ARTICLES

Literature-Based Generation of Hypotheses on Chemical Composition Using Database Co-occurrence of Chemical Compounds

Boris L. Milman*

D. I. Mendeleyev Institute for Metrology, 19 Moskovsky pr., Saint Petersburg 198005, Russia

Received September 21, 2004

Candidates for identification of unknown constituents in a sample to be chemically analyzed are hypothetical. It is proposed to generate these hypotheses according to the co-occurrence of different chemical compounds with a known sample constituent in the chemical literature. The efficiency of the co-occurrence approach for predicting chemical compositions was tested for 67 impurities in 17 chemical/pharmaceutical products. The relative co-occurrence of impurity compounds and these products in the Chemical Abstracts Service database was evaluated and compared with corresponding values for several reference groups of probability sampled compounds from the literature. Almost all impurities (97%) and only ≤8% randomly sampled compounds co-occurred with these chemical products. Mean and median values of relative co-occurrence for impurities are much higher than those of probability sampled compounds which co-occurred with the products. For the combination of impurities and the probability sample of 396 interfering compounds, the power to predict the chemical composition using the highest co-occurrences is 0.49−0.59. The co-occurrence value can also be considered as an "empiric" indicator of chemical similarity useful to generate new hypotheses on relationships both between compounds and between compounds and their properties.

INTRODUCTION

The determination of unknown constituents in a sample/matrix is the most complex problem in chemical analysis. To solve this problem using appropriate analytical techniques, spectral data libraries, and reference materials, an analyst needs hypotheses on particular chemical compounds presented in a sample. The identification procedure of unknown is considered as screening and testing of such hypotheses.[1−4]

Candidates for identification as corresponding hypotheses can be selected according to the high rates of their occurrence and co-occurrence in the chemical literature[2,4] (the related terms "citation" and "co-citation"/"co-reference", respectively, were previously used[2,4]). Co-occurrence is the common occurrence of the names of two compounds (one of them is the known abundant/main component of the sample) or a pair consisting of a compound and a matrix in the same document (article, patent, database record, etc.). Occurrence and co-occurrence rates are the numbers of such documents recording the name of the compound or the pair of compounds (or the compound and the matrix), respectively. The hypotheses that compounds selected by their high occurrence and/or co-occurrence rates are actually contained in the sample and are then tested under chemical analysis.[1,2,4]

Citation (occurrence) and cocitation (co-occurrence) analysis of text constituents, i.e., bibliographic references, words/terms, and author names, is widely used in the science of science, information science, sociology, etc. to explore intellectual and social structure of science.[5−12] Coword networks are also used for the generation of advanced hypotheses and the discovery of new relationships between phenomena in biomedicine.[13−16] Some words may be names of chemical compounds depicting research specialties[9−11] or having biological activity.[13−15]

In our previous work,[2,4] the occurrence/co-occurrence behavior of constituents of three samples in 1987 and 1997 was studied and discussed. In this article, the approach of setting up identification hypotheses based on co-occurrence data will be tested for (a) a greater number of matrices and (b) a longer time covered by the full Chemical Abstracts Service (CAS) database. Co-occurrence of organic impurities determined in 17 chemical/pharmaceutical products[17−30] with these particular matrix compounds vs co-occurrence rates for several reference compound groups is studied. These data are used for the reverse prediction of the impurity composition of these chemical products. Then the reasons for different pairs of chemical compounds co-occurring, i.e., corresponding kinds of relationships between compounds, are explored. The eventual purpose of this research is to increase the reliability of conclusions obtained by using co-occurrence data in chemistry.

METHODS

All chemical compounds or documents containing their names were probability (systematically or randomly) sampled. (In this article, the term "sample" refers to both "chemical sample" and "statistical sample".)

Relevant articles on impurities in organic compounds were retrieved from the Science Citation Index Expanded database (Thomson ISI, Philadelphia, PA) for the decade of 1994−2003 using the search term combination "impurities AND determination". Every fifth record was sampled. The articles on metal and gases were omitted. The actual sample

* Corresponding author e-mail: bmilman@mail.rcom.ru.

10.1021/ci049716u CCC: $30.25
Published on Web 07/29/2005

Title

Chromatographic separation process

Inventor Name

Knox, John H.; Pryde, Andrew

Patent Assignee

Shandon Southern Products Ltd., UK

Publication Source

Ger. Offen., 17 pp.

Identifier-CODEN

GWXXBX

Patent Information

PATENT NO.	KIND	DATE	APPLICATION NO.	DATE
DE 2605789	A1	19760826	DE 1976-2605789	19760213 <--
DE 2605789	C2	19820701		
GB 1540715	A	19790214	GB 1975-6133	19760211

Priority Application Information

GB 1975-6133 19750213

Abstract

Liq. column chromatog. is carried out with SiO2 particles that have had their surface silanol groups converted substantially entirely to trimethyl- or triethylsilyl groups. Thus, SiO2 packing material 3.5 g was dried 4 h at 125° under vacuum and suspended in dry Decalin 25 mL. Trimethylsilylimidazole 2 mL was added to the suspension. The mixt. was stirred 16 h at 100° under N2. Me2CO was added to dissolve the imidazole and the suspension was filtered. The material was washed with Me2CO 200 mL and dried 24 h at 125° under vacuum. The product contained C 5.4 and H 1.6%. The product was placed in a column (diam. 5, length 125 mm) that was used as part of a high-pressure liq. chromatog. for sepg. **tetracyclines**. Quant. sepn. was achieved in <10 min of mixts. contg. 4-epitetracycline, **tetracycline**, 7-chlorotetracycline, 4-epianhydrotetracycline, and **anhydrotetracycline**. Vitamins B2, B3, and B6 were sepd. also.

International Patent Classification

B01D015-08

Document Type

Patent

Language

German

Accession Number

1977:487052 CAPLUS

Document Number

87:87052

Figure 1. The typical abstract recording names of tetracycline and anhydrotetracycline. The latter is the impurity in the former determined in other research. Here both compounds are components of the same mixture.

contained 14 articles[17−30] covering 17 organic products (main components) and 67 impurities in them as a whole. These products are drugs (oxiracetam,[17] alevodopa and carbidopa,[18] almokalant,[20] ciprofloxacin,[21] tetracycline,[23] isosorbide 5-mononitrate,[24] ropinirole,[25] acetylsalicylic and ascorbic acids,[26] paracetamol,[26] nevirapine,[27] oxytetracycline,[28] and sildenafil);[30] a pharmaceutical synthetic precursor (2-acetylbenzothiophene);[22] a contrast enhancing agent for magnetic resonance (mangafodipir trisodium);[29] and a pesticide (chlortoluron).[19]

For each impurity, rates of occurrence and co-occurrence with the corresponding main component were evaluated as the number of search answers in the Chemical Abstracts Plus database (CAS, Columbus, OH) with the use of a demo version of STN Easy network. The names of corresponding compounds or their CAS numbers (for compounds not having short brand names) were entered as the search terms. Figure 1 shows one of the retrieved documents as the example. Then the relative co-occurrence rate (RCR) was calculated as the ratio of the count of the impurity co-occurrences with the corresponding chemical product to the count of impurity occurrences for the period from the earliest publication year to 2003.

Example. Anhydrotetracycline was detected as one of the impurities in tetracycline.[23] The search (May, 2004) for "anhydrotetracycline" and "anhydrotetracycline AND tetracycline" led to 392 and 346 answers/documents, respectively. The RCR is 346/392 = 0.88.

To conclude with significance, the occurrence and RCR values for impurities were compared with those for three reference groups of compounds (Figure 2). One of them represents all known chemical compounds; two other groups comprised compounds obviously cocited with the products under consideration.

Reference group I contained 100 chemical compounds randomly sampled from the set of all known compounds as covered by the *CAS Registry Handbook* for 1965−2002.[31] In this procedure, the pages of the handbook and compound locations on the pages were randomly chosen. The occurrence rates and RCRs were calculated for the subset of 59 organic compounds contained in the group by the same method; co-occurrence of each reference compound with each of the 17 chemical products sampled in the first stage was accounted.

Reference group II comprised organic compounds which co-occurred with the same chemical products in 1994−2003. These compounds, except the impurities, were extracted from the titles and abstracts of 10 reference documents retrieved by the program for each product and for each year from the time period (the demo search engine retrieves one document

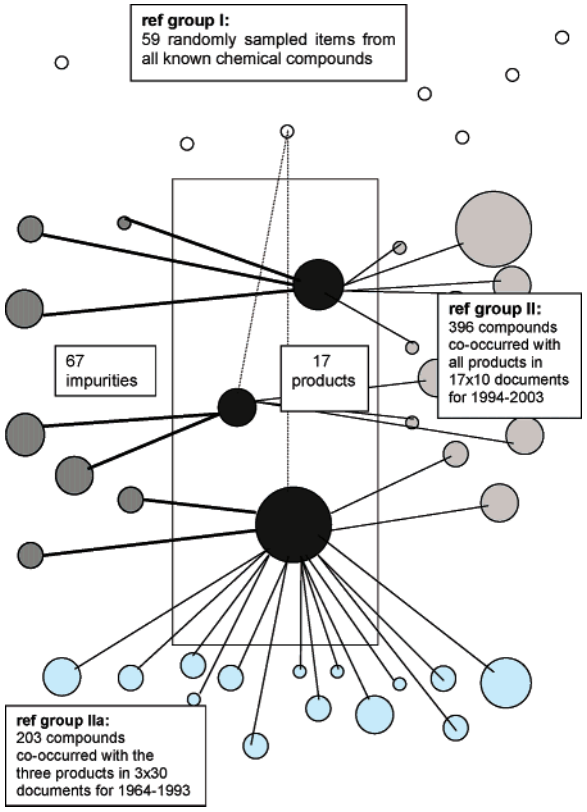


Figure 2. The chart of different groups of compounds.

per query). The reference data sampled in this way consisted of 170 CAS records/documents mentioning co-occurring compounds. Their relative co-occurrence rates with corresponding products were also calculated.

To form reference group IIa, three products, tetracycline, nevirapine, and 2-acetylbenzothiophene, representing impurities with high, medium, and low RCRs, were selected. For each of these products, 30 additional documents of the 1964−1993 period were chosen by the same manner. Names of 203 organic compounds co-occurring with the three products except the impurities were extracted from these documents and included in reference group IIa of the co-occurring compounds.

Hence, all the impurities and subset of three typical items from them were compared in occurrence and co-occurrence values to one and two reference groups of the co-occurring compounds, respectively (Figure 2).

For all possible pairs of chemical compounds, products, impurities, and items from group II, reasons for co-occurrence in the same documents were found out. The reasons are features of a relationship between compounds, e.g. belonging to the same reaction, co-occurrence in the same formulations, similarity in properties, etc.

Example. The patent in Figure 1 was retrieved for the search term combination "tetracycline AND anhydrotetracycline". These two compounds and several other drugs are available in the same mixture chemically analyzed. Thus, the reason for their pair co-occurrence in the document is "the same sample/mixture/solution/matrix".

The compound distribution in occurrence and co-occurrence rates was estimated by using the following parameters: mean value, standard deviation, confidence interval, median, and quartile values. The significance of difference between the mean values was estimated for a normal distribution, a two-sided test, and the probability of 0.05. The fraction, w%, of compounds from some or other group having an occurrence/co-occurrence rate above or below the particular value can also be calculated. A sampling error Δw due to the limited sample size was estimated by the formula:[32] $\Delta w = 2\cdot\sqrt{[w(100-w)]/n}$%, where n is the sample size; the probability is 0.95.

Table 1. Statistic Parameters for the Occurrence Rate[a]

parameter	main components	impurities	reference group I	reference group II
mean	**7785**	**1086** (1299)	**18** (67)	**24887** (25528)
n	**17**	**67** (56)	**59** (15)	**396** (386)
25% quartile	**207**	**5** (10)	**1** (2)	**283** (333)
median	**1000**	**26** (41)	**1** (4)	**2214** (2420)
75% quartile	**7315**	**107** (140)	**2** (6)	**15897** (16893)

[a] Data excluding null or unit count of occurrences are in parentheses; differences between some means are significant.

The power of the co-occurrence approach to predict the chemical composition of a mixture to be chemically analyzed was estimated in two ways. First, the fraction of compounds which co-occurred with the chemical products is compared for impurities to randomly sampled compounds (group I). Second, the RCR values as rates of the co-occurrence approach efficiency were used to compare impurities and different co-occurring probability sampled compounds (group II and also IIa). In this case, impurities and compounds of the group II were combined and ranked by their RCRs. In this hit list, originally composed from m impurities and some number of other compounds, the highest m RCR values were selected. In this subset of highest RCRs, m_{high} and $(m - m_{high})$ values are related to impurities and highly co-occurring foreign compounds ("candidates for impurity"), respectively. The predicting power was calculated as m_{high}/m. The indicator shows the fraction of impurities truly predicted by their RCR values. The former can be calculated both for impurities of the particular main component and for all impurity compounds as a whole.

Example. In nevirapine, 11 impurities were detected.[27] During the information retrieval, 34 co-occurring compounds were selected. The set of 11 highest RCRs was comprised from values of five impurities and six candidates for impurities. Hence, m is 11, m_{high} is 5, and the predicting power is $5/11 = 0.45$.

All calculations were made using standard commercial software.

RESULTS AND DISCUSSION

Differences in Occurrence and Co-Occurrence Rates. The parameters of the distribution of products, impurities, and compounds of the reference groups over occurrence and co-occurrence are given in Tables 1 and 2 and in Figure 3.

Table 1 demonstrates that randomly sampled compounds (group I) are rare; on average, they are occurred only 1/60 times as much as impurities. Reference compounds I have also very low co-occurrence. Only two randomly sampled compounds of group I (≤8% taking into account the sample error) and all but two impurities (97%, the sample error 4%) co-occurred with the main components. The mean RCR of this group is 5400 times as low as that of impurities (Table

Table 2. Statistic Parameters for Relative Co-Occurrence Rate[a]

parameter	impurities	reference group I	reference group II
mean	**0.42** (0.37)	**0.000078** (0.00028)	**0.11** (0.089)
confidence interval,level 0.95	± 0.09 (0.09)	± 0.00012 (0.00043)	± 0.02 (0.018)
n	**67** (56)	**918** (255)	**396** (386)
25% quartile	**0.060** (0.061)	**0** (0)	**0.0012** (0.0011)
median	**0.33** (0.25)	**0** (0)	**0.0070** (0.0062)
75% quartile	**0.80 (**0.60)	**0** (0)	**0.11** (0.10)

[a] Data excluding null or unit count of occurrences are in parentheses; differences between all means are significant.

Table 3. Reasons of Co-Occurrence

feature	example	%
similarity in properties, activity, structure, use	drugs/pharmaceuticals of the same class	58
the same reaction system	reactants and products belonging to the same reaction	13
the same sample/mixture/solution/matrix	the same formulation; impurities of the same product; pollutions in same water sample	17
other/hidden		12

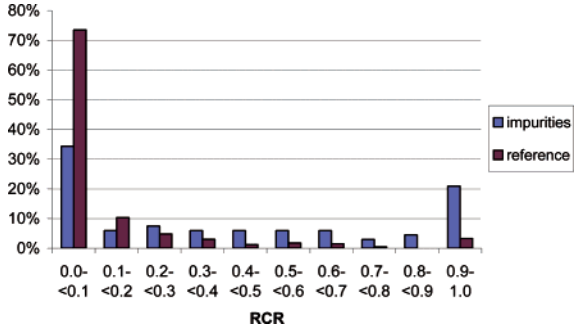


Figure 3. The distribution of impurities and reference compounds II over RCR. The difference between the two groups in percentages is significant for the minimum and maximum RCR.

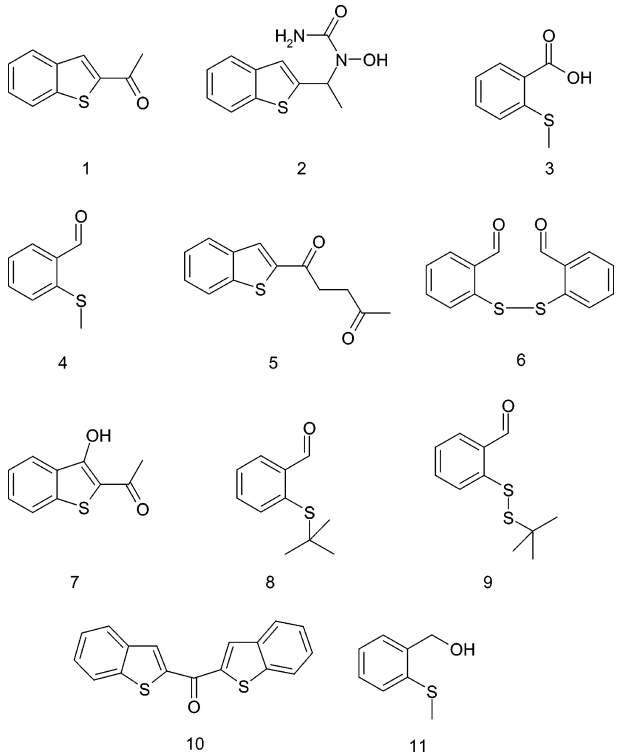
2). This means that the probability of an accidental co-occurrence is very low; a fact of co-occurrence reflects some kind of regular relationship between co-occurring compounds.

However, the occurrence rate of impurities is lower than that of the main components and especially those items from reference group II. The latter has a very high occurrence rate because the group contains such common compounds as ethylene (occurrence count is $4.7 \cdot 10^5$), glucose ($3.6 \cdot 10^5$), ethanol ($2.1 \cdot 10^5$), methanol ($1.6 \cdot 10^5$), etc. It may be concluded that impurity compounds cannot be differentiated from some other groups by the highest occurrence value.

In contrast, the RCR indicator does that. Impurities have much higher values of relative co-occurrence than probability sampled compounds (group II): mean values are 0.42 ± 0.09 and 0.11 ± 0.02 and medians are 0.33 and 0.0070, respectively (Table 2). Removal of rare compounds (occurrence count is 0 or 1) from all groups compared retains this regularity (Table 2).

Reasons for Co-Occurrence. Most co-occurrences can be explained by a few general reasons (Table 3). According to this classification, the articles[17−30] and other publications on determination of impurities in chemical products refer to pairs consisting of a main component and an impurity as constituents of the same "sample/mixture/solution/matrix" (Table 3). However, such analytical articles are rather few. Other reasons influence co-occurrence of impurities and the main components of products/materials to a higher extent.

First, an impurity is often a synthetic precursor or a decomposition product of a material. In both cases, impurities and matrix compounds belong to the same reaction system. Second, this pair of compounds may be structurally related. This also leads to similar properties or use; it is another kind of feature determining co-occurrence (Table 3). For example, 2-acetylbenzothiophene **1**, the material for the synthesis of drug zileuton **2**, contains nine impurities **3**−**11** which resemble **1** in structure: the common substructure is the benzene ring bonded to the sulfur atom. This reason is the most popular one resulting in relatively frequent co-occurrence main compounds and substances of similar structure, e.g. both impurities and members of reference groups II and IIa. "Foreign" similar compounds interfere with "useful" relationships of main compounds with impurities. The co-occurrence rate can be also considered as another indicator of chemical similarity[33] estimated on "an empiric base".



Reverse Prediction of Impurity Compositions. The predicting power of chemical composition using co-occur-

Table 4. Overall Statistical Data for Three Chemical Products[a]

	RCR		
parameter	impurities	reference subgroup II	reference group IIa
mean	**0.43**[b] (0.35[c])	**0.20** (0.18)	**0.28** (0.09)
confidence interval, level 0.95	±0.16 (0.16)	±0.05 (0.04)	±0.06 (0.03)
n	**24** (20)	**98** (95)	**203** (162)
25% quartile	**0.08** (0.06)	**0.02** (0.02)	**0.001** (0.0003)
median	**0.25** (0.19)	**0.12** (0.10)	**0.02** (0.0020
75% quartile	**0.87** (0.98)	**0.31** (0.29)	**0.50** (0.10)
candidates for impurities		**10 (7)**	**43 (9)**

[a] Data excluding unit count of occurrences are in parentheses. [b] The difference between impurities and subgroup II in mean is significant. [c] The differences between impurities and both reference sets in mean are significant.

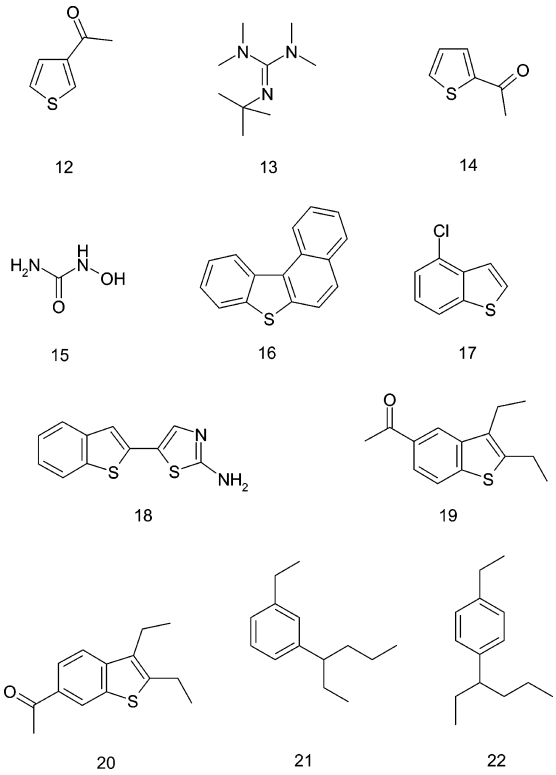
rence can be estimated first for the impurities and interfering compounds of group II. For the sum of impurities, this value is 0.49. If this indicator is originally calculated for each main component containing its "own" impurities and then averaged, a higher value (0.59) is observed. We consider this result as very promising because a predicting approach with the only parameter is used.

In the general case, predicting power depends strongly on the sampled groups of compounds because the two compared samples of data may differently represent a "set of impurities" and a "set of all other co-occurring compounds". It is demonstrated by different groups of co-occurring compounds for three materials (subgroup of II and group IIa). In the case of IIa, the three times increase of the number of reference documents increases the sum of co-occurring compounds and candidates for identification as impurities, i.e., highly co-occurring items 2.1 and 4.3 times, respectively (see Table 4). For two of the three products, real impurities are replaced by candidates for impurities at the top of the RCR lists. The large majority of additional high RCRs are related to rare compounds (unit count of occurrence and co-occurrence). If rare compounds are not considered, the number of candidates for impurities is increased only 1.3 times (Table 4).

On average, compounds belonging to group IIa have higher RCRs than items from subset II (Table 4). This results in a relatively low predicting power (0.15) for the combined list of interfering compounds II+IIa. This high level of "information noise" is due to the presence of rare compounds. If the latter are not taken into account, the predicting power raises approximately twice (0.28).

Approaches to Hypotheses Generation on Chemical Composition. In general, the list of all compounds co-occurring with the product analyzed in the database is formed and ranked. Automatic record processing is needed to speed up these operations in solution of real-world analytical problems. This list is essentially one of the hypotheses on compounds possibly available in the sample. The hypotheses are tested by turns starting from the maximum RCR with the use of an appropriate analytical technique. Given the probability that the substance contained in the sample is higher than the randomly chosen value, testing hypotheses may be stopped relatively far from the bottom list line (lowest RCR). At the worst, all co-occurring substances should be screened.

In many cases, the testing procedure itself may be not very complicated because even a part of the experimental data or simple prior considerations are often sufficient for a true conclusion. The impurities in 2-acetylbenzothiophene having the lowest mean RCR as compared with those in other products, i.e., the most difficult to be predicted, are an example. Table 5 containing the features which enable us to differentiate between these impurities and numerous candidates for identification (compounds **12**−**17** are candidates



mostly co-occurring with the product) demonstrates this. In most cases, integer molecular mass measured by low resolution mass spectrometry (this technique, e.g. see the book in ref 34) makes it possible to accept/reject identification hypotheses. To differentiate impurity **5** from candidates for identification of **18**−**20** having the same integer molec-

Table 5. Rejection of Compounds Cocited with 2-Acetylbenzothiophene as Candidates for Identification

feature to discriminate	method	no. of compds
molecular weight	low resolution mass spectrometry	27
chlorine/bromine isotope pattern *or* molecular weight	low resolution mass spectrometry	4
fragmentation pattern *or* molecular weight	low resolution mass spectrometry high resolution mass spectrometry	3
molecular weight *or* solubility	low resolution mass spectrometry general consideration	2

ular mass (132), full spectra to uncover fragmentation patterns or high-resolution mass spectrometry for measuring different fractional mass are necessary [monoisotopic mass: 232.06 (**5**), 232.01 (**18**), 232.09 (**19** and **20**)]. Some hypotheses can be a priori rejected. At least two co-occurring compounds (**21** and **22)** are nonpolar, and their low solubility in polar solvents does not fit the experimental conditions used for determining impurities in 2-acetylbenzothiophene.[22]

CONCLUSION

The list of compounds co-occurring with a chemical product in a comprehensive chemical documentary database contains almost all impurities determined in the product, with a significant chance for the impurity compounds to be present in the upper part of the list. The efficiency and productivity of the co-occurrence approach in chemical analysis can be supposed to depend on (a) the availability of special software for processing texts, (b) the eventual compound list size also related to data cost, (c) the co-occurrence rates of compounds identified, and (d) the amount and complexity of data on compounds (features, properties) necessary to screen list items as identification hypotheses. A co-occurrence is also related to the similarity of compounds in structure/property/use. Therefore this "empiric" indicator may be applicable to generate new hypotheses on relationships both between chemical compounds and between compounds and their properties.

ACKNOWLEDGMENT

The research is partly supported by Exchange Visitor Program No. G-3-0334 of the U.S. Department of Commerce, National Institute of Standards and Technology. The author also thanks Dr. Gary Mallard (NIST) for the discussion of the article and valuable remarks and Dr. Zvi Boger (OPTIMAL-Industrial Neural Systems Ltd., Israel) for the idea to use the demo version of STN Easy network for counting citations.

REFERENCES AND NOTES

(1) Milman, B. L.; Konopelko, L. A. Identification of Chemical Substances by Testing and Screening of Hypotheses − I. General. *Fresenius J. Anal. Chem.* **2000**, *367*, 621−628.

(2) Milman, B. L.; Kovrizhnych, M. A. Identification of Chemical Substances by Testing and Screening of Hypotheses − II. Determination of Impurities in *n*-Hexane and Naphthalene. *Fresenius J. Anal. Chem.* **2000**, *367*, 629−634.

(3) Hartstra, J.; Franke, J. P.; de Zeeuw, R. A. How to Approach Substance Identification in Qualitative Bioanalysis. *J. Chromatogr. B* **2000**, *739*, 125−137.

(4) Milman, B. L. A Procedure for Decreasing Uncertainty in the Identification of Chemical Compounds Based on Their Literature Citation and Cocitation. Two Case Studies. *Anal. Chem.* **2002**, *74*, 1484−1492.

(5) Small, H. Co-citation in the Scientific Literature: A New Measure of the Relationship Between Two Documents. *J. Am. Soc. Inf. Sci.* **1973**, *24*, 265−269.

(6) Small, H.; Sweeney, E. Clustering the Science Citation Index Using Co-citation I. A Comparison of Methods. *Scientometrics* **1985**, *7*, 391−409.

(7) Small, H.; Sweeney, E.; Greenlee, E. Clustering the Science Citation Index Using Co-citation II. Mapping Science. *Scientometrics* **1985**, *8*, 311−340.

(8) Milman, B. L.; Gavrilova, Yu. A. Analysis of Citation and Co-citation in Chemical Engineering. *Scientometrics* **1993**, *27*, 53−74.

(9) Law, J.; Bauin, S.; Courtial, J.-P.; Whittaker, J. Policy and the Mapping of Scientific Change: A Co-word Analysis of Research into Environmental Acidification. *Scientometrics* **1988**, *14*, 251−264.

(10) Peters, H. P. F.; Hartmann, D.; Van, Raan A. F. J. Monitoring Advances in Chemical Engineering. *Informetrics 87/88* **1988**, 175−195.

(11) Milman, B. L.; Gavrilova, Yu. A. Science News in Business Journals as the Source of Information on Applied and Strategic Research and Science Policy (In Russian). *Sci. Technol. Inf. 1 (Moscow)* **1994**, (7), 17−26.

(12) Wolfram, D. *Applied Informetrics for Information Retrieval Research*; Library Unlimited: Westport, 2003.

(13) Smalheiser, N. R.; Swanson, D. R. Using ARROWSMITH: a Computer-Assisted Approach to Formulating and Assessing Scientific Hypotheses. *Comput. Methods Programs Biomed.* **1998**, *57*, 149−153.

(14) Weeber, M.; Klein, H.; de Jong-van den Berg, L. T. W.; Vos, R. Using Concepts in the Literature-Based Discovery: Simulating Swanson's Raynaud-Fish Oil and Migraine-Magnesium Discoveries. *J. Am. Soc. Inform. Sci. Technol.* **2001**, *52*, 548−557.

(15) Wren, J. D.; Bekeredjian, R.; Stewart, J. A.; Shohet, R. V.; Garner, H. R. Knowledge Discovery by Automated Identification and Ranking of Implicit Relationships. *Bioinformatics* **2004**, *20*, 389−398.

(16) Jenssen, T. K.; Oberg, L. M. J.; Andersson, M. L.; Komorowski, J. Methods for Large-Scale Mining of Networks of Human Genes. http://www.siam.org/meetings/SDM01/pdf/sdm01_10.pdf.

(17) Gagliardi, L.; D Orsi, D.; Cavazzutti, G.; Tonelli, D.; Zappoli, S. HPLC Determination of Oxiracetam, its Impurities, and Piracetam in Pharmaceutical Formulations. *Anal. Lett.* **1994**, *27*, 879−85.

(18) Kafil, J. B.; Dhingra, B. S. Stability-Indicating Method for the Determination of Levodopa, Levodopa-Carbidopa and Related Impurities. *J. Chromatogr. A* **1994**, *667*, 175−81.

(19) Toscano, Q. L. C.; Romero, M. A. A. Determination of Some Impurities in Technical Grade Chlortoluron by High Performance Liquid Chromatography (in Spanish). *Afinidad* **1994**, *51*, 364−8.

(20) Klick, S. Evaluation of Different Injection Techniques in the Gas Chromatographic Determination of Thermolabile Trace Impurities in a Drug Substance. *J. Pharm. Biomed. Anal.* **1995**, *13*, 563−6.

(21) Husain, S.; Khalid, S.; Nagaraju, V.; Rao, R. N. High-Performance Liquid Chromatographic Separation and Determination of Small Amounts of Process Impurities of Ciprofloxacin in Bulk Drugs and Formulations. *J. Chromatogr. A* **1995**, *705*, 380−384.

(22) Shada, D. M.; Wong, C. F.; Elrod, L.; Morley, J. A.; Gay, C. M. Determination of 1-Benzo[*b*]thien-2-ylethanone and Related Impurities by High Performance Liquid Chromatography. *J. Pharm. Biomed. Anal*. **1996**, *14*, 501−510.

(23) Tjornelund, J.; Hansen, S. H**.** Determination of Impurities in Tetracycline Hydrochloride by Nonaqueous Capillary Electrophoresis. *J. Chromatogr. Sci. A* **1996**, *737*, 291−300.

(24) Marinkovic, V. D.; Milojkovic, S. S.; Nedeljkovic, J. M.; Comor, J. J.; Agbaba, D.; Zivanov-Stakic, D.; Gas Chromatography Mass Spectrometry Determination of Isosorbide 5-Mononitrate and Related Impurities in Raw Materials and Dosage Formulations. *J. Pharm. Biomed. Anal.* **1997**, *16*, 425−429.

(25) Coufal, P.; Stulik, K.; Claessens, H. A.; Hardy, M. J.; Webb, M. Determination of the Dissociation Constants of Ropinirole and Some Impurities and Their Quantification Using Capillary Zone Electrophoresis. *J. Chromatogr. Sci. B* **1998**, *720*, 197−204.

(26) Krzek, J.; Starek, M. Densitometric Determination of Active Constituents and Impurities in Complex Analgesic and Antipyretic Pharmaceuticals. *J. Planar Chromatogr. − Mod. TLC* **1999**, *12*, 356−360.

(27) Li, Q. C.; Tougas, T.; Cohen, K.; Lee, R.; Meagan, P.; Corson, M.; Muchnick, T. Validation of a High-Performance Liquid Chromatography Method for the Assay of and Determination of Related Organic Impurities in Nevirapine Drug Substance. *J. Chromatogr. Sci.* **2000**, *38*, 246−254.

(28) Hernandez, M.; Borrull, F.; Calull, M. Determination of Oxytetracycline and Some Impurities in Plasma by Non-Aqueous Capillary Electrophoresis Using Solid-Phase Extraction. *Chromatographia* **2000**, 52, 279−284.

(29) Gjerde, H.; Snotun, A.; Hem, H.; Larsen, K. H.; Dohl, J. Determination of Mangafodipir Trisodium and Related Impurities in Bulk Substance and Pharmaceutical Formulation by Ion-Pair High-Performance Liquid Chromatography. *J. Pharm. Biochem. Anal.* **2001**, *25*, 109−114.

(30) Nagaraju, V**.;** Sreenath, D.; Rao, J. T.; Rao, R. N. Separation and Determination of Synthetic Impurities of Sildenafil (Viagra) by Rreversed-Phase High-Performance Liquid Chromatography. *Anal. Sci.* **2003**, *19*, 1007−1011.

(31) *Chemical Abstracts Service Registry Handbook;* American Chemical Society: Columbus, OH, 1965−2002.

(32) Thompson, S. K. *Sampling;* Wiley: New York, 1992.

(33) Rouvray, D. H. Definition and Role of Similarity Concepts in the Chemical and Physical Sciences. *J. Chem. Inf. Comput. Sci.* **1992**, *32*, 580−586.

(34) Chapman, J. R. *Practical Organic Mass Spectrometry: A Guide for Chemical and Biochemical Analysis*; Wiley: Chichester, 1995.

CI049716U