JOURNAL OF CHEMICAL INFORMATION AND MODELING

Article

pubs.acs.org/jcim

Automated Extraction of Information on Chemical−P-glycoprotein Interactions from the Literature

Shuya Yoshida,[†] Fumiyoshi Yamashita,*[,†] Atsushi Ose,[‡] Kazuya Maeda,[§] Yuichi Sugiyama,[∥] and Mitsuru Hashida[†,⊥]

[†]Department of Drug Delivery Research, Graduate School of Pharmaceutical Sciences, and [⊥]Institute for Integrated Cell-Material Sciences, Kyoto University, 46-29 Yoshidashimoadachi-cho, Sakyo-ku, Kyoto 606-8501, Japan

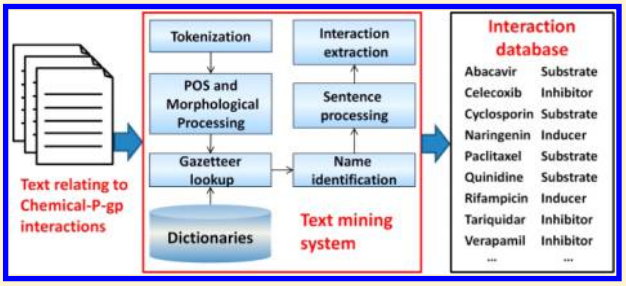
[‡]Development Planning, Clinical Development Center, Asahi Kasei Pharma Corporation, 1-105 Kanda Jinbocho, Chiyoda-ku, Tokyo 101-8101, Japan

[§]Department of Molecular Pharmacokinetics, Graduate School of Pharmaceutical Sciences, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033, Japan

[∥]Sugiyama Laboratory, RIKEN Innovation Center, Research Cluster for Innovation, RIKEN, 1-6 Suehiro-cho, Tsurumi-ku, Yokohama 230-0045, Japan

Supporting Information

ABSTRACT: Knowledge of the interactions between drugs and transporters is important for drug discovery and development as well as for the evaluation of their clinical safety. We recently developed a text-mining system for the automatic extraction of information on chemical−CYP3A4 interactions from the literature. This system is based on natural language processing and can extract chemical names and their interaction patterns according to sentence context. The present study aimed to extend this system to the extraction of information regarding chemical−transporter interactions. For this purpose, the key verb list designed for cytochrome P450 enzymes was replaced with that for known drug transporters. The performance of the system was then tested by examining the accuracy of information on chemical−P-glycoprotein (P-gp) interactions extracted from randomly selected PubMed abstracts. The system achieved 89.8% recall and 84.2% precision for the identification of chemical names and 71.7% recall and 78.6% precision for the extraction of chemical−P-gp interactions.



INTRODUCTION

Transporters are expressed in various tissues and mediate the cellular uptake and efflux of a broad variety of endogenous and exogenous compounds, including drugs. The importance of drug transporters in modulating drug absorption, distribution, metabolism, and excretion (ADME) is increasingly recognized, as well as the relevance thereof to drug−drug interactions.[1] Knowledge of the interactions between chemicals and transporters is therefore critical for drug discovery and development as well as for the evaluation of their clinical safety.

There are several articles devoted to databases and models of chemical−transporter interactions.[2−4] In the majority of these studies, information on transporters was collected by a search of available literature comprising research papers, review articles, pharmacology textbooks, and other relevant publications. However, because of the exponential growth of biomedical literature and the resulting increase of available data, it has become more difficult for researchers to keep up with publication outputs without the assistance of computers. There is therefore an increasing need for systems that enable researchers to automatically collect and organize information from these publications.

We recently developed a text-mining system based on natural language processing (NLP), for the automatic extraction of information from the literature on chemical−CYP3A4 interactions (i.e., between CYP3A4 and substrates, inducers, and inhibitors).[5] The system identified chemicals and CYP3A4 name variants in the text according to a combination of name dictionaries and context features, and extracted information on chemical−CYP3A4 interactions based on the order of three keywords: chemical names, CYP3A4 name variants, and key verbs. Although this is a simple pattern-matching method, the system achieved 87.4% recall and 92.3% precision for the identification of chemical names and 85.2% recall and 92.0% precision for the extraction of chemical−CYP3A4 interactions. More recently, we extended this system for application to different CYP isoforms.[6]

On the basis of its system architecture, this NLP system was also expected to perform well when extracting information on the interactions of chemicals with transporters. To validate the applicability of the text-mining system, a "verb" dictionary

Received: May 29, 2013
Published: September 6, 2013

dx.doi.org/10.1021/ci4003188 | J. Chem. Inf. Model. 2013, 53, 2506−2510

suitable for transporters was newly developed and substituted for the original dictionary designed for CYP enzymes. The performance of the system was then tested by extracting information on chemical−P-glycoprotein (P-gp) interactions from the PubMed database. P-gp recognizes a large number of chemicals and is expressed in many tissues, including the apical surface of mature enterocytes, the canalicular membranes of hepatocytes, and in kidney and brain endothelial cells.[7,8] Because of this broad substrate specificity, inhibition or induction of P-gp can result in unexpected drug−drug interactions.[9,10] Thus, it is important in terms of drug safety and efficacy to collect information on chemical−P-gp interactions.

MATERIALS AND METHODS

Text-Mining of Chemical−P-gp Interactions. The automated extraction of information on chemical−P-gp interactions consisted of three major steps. First, the chemical and P-gp name variants were identified in the text. Second, sentences containing any names of chemicals and P-gp variants were transformed into simple clauses, each of which contained a single event. Finally, information on chemical−P-gp interactions was extracted from the clauses using a pattern-matching method. Because details of each step were reported in our previous article,[5] only a brief summary is provided herein. The P-gp name dictionaries containing name variants and the chemical name dictionary comprising approximately 100 000 entries were created by extracting the names of chemicals and proteins from the Medical Subject Headings (MeSH) database of the US National Library of Medicine. Because simple dictionary-based approaches overlook a number of chemical names or fall into partial match problems in the identification of chemical names, a context-based approach comprising a set of pattern-matching rules was adopted together with a dictionary-based one. Specific term dictionaries containing words that are related to chemical names (concentration, effect, mg/mL, etc.) were used for this purpose. To create a chemical name dictionary, an initial scan of an entire text was performed by pattern matching, applying the above rules. The entire text was then rescanned to identify chemical and P-gp names, and sentences containing both of these were subjected to sentence processing. Each sentence was divided into noun and verb phrases based on part-of-speech (POS) tags. Simple clauses expressing a single event involving chemical and P-gp names were then reconstructed by considering the voice of each verb phrase and the syntactic structure. When a sentence was complex, noun phrases paired with each verb phrase were identified by a set of pattern-matching rules.

A major modification to the previous text-mining system was that a key verb dictionary was reconstructed to fit the analysis of chemical−P-gp interactions (Table 1). The mode of interaction of compounds with P-gp was specified by the semantics of the verbs. The list included verbs indicating chemical−P-gp interactions (e.g., transport, induce, inhibit) along with the verbal nouns (e.g., transport, induction, inhibition) derived from these verbs. Chemical-denoting nouns (e.g., substrate) were also added to the list. The pattern-matching analysis was performed based on the order of appearance of these three types of keywords (chemical names, P-gp names, and key verbs) in a clause and by taking prepositions and coordinating conjunctions into account. Several patterns were created in each case depending on whether chemical and P-gp names existed in the same phrase or in different phrases of a simple clause. In addition, negative elements (e.g., "not", "no", "n't", "unable", and "unlikely") were identified to check if the clause implied a negative context. The overall architecture of our system was implemented on the general architecture for text engineering (GATE) platform[11] developed at the University of Sheffield, United Kingdom. For tokenization, POS tagging, and morphological processing, we used the tools provided by GATE without any modifications. A typical example of the chemical−P-gp interaction extraction process is shown in Table 2.

Table 1. Verbs and Verbal Nouns Used for the Classification of Data Sets into Three Categories: Substrates, Inducers, and Inhibitors of Transporters

category	verbs and verbal nouns
substrates	accept, accumulate, mediate, recognize, substrate, stimulate, transport, uptake
inducers	induce, inducer, up-regulate, upregulate
inhibitors	down-regulate, downregulate, inhibit, inhibitor

RESULTS

Extraction of Information on Chemical−P-gp Interactions from PubMed Abstracts. Full-text PubMed abstracts of articles regarding P-gp published between November 1976 and March 2013 were searched by entering the key phrases comprising P-gp name variants (i.e., ATP-binding cassette sub-family B member 1 OR ABCB1 OR Multidrug resistance 1 OR MDR1 OR hMDR1 OR P-glycoprotein OR P-gp OR Pgp) in the search box. The number of abstracts retrieved was 21 509. This set of abstracts was then subjected to the text-mining algorithm. The texts were analyzed in a sentence-by-sentence manner, and the sentences that included both chemical names and P-gp names were further subjected to chemical−P-gp interaction extraction analysis. In cases where a verb or verbal noun in the key verb list (Table 1) was detected by pattern matching, the record comprising the set of chemical/P-gp/verb keywords was stored in the database. The numbers of records were 3244, 669, and 2939 for substrate, inducer, and inhibitor categories, respectively. After records containing duplicate interactions were deleted, the accuracy of the information mined was cross-checked through literature review. As a result, 748 curated chemical−P-gp interactions (334 substrates, 104 inducers, and 310 inhibitors) were obtained. Detailed information is available in the Supporting Information (Table S1).

Evaluation of the Text-Mining System. To confirm the validity of the text-mining system, 200 abstracts were randomly selected from the entire abstract corpus, and recall and precision were calculated for both the chemical name identification and the interaction extraction procedures. All chemical names described in the text (with duplication allowed) were subjected to evaluation of the recall and precision for chemicals. On the other hand, the recall and precision for chemical−P-gp interactions were evaluated as follows. In our text-mining system, as mentioned before, information on chemical−P-gp interactions can be extracted from sentences containing both chemical and P-gp names. In order to avoid double consideration of the error of chemical name extraction, only records with chemical name correctly extracted were subjected to evaluation of recall and precision for chemical−P-gp interactions.[6] In this time, the records that

dx.doi.org/10.1021/ci4003188 | J. Chem. Inf. Model. 2013, 53, 2506−2510

Table 2. Typical Example of the Interaction Extraction Process[a]

step no.	process	result
0		Rifampicin can induce P-gp in hepatic and intestinal cells through the activation of nuclear receptors
1	part-of-speech and morphological processing	Rifampicin/$_{NN}$ can/$_{MD}$ induce/$_{VB}$ P-gp/$_{NNP}$ in/$_{IN}$ hepatic/$_{JJ}$ and/$_{CC}$ intestinal/$_{JJ}$ cells/$_{NNS}$ through/$_{IN}$ the/$_{DT}$ activation/$_{NN}$ of/$_{IN}$ nuclear/$_{JJ}$ receptors/$_{NNPS}$
2	Gazetteer lookup and name identification	**Rifampicin** can induce **P-gp** in hepatic and intestinal cells through the activation of nuclear receptors
3	noun group creation	Rifampicin *can induce* P-gp in hepatic and intestinal cells through the activation of nuclear receptors
4	noun phrase creation	[Rifampicin] (can induce) [P-gp] in [hepatic and intestinal cells] through the [activation] of [nuclear receptors]
5	noun phrase group creation	**Rifampicin** {can induce} **P-gp** in **hepatic and intestinal cells** through the **activation** of **nuclear receptors**
6	clause reconstruction	Rifampicin induce P-gp
7	interaction extraction	Rifampicin induce P-gp (inducer); rule: (Chem) (keyVerb, induce) (P-gp)

[a]Step 1: subscript text indicates the part of speech. Definitions: CC, coordinating conjunction; DT, determiner; IN, preposition; JJ, adjective; MD, modal; NN, noun; NNS, noun plural; NNP, proper noun; NNPS, proper noun plural; VB, verb, base form. Step 2: bold text indicates the chemical or P-gp names as listed in the dictionaries. Step 3: underlined text indicates noun groups, and underlined italics indicate verb groups. Step 4: square brackets [ ] indicate noun phrases, and parentheses () indicate verb phrases. Step 5: bold text indicates noun phrase groups, and curly brackets { } indicate verb phrase groups. Step 6: clauses involving relationships between chemicals and P-gp are reconstructed. Step 7: chemical−P-gp interactions are extracted by pattern matching based on the order of keywords.

Table 3. Frequency of Dual Substrates of P-gp and CYP Isoforms

	CYP1A2	CYP2C9	CYP2C19	CYP2D6	CYP2E1	CYP3A4	non-CYP3A4[a]
number of CYP substrates[b]	147	95	86	159	95	283	266
number of P-gp/CYP dual substrates[c]	23	19	15	37	7	78	24
Jaccard−Tanimoto similarity indices[d]	0.050	0.046	0.037	0.081	0.017	0.145	0.041

[a]Sum of the number of substrates listed in the database for the 5 non-CYP3A4 isoforms (CYP1A2, 2C9, 2C19, 2D6, and 2E1), without duplication. This category does not include chemicals that are also CYP3A4 substrates: for example, chemicals that are metabolized by both CYP2D6 and CYP3A4 were not regarded as a non-CYP3A4 substrate, and therefore, the number of dual substrates for non-CYP3A4 could be less than that for CYP2D6. [b]Number of CYP substrates obtained by curation of our previous data set (ref 6). The curated data set of CYP substrates was given in Supporting Information Table S2. [c]Number of dual substrates of CYP and P-gp. The total number of P-gp substrates was 334. [d]The Jaccard−Tanimoto similarity index (JT) was calculated using the following equation:

$$JT = |A \cap B|/|A \cup B|$$

where A and B are sets of P-gp and CYP substrates, respectively.

were not related to any of substrates, inducers and inhibitors were excluded from the evaluation (regarded as not extracted). The recall was the ratio of the number of relevant items retrieved to the number of relevant items in the collection, while the precision was the ratio of the number of relevant items retrieved to the number of total items retrieved. The system achieved 89.8% recall and 84.2% precision for the identification of chemical names, and 71.7% recall and 78.6% precision for the extraction of chemical−P-gp interactions.

Comparison with the TP-Search Database. The obtained data set was compared with that of the transporter database, TP-search.[4] TP-search stores 747 records of P-gp-related interactions (280 substrates, 38 inducers, and 429 inhibitors). The number of interactions in common between the TP-search and our database was 222 (105 substrates, 15 inducers, and 102 inhibitors). In other words, 526 of the 748 interactions found in the present analysis (229 substrates, 89 inducers, and 208 inhibitors) were not listed in the TP-search database. Reasons for these inconsistencies will be discussed later.

Evaluation of the Co-occurrence between P-gp Substrates and CYP3A4 Substrates. It has been noted that P-gp and CYP3A4 have significant overlap in their substrate specificities and can function cooperatively as a barrier to intestinal drug absorption. Using the data set obtained by the text-mining system in a previous study,[6] co-occurrence between P-gp and CYP3A4 substrates was investigated. First, accuracy of the information for all the compounds listed in the databases was cross-checked through literature review. The number of substrates for P-gp, CYP1A2, CYP2C9, CYP2C19, CYP2D6, CYP2E1 and CYP3A4, were 334, 147, 95, 86, 159, 95 and 283, respectively (details are provided in the Supporting Information). Non-CYP3A4 substrates totaled 266; comparable in number to CYP3A4 substrates. Table 3 summarizes the analysis of commonality in substrates between P-gp and each CYP isoform based on Jaccard−Tanimoto similarity indices.[12] This is a widely used similarity measure for two binary variables, which is calculated as $|A \cap B|/|A \cup B|$. The result clearly indicates that P-gp shares more substrates with CYP3A4 than with other CYP isoforms.

DISCUSSION

In recent years, many NLP-based text-mining systems have been developed for extracting data on interactions relevant to the biomedical field.[13−17] However, almost all of these studies have focused on extracting information involving genes and proteins. Because of differences in linguistic expression, it is generally difficult to adopt these NLP systems and directly apply them to other fields, such as the interactions between exogenous chemicals and biomolecules. Besides our research, a statistics-based approach that is usually performed using machine-learning methods has been employed, and achieved a high level of accuracy in chemical name identification and interaction extraction.[18] In general, a large and highly annotated

corpus is required for applying statistics-based text-mining approaches. However, corpuses matched with a specialized purpose are not always available in the public domain, thereby limiting the application of such approaches.[19,20] Semisupervised machine learning systems that perform bootstrapping from protein-chemical seed pairs are under investigation.[21] Conversely, our system employed a rule-based approach that uses pattern matching for extraction of interactions. The challenge when using this approach is the difficulty involved in compiling a rule set. We have already developed the rule set for CYP interaction studies.[5,6] Therefore, by only changing the key verb and name lists, we were able to apply this system to extracting information related to chemical−P-gp interactions. This system was shown to achieve satisfactory performance in both chemical name identification and interaction extraction.

To evaluate the applicability of the present text-mining system, we compared the data obtained by our method with P-gp-related records listed in the TP-search database.[4] TP-search is one of the most prominent transporter databases, containing information on more than 75 membrane transporters and their substrates, inducers, and inhibitors. This information has been comprehensively compiled by expert transporter researchers and summarized to form a web-accessible database. The present system detected 526 chemical−P-gp interactions (229 substrates, 89 inducers, and 208 inhibitors) that were not registered in the TP-search database, although this number does include a large number of interactions, i.e., 271 interactions, that were obtained from abstracts published after the last update of TP-search (June 26, 2007). However, 525 interactions (175 substrates, 23 inducers, 327 inhibitors) in the TP-search database could not be detected by the present system. One of the reasons why these interactions were not detected is that the text-mining analysis was only conducted on a corpus of PubMed abstracts. Of the missing chemical names, 37% did not exist in this corpus. Another concern is that information extraction cannot be done unless both chemical name and P-gp variant name are included in the sentence. It is still challenging to extract information from plural sentences using association processing.

As well as P-gp, CYP3A4 is known to show broad substrate specificity and limit oral absorption of various drugs. Due to the overlap in substrate specificity between P-gp and CYP3A4, the concept of "intestinal recycling" has been proposed.[22−25] The transporter function of P-gp may allow CYP3A4 to have repeated and prolonged access to substrate molecules in the intestine, thereby significantly reducing absorption into the bloodstream. A recent in silico simulation study indicated that physiological distribution patterns of CYP3A4 and P-gp (that is, expression of the former is higher in the proximal region of the intestine, while the latter in the distal region) are suitable for limiting oral absorption of dual P-gp/CYP3A4 substrates.[26] However, degree of overlap in substrate specificity between P-gp and CYP3A4 has not yet been comprehensively investigated, despite a previous study that employed a massive literature search.[27] To address this issue, we curated the data set of substrates for CYP isoforms that were previously collected using the text-mining system,[6] and compared it with the data set of P-gp substrates. The Jaccard−Tanimoto similarity index for frequency of dual P-gp/CYP3A4 substrates was significantly higher than that for other combinations (Table 3). The present result therefore clearly supports overlapped substrate specificity between P-gp and CYP3A4.

In conclusion, we were able to apply a text-mining system to the automatic extraction of information on chemical−transporter interactions. Using this system, we successfully collected information on chemical−P-gp interactions, some of which are not currently listed in databases. Furthermore, this system could be applied in future to other transporters without any modifications, as there should be few syntactical differences between P-gp and other transporters. This system could therefore assist researchers in the field of transporter proteins, by providing a tool to facilitate manual literature search processes.

■ ASSOCIATED CONTENT

Supporting Information

Curated information on P-gp interactants (Table S1) and the substrates of each CYP isoform (Table S2). This material is available free of charge via the Internet at http://pubs.acs.org. The text-mining system developed can be downloaded from http://dds.pharm.kyoto-u.ac.jp/downloads/text-mining.

■ AUTHOR INFORMATION

Corresponding Author

*Tel.: +81-75-753-4535. Fax: +81-75-753-4575. E-mail: yama@pharm.kyoto-u.ac.jp.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

This research was supported in part by a Grant-in-Aid for Scientific Research on Innovative Areas (23136102) from the Ministry of Education, Culture, Sports, Science & Technology, Japan.

■ REFERENCES

(1) Hyo-Kyung, H. Role of Transporters in Drug interactions. *Arch. Pharm. Res.* **2011**, *34*, 1865−1877.

(2) Vasanthanathan, P.; Norbert, H.; Gerhard, F. E. Fingerprint-based in silico models for the prediction of P-glycoprotein substrates and inhibitors. *Bioorg. Med. Chem.* **2012**, *20*, 5388−5395.

(3) Bikadi, Z.; Hazai, I.; Malik, D.; Jemnitz, K.; Veres, Z.; Hari, P.; Ni, Z.; Loo, T. W.; Clarke, D. M.; Hazai, E.; Mao, Q. Predicting P-Glycoprotein-Mediated Drug Transport Based On Support Vector Machine and Three-Dimensional Crystal Structure of P-glycoprotein. *PLoS One* **2011**, *6*, e25815.

(4) Ozawa, N.; Shimizu, T.; Morita, R.; Yokono, Y.; Ochiai, T.; Munesada, K.; Ohashi, A.; Aida, Y.; Hama, Y.; Taki, K.; Maeda, K.; Kusuhara, H.; Sugiyama, Y. Transporter database, TP-Search: A Web-accessible comprehensive database for research in pharmacokinetics of drugs. *Pharm. Res.* **2004**, *21*, 2133−4.

(5) Feng, C. L.; Yamashita, F.; Hashida, M. Automated extraction of information from the literature on chemical-CYP3A4 interactions. *J. Chem. Inf. Model.* **2007**, *47*, 2449−2455.

(6) Yamashita, F.; Feng, C.; Yoshida, S.; Itoh, T.; Hashida, M. Automated information extraction and structure-activity relationship analysis of cytochrome P450 substrates. *J. Chem. Inf. Model.* **2011**, *51*, 378−385.

(7) Ambudkar, S. V.; Kimchi-Sarfaty, C.; Sauna, Z. E.; Gottesman, M. M. P-glycoprotein: from genomics to mechanism. *Oncogene* **2003**, *22*, 7468−7485.

(8) Seelig, A. A general pattern for substrate recognition by P-glycoprotein. *Eur. J. Biochem.* **1998**, *251*, 252−261.

(9) Rameis, H. Quinidine-digoxin interaction: are the pharmacokinetics of both drugs altered? *Int. J. Clin. Pharmacol. Ther. Toxicol.* **1985**, *23*, 145−153.

(10) Westphal, K.; Weinbrenner, A.; Zschiesche, M.; Franke, G.; Knoke, M.; Oertel, R.; Fritz, P.; von Richter, O.; Warzok, R.; Hachenberg, T.; Kauffmann, H. M.; Schrenk, D.; Terhaag, B.; Kroemer, H. K.; Siegmund, W. Induction of P-glycoprotein by rifampin increases intestinal secretion of talinolol in human beings: a new type of drug/drug interaction. *Clin. Pharmacol. Ther.* **2000**, *68*, 345−355.
(11) Cunningham, H.; Tablan, V.; Roberts, A.; Bontcheva, K. Getting more out of biomedical documents with GATE's full lifecycle open source text analytics. *PLoS Comput. Biol.* **2013**, *9*, e1002854.
(12) Todeschini, R.; Consonni, V.; Xiang, H.; Holliday, J.; Buscema, M.; Willett, P. Similarity coefficients for binary chemoinformatics data: overview and extended comparison using simulated and real data sets. *J. Chem. Inf. Model.* **2012**, *52*, 2884−901.
(13) Ono, T.; Hishigaki, H.; Tanigami, A.; Takagi, T. Automated extraction of information on protein-protein interactions from the biological literature. *Bioinformatics* **2001**, *17*, 155−161.
(14) Ding, J.; Berleant, D.; Nettleton, D.; Wurtele, E. Mining MEDLINE: abstracts, sentences, or phrases? *Pac. Symp. Biocomput.* **2002**, *7*, 326−337.
(15) Pustejovsky, J.; Castano, J.; Zhang, J.; Kotecki, M.; Cochran, B. Robust relational parsing over biomedical literature: extracting inhibit relations. *Pac. Symp. Biocomput.* **2002**, *7*, 362−373.
(16) Horn, F.; Lau, A. L.; Cohen, F. E. Automated extraction of mutation data from the literature: application of MuteXt to G protein-coupled receptors and nuclear hormone receptors. *Bioinformatics* **2004**, *20*, 557−568.
(17) Hu, Z. Z.; Narayanaswamy, M.; Ravikumar, K. E.; Vijay-Shanker, K.; Wu, C. H. Literature mining and database annotation of protein phosphorylation using a rule-based system. *Bioinformatics* **2005**, *21*, 2759−2765.
(18) Jiao, D.; Wild, D. J. Extraction of CYP Chemical Interactions from Biomedical Literature Using Natural Language Processing Methods. *J. Chem. Inf. Model.* **2009**, *49*, 263−269.
(19) Krauthammer, M.; Nenadic, G. Term identification in the biomedical literature. *J. Biomed. Inf.* **2004**, *37*, 512−26.
(20) Meystre, S. M.; Savova, G. K.; Kipper-Schuler, K. C.; Hurdle, J. F. Extracting information from textual documents in the electronic health record: a review of recent research. *Yearb. Med. Inf.* **2008**, *3*, 128−144.
(21) Yan, Y.; Kim, J. K.; Croset, S.; Rebholz-Schuhmann, D. Finding small molecule and protein pairs in scientific literature using a bootstrapping method. In *BioNLP2012 Proceedings of the 2012 Workshop on Biomedical Natural Language Processing*, Montréal, Canada, June 8, 2012; pp 172−175.
(22) Benet, L. Z.; Wu, C. Y.; Hebert, M. F.; Wacher, V. J. Intestinal drug metabolism and antitransport processes: a potential paradigm shift in oral drug delivery. *J. Controlled Release* **1996**, *39*, 139−143.
(23) Watkins, P. B. The barrier function of CYP3A4 and P-glycoprotein in the small bowel. *Adv. Drug Delivery Rev.* **1997**, *27*, 161−170.
(24) Zhang, Y.; Benet, L. Z. The gut as a barrier to drug absorption: combined role of cytochrome P450 3A and P-glycoprotein. *Clin. Pharmacokinet.* **2001**, *40*, 159−168.
(25) Ito, K.; Kusuhara, H.; Sugiyama, Y. Effects of intestinal CYP3A4 and P-glycoprotein on oral drug absorption−theoretical approach. *Pharm. Res.* **1999**, *16*, 225−231.
(26) Watanabe, T.; Maeda, K.; Nakai, C.; Sugiyama, Y. Investigation of the effect of the uneven distribution of CYP3A4 and P-glycoprotein in the intestine on the barrier function against xenobiotics: A simulation study. *J. Pharm. Sci.* **2013**, *102*, 3196−3204.
(27) Wacher, V. J.; Wu, C. Y.; Benet, L. Z. Overlapping substrate specificities and tissue distribution of cytochrome P450 3A and P-glycoprotein: implications for drug delivery and activity in cancer chemotherapy. *Mol. Carcinog.* **1995**, *13*, 129−134.