Journal of the American Medical Informatics Association Advance Access published October 24, 2015

Cheminformatics-aided pharmacovigilance: application to Stevens-Johnson Syndrome

RECEIVED 27 March 2015
REVISED 6 July 2015
ACCEPTED 11 July 2015

Yen S Low[1,2], Ola Caster[3,4], Tomas Bergvall[3], Denis Fourches[1], Xiaoling Zang[1], G Niklas Norén[3,5], Ivan Rusyn[2], Ralph Edwards[3], Alexander Tropsha[1]

AMIA INFORMATICS PROFESSIONALS. LEADING THE WAY. OXFORD UNIVERSITY PRESS

ABSTRACT

Objective Quantitative Structure-Activity Relationship (QSAR) models can predict adverse drug reactions (ADRs), and thus provide early warnings of potential hazards. Timely identification of potential safety concerns could protect patients and aid early diagnosis of ADRs among the exposed. Our objective was to determine whether global spontaneous reporting patterns might allow chemical substructures associated with Stevens-Johnson Syndrome (SJS) to be identified and utilized for ADR prediction by QSAR models.

Materials and Methods Using a reference set of 364 drugs having positive or negative reporting correlations with SJS in the VigiBase global repository of individual case safety reports (Uppsala Monitoring Center, Uppsala, Sweden), chemical descriptors were computed from drug molecular structures. Random Forest and Support Vector Machines methods were used to develop QSAR models, which were validated by external 5-fold cross validation. Models were employed for virtual screening of DrugBank to predict SJS actives and inactives, which were corroborated using knowledge bases like VigiBase, ChemoText, and MicroMedex (Truven Health Analytics Inc, Ann Arbor, Michigan).

Results We developed QSAR models that could accurately predict if drugs were associated with SJS (area under the curve of 75%–81%). Our 10 most active and inactive predictions were substantiated by SJS reports (or lack thereof) in the literature.

Discussion Interpretation of QSAR models in terms of significant chemical descriptors suggested novel SJS structural alerts.

Conclusions We have demonstrated that QSAR models can accurately identify SJS active and inactive drugs. Requiring chemical structures only, QSAR models provide effective computational means to flag potentially harmful drugs for subsequent targeted surveillance and pharmacoepidemiologic investigations.

Keywords: pharmacovigilance, cheminformatics, QSAR, Stevens-Johnson Syndrome, adverse drug reactions

BACKGROUND AND SIGNIFICANCE

Pharmacovigilance, the detection of adverse drug reactions (ADRs), relies on the surveillance of spontaneous reports submitted by health care practitioners and pharmaceutical manufacturers.[1–3] However, warning signals require a sufficient number of ADR reports to accumulate, inadvertently exposing more patients to potentially harmful drugs.[4]

Growing digitization of health care data offers new opportunities for improving ADR detection. Indeed, it is feasible to detect ADRs using electronic health records,[5] biomedical literature,[6] drug labels,[7] bioassays,[8] and Quantitative Structure-Activity Relationships (QSAR) models.[9] The latter approach that correlates chemical descriptors of molecules with their chemical activity was introduced by Hansch *et al.*[10] In early QSAR studies, few descriptors (eg, electronic,[11] hydrophobic,[12] and steric[13]) were used. Modern QSAR models now employ numerous diverse chemical descriptors (calculated, for example, by Dragon software [TALETE srl, Milano, Italy][14]) that represent physicochemical properties, substructural fragments (eg, presence of chemical functional groups, In Silico Design and Data Analysis [ISIDA][15] fragments), molecular signatures (eg, Molecular ACCess System [MACCS][16] fingerprints), and abstract mathematical derivations based on quantum theory (eg, orbital energies).[17] Often, machine learning methods are used to correlate multiple chemical descriptors to the compounds' activities. Quantitative Structure-Activity Relationships modeling has been widely used for predicting drug potencies and chemical toxicities, including ADRs.[9,18–21]

Requiring only chemical structures as input, QSAR modeling allows *in silico* prediction of drug effects before drugs are released to the market and initiates targeted surveillance, protecting patients from unnecessary exposure and hastening the diagnosis of ADRs. Moreover, QSAR models can provide insight into the mechanisms underlying the ADR and guide safer drug design.

In this study, we have focused on Stevens-Johnson Syndrome (SJS) because of its medical severity[22] and well-established structure-activity relationships linking drug classes such as sulfonamide antibiotics, penicillin, and quinolones to SJS.[18,22,23] Because these drugs are widely used, many patients may be unnecessarily at risk of SJS, especially predisposed populations with certain human leukocyte antigen subtypes.[24–26] In SJS, the skin and mucous membranes separate into large blisters, leaving denuded, hemorrhagic areas over the whole body with a mortality rate of up to 30% to 40%.[22] Although exact pathogenesis remains to be established, SJS is often drug-induced and immune-mediated and may manifest as a hypersensitivity reaction to drugs.[27] Implicated drugs have distinct molecular structures such as sulfonamide and penicillin, leading researchers to question if there is a chemical basis for drug-induced SJS.[18,22,23] Understanding the chemical basis would help identify toxicophore alerts to guide prescription.

Correspondence to Alexander Tropsha, Division of Chemical Biology and Medicinal Chemistry, Eshelman School of Pharmacy, University of North Carolina, 100K Beard Hall, Campus Box 7568, Chapel Hill, NC 27599-7568, USA; alex_tropsha@unc.edu. For numbered affiliations see end of article.

©The Author 2015. Published by Oxford University Press on behalf of the American Medical Informatics Association.
This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (http://creativecommons.org/licenses/by-nc/4.0/), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.com

RESEARCH AND APPLICATIONS
Downloaded from http://jamia.oxfordjournals.org/ by guest on May 24, 2016

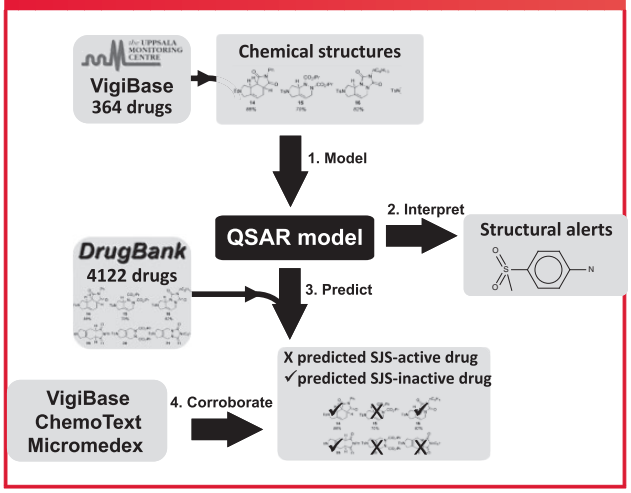
Previously, QSAR models with 69% to 73% prediction accuracy have been reported for a dataset of 110 drugs obtained from the US Food and Drug Administration Adverse Event Reporting System.[18] This study expands prior work by drawing upon a larger database of global spontaneous reports (VigiBase; Uppsala Monitoring Center, Uppsala, Sweden) and a larger set of 364 drugs for QSAR modeling. Our objectives were to develop, validate, and interpret QSAR models that could more accurately predict drugs' association with SJS. Although this study is specific to SJS, the general methodological workflow combining QSAR modeling, virtual screening of drug databases, and validation of predictions by focused exploration of existing knowledge bases-can be applied to many ADRs.

METHODS

Overview

Figure 1 presents a general methodological workflow integrating the development, interpretation, and validation of QSAR model for SJS. First, QSAR models are developed using a diverse set of drugs associated with SJS according to VigiBase®,[28] the World Health Organization (WHO) global database of suspected ADRs, maintained and analyzed by the Uppsala Monitoring Centre. As of February 2012, VigiBase contained 7 014 658 reports from 107 countries, covering approximately 20 000 drugs (ie, generic substances) and 2000 ADRs coded according to the WHO Drug Dictionary Enhanced™ and the WHO-Adverse Reactions Terminology™, respectively. Second, QSAR models are interpreted for important chemical features to detect structural alerts, which are chemical substructures characteristic of SJS-active drugs. Third, these models are used to screen DrugBank[29] for potential SJS-active drugs. Fourth, predictions are checked for either the evidence of SJS or lack thereof using VigiBase,[28] ChemoText,[30] and Micromedex (Truven Health Analytics Inc, Ann Arbor, Michigan).[31]

Figure 1: Schematic workflow showing the use of multiple data sources for developing, interpreting, and validating QSAR models that classify drugs as SJS-active or inactive. First, VigiBase provided 364 drugs whose chemical structures were used as variables for QSAR modeling. Second, QSAR models provided structural alerts for interpretation. Third, QSAR models predicted potential SJS actives and inactives in DrugBank. Fourth, the predicted actives and inactives were evaluated for evidence of SJS activity or lack thereof in VigiBase, ChemoText, and Micromedex.
Abbreviations: QSAR, Quantitative Structure-Activity Relationships; SJS, Stevens-Johnson syndrome.



SJS-active and SJS-inactive drugs

To develop the QSAR models, a reference set of drugs was extracted based on their reporting correlations with SJS in VigiBase. A drug was defined as *active* if it had higher-than-expected reporting with SJS, as indicated by a positive coefficient in a shrinkage regression model for the reporting of SJS in VigiBase.[32] By considering all 20 000 drugs simultaneously, regression is more conservative than standard disproportionality analysis and minimizes false inclusion of innocent bystanders coreported with true SJS actives. Drugs were defined as *inactive* if they had no or minimal reporting correlation with SJS, as specified by the following criteria: (1) for drugs with less than 1000 reports in total, inactives must never have been reported with SJS; for drugs with at least 1000 reports in total, inactives must have disproportionately few SJS reports as indicated by a negative Information Component (IC) 95% credibility interval[33,34]; and (2) never be the sole suspect drug in any SJS report. The *sole suspect* criterion minimizes the risk of including "inactive" drugs that have weak overall correlation to SJS in the database but may have strong implications for a causal link in 1 or a few reports.

Chemical structures

Of the 436 drugs extracted from VigiBase (excluding mixtures and biologics), chemical structures were retrieved and curated to ensure that drug structures were correctly represented and standardized prior to model development.[35] After removing salts, metal-containing compounds, large molecules (molecular weight > 2000 daltons), and structural duplicates (using ChemAxon v.5.0, Budapest, Hungary; and Pipeline Pilot Student Edition v.6.1.5, Accelrys, San Diego, California), 194 actives and 170 inactives remained for QSAR modeling (supplementary table S1).

We used 3 different sets of chemical descriptors: Dragon, ISIDA substructural fragments, and MACCS fingerprints. Dragon descriptors (v.5.5),[14] known for their comprehensive characterization of chemicals structures, include constitutional, functional groups, atom-centered fragments, molecular properties, and 2-dimensional frequency fingerprints. To generate ISIDA[15] fragment descriptors, each molecular structure was split into substructural fragments containing 2 to 6 atoms in linear sequence. Fragment descriptors were binarized, depending on whether the fragments were present or absent in a drug. The third type of descriptors, MACCS fingerprints, were a predefined set of 166 binary hash representations compressing numerous descriptors.[16]

All subsequent analyses were performed in R (v.2.14). Continuous descriptors (Dragon) were autoscaled to z scores. Descriptors were excluded if they were invariant (< 0.001 standardized standard deviation; > 99% constant values) or intercorrelated (if pairwise $r^2 > 0.99$, randomly remove 1 of the 2 descriptors) such that 354 Dragon, 138 MACCS, and 1091 ISIDA descriptors remained for modeling (supplementary table S2).

Step 1: QSAR modeling and evaluation

For each of the 3 descriptor sets, 2 classification methods (Random Forest [RF][36] and support vector machines [SVMs][37]) were used to build QSAR models. All models were evaluated by external 5-fold cross validation[38] whereby the entire dataset was divided randomly into 5 equal parts. Each individual part was systematically left out as an external validation set, whereas the remaining 80% of compounds in the dataset were used for model development. We used the default

Downloaded from http://jamia.oxfordjournals.org/ by guest on May 24, 2016

parameters for RF (randomForest R package version 4.6-5) as RF are known to perform well even without parameter tuning.[39] SVM modeling parameters (ie, cost and gamma) were tuned for minimum mean error by additional internal 5-fold cross validation within the 80% training set. This tuned model was externally validated with the corresponding 20% external set, which was never used for parameter tuning and modeling. Reported prediction accuracies are based on the external sets only.

Models were assessed by specificity, sensitivity, balanced accuracy, area under the curve (AUC), and coverage. Balanced accuracy is the average of specificity and sensitivity. Coverage measures the fraction of test drugs that can be reliably predicted by the QSAR model, depending on the test drugs' chemical similarity to drugs in the training set. The chemical space bound by the training set molecules defines the limits of the extrapolation region (termed *application domain*) of the QSAR model.[40,41] For more reliable predictions, the application domain can be tightened by setting a high minimum chemical similarity threshold between the test drugs and training drugs. In this study, the application domain was set to mean interchemical distance plus half a standard deviation. Here, coverage refers to the fraction of drugs in the external set that are within the above defined applicability domain. Additionally, precision was used to assess structural alerts. Standard errors of all metrics were calculated by bootstrapping[42] with 1000 trials.

Additional internal validation with *y*-randomization test ensured that models were robust and not due to chance correlations.[43] After permuting the *y* activity labels in the modeling sets, models were rebuilt following the same procedures as outlined above for nonrandomized data. This process was repeated 30 times to generate a null distribution of *y*-randomized model accuracy for comparison under a 1-tailed, 1 sample *t* test.

Step 2: QSAR model interpretation

Model interpretation involved identifying key chemical predictors of SJS in terms of *f* most important individual ISIDA fragments and fused substructures reconstituted from the fragments.

Important chemical fragments

To determine the minimal subset of *f* most important chemical fragments predictive of SJS, ISIDA fragments were ranked by RF conditional importance[44] in the ISIDA-RF model. Because the fragment ranking varied slightly across 5 models generated with 5-fold external cross validation, only *f* fragments that were consistently among the top 10, 25, 50, 75, and 150 fragments in all 5 models were selected for rebuilding reduced RF models. For each value of *f*, the reduced RF model's out-of-bag (OOB_f) error[36] was compared with that of the full RF model (OOB_{full}) incorporating all 1091 fragments. Optimal *f* (ie, f_{min}) was defined as *f* with the minimum OOB_f error less than or equal to OOB_{full} error.

Structural alerts identified from co-occurring fragments (Method 1)

The fused structural alerts were reconstituted from clusters of fragments that co-occurred more frequently in actives than in inactives. All possible pairs of *f* fragments were tested for higher-than-expected co-occurrence in actives compared with inactives by a 2-tailed Fisher exact test. A fragment pair was said to have significant co-occurrence when its $P < 0.1$ after adjustment for multiple testing by permutation (figure 2a). Specifically for each fragment *i*, its pairwise co-occurrence with each of 1000 noise fragments (randomly present or absent) generated Fisher test values $t_{i,\text{noise1}}, t_{i,\text{noise2}}, \ldots, t_{i,\text{noise1000}}$, forming a permutation null distribution D_i. To adjust the test value of the pairwise co-occurrence of fragments *i* and *j*, $t_{i,j}$ was compared against the relevant null distributions D_i and D_j such that the larger of its quantiles along the null distributions, max (q_i, q_j), was taken as the adjusted *P* value.

Co-occurring fragments were represented by a network where fragment nodes were connected if they co-occurred significantly (figure 2b). Frequently co-occurring fragments formed densely connected subnetworks known as *communities* in network analysis. Communities were detected by the walktrap algorithm, which stochastically agglomerated the fragment nodes such that they were disproportionately more connected with nodes inside the community than outside.[45] Within a distinctly colored community, the fragments co-occurred more frequently in drugs of 1 class vs the other. Hence, they could be assembled into a larger substructure as structural alerts for a certain class of drugs.

Structural alerts identified by maximal common substructures analysis (Method 2)

The maximal common substructures (MCS)[46] method provided a second set of structural alerts for comparison with those obtained by co-occurring fragments (Method 1). MCS extracted the largest substructures being more frequently associated with actives than with inactives. For MCS of reasonable utility, we set the following criteria: size ≥ 8 atoms, frequency ratio ≥ 2, and derived from ≥ 6 molecules.

Step 3: QSAR model application: prediction of SJS-active and inactive drugs in DrugBank

We used our best QSAR model (RF model of Dragon descriptors) to virtually screen the DrugBank library of 4122 drugs[29] for potential SJS-active drugs after excluding drugs used for modeling. The same chemical curation and descriptor treatment procedures used earlier for the modeling dataset were applied to DrugBank. Compounds were ranked by virtual screening prediction scores. Recall that in RF, an ensemble of models is generated. The prediction score is the average of the classification labels (0 or 1) generated by individual models; thus, the closer the average prediction is to 1, the higher is the probability that a compound is active.

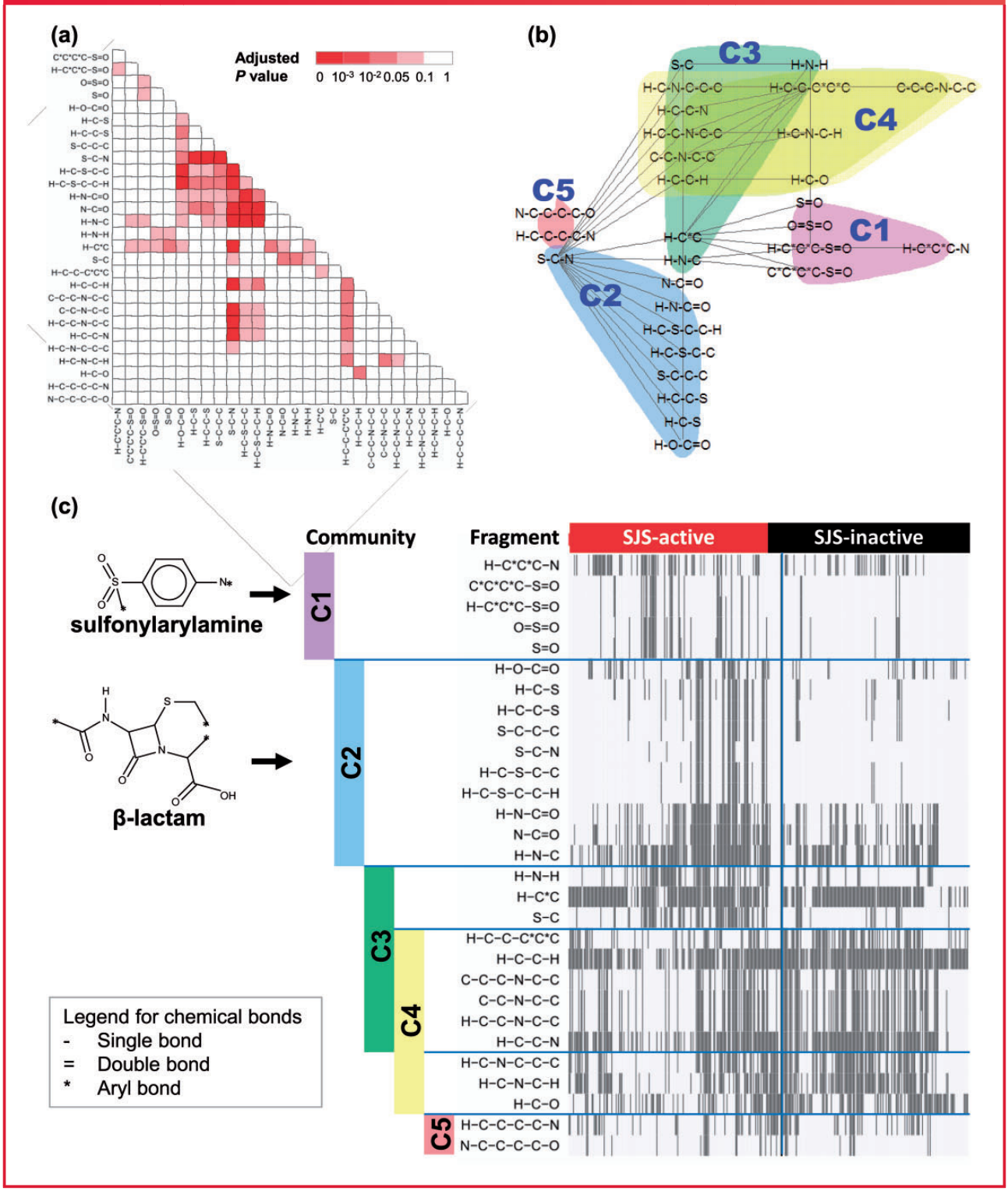
Step 4: Corroboration of model predictions using knowledge bases

The following subsets of drugs were evaluated for evidence of SJS (or lack thereof) in knowledge bases: 10 compounds most likely to be active (based on their prediction scores being closest to 1); 10 compounds most likely to be inactive (predictions closest to zero); 2 positive controls with known association with SJS (sulfamethoxazole and amoxicillin); and 2 negative controls with zero or minimal association with SJS (progesterone and vardenafil). Predicted actives and inactives were ranked according to their mean predicted value (average of 5 predictions from 5 models from 5-fold cross validation, supplementary table S3).

The 3 knowledge bases VigiBase,[28] ChemoText,[30] and Micromedex[31] reflect the association between the predicted drug and SJS in various data sources—namely, spontaneous ADR reports, the biomedical literature, and a curated knowledge source, respectively. VigiBase, a repository of global spontaneous ADR reports, provided an IC value that measured if each drug was linked to a disproportionate number of spontaneous SJS reports.[28] ChemoText, a chemocentric database of MeSH (Medical Subject Headings) annotations sourced from PubMed,[30] provided the number of human studies coannotating the drug of interest and "Stevens-Johnson syndrome" (also included MeSH synonyms and related terms "erythema multiforme" and "epidermal necrolysis, toxic"). Micromedex,[31] an evidence-based resource referenced by clinicians as an industry standard, was searched for co-mentions of SJS and related hypersensitivity.

Downloaded from http://jamia.oxfordjournals.org/ by guest on May 24, 2016

Figure 2: Results of co-occurrence analysis of ISIDA chemical fragments. **(a)** Adjusted *P* values show the association between pairwise co-occurrence of any 2 fragments and SJS inducing activity (from a 2-sided Fisher exact test). **(b)** Distinctly colored communities of co-occurring fragments detected by the walktrap community algorithm. Fragment nodes are connected if significantly co-occurring ($P < 0.1$). **(c)** Heat map shows the joint presence of co-occurring fragments within a community (eg, purple C1, corresponding to sulfonylarylamine reconstituted from 5 co-occurring fragments, is more frequently present among SJS-active drugs).
Abbreviations: SJS, Stevens-Johnson syndrome; ISIDA, In Silico Design and Data Analysis.



Downloaded from http://jamia.oxfordjournals.org/ by guest on May 24, 2016

RESULTS

The reference set extracted from VigiBase consisted of 194 active and 170 inactive drugs (supplementary table S1). The actives (table 1) had more SJS reports in VigiBase than inactives (mean = 104 vs 1.52, respectively), more ADR reports overall (mean = 6953 vs 3505, respectively), and were disproportionately drawn from several Anatomical Therapeutic Chemical (ATC) groups such as the anti-infectives (J) and musculoskeletal system (M) groups. At the same time, inactives were disproportionately drawn from ATC groups such as genitourinary system and sex hormones (G) and nervous system (N).

Step 1: QSAR model performance

QSAR models were built using 3 sets of chemical descriptors (Dragon, ISIDA, and MACCS; supplementary table S2) and 2 classification methods (RF[36] and SVM[37]). The 6 models and their consensus (single-vote average of 6 predictions) showed high accuracy characterized by AUC values of 75% to 81% (table 2). Coverage, which is defined as a fraction of drugs within the QSAR model applicability domain, was generally high for all models (97%–100%) although a few macrolides (eg, bleomycin) were too structurally dissimilar from other compounds to be predicted reliably. The *y*-randomization test showed that all models were unlikely to be fitted by chance ($P = 0.03$).

Step 2: Model interpretation (structural alerts)

We focused interpretation on the RF model built with ISIDA fragment descriptors, since many of these descriptors corresponded to chemical functional groups. We progressively rebuilt RF models using *f* most important[44] fragments to find the fewest number of fragments yielding the model with OOB error less than or equal to that of the full model using all 1091 fragments. We found empirically that this objective was met with $f = 29$ whereby the 29 fragments were the common overlap among the 5 sets of top 50 fragments from 5 models developed with 5-fold cross-validation (supplementary figure S1). These 29 fragments were used for subsequent analysis to identify structural alerts associated with SJS.

Although each of these 29 discriminatory fragments could *individually* serve as an indicator for SJS activity (or lack thereof), we hypothesized that some of them that occur frequently within the same drugs can be fused to generate larger substructures with potentially higher specificity as structural alerts for SJS. These *fused* structural alerts were uncovered by looking for significant pairwise co-occurrences in actives vs inactives using the Fisher exact test (figure 2a). The co-occurrences were also elucidated by considering a network of fragment nodes, connected whenever a pair co-occurred significantly (figure 2b). In the network, clusters of co-occurring fragments formed densely connected subnetworks (ie, communities), which were identified by the walktrap community detection method.[45] Within each community, some co-occurring fragments could be manually assembled into a larger substructure as an indicator for either SJS class. Of the 5 communities identified, 2 contained fragments that could be assembled into larger substructures, forming a novel structural alert for SJS activity (communities C1 and C2, figures 2b and c). The first community (C1, purple) consisted of 5 fragments corresponding to arylamines, sulfonylarenes, and sulfones that were assembled into a sulfonylarylamine structural alert, the substructure incorporating all of these fragments. The second community (C2, blue) formed a β-lactam substructure. The green and yellow communities were composed of aliphatic chains and secondary amines that more frequently occurred in inactives than in actives, forming a *safe* substructure (C3, C4, green and yellow, respectively). The remaining community (C5, pink) contained only 2 fragments, too small for any meaningful interpretation.

Table 1: Properties of SJS-active and inactive drugs used for QSAR modeling

	SJS-active drugs ($n = 194$)	SJS-inactive drugs ($n = 170$)
No. of SJS reports[a], mean (SD)	104 (262)	1.52 (4.10)
No. of ADR reports[a], mean (SD)	6953 (9849)	3505 (5416)
Anatomical Therapeutic Chemical (ATC) classification[b], mean (%)		
A: Alimentary tract and metabolism	26 (13)	14 (8)
B: Blood and blood forming organs	1 (1)	9 (5)
C: Cardiovascular system	18 (9)	25 (15)
D: Dermatologicals	24 (12)	12 (7)
G: Genitourinary system and sex hormones	9 (5)	21 (12)
H: Systemic hormonal preparations, excluding sex hormones and insulins	4 (2)	5 (3)
J: Anti-infectives for systemic use	81 (42)	5 (3)
L: Antineoplastic and immunomodulating agents	7 (4)	24 (14)
M: Musculoskeletal system	35 (18)	4 (2)
N: Nervous system	25 (13)	48 (28)
P: Antiparasitic products, insecticides, and repellents	5 (3)	1 (1)
R: Respiratory system	15 (8)	22 (13)
S: Sensory organs	38 (20)	13 (8)
V: Various	26 (13)	14 (8)

Abbreviations: SJS, Stevens-Johnson syndrome; QSAR, Quantitative Structure-Activity Relationship; ADR, adverse drug reaction.
[a]Significant difference ($P < .01$) by the Welch *t* test for unequal variances.
[b]Signficance was not determined for ATC because drugs could belong to multiple ATC.

Both structural alerts uncovered by the above co-occurrence analysis were consistent with those obtained from the second approach to identify larger significant fragments based on the MCS method. This concordance provides additional evidence that co-occurrence analysis is a valid method to derive structural alerts. However, MCS discovered 2 additional structural alerts, fluoroquinolones and tetracyclines (figures 3.3–3.4b) that were present only in a few drugs. Because of their rarity, their key-related fragments (eg, fluorinated groups, quinones) were not found among the 29 most important fragments analyzed for co-occurrences. Thus, co-occurrence analysis may be better suited for detecting substructures with the occurrence above some minimum frequency.

Substructures known to be associated with actives[22,47] are shown in figure 3. Such substructures were inferred from drug classes implicated with SJS such as sulfonamide antibiotics, penicillin, quinolones,

RESEARCH AND APPLICATIONS Downloaded from http://jamia.oxfordjournals.org/ by guest on May 24, 2016

Table 2: Performance of QSAR models[a]

Descriptors	Method	Specificity	Sensitivity	Balanced Accuracy	Area Under Curve	Coverage
354 Dragon	RF	0.71 (0.03)	0.77 (0.04)	0.74 (0.02)	0.81 (0.02)	0.97
354 Dragon	SVM	0.72 (0.03)	0.71 (0.04)	0.71 (0.02)	0.78 (0.02)	0.97
1091 ISIDA	RF	0.69 (0.03)	0.74 (0.04)	0.71 (0.02)	0.77 (0.02)	0.98
1091 ISIDA	SVM	0.68 (0.03)	0.71 (0.03)	0.69 (0.03)	0.75 (0.03)	0.98
138 MACCS	RF	0.74 (0.03)	0.72 (0.03)	0.73 (0.02)	0.80 (0.02)	1.00
138 MACCS	SVM	0.71 (0.03)	0.71 (0.03)	0.71 (0.02)	0.77 (0.03)	1.00
Consensus	–	0.73 (0.03)	0.74 (0.03)	0.73 (0.02)	0.79 (0.02)	1.00

Abbreviations: RF, Random Forest; SVM, support vector machines; ISIDA, In Silico Design and Data Analysis; MACCS, Molecular ACCess System; NA, not applicable.
[a]Results are presented as mean (with standard errors in parentheses) unless otherwise indicated.

and tetracyclines.[22,47] Our systematic chemical analysis found larger, more specific substructures (figure 3, right column) that were more likely to yield true positives (ie, higher precision). For example, the sulfonylarylamine structural alert (figure 3.1b) correctly identified drugs positively associated with SJS all 20 times it was present in a drug, unlike the sulfonamide structural alert (figure 3.1a) which falsely predicted the SJS activity for some sulfonamide drugs. In minimizing false positives, more precise structural alerts could spare drugs from wrongful association with SJS and leave more drug options available for use.

Step 3: Model application: prediction of SJS for drugs in DrugBank

We used the best QSAR model (Dragon-RF) for the virtual screening of DrugBank, assessing 4122 drug structures for potential SJS activity (supplementary table S3). Among the 10 most likely SJS-active drugs (excluding experimental drugs), 8 contained either the sulfonylarylamine or β-lactam with adjacent sulfur structural alert (figure 4). Among the 10 most likely inactives, etonogestrel, mestranol, and rapacuronium were chemically similar to many steroidal inactives in our reference set such as progesterone.

Step 4: Corroboration of model predictions of SJS-active and inactive drugs

We checked VigiBase and the literature (ChemoText[30] and MicroMedex[31]) for reports of SJS associated with the predicted SJS actives and inactives (table 3). Between predicted actives and predicted inactives, the former was associated with a higher number of SJS reports and higher IC values indicative of higher-than-expected SJS reporting in VigiBase and more instances of SJS in ChemoText and Micromedex. Despite the evidence in VigiBase, these predicted drugs were not included in our reference set for modeling as they were not obvious candidates for actives and inactives due to coreporting with other comedications and low usage (evident by a few ADR reports). When only a few ADR reports are available, we should not rely on the IC as the only indication of SJS, because their 95% credibility intervals are very wide. Nevertheless, the general trends in the IC values and other data sources showed that the predicted actives were associated with more SJS instances than predicted inactives, supporting our models' predictions.

DISCUSSION

To meet our objectives of developing, interpreting, and applying QSAR models of SJS as well as validating predictions made with these models, we have extracted spontaneous reports of SJS as primary data and generated QSAR classifiers that predicted SJS active and inactive drugs from chemical structures (AUC of 75%−81%; table 2), improving upon previous linear models (69%−73% accuracy) which used spontaneous reports from the United States only.[18] Models built with ISIDA fragment descriptors identified the most predictive fragments from which we further created new larger structural alerts associated with the active class (figure 2). Although these larger alerts were less prevalent, their precision in identifying actives was higher than previous smaller structural alerts (figure 3).

The additional chemical features encapsulated in our larger structural alerts offered important mechanistic clues. For example, although it is known that sulfonamides alone do not induce SJS,[48] sulfonamides antibiotics have long been implicated with SJS.[22] Instead, studies have attributed immunogenic reactions related to SJS to an arylamine group within the sulfonylarylamine[49] structural alert (figure 3.1b). The purported mechanism involves the metabolic transformation of the arylamine group into a reactive nitroso metabolite that covalently binds to cellular macromolecules to initiate an immune response consistent with the hapten hypothesis.[48–50] Arylamines are generally rare among drugs due to their reactivity. Exceptions are drugs such as sulfonamide antibiotics, which contain a sulfone group (SO_2) in the electron-withdrawing *para*-position to stabilize the arylamine against overt toxicity but not exculpate it from metabolizing into the nitroso culprit.[51]

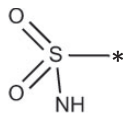
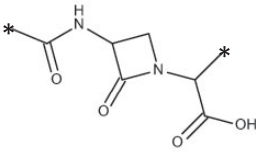
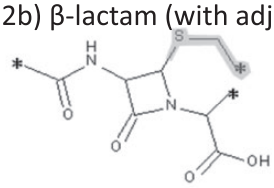
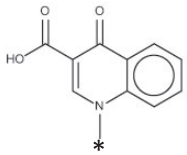
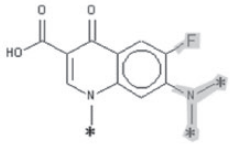
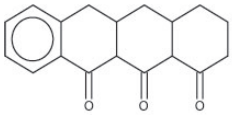
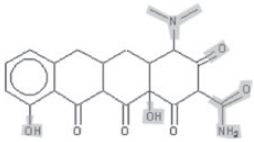
The other structural alert, β-lactam with adjacent sulfur (figure 3.2b), suggests that the additional sulfur atom may be necessary for SJS activity. By incorporating the adjacent sulfur atom into an extended alert, precision increases to 100% such that all β-lactam antibiotics containing this moiety are actives. Conversely, analogs without the adjacent sulfur atom such as latamoxef were inactives.

Our third structural alert refers to a fluoroquinolone (figure 3.3b) instead of quinolone as a known alert. However, because all the quinolones were also fluoroquinolones in the data set used for our study, we could not conclude that fluoroquinolone was a better alert. We note that such a distinction between the 2 may be irrelevant as most unfluorinated quinolones have been discontinued in favor of the more efficacious fluoroquinolones.[52]

Our fourth structural alert refers to tetracycline antibiotics instead of the more general four-ring system present in both tetracycline antibiotics and anthracyclines. In our study, all 6 tetracycline antibiotics were active, while all 3 anthracyclines were inactive. Their structural

Downloaded from http://jamia.oxfordjournals.org/ by guest on May 24, 2016

Figure 3: Structural alerts for SJS activity. Left column shows previously inferred substructures. Right column shows structural alerts uncovered in this study. Structural differences are highlighted in gray.
Abbreviation: SJS, Stevens-Johnson syndrome.

Previously reported structural alerts		Our augmented structural alerts	
3.1a) Sulfonamide 	29:5 (Toxic:Nontoxic) Precision=0.85	3.1b) Sulfonylarylamine	20:0 1.00
3.2a) β-lactam 	25:1 0.96	3.2b) β-lactam (with adjacent sulfur) 	24:0 1.00
3.3a) Quinolone 	6:1 0.86	3.3b) Fluoroquinolone 	6:1 0.86
3.4a) Tetracycline/anthracycline 	6:3 0.67	3.4b) Tetracycline 	6:0 1.00

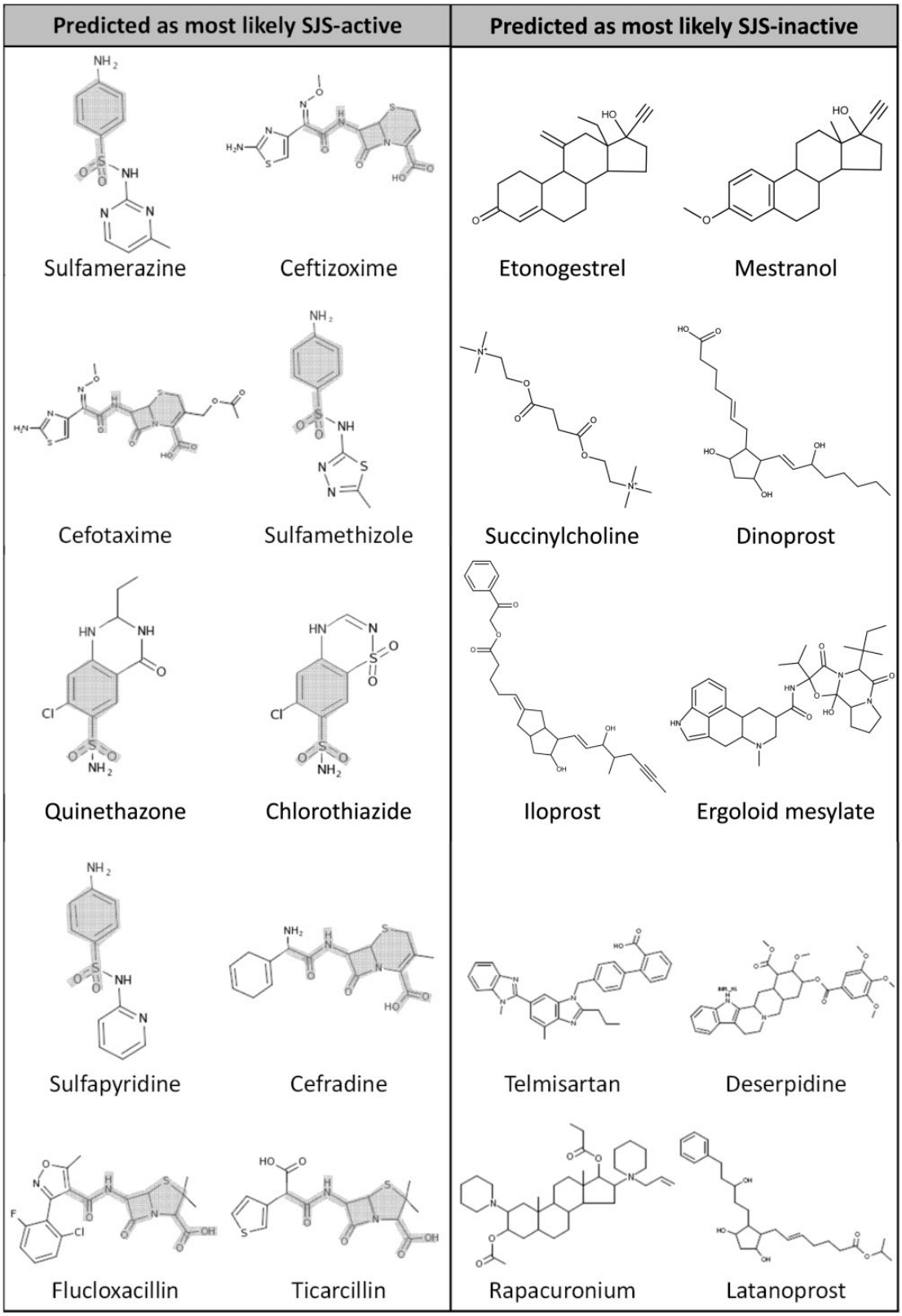
differences lie in the presence of a dimethylamine group and absence of a sugar ring in tetracycline antibiotics. By using a more refined structural alert that can differentiate the SJS-inducing tetracycline antibiotics from the noninducing anthracyclines, we improved the precision to 100%. Other substructures such as the aromatic ring have been suggested as a structure alert for anticonvulsants in a previous study.[53] However, we did not find this trend in our study while using our expanded set of drugs that included non-anticonvulsants.

In addition to identifying new alerts, we have also demonstrated the practical utility of QSAR models for identifying potential SJS-active drugs by virtual screening of the DrugBank library of 4122 drugs. Drugs predicted with high confidence were chemically similar to drugs in our reference set used for training QSAR models (figure 4). In addition, predicted actives were associated with more SJS reports than predicted inactives (table 3). The chemical structures found to be associated with SJS may be considered for inclusion in predictive models for postmarketing safety surveillance that simultaneously account for multiple aspects of strength of evidence.[54] Improved model prediction will prioritize drugs for targeted monitoring such that patients can be better monitored and more data can be collected to further improve the model.

A study limitation stems from the underreporting and reporting bias inherent in spontaneous reporting systems. Nevertheless, VigiBase remains the largest source of spontaneous ADR reports, thus providing the largest reference set for the analysis. Another limitation is that predictions by our QSAR models were not further validated against a clinically defined gold standard. Instead, only the 10 most likely active and inactive drugs were corroborated by evidence in existing databases. We note that the spontaneous reports may have contained additional information (eg, dose, route of administration) beneficial for prediction that we did not utilize. Using chemical structures alone might have curtailed predictivity (up to 81% AUC) as unaccounted nonchemical factors (eg, pharmacogenomics and metabolism[25,55]) may also contribute towards SJS activity. The relationship between chemical structures and toxicity is often more circuitous than QSAR models assume, involving many nonchemical factors that depend on the biological host (eg, pharmacogenomics affecting the pharmacokinetics of the drug) and the ADR of interest. Generally, QSAR models are more successful at predicting direct chemical-induced outcomes (eg, mutagenicity due to chemical adduction to DNA) than outcomes farther downstream of chemical-initiating events (eg, carcinogenicity that could arise from multiple modes of action affecting cell growth and repair etc).[56–59] Therefore, in many other ADRs

RESEARCH AND APPLICATIONS
Downloaded from http://jamia.oxfordjournals.org/ by guest on May 24, 2016

Figure 4: Most likely SJS actives and inactives predicted by QSAR model (Dragon-RF). Structural alerts, if any, were highlighted within the predicted drugs.
Abbreviations: SJS, Stevens-Johnson syndrome; QSAR, Quantitative Structure-Activity Relationships; RF, Random Forest.



Downloaded from http://jamia.oxfordjournals.org/ by guest on May 24, 2016

Table 3: Mostly likely SJS-active and inactive drugs in DrugBank (as predicted by Dragon-RF model)

				VigiBase			Chemo Text	Micromedex
DrugBank Identification	Predicted Value	standard deviation	Name	SJS Reports	All ADR Reports	IC[a]	SJS Articles[b]	SJS
Predicted Inducers (from DrugBank)								
DB01581	0.978	0.010	Sulfamerazine	0	1	−0.01	2	No
DB01332	0.967	0.006	Ceftizoxime	2	748	−0.26	0	Yes
DB00493	0.966	0.016	Cefotaxime	40	7550	0.66	15	Yes
DB00576	0.964	0.007	Sulfamethizole	5	490	1.37	5	Yes
DB01325	0.963	0.007	Quinethazone	0	25	−0.22	0	No
DB00880	0.959	0.040	Chlorothiazide	2	800	−0.34	1	Yes
DB00891	0.955	0.011	Sulfapyridine	0	29	−0.25	2	No
DB01333	0.951	0.012	Cefradine	3	994	−0.12	1	No[c]
DB00301	0.937	0.020	Flucloxacillin	29	5272	0.71	3	No
DB01607	0.937	0.006	Ticarcillin	2	338	−0.11	0	Yes
Predicted Noninducers (from DrugBank)								
DB00294	0.066	0.031	Etonogestrel	1	4443	−3.35	0	No[c]
DB01357	0.084	0.036	Mestranol	0	26	−0.23	2	No
DB00202	0.099	0.144	Succinylcholine	4	3581	−1.46	0	No
DB01160	0.100	0.080	Dinoprost	0	161	−1.05	0	No
DB01088	0.103	0.020	Iloprost	1	1518	−1.88	0	No
DB01049	0.109	0.037	Ergoloid mesylate	2	171	1.23	0	No
DB00966	0.110	0.023	Telmisartan	8	4845	−0.96	0	No[c]
DB01089	0.120	0.044	Deserpidine	0	9	−0.08	0	No
DB04834	0.123	0.079	Rapacuronium	0	112	−0.80	0	No
DB00654	0.124	0.042	Latanoprost	3	5423	−2.40	1	No[c]
Known Inducers (from reference set)								
NA	NA	NA	Sulfamethoxazole	37	971	1.43	104	Yes
NA	NA	NA	Amoxicillin	648	48501	1.36	44	Yes
Known Noninducers (from reference set)								
NA	NA	NA	Progesterone	0	3825	−4.72	0	No[c]
NA	NA	NA	Vardenafil	0	4506	−4.95	0	No

Abbreviations: SJS, Stevens-Johnson syndrome; RF, Random Forest; SD, standard deviation; IC, information component; ADR, adverse drug reaction; NA, not applicable.

[a]Information component (IC) is a disproportionality frequency measuring the number of SJS reports lower than or higher than expected in VigiBase.

[b]Number of articles in Medline matching the search terms [drugname] AND "Stevens-Johnson syndrome" [MeSH] OR "erythema multiforme" [MeSH] OR "epidermal necrolysis, toxic" [MeSH] Filters: Humans (as of Februrary 2013).

[c]Hypersensitivity reaction although SJS was not explicitly mentioned.

where nonchemical factors play larger roles, QSAR modeling may present even less optimistic results than our case of SJS, which is known to have a strong underlying chemical basis. Possible solutions may entail the development of hybrid QSAR models that incorporate nonchemical factors as additional variables[60,61] or modeling more specific ADRs (eg, building separate QSAR models specifically for mutagenicity and carcinogenicity).[58,59]

CONCLUSIONS

Using drug chemical structures and the largest database of ADR reports in the world, we have developed accurate and interpretable QSAR models for predicting drugs' association with SJS. Because QSAR models require only drug chemical structures, they enable efficient virtual screening of large drug libraries to prioritize potentially harmful drugs for focused surveillance and in-depth

RESEARCH AND APPLICATIONS Downloaded from http://jamia.oxfordjournals.org/ by guest on May 24, 2016

pharmacoepidemiologic evaluation, thereby limiting patient exposure to such medications.

CONTRIBUTORS

RE and AT made equal contributions as senior authors. YSL, AT, IRE, OC and GNN designed the study. YSL, OC, and XZ performed the analysis. YSL, OC, TB, DF, XZ, GNN, IR, RE, and AT wrote and approved the manuscript.

COMPETING INTERESTS

None.

FUNDING

The work has been supported in part by National Institutes of Health research grants GM066940 and R21GM076059 awarded to collaborators from the University of North Carolina at Chapel Hill.

DATA SHARING

Supplementary materials include list of drugs used for QSAR modeling (supplementary table S1), list of chemical descriptors used for QSAR modeling (supplementary table S2), QSAR-based predictions of DrugBank (supplementary table S3), and a figure showing out-of-bag error of random forest model using various most important fragments (supplementary figure S1).

ACKNOWLEDGEMENTS

We thank Johanna Strandell for VigiBase support, Nancy Baker for ChemoText support, and Stacie Dusetzina for pharmacoepidemiological consultation and manuscript advice.

SUPPLEMENTARY MATERIAL

Supplementary material is available online at http://jamia.oxfordjournals.org/.

REFERENCES

1. Wilson AM, Thabane L, Holbrook A. Application of data mining techniques in pharmacovigilance. *Br J Clin Pharmacol* 2003;57: 127–134. doi:10.1046/j.1365-2125.2003.01968.x.
2. Almenoff JS, Pattishall EN, Gibbs TG, *et al*. Novel statistical tools for monitoring the safety of marketed drugs. *Clin Pharmacol Ther* 2007;82:157–166. doi:10.1038/sj.clpt.6100258.
3. Harpaz R, DuMouchel W, Shah NH, *et al*. Novel data-mining methodologies for adverse drug event discovery and analysis. *Clin Pharmacol Ther* 2012;91:1010–1021. doi:10.1038/clpt.2012.50.
4. Trifirò G, Pariente A, Coloma PM, *et al*. Data mining on electronic health record databases for signal detection in pharmacovigilance: which events to monitor? *Pharmacoepidemiol Drug Saf*. 2009;18:1176–1184. doi:10.1002/pds.1836.
5. Brown JS, Kulldorff M, Chan KA, *et al*. Early detection of adverse drug events within population-based health networks: application of sequential testing methods. *Pharmacoepidemiol Drug Saf* 2007;16:1275–1284. doi:10.1002/pds.1509.
6. Shetty KD, Dalal SR. Using information mining of the medical literature to improve drug safety. *J Am Med Inform Assoc* 2011;18:668–674. doi:10.1136/amiajnl-2011-000096.
7. Campillos M, Kuhn M, Gavin A-C, Jensen LJ, Bork P. Drug target identification using side-effect similarity. *Science* 2008;321:263–266. doi:10.1126/science.1158140.
8. Pouliot Y, Chiang AP, Butte AJ; Nature Publishing Group. Predicting adverse drug reactions using publicly available PubChem BioAssay data. *Clin Pharmacol Ther* 2011;90:90–99. doi:10.1038/clpt.2011.81.
9. Matthews EJ, Kruhlak NL, Weaver J, *et al*. Assessment of the health effects of chemicals in humans: II. Construction of an adverse effects database for QSAR modeling. *Curr Drug Discov Technol* 2004;1:243–254. doi:10.2174/1570163043334794.
10. Hansch C, Maloney P, Fujita T, *et al*. Correlation of biological activity of phenoxyacetic acids with Hammett substituent constants and partition coefficients. *Nature* 1962;194:178–180. doi:10.1038/194178b0.
11. Hammett LP. The effect of structure upon the reactions of organic compounds. Benzene derivatives. *J Am Chem Soc* 1937;343:96–103. doi:10.1021/ja01280a022.
12. Collander R, Lindholm M, Haug CM, *et al*. The partition of organic compounds between higher alcohols and water. *Acta Chemica Scandinavica* 1951;5:774–780. doi:10.3891/acta.chem.scand.05-0774.
13. Taft RW. Linear free energy relationships from rates of esterification and hydrolysis of aliphatic and ortho-substituted benzoate esters. *J Am Chem Soc* 1952;74:2729–2732. doi:10.1021/ja01131a010.
14. Todeschini R, Consonni V. *Handbook of Molecular Descriptors*. Weinheim, Germany: Wiley-VCH Verlag GmbH; 2000.
15. Varnek A, Fourches D, Hoonakker F, *et al*. Substructural fragments: an universal language to encode reactions, molecular and supramolecular structures. *J Comput Aided Mol Des* 2005;19:693–703. doi:10.1007/s10822-005-9008-0.
16. Durant JL, Leland BA, Henry DR, *et al*. Reoptimization of MDL keys for use in drug discovery. *J Chem Inf Comput Sci* 2002;42:1273–1280. doi:10.1021/ci010132r.
17. Selassie C, Verma RP. History of quantitative structure–activity relationships. In: Abraham DJ, ed. *Burger's Medicinal Chemistry and Drug Discovery*. 7th ed. Hoboken, NJ: John Wiley & Sons, Inc; 2010:1–96. doi:10.1002/0471266949.bmc001.pub2.
18. Shirakuni Y, Okamoto K, Uejima E, *et al*. A practical estimation method for analyzing adverse drug reactions using data mining. *Drug Inf J* 2012;47:235–241. doi:10.1177/0092861512460759.
19. Rodgers AD, Zhu H, Fourches D, *et al*. Modeling liver-related adverse effects of drugs using knearest neighbor quantitative structure-activity relationship method. *Chem Res Toxico*. 2010;23:724–732. doi:10.1021/tx900451r.
20. Liu Z, Kelly R, Fang H, *et al*. Comparative analysis of predictive models for nongenotoxic hepatocarcinogenicity using both toxicogenomics and quantitative structure-activity relationships. *Chem Res Toxicol* 2011;24:1062–1070. doi:10.1021/tx2000637.
21. Gatnik MF, Worth A. *Review of Software Tools for Toxicity Prediction*. Luxembourg; Publications Office of the European Union, 2010.
22. Roujeau JC, Kelly JP, Naldi L, *et al*. Medication use and the risk of Stevens-Johnson syndrome or toxic epidermal necrolysis. *N Engl J Med* 1995;333:1600–1607. doi:10.1056/NEJM199512143332404.
23. Mockenhaupt M, Viboud C, Dunant A, *et al*. Stevens-Johnson syndrome and toxic epidermal necrolysis: assessment of medication risks with emphasis on recently marketed drugs. The EuroSCAR-study. *J Invest Dermatol* 2008;128:35–44. doi:10.1038/sj.jid.5701033.
24. Génin E, Schumacher M, Roujeau J-C, *et al*. Genome-wide association study of Stevens-Johnson Syndrome and Toxic Epidermal Necrolysis in Europe. *Orphanet J Rare Dis* 2011;6:52. doi:10.1186/1750-1172-6-52.
25. Lonjou C, Borot N, Sekula P, *et al*. A European study of HLA-B in Stevens-Johnson syndrome and toxic epidermal necrolysis related to five high-risk drugs. *Pharmacogenet Genomics* 2008;18:99–107. doi:10.1097/FPC.0b013e3282f3ef9c.
26. Chung W-H, Hung S-I, Hong H-S, *et al*. Medical genetics: a marker for Stevens-Johnson syndrome. *Nature* 2004;428:486. doi:10.1038/428486a.
27. Reilly TP, Ju C. Mechanistic perspectives on sulfonamide-induced cutaneous drug reactions. *Curr Opin Allergy Clin Immunol* 2002;2:307–315.
28. Lindquist M. VigiBase, the WHO global ICSR database system: basic facts. *Drug Inf J* 2008;42:409–419.
29. Wishart DS, Knox C, Guo AC, *et al*. DrugBank: a knowledgebase for drugs, drug actions and drug targets. *Nucleic Acids Res* 2008;36:D901–D906. doi:10.1093/nar/gkm958.
30. Baker NC, Hemminger BM. Mining connections between chemicals, proteins, and diseases extracted from Medline annotations. *J Biomed Inform* 2010;43:510–519. doi:10.1016/j.jbi.2010.03.008.
31. Micromedex Healthcare Series Databses: DRUGDEX System. Greenwood Village, CO: Truven Health Analytics; 2012. http://www.thomsonhc.com/ Accessed August 10, 2012.

Downloaded from http://jamia.oxfordjournals.org/ by guest on May 24, 2016

32. Caster O, Norén GN, Madigan D, *et al*. Large-scale regression-based pattern discovery: the example of screening the WHO global drug safety database. *Stat Anal Data Min* 2010;3:197–208. doi:10.1002/sam.10078.
33. Bate A, Lindquist M, Edwards IR, *et al*. A Bayesian neural network method for adverse drug reaction signal generation. *Eur J Clin Pharmacol* 1998;54:315–321.
34. Norén GN, Hopstadius J, Bate A. Shrinkage observed-to-expected ratios for robust and transparent large-scale pattern discovery. *Stat Methods Med Res* 2013;22:57–69. doi:10.1177/0962280211403604.
35. Fourches D, Muratov E, Tropsha A. Trust, but verify: on the importance of chemical structure curation in cheminformatics and QSAR modeling research. *J Chem Inf Model* 2010;50:1189–1204. doi:10.1021/ci100176x.
36. Breiman L. Random forests. *Mach Learn*.2001;45:5–32. doi:10.1023/A:1010933404324.
37. Vapnik VN. *The Nature of Statistical Learning Theory*. New York: Springer; 2000.
38. Tropsha A, Golbraikh A. Predictive QSAR modeling workflow, model applicability domains, and virtual screening. *Curr Pharm Des* 2007;13:3494–3504.
39. Svetnik V, Liaw A, Tong C, *et al*. Random forest: a classification and regression tool for compound classification and QSAR modeling. *J Chem Inf Comput Sci* 2003;43:1947–1958. doi:10.1021/ci034160g.
40. Tropsha A, Gramatica P, Gombar VK. The importance of being earnest: validation is the absolute essential for successful application and interpretation of QSPR models. *Qsar Comb Sci* 2003;22:69–77. doi:10.1002/qsar.200390007.
41. Tetko I V, Sushko I, Pandey AK, *et al*. Critical assessment of QSAR models of environmental toxicity against Tetrahymena pyriformis: focusing on applicability domain and overfitting by variable selection. *J Chem Inf Model* 2008;48:1733–1746. doi:10.1021/ci800151m.
42. Efron B, Tibshirani R. Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy. *Stat Sci* 1986;1:54–75. doi:10.1214/ss/1177013815.
43. Wold S, Eriksson L. Statistical validation of QSAR results. In: van de Waterbeemd H, ed. *Chemometrics Methods in Molecular Design*. Weinheim, Germany: VCH; 1995:309–318.
44. Strobl C, Boulesteix A-L, Kneib T, *et al*. Conditional variable importance for random forests. *BMC Bioinformatics* 2008;9:307. doi:10.1186/1471-2105-9-307.
45. Pons P, Latapy M. *Computing communities in large networks using random walks*. arXiv Journal: Physics and Society 2005;1–20.
46. Chakravarti SK, Saiakhov RD, Klopman G. Optimizing predictive performance of CASE Ultra expert system models using the applicability domains of individual toxicity alerts. *J Chem Inf Model* 2012;52:2609–2618. doi:10.1021/ci300111r.
47. Roujeau J-C, Bricard G, Nicolas J-F. Drug-induced epidermal necrolysis: Important new piece to end the puzzle. *J Allergy Clin Immunol* 2011;128:1277–1278. doi:10.1016/j.jaci.2011.10.015.
48. Toler SM, Rodriguez I. Not all sulfa drugs are created equal. *Ann Pharmacother* 2004;38:2166–2167. doi:10.1345/aph.1E206.
49. Brackett CC, Singh H, Block JH. Likelihood and mechanisms of cross-allergenicity between sulfonamide antibiotics and other drugs containing a sulfonamide functional group. *Pharmacotherapy* 2004;24:856–870. doi:10.1592/phco.24.9.856.36106.
50. Naisbitt DJ, Hough SJ, Gill HJ, *et al*. Cellular disposition of sulphamethoxazole and its metabolites: implications for hypersensitivity. *Br J Pharmacol* 1999;126:1393–1407. doi:10.1038/sj.bjp.0702453.
51. Uetrecht J. N-oxidation of drugs associated with idiosyncratic drug reactions. *Drug Metab Rev* 2002;34:651–665. doi:10.1081/DMR-120005667.
52. King DE, Malone R, Lilley SH. New classification and update on the quinolone antibiotics. *Am Fam Physician* 2000;61:2741–2748.
53. Handoko KB, van Puijenbroek EP, Bijl AH, *et al*. Influence of chemical structure on hypersensitivity reactions induced by antiepileptic drugs: the role of the aromatic ring. *Drug Saf* 2008;31:695–702.
54. Caster O, Juhlin K, Watson S, *et al*. Improved statistical signal detection in pharmacovigilance by combining multiple strength-of-evidence aspects in vigiRank. *Drug Saf* 2014;37:617–628. doi:10.1007/s40264-014-0204-5.
55. Wei C-Y, Ko T-M, Shen C-Y, *et al*. A recent update of pharmacogenomics in drug-induced severe skin reactions. *Drug Metabol Pharmacokinet* 2012:132–141. doi:10.2133/dmpk.DMPK-11-RV-116.
56. Stouch TR, Kenyon JR, Johnson SR, *et al*. In silico ADME/Tox: why models fail. *J Comput Aided Mol Des* 2003;17:83–92.
57. Penzotti JE, Landrum GA, Putta S. Building predictive ADMET models for early decisions in drug discovery. *Curr Opin Drug Discov Devel* 2004;7:49–61.
58. Cherkasov A, Muratov EN, Fourches D, *et al*. QSAR Modeling: where have you been? Where are you going to? *J Med Chem* 2014;57:4977–5010. doi:10.1021/jm4004285.
59. Benigni R, Bossa C. Mechanisms of chemical carcinogenicity and mutagenicity: a review with implications for predictive toxicology. *Chem Rev* 2011;111:2507–2536. doi:10.1021/cr100222q.
60. Rusyn I, Sedykh A, Low Y, *et al*. Predictive modeling of chemical hazard by integrating numerical descriptors of chemical structures and short-term toxicity assay data. *Toxicol Sci* 2012;127:1–9. doi:10.1093/toxsci/kfs095.
61. Low YS, Sedykh AY, Rusyn II, *et al*. Integrative approaches for predicting in vivo effects of chemicals from their structural descriptors and the results of short-term biological assays. *Curr Top Med Chem* 2014;14:1356–1364. doi:10.2174/1568026614666140506121116.

AUTHOR AFFILIATIONS

[1]Division of Chemical Biology and Medicinal Chemistry, Eshelman School of Pharmacy, University of North Carolina, Chapel Hill, North Carolina, USA

[2]Department of Environmental Sciences and Engineering, Gillings School of Public Health, University of North Carolina, Chapel Hill, North Carolina, USA

[3]Uppsala Monitoring Centre, Uppsala, Sweden

[4]Department of Computer and Systems Sciences, Stockholm University, Kista, Sweden

[5]Department of Mathematics, Stockholm University, Stockholm, Sweden

Downloaded from http://jamia.oxfordjournals.org/ by guest on May 24, 2016