J Comput Aided Mol Des (2007) 21:139–144
DOI 10.1007/s10822-007-9109-z

ORIGINAL PAPER

Enhancement of chemical rules for predicting compound reactivity towards protein thiol groups

James T. Metz · Jeffrey R. Huth · Philip J. Hajduk

Received: 20 November 2006 / Accepted: 16 January 2007 / Published online: 6 March 2007

Abstract Non-specific chemical modification of protein thiol groups continues to be a significant source of false positive hits from high-throughput screening campaigns and can even plague certain protein targets and chemical series well into lead optimization. While experimental tools exist to assess the risk and promiscuity associated with the chemical reactivity of existing compounds, computational tools are desired that can reliably identify substructures that are associated with chemical reactivity to aid in triage of HTS hit lists, external compound purchases, and library design. Here we describe a Bayesian classification model derived from more than 8,800 compounds that have been experimentally assessed for their potential to covalently modify protein targets. The resulting model can be implemented in the large-scale assessment of compound libraries for purchase or design. In addition, the individual substructures identified as highly reactive in the model can be used as look-up tables to guide chemists during hit-to-lead and lead optimization campaigns.

Keywords Bayesian classifier · Compound reactivity · ALARM-NMR · Pipeline pilot

Electronic supplementary material The online version of this article (doi:10.1007/s10822-007-9109-z) contains supplementary material, which is available to authorized users.

J. T. Metz · J. R. Huth · P. J. Hajduk (✉)
Pharmaceutical Discovery Division, Abbott Laboratories, R46Y, AP-10, 100 Abbott Park Road, Abbott Park, IL 60064-3500, USA
e-mail: philip.hajduk@abbott.com

Introduction

The ability of a small organic compound to covalently modify (e.g., oxidize or form a covalent adduct with) a protein is a surprisingly common phenomenon in drug discovery [1–3]. Against certain targets, such as cysteine proteases, a covalent mechanism can be exploited during drug design, provided that sufficient selectivity can be achieved [4]. Targeting cysteine residues for covalent attack can also enhance the specificity of existing inhibitors, as has recently been demonstrated for protein kinases [5]. However, in the vast majority of cases, compound reactivity is generally avoided due to increased risks for organ toxicity—especially in the liver [6, 7]. In order to address this issue, a number of experimental assays have been developed to identify reactive compounds or reactive metabolites, including fluorescent- and NMR-based experiments [2, 3]. Such assays are of sufficient throughput for testing thousands of compounds for their ability to covalently modify protein targets.

In addition to experimental assays, a number of computational approaches have been developed that predict compound reactivity [8]. Such algorithms are especially useful for evaluating external compounds for purchase or proposed compounds that have not yet been synthesized—for which an experimental test cannot be performed. These computational approaches typically involve the identification of certain groups or substructures, such as quinones, that are known to be highly reactive. In our initial description of ALARM-NMR [3], we reported on the use of a group contribution model for predicting compound reactivity based on a test set of ~3,500 compounds. Reactivity in this model was defined as the ability to covalently modify a

cysteine in the La antigen, which has been demonstrated to contain a highly reactive cysteine that can serve as a surrogate protein for measuring general, non-specific compound reactivity [3]. Statistical analyses of the experimental data revealed that certain substructures or functional groups (e.g., quinones, sulfhydryls, alkyl halides, etc) are associated with high frequencies of chemical reactivity (e.g., compounds containing these substructures are highly likely to react with protein thiol groups). In addition to these known structures, a number of new substructures were identified that also exhibit reactivity with protein thiols and can be utilized in computational analyses of compound libraries. All of this data was included in the development of an in silico model for flagging compounds that have a high probability of reacting with protein thiol groups.

Here we report on an enhancement of these rules using a Bayesian classification model and a dataset of more than 8,800 compounds. It was found that models built using the initial dataset of 3,500 compounds exhibited reasonable overall prediction accuracies for the new data (~77%). However, the initial model did not accurately predict many of the reactive compounds in the new dataset, exhibiting a true-positive prediction accuracy of only 44%. Examination of the data revealed that the poor true-positive prediction rate was due to the presence of a large number of new substructures that are highly correlated with reactivity towards proteins but were not present in the original compound set. Thus, a new Bayesian classification model was constructed using all of the data, which yielded more than 100 new chemical groups that are associated with covalent modification of protein thiol groups. These substructures can be incorporated into computational algorithms for assessing compound reactivity and improve library enhancement initiatives and the hit-to-lead and lead optimization processes.

Results and discussion

In silico prediction of compound reactivity using a Bayesian classification model

Our original analysis of compound reactivity [3] using a 3,500 compound dataset consisted of the following multi-step process: (1) collation of the ALARM-NMR data, (2) generation of SMILES strings, (3) substructure/functional group identification using a modified RECAP procedure [9], (4) generation of substructure fingerprints for each compound using Daylight tools [10], and (5) calculations in Microsoft Excel for construction of the group contribution model. All of these steps were performed manually and, therefore, this analysis did not lend itself to rapid re-evaluation as more data became available. As an alternative, we wanted to take advantage of the powerful workflow and data analysis capabilities within Pipeline Pilot [11]. A workflow was constructed that incorporated all of the steps in the original analysis, except that substructure identification was performed using extended connectivity fingerprints (ECFPs) [12] and reactivity predictions were made using a Bayesian classification model [13–15]—both of which are available components within the Pipeline Pilot suite of analysis tools. Extended connectivity fingerprints are a class of 2D descriptors that define atoms by their ''neighborhood'' of surrounding atoms up to a defined number of bonds. Thus, ECFP_6 contains all substructures, around each atom, up to a maximum width of six bonds. This results in a large list of substructures that are similar but not identical to those identified using the RECAP procedure [9].

In order to demonstrate that the Bayesian classification model with extended connectivity fingerprints produced results similar to the previous analysis, we re-analyzed the initial 3,500 compound dataset using the new procedure. As expected, the Bayesian model performed comparably to the original group contribution model (see Table 1), correctly classifying ~74% (376/505) of the known reactive compounds, with an overall predictive accuracy of 78% (2,697/3,469). The significant ECFP_6 substructures identified in the Bayesian model were also similar to the reactive groups identified in the previous analysis, as shown in Table 2. Importantly, the Bayesian model produces probabilities associated with the propensity for compound reactivity, which are analogous to the Thiol Reactivity Indices (TRIs) reported in the original work [3]. As a cumulative Bayesian score (the sum over all Bayesian scores for each fingerprint in the compound) greater than or equal to 1.0 was classified as reactive in this model, large positive values (e.g., > 0.6) for any individual chemical fingerprint can be regarded as highly correlated with chemical reactivity (see below).

Performance with naïve data

This Bayesian model was then used to predict compound reactivity for a set of ~5,000 new compounds that were not part of the training set (see Table 1). The overall predictive accuracy with the naïve dataset was 77% (4,142/5,385), which, on the surface, compares extremely well with the overall accuracy of 78% achieved with the training set. However, a closer

Table 1 Contingency tables of experimental compound reactivity and predicted reactivity from Bayesian classification models using (A) the original 3,469 compound training set, (B) a 5,385 naïve compound prediction set (not used in A), and (C) the entire 8,838 compound set. The number of compounds in C (8,838) is 16 less than the sum of A and B (8,854) because 16 of the compounds in the original dataset were reclassified with respect to their ability to covalently modify the La protein, and hence were removed from the analysis

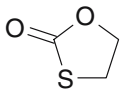
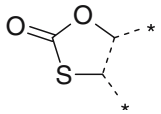
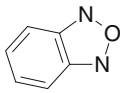
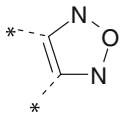
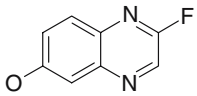
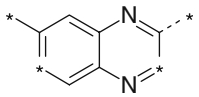
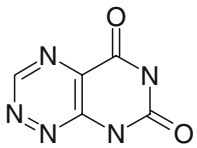
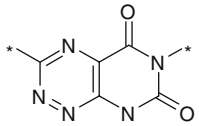
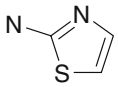
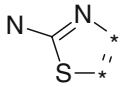
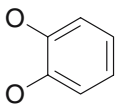
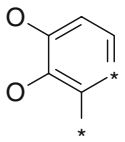
		Experimental		Totals
		Non-reactive	Reactive	
A. Original Data				
Predicted	Non-reactive	2,321	129	2,450
	Reactive	643	376	1,019
	Totals	2,964	505	3,469
Overall Accuracy: 78%				
True Positive Accuracy: 74%				
True Negative Accuracy: 78%				
B. Naïve data				
Predicted	Non-reactive	3,929	274	4,203
	Reactive	969	213	1,182
	Totals	4,898	487	5,385
Overall Accuracy: 77%				
True Positive Accuracy: 44%				
True Negative Accuracy: 80%				
C. All Data				
Predicted	Non-reactive	6,589	294	6,883
	Reactive	1,263	692	1,995
	Totals	7,852	986	8,838
Overall Accuracy: 82%				
True Positive Accuracy: 70%				
True Negative Accuracy: 84%				

inspection of the data reveal that this predictive accuracy is biased by the disproportionate number of non-reactive compounds in the dataset (90% of the compounds are non-reactive). Thus, the overall accuracy is dominated by the true-negative rate of 80% (3,929/4,898). The true-positive rate for the naïve data is only 44% (213/487), as compared to ~74% in the training set. This significant decrease in true-positive performance is indicative that the naïve data contained many compounds that were outside of the applicability domain [16] for this Bayesian model of compound reactivity. Thus, the 44% of correctly classified reactive compounds contained substructures that were already present at high abundance in the original training set. This is a major limitation of any group contribution model, in that the prediction capability will be limited by the chemical diversity contained in the training set and the similarity of the naïve compounds to the reference compounds. Interestingly, the false-positive rate of 20% (969/4,898) for the naïve data compares very well with the false-positive rate of 22% (643/2,964) achieved for the training set. This again indicates that the decrease in true-positive performance with the naïve data is the result of new chemical matter present in the dataset, and not the inappropriate weighting of existing substructures.

Identification of additional reactive substructures

We then constructed a new Bayesian model using the entire set of 8,838 compounds as the training set (see Table 1). The overall accuracy for this model was 82% (7,281/8,838). Importantly, the true-positive rate for this model was 70% (692/986)—a substantial improvement over the results with the previous model. The false positive rate for this model is 16% (1,263/7,852)—consistent with the low false-positive rate observed in the analyses described above. A comparison of the ECFP_6 fingerprints with high Bayesian scores in the old and new models reveals that the increase in true-positive performance is due primarily to the identification of more than 100 new substructures that are highly associated with chemical reactivity. A subset of these new substructures is shown in Fig. 1. For example, more than 50% of compounds that contain benzoxanthian, phenylimine, or divinylketone substructures were reactive in the ALARM-NMR assay, and were accordingly assigned Bayesian scores greater than 1.0 in the model. However, these structures were not statistically represented in the original dataset and could not be accurately classified with the initial model. Interestingly, many of the new substructures are similar but not identical to those identified in the original work.

Table 2 Comparison of a subset of common substructures identified in the original group contribution model with the ECFP_6 fingerprints from the Bayesian classification analysis

Name	Substructure[a]	F (%)[b]	TRI[c]	ECFP_6[d]	Bayesian Score
2-oxo-1,3-oxathiolane		85	0.30		1.76
benzofurazan		48	0.30		0.93
quinoxaline		47	0.30		0.98
xanthine		37	0.30		0.62
aminothiazole		30	0.30		0.92
o-catechol		22	0.30		1.13
sulfoxide		26	0.30		1.10
thioamide		24	0.30		1.01

[a] Substructure associated with high rates of chemical reactivity towards protein thiol groups identified in the original analysis of ALARM NMR data [3]

[b] Percentage of compounds containing this substructure that were experimentally classified as chemically reactive

[c] Thiol Reactivity Index as previously described [3]

[d] Extended connectivity fingerprint [12] identified in the Bayesian classification model. The asterisks denote any atom and the dashed lines denote any bond type

This suggests, at least for chemically reactive moieties, that the applicability domain for this model can be expanded at least in part by considering not only the exact substructures associated with chemical reactivity but also highly related structures. This possibility is currently being explored within our group.

The correlation between the normalized Bayesian probabilities for all 3,665 fingerprints and the percentage of compounds containing each substructure that were experimentally classified as chemically reactive is shown in Fig. 2. Of note is the fact that this correlation is not linear at least in part because of the Laplacian correction implemented in the Bayesian analysis that adjusts the probability estimates for each fingerprint to account for different sampling frequencies [13, 15]. It can be seen from this plot that Bayesian probabilities in excess of 1.0 correspond to experimental frequencies of chemical reactivity with protein thiol groups in the range of 30–100%—substantially higher than the average frequency of ~11% (red dashed line in Fig. 2). A Bayesian score of ~0.6 is consistent with an approximately two-fold increase in the probability that a compound containing that substructure will be chemically reactive. Thus, even in the absence of a sophisticated classification model, compounds that contain substructures with associated

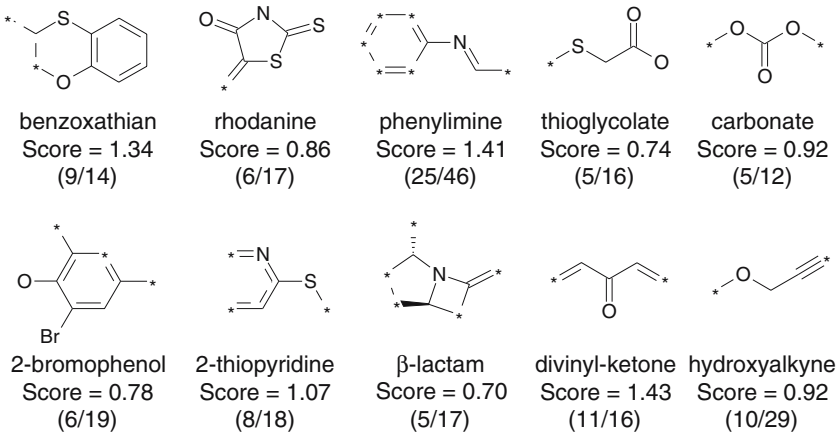


Fig. 1 A subset of new fingerprints associated with high rates of chemical reactivity towards protein thiol groups. Listed are the substructures (where the asterisks denote any atom and the dashed lines denote any bond type), a common name, and the Bayesian score. The number of reactive and total molecules containing this substructure in the 8,838 compound dataset is given in parentheses

Bayesian scores greater than 1.0 have a high likelihood of reacting with protein thiol groups, while those that contain substructures with scores between 0.6 and 1.0 should be carefully examined. An extended list of 175 substructures that were assigned Bayesian scores greater than 0.6 in the final model is given in Table S1 (Supplemental Material), along with the experimental frequencies of observed reactivity. It is important to stress that the 175 substructures listed in Table S1 contain new information that is not captured by conventional nuisance alerts [8]. In fact, only 70 of the 175 structures listed in Table S1 were already included as part of our corporate in silico nuisance alert protocols.

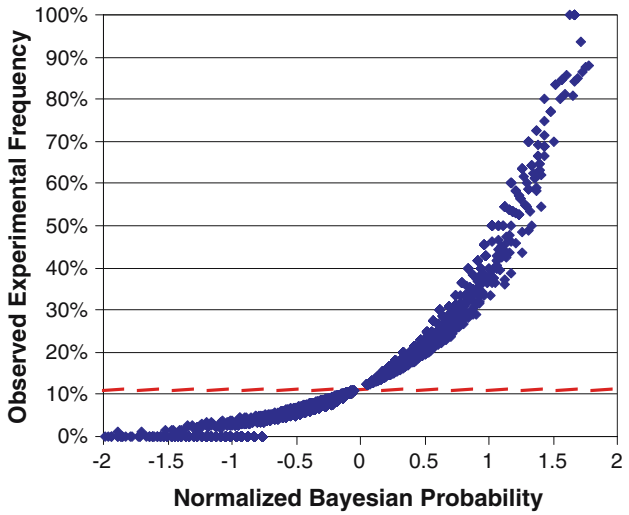


Fig. 2 Plot of the normalized Bayesian probabilities for the 3,665 fingerprints described in the text against the percentage of compounds containing each substructure that were experimentally classified as chemically reactive. The average frequency for observed chemical reactivity (11.2%) is shown with a red dashed line

Summary

Chemical reactivity continues to be a significant source of false positive activity in drug research. It is also clear from the above analysis that, in our ongoing application of ALARM NMR, we continue to find new chemical moieties that exhibit high rates of chemical reactivity with protein thiol groups. While such experimental approaches for assessing chemical reactivity are reliable, they cannot be used to routinely assess entire corporate compound repositories or for evaluating compounds for purchase or synthesis. Thus, the cheminformatic approaches for predicting compound reactivity described here are important tools that can aid in compound triage. The list of substructures given in Table S1 can serve as a simple look-up table for the research scientist to visually inspect the propensity for a given compound to react with protein thiol groups. However, for large numbers of compounds, the Bayesian classification model offers a rapid and comprehensive assessment of potential reactivity. While the classification model described here has an acceptable false positive rate (16%), it is imperative that it be continuously updated with new data in order to capture the largest fraction of reactive groups. The construction and utilization of group contribution models in the Pipeline Pilot environment enables the rapid and facile incorporation of new data to create enhanced models that can improve Discovery research.

Methods

A Bayesian classifier model [13] of the original dataset [3] was created using several components in Pipeline Pilot 5.1.0 in the following steps. A list of Abbott compounds from the original 3,500 compound dataset

and experimental activities were read from an Excel spreadsheet. Structures were classified as either ALARM reactive or non-reactive in the corporate database. Structures were then retrieved from the Abbott corporate database and desalted. Compounds were removed for which the reactivity classification was inconsistent for replicate experiments in the Abbott database. This resulted in an initial dataset of 3,469 compounds. Extended connectivity fingerprints (ECFP_6) [12] were generated and used as descriptors for the initial dataset. Initially, 409,948 fingerprints were calculated over all 8,838 compounds in the dataset. Removal of duplicate fingerprints and those with low occurrence (less than 10 examples in the dataset) resulted in 3,851 fingerprints. Finally, 186 non-informative features (those with normalized estimates of 0 ± 0.05) were removed, resulting in a final filtered list of 3,665 fingerprints. These data were then used to construct a Bayesian classifier model as implemented within Pipeline Pilot [13]. The performance of the initial model is shown in Table 1. The overall accuracy is 78%, while the classification accuracy for the reactive group is 74%.

An up-dated set of Abbott compounds with experimentally determined ALARM reactivity classifications were then downloaded from the Abbott corporate database (8,838 compounds). The downloaded set was then compared to the 3,469 compound initial set and new compounds were then desalted, checked for experimental consistency and the resulting 5,385 compounds which represented new structures were separated as a prediction set. The performance of the initial model against the 5,385 compound prediction set is shown in Table 1. The overall accuracy is 77%, while the classification accuracy for the reactive group is 44%.

Finally, all 8,838 compounds were used to create the final Bayesian classifier model. It should be noted that 16 of the compounds in the original dataset were reclassified with respect to their ability to covalently modify the La protein, and hence were removed from the analysis. The performance of the final model is shown in Table 1. The overall accuracy is 82%, while the classification accuracy for the reactive group is 70%. Fragment smiles generated from the ECFP_6 fingerprints were passed into a first occurrence filter component to determine which substructures were learned going from the initial to the final model. Selected new fragment smiles with high Bayesian scores are shown in Fig. 1.

Acknowledgments JTM would like to thank the support staff at SciTegic for many helpful discussions, suggestions, and corrections to Pipeline Pilot protocols.

References

1. Rishton GM (1997) Drug Discov Today 2:382
2. Epps DE, Raub TJ, Hezdy FJ (1995) Anal Biochem 227:342
3. Huth JR, Mendoza R, Olejniczak ET, Johnson RW, Cothron DA, Liu Y, Lerner CG, Chen J, Hajduk PJ (2005) J Am Chem Soc 127:217
4. Vicik R, Busemann M, Baumann K, Schirmeister T (2006) Curr Top Med Chem 6:331
5. Cohen MS, Zhang C, Shokat KM, Taunton J (2005) Science 308:1318
6. Evans DC, Watt AP, Nicoll-Griffith DA, Baillie TA (2004) Chem Res Toxicol 17:3
7. Zhou S, Chan E, Duan W, Huang M, Chen YZ (2005) Drug Metab Rev 37:41
8. Roche O, Schneider P, Zuegge J, Guba W, Kansy M, Alanine A, Bleicher K, Danel F, Gutknecht EM, Rogers-Evans M, Neidhart W, Stalder H, Dillon M, Sjogren E, Fotouhi N, Gillespie P, Goodnow R, Harris W, Jones P, Taniguchi M, Tsujii S, von der Saal W, Zimmermann G, Schneider G (2002) J Med Chem 45:137
9. Lewell XQ, Judd DB, Watson SP, Hann MM (1998) J Chem Inf Comput Sci 38:511
10. Daylight, http://www.daylight.com
11. Pipeline Pilot, SciTegic, Inc. 10188 Telesis Court, Suite 100, San Diego, CA 92121, USA, http://www.scitegic.com/products_services/pipeline_pilot.htm
12. Hassan M, Brown RD, Varma-O'brien S, Rogers D (2006) Mol Divers 10:283
13. Xia X, Maliski EG, Gallant P, Rogers D (2004) J Med Chem 47:4463
14. Glick M, Klon AE, Acklin P, Davies JW (2004) J Biomol Screen 9:32
15. Glick M, Jenkins JL, Nettles JH, Hitchings H, Davies JW (2006) J Chem Inf Model 46:193
16. Dimitrov S, Dimitrova G, Pavlov T, Dimitrova N, Patlewicz G, Niemela J, Mekenyan O (2005) J Chem Inf Model 45:839