Skin Sensitization Study by a New Qualitative Structure-toxicityRelationships (QSTR) Approach: K-step Yard Sampling (KY) Methods

ORIGINAL ARTICLE

Skin Sensitization Study by a New Qualitative Structure-toxicity Relationships (QSTR) Approach: K-step Yard Sampling (KY) Methods

Kazuhiro SATO[1], Tomohiro UMEMURA[1], Taro TAMURA[1], Yukinori KUSAKA[1], Kohji AOYAMA[2], Atsushi UEDA[3], Kohichi HARADA[4], Keiko MINAMOTO[3], Takemi OTSUKI[5], Kunihiko YAMASHITA[6], Masahiro HAITSUKA[6], Tatsuya TAKESHITA[7], Eiji SHIBATA[8], Kunio DOBASHI[9], Satomi KAMEO[10], Muneyuki MIYAGAWA[11], Masaaki KANIWA[12], Takahiko YOSHIDA[13], Tetsuhito FUKUSHIMA[14] and Kohtaro YUTA[15,16]

[1]Department of Environmental Health, School of Medicine, University of Fukui, Fukui, Japan, [2]Department of Environmental Medicine, Kagoshima University Graduate School of Medical and Dental Sciences, Kagoshima, Japan, [3]Department of Environmental Health, Faculty of Medical and Pharmaceutical Sciences, Kumamoto University, Kumamoto, Japan, [4]Kumamoto University, Graduate School of Health Sciences, School of Health Sciences, [5]Department of Hygiene, Kawasaki Medical School, Kurashiki, Japan, [6]Daicel Corporation, Ltd., Himeji, Japan, [7]Department of Public Health, School of Medicine, Wakayama Medical University, Wakayama, Japan, [8]Department of Hygiene, Aichi Medical University, Nagakute, Japan, [9]School of Health Sciences, Faculty of Medicine, Gunma University, Maebashi, Japan, [10]Department of Public Health, Gunma University Graduate School of Medicine, Maebashi, Japan, [11]National Institute of Occupational Safety and Health, Kawasaki, Japan, [12]Division of Medical Devices, National Institute of Health Sciences, Tokyo, Japan, [13]Deaprtment of Health Science, School of Medicine, Asahikawa Medical University, Asahikawa, Japan, [14]Department of Hygiene and Preventive Medicine, School of Medicine, Fukushima Medical University, Fukushima, JAPAN, [15]Fujitsu Limited, Tokyo, Japan, [16]In Silico Data Ltd. , Narashino, Japan

Skin sensitization is also related to the dental field. In addition, the regeneration of skin and mucous membranes of the lips may be consulted from the patient. *In silico* assessment of skin sensitization is increasingly needed owing to the problems concerning animal welfare, as well as excessive time consumed and cost involved in the development and testing of new chemicals. We could perfectly classify skin sensitizers (positive/negative) using a newly developed K-step Yard sampling (KY) methods (U.S. Patent No. 7725413, 2010). Therefore, the KY methods could be applied to qualitative structure-toxicity relationships (QSTR) study on classifying and predicting samples.

A total of 593 compounds (419 positive sensitizers and 174 negative sensitizers) were used in this study. Parameters were generated from 2-D and 3-D structures of compounds. All of the 1015 parameters generated were reduced by various feature selection methods. KY methods were performed using ADMEWORKS/ModelBuilder software. All 593 compounds were perfectly classified by 3 steps. Discriminant function of each step was a linear dicriminant function, the Iterative Least Squares linear discriminant (TILSQ). KY methods were referred to as a meta-algorithm approach because it requires ordinary data analysis methods to generate discriminant functions.

KY methods were the repetition of removal of gray zone of samples and reclassification of them to attain no gray zone (100% classification) at final step. This methods always attain perfect classification at final step, even though samples are large number, large of structural diversity or highly overlapped on the sample space.

KY methods are promising tool in QSTR technology.

Keywords: *skin sensitization, qualitative structure-toxicity relationships (QSTR), K-step Yard sampling (KY) methods, animal study*

INTRODUCTION

Skin sensitization is also related to the dental field. In addition, the regeneration of skin and mucous membranes of the lips may be consulted from the patient. In occupational health, occupational skin disorders are the most common non-traumatic occupational condition. Among them, contact dermatitis is by far the most common form of occupational skin illness [1].

Under the new European Union (EU) Registration, Evaluation, and Authorization of Chemicals (REACH) rules, all chemicals in the EU that are produced or imported in quantities of more than 1 ton per annum will need to be assessed as potential human and environmental hazards, for example, in terms of their carcinogenicity and human sensitivity to such chemicals will also need to be determined. REACH calls for increased use of hazard assessment alternatives such as *in vitro* methods and quantitative and qualitative structure-toxicity relationships (QSTRs) [2]. Since no *in vitro* replacement is currently available for sensitization, the use of QSTR approaches presents an attractive alternative [3]. One of the most difficult subjects in QSTR research is computer classification and prediction of chemical toxicity of compounds. This is because 1) there is large structural diversity among samples, 2) the sample number is enormously large, and 3) high classification and prediction rate are required. Non-linear discriminant functions, such as neural network (NN), Support Vector Machine (SVM) and AdaBoost, sometimes provide higher classification rate than that of linear methods. However, non-linear method is often accompanied by over-fitting, which lowers prediction rate significantly.

Previously, we made the QSTR model for skin sensitization, which is statistically based [4]. This study revised the previous prediction models using more extensive compounds, by the newly developed binary classification method, K-step Yard sampling (KY) methods [5]. This KY methods are new approaches to classify samples in QSTR technology and always attain perfect classification even though samples are large number, large structural diversity or highly overlapped on the sample space. The KY methods can be applied to a linear and non-linear discriminant function. The KY methods could classify a set of Ames test samples (6,965 compounds, 2932 posi, 4033 nega) into two classes (Positive/ Negative) correctly by 23 steps (data are not shown).

In this paper, we illustrates these new methods, these value, and our QSTR model for skin sensitization.

MTERIALS AND METHODS

1 Chemicals

Positive data for skin sensitization are from a list of compounds as allergen, that is, Sh and Sah [6] and dictionary of contact allergens [7] . The criteria are

based on human epidemiological studies, case reports or validated animal studies (guinea pig maximization test, Buhler guinea pig test or mouse local lymph node assay) for skin sensitization. On the other hand, negative data for a list of negative compounds are belonging to the group defined as 'not classified' for skin sensitization by the Japanese Globally Harmonized System of Classification and Labeling of Chemicals (GHS) Inter-ministerial Committee [8], which means that these chemicals are reported as non-skin sensitizers. However, inorganic chemicals, organic metal chemicals and polymers are special compounds, and can not be analyzed with general organic compounds in computational chemistry. Therefore, we deleted these chemicals and finally assessed 419 positive compounds and 174 negative (total 593) compounds.

2 Parameters and discriminant function

A total of 593 compounds (419 positive and 174 negative) were used for analysis. Parameters were generated from 2-D and 3-D structures of the compounds. All of the 1,015 generated parameters were reduced by various feature selection (e.g. removing low appearance parameter, high correlation or multicolinearity) methods. The K-step Yard sampling (KY) methods were applied. In this case, the Iterative Least Squares linear discriminate functions (TILSQ) was used for generating discriminant functions.

3 K-step Yard sampling (KY) methods

Existing binary classifiers generate only one discriminant function in order to classify a sample set into two classes (Figure 1).

The same is true in the case newly developed non-linear classification methods, such as Neural Network (NN), Support Vector Machine (SVM) and AdaBoost. These non-linear methods sometimes provide a higher classification rate than those of linear methods. However, if the samples are highly overlapped one another, linear and non-linear methods can not classify the samples correctly into two classes (Figure 2).

In the process of the KY methods, two different types of discriminant functions were created to determine positive, negative and gray zones (Figure 3). One of the discriminant functions are called as all negative (AN) model, the other as all positive (AP) model. The AN model

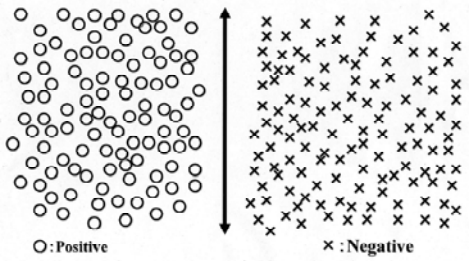

Fig. 1 Linear discriminant function to classify all samples.

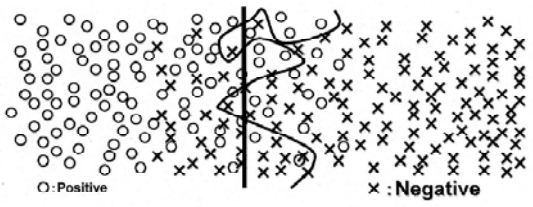

Fig. 2 Highly overlapped samples could not be classified completely by linear and non-linear discriminant functions.

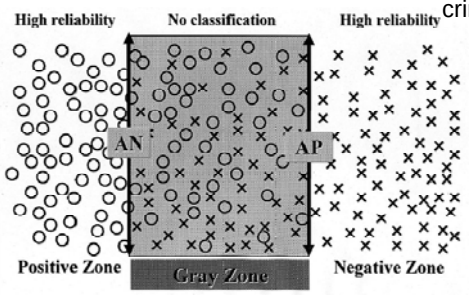

Fig. 3 Classification results by all positive (AP) and all negative (AN) discriminant functions. Samples in positive zone and negative zone had high reliability of classification. Gray zone was not classified.

classified all negative samples in the sample set correctly and the AP model classified all positive samples correctly. The samples which were classified as negative samples by AN model and positive samples by AP model, belonged to the gray zone. KY methods focused on the both sides of a sample space and found that there were special areas, which included only correctly classified samples. These two areas have been defined as positive zone and the others as negative zone. The third zone was named as gray zone. All samples included in the positive zone belonged to a positive class, while all samples in the negative zone belonged to a negative class. On the other hand, the samples included in the gray zone could not be determined whether they belonged to a positive or negative class since they were highly overlapped (Figure 3). If the gray zone (1) was determined by AN1 and AP1 discriminant functions, the gray zone (1) could be extracted and classified by AN2 and AP2 models to build a new sample set. If a new gray zone (2) was determined with respect to the new sample set, a further new sample set can be built as shown in Fig. 4. Repeating these steps, all samples in the original sample set can be classified correctly (Figure 5). This is the basic concept of K-step Yard sampling (KY) methods. The AN model and the AP models can be generated based on any conventional linear and non-linear discriminant function. Therefore, KY methods can be categorized as a meta-algorithm approach.

All data analysis were performed using ADMEWORKS / ModelBuilder software (Fujitsu Kyushu Systems Limited, Japan) [4].

4 Validation study and comparison with TOPKAT and Derek

Using the Interagency Coordinating Committee on the Validation of Alternative Methods (ICCVAM) data set [9], 165 chemicals were chosen and evaluated. Inorganic salts, natural products, organic metals were excluded as they can not be analyzed with general organic compounds. The results were compared with those of Toxicity Prediction Komputer-Assisted Technology (Acclyrs Inc., CA, USA; TOPKAT) and Deductive Estimation of Risk from Existing Knowledge for Windows (DfW. LHASA Ltd., Leeds, UK; Derek) [10].

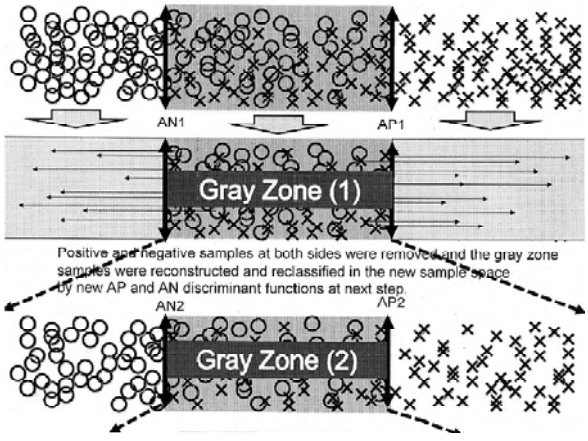

Fig. 4 Improvement of classification rate by K-step Yard sampling (KY) methods. Correctly classified positive and negative samples at both sides were removed and the gray zone samples were reconstructed and reclassified in the new sample space by new discriminant functions at next step.

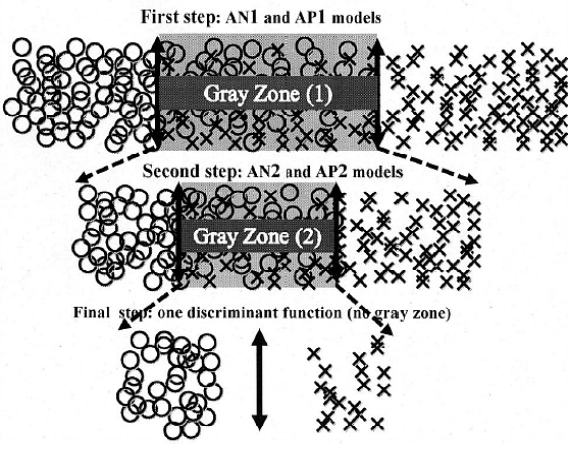

Fig. 5 Meta-algorithm: repetition of reclassification of gray zone. (K-step Yard sampling (KY) methods). High reliability zone (correctly classified samples) were removed and gray zone was reclassified and the sample space was reconstructed by new discriminant functions at next step. All samples were correctly classified at final step.

Table 1 Data flow and structure of KY methods on the skin sensitization data set.

Step 1: 2 models (AP1 model: 81 parameters and AN1 model: 69 parameters) Total compounds 593 (Positive: 419, Negative: 174) Correctly classified compounds 275
Step 2: 2 models (AP2 model: 20 parameters and AN2 model: 86 parameters) Total compounds (Gray zone 1) 318 (Positive: 232, Negative: 86) Correctly classified compounds 211
Step3: 1 model (28 parameters) Total compounds 107 (Gray zone 2) (Positive: 55, Negative, 52) Correctly classified compounds 107 (perfect classification)

RESULS

In this study, all discriminant functions (AN, AP and final one discriminant function) were generated by TILSQ. At step1, the starting 593 compounds (419 positive sensitizers and 174 negative sensitizers) were classified into three groups. Total 275 compounds were correctly classified for positive and negative classes by AN1 and AP1 models. Samples located on the gray zone (1) (318 compounds) were reclassified by AN2 and AP2 models at step2. 211 compounds were correctly classified. Gray zone (2) (107 compounds) were reclassified by ordinary method (TILSQ) at step 3. All 593 compounds were perfectly classified by 3 steps (Table 1).

The correct classification of QSTR prediction for guinea pig data and murine local lymph node assay (LLNA) data were 68.3% (sensitivity 69.7%, specificity 54.5%) and 61.2% (sensitivity 60.7%, specificity 62.8%).

DISCUSSION

Since the implementation of Animal Welfare Guideline 86/609/EC in 1986, it is the declared policy of EU institutions to support the development and use of alternative methods of testing chemicals, that is of “any method that can be used to replace, reduce or refine the use of animal experiments in biomedical research, testing or education.”[11] However, no *in vivo* replacement is currently available for testing skin sensitization in compliance with REACH system [3, 12].

Therefore, several QSTR-related systems have been developed for skin sensitization. These are Toxicity Prediction Komputer-Assisted Technology (Acclyrs Inc., CA, USA; TOPKAT) and Multi Computer - Automated Structure Evolution (MultiCASE Inc., Ohio, USA: M-CASE), which are both statistically based, Deductive Estimation of Risk from Existing Knowledge (Dereck) for Windows (DfW. LHASA Ltd., Leeds, UK), which is knowledge-based and Times Metabolism Stimulator for Skin Seneitization (LMC, University of Bourgas, Bulgaria; TIMES-SS) which is a hybrid [13, 14]. Our previous QSTR model for skin sensitization is also statistically based [4]. All models are required high classification and prediction rate, e.g., over 90%.

In this study, the KY methods could be applied to this QSTR study. All 593 skin sensitization compounds were classified perfectly by total 3 steps. These methods could classify a set of Ames test samples (6,965 compounds, 2932 posi, 4033 nega) into two classes (Positive/Negative) correctly by 23 steps (almost TILSQ and some AdaBoost, data are not shown). If only linear discriminant function is performed, this new methods always attain perfect classification rate without over-fitting, which causes lower prediction rate. This QSTR model applied only TILSQ, linear discriminant function. The correct clas-

sification of QSTR prediction for guinea pig data and LLNA data were 68.3% (sensitivity 69.7%, specificity 54.5%) and 61.2% (sensitivity 60.7%, specificity 62.8%). That of TOPKAT for guinea pig test was 73.3% (75.6%, 65.2%) and those of Derek were 82.9% (92.7%, 47.8%) and 73% (87.1%, 32.6%) [10]. Our results of sensitivity were lower than those of TOPKAT and Derek. One of the reasons might be that our criteria of sensitizer are composed of animal and human data. However, GPMT, BA, and LLNA have only a 72-73% total accuracy of predicting actual human skin sensitizer [9, 15]. Our results might not be necessarily low, but appropriate. More research is needed to improve prediction rate. This QSTR model applied only TILSQ, linear discriminant function. We concluded that this QSTR system is thought to be applicable to initial prediction of skin sensitizing ability of untested chemicals and that the KY methods were promising tool in QSTR technology (classifying and predicting toxicity compounds).

ACKNOWLEDGEMENTS

Financial support for this research was provided by a Grant-in-Aid from the Ministry of Health, Labour and Welfare, Japan (H20-Labour-009).

REFERENCES

1) Kanerva L, Elsner LP, Wahlberg JE, Maibach HI (eds). Handbook of occupational dermatology. 2000, Springer, Berlin, Germany.
2) EU Regulation (EC) 1907/2006 of the European Parliament and the European Council of 18 December 2006 concerning the Registration, Evaluation, Authorization and Restriction of Chemicals (REACH), establishing a European Chemical Agency, amending Regulation 1999/45/EC and repeating Council Regulation (EEC) No 93/793 and Communication Regulation (EC) No 1488/94 as well as Council Directive 76/769/EEC and Commission Directive 91/155/EEC, 93/677/EEC and Commission Directive 91/155/EEC, 93/677/EEC, 93/105/EEC 93/105/EEC and 2000/21/EEC, Off J Eur Comm L 396, 30 December 2006, pp1-849.
3) Patlewicz G, Aptula AO, Roberts DW, Uriarte E. A minireview of available skin sensitization (Q)SARs/expert systems. QSAR Comb. Sci. 2008; 27: 60-76.
4) Sato K, Umemura T, Tamura T, Kusaka Y, Aoyama K, Ueda A, Harada K, Minamoto K, Otsuki T, Yamashita K, Takeshita T, Shibata E, Dobashi K, Kameo S, Miyagawa M, Kaniwa M, Endo Y, Yuta K. Skin sensitization study by quantitative structure-activity relationships (QSAR). AATEX 2009;14:940-946.
5) K-step Yard sampling (KY) methods: Kohtarou Yuta, US Patent No.:US 7725413 B2 May 25, 2010.
6) Deutsche Forschungsgemeinschaft (DFG). IV Sensitizing substances. In Lst of MAK and BAT Values. Ailey-VCH, 2008, Weinheim, pp.158-173.
7) Coz CJL, Lepoittevin JP. Dictionary of contact allergens: chemical sources, sources and references. In Frosch P.J., Menne T., Lepoittevin J.P. eds. Contact Dermatiis 4th ed. 2006, Springer-Verlag, Berlin, pp.943-1105.
8) The Japanese GHS Inter-ministerial Committee. Results of the classification, in NITE Information about the status of the implementation of GHS in Japan, http://www.safe.nite.go.jp/english/ghs_index.html#results accessed on February 10th, 2012
9) Haneke KE, Tice RR, Carson BL, Margolin BH, Stokes WS. ICCVAM evaluation of the murine local lymph node assay. Data analyses completed by the National Toxicology Program Interagency Center for the Evaluation of Alternative Toxicological Methods. Regul Toxicol Pharmacol 2001;34:274-286.
10) Fedorowicz A, Singh H, Soderholm S, Demchuk E. Structure-activity models for contact sensitization. Chem Res Toxicol 2005;18:954-969.
11) Lilienblum W, Dekant W, Foth H, Gebel T, Hengstler JG, Kahl R, Kramer PJ, Schweinfurth H, Wollin KM. Alternative methods to safety studies in experimental animals: role in the risk assessment of chemicals under the new European Chemicals Legislation (REACH). Arch Toxicol 2008;82:211-236.
12) Grindon C, Combes R, Cronin MT, Roberts DW, Garrod JF. An integrated decision-tree testing strategy for skin sensitisation with respect to the requirements of the EU REACH legislation.

Altern Lab Anim 2008;36(Suppl 1):75-89.
13) Patle wicz G, Aptula AO, Uriarte E, Roberts DW, Kern PS, Gerberick GF, Kimber I, Dearman RJ, Ryan CA, Basketter DA. An evaluation of selected global (Q)SARs/expert systems for the prediction of skin sensitisation potential. SAR QSAR Environ Res. 2007;18: 515-541.
14) Patlewicz G, Dimitrov SD, Low LK, Kern PS, Dimitrova GD, Comber MI, Aptula AO, Phillips RD, Niemelä J, Madsen C, Wedebye EB, Roberts DW, Bailey PT, Mekenyan OG. TIMES-SS--a promising tool for the assessment of skin sensitization hazard. A characterization with respect to the OECD validation principles for (Q)SARs and an external evaluation for predictivity. Regul Toxicol Pharmacol 2007;48:225-239.
15) The murine local lymph node assay. The results of an independent peer review evaluation coordinated by the Interagency Coordinating Committee on the Valuation of Alternative Methods (ICCVAM) and the National Toxicology Program Interagency Center for the Evaluation of Alternative Toxicological Methods (NICEATM). NIH Publications no.99-4494, 1999, NIH, Research Triangle Park, NC.

(Received, February 13, 2012/
Accepted, March 27, 2012)

Corresponding author:
Associate Professor Kazuhiro Sato, Ph.D.,
Department of Environmental Health,
School of Medicine, University of Fukui,
Shimoaizuki 23-3, Matsuoka, Eiheiji-cho,
Yoshida-gun, Fukui 910-1193, Japan
Tel.:+81-776-61-8338
Fax:+81-776-61-8107
E-mail: satokazu@u-fukui.ac.jp