. W.; Hill, P. D.; Couturier, M.; Young, G. R. A Scalable Synthesis of CE-157119 HCl Salt, an SRI/5-HT2A Antagonist. *Org. Process Res. Dev.* **2012**, *16*, 1805–1810.

(132) Barcan, G. A.; Conde, J. J.; Mokhallalati, M. K.; Nilson, M. G.; Xie, S.; Allen, C. L.; Andemichael, Y. W.; Calandra, N. A.; Leitch, D. C.; Li, L.; Morris, M. J. Nucleophilic Aromatic Substitutions of 2-Halo-5-(Sulfamoyl)Benzoic Acids and N,O-Bis-Alkylation via Phase Transfer Catalysis: Synthesis of RoRγ Inverse Agonist GSK2981278A. *Org. Process Res. Dev.* **2019**, *23*, 1396–1406.

(133) Lu, Y.; Cole, K. P.; Fennell, J. W.; Maloney, T. D.; Mitchell, D.; Subbiah, R.; Ramadas, B. An Alternative Indazole Synthesis for Merestinib. *Org. Process Res. Dev.* **2018**, *22*, 409–419.