



Fragrance Allergens: Classification and Ranking by QSAR

J. J. HOSTÝNEK*†‡ and P. S. MAGEE‡§

†Euromerican Technology Resources, Inc., Lafayette, CA 94549,

‡UCSF School of Medicine, Department of Dermatology, San Francisco, CA 94143
and §BIOSAR Research Project, Vallejo, CA 94591, USA

(Accepted 7 March 1997)

Abstract—Quantitative structure–activity relationship (QSAR) models which predict both skin penetration and cell mediated immunity for small molecular weight non-electrolytes developed earlier were validated on 74 known allergens and non-allergens chosen among fragrance chemicals in common use to test discriminating and grading power. While the test set used for classification was based on experience in humans exclusively, the rank model was tested for sensitization potency including guinea pig data also. In the classification test, 12 of 74 compounds fell in the indeterminate range and were non-classifiable by the present QSAR model. On the remaining 62 compounds the model performs with 90% sensitivity and 100% specificity at 92% concordance. The rank model correctly grades 65 of 74 compounds (88% concordance), with 60% specificity based on exact prediction of non-allergens (NON), and 95% sensitivity on allergens (ACD) allowing for a variance of $\pm$ one level among weak, moderate and severe ratings. In combination, the two models perform with 93% overall concordance on the test set of 74 compounds.
© 1997 Published by Elsevier Science Ltd

Abbreviations: ACD = allergic contact dermatitis; allergens; FCA = Freund's complete adjuvant; GPMT = guinea pig maximization test; HBA = hydrogen bond acceptor; HBD = hydrogen bond donor; ICDRG = International Contact Dermatitis Research Group; IFRA = International Fragrance Research Association; MR = molar refraction; NACDG = North American Contact Dermatitis Group; NON = non allergens; QSAR = quantitative structure–activity relationship; RIFM = Research Institute of Fragrance Manufacturers; Y-est = Y-estimate (model score).

INTRODUCTION

Increasingly, published data provide evidence that an important part of cutaneous cosmetic intolerance is due to their fragrance component, the incidence of sensitization ranging from 7% to 18% of a normal, asymptomatic test population (Dooms-Goossens, 1993; Dooms-Goossens *et al.*, 1992; Fisher, 1995; Guin and Berry, 1980; Larsen *et al.*, 1992; Maibach, 1991; Nielsen and Menne, 1992). It is understandable therefore that the fragrance and cosmetics industries, health professionals as well as regulatory agencies, assign high priority to the development of methods predictive of allergy that are robust and minimize use of experimental animals. Computers and statistical software now make it possible to move conventional animal evaluation programs away from unavoidable experimental and biological variables onto a more objective plane, where physicochemical properties and chemical reactivity can be evaluated individually as factors determining events such as skin penetration and allergic sensitization. Quantitative structure–

activity relationships (QSAR) is one such alternative to animal experimentation which objectively describes the interaction between chemicals and biological systems (Hansch and Leo, 1995; Hansch *et al.*, 1995), and is now increasingly used in the prediction of a variety of toxicological endpoints, including allergic contact dermatitis (ACD) (Cronin and Dearden, 1995).

Biologically-based, comprehensive QSAR models combining both structural alerts with barrier diffusion criteria have been developed through systematic analysis of known contact allergens, which identify both the features that make a chemical an allergen, as well as the physicochemical properties which determine penetration of a compound through the stratum corneum (Hostýnek, 1995; Magee *et al.*, 1994a). Development of these models seek to encompass the several characteristics known to be necessary attributes for immunogenic activity. Such determining factors are: (a) molecular size and hydrogen bonding critical for diffusivity through the stratum corneum; (b) metabolism of prohaptens in the skin; (c) lipophilic substructures which have affinity for cell membrane lipids; and (d) electrophilic and nucleophilic haptens which bind to native

*Author for correspondence at: 3644 Happy Valley Rd., Lafayette, CA 94549, USA.

proteins, as well as free radical generators which modify the structure of proteins, and by forming covalent bonds yield non-self proteins. Adequate factoring of all components however, poses significant problems of a statistical nature. To cover the diverse functional groups present among electrophiles, nucleophiles and potential radical generators and, most importantly, the spatial relations within molecules which often determine reactivity, a comprehensive database would necessarily consist of hundreds of compounds, as documented in the recent predictive systems DEREK (Ridings *et al.*, 1996), TOPKAT (Enslein *et al.*, 1996) and MULTICASE (Graham *et al.*, 1996). In these so-called expert systems, where at their present stage prediction rests on a set of rulebases, the larger the data set the more comprehensive is the analysis for toxophores statistically associated with protein reactivity. In our model development, which combines structural alerts for protein reactivity with criteria for diffusivity and metabolism, increasing sample size causes the importance of reactivity descriptors to outweigh the physicochemical parameters which account for the barrier permeation process, to the point where their statistical relevance disappears altogether. Thus, we could establish empirically (Magee *et al.*, 1994a) that the number of 60–80 compounds represents an optimal trade-off, where size, polarity and hydrogen bond formation are still adequately important in describing translation of non-electrolytes through a biological membrane such as the skin. To our knowledge, the QSAR model developed earlier using stepwise regression, discriminant and principal component analysis (Magee *et al.*, 1994a) and which we propose to test here, incorporates all the above prerequisites for adequate ACD prediction to a near-optimal degree. It lies in the nature of statistics mentioned that some exceptional biophores will occasionally remain outside that model.

Considering the prominence that fragrances have attained in the aetiology of contact hypersensitivity among the general population, demonstration of robust predictive performance of the QSAR model as applied to fragrance chemicals of well-documented sensitization potential could demonstrate its utility in the prediction of ACD potential of new, as-yet untested structures. With that intent, we proceeded to test the existing discriminant model in the classification of proven human allergens and non-allergens among a number of compounds frequently used in perfumery, and also evaluated the performance of a newly developed rank model in grading for allergenic potency.

MATERIALS AND METHODS

Criteria for compound selection

It is the primary objective of this study to establish the utility of QSAR to predict sensitization in man by

compounds used in perfumery, and thus it was of paramount importance for the validation of the previously constructed models to select compounds for which sensitizing properties had been demonstrated in humans. A reliable source for a list of sensitizers is IFRA, the International Fragrance Research Association. With headquarters in Geneva, Switzerland, IFRA is composed of the national associations of over 100 fragrance manufacturers in 15 countries. Its Technical Advisory Committee reviews all available data on fragrance ingredients and, when necessary, issues recommendations for the safe use of these materials, including restrictions on use levels and advisories not to use substances at all when warranted by safety considerations. Thus, to date, the use of nearly 100 fragrance raw materials has been either restricted or prohibited, in greatest part due to their potential for sensitization (Ford, 1991). Such risk assessment is based on information gathered by RIFM (Research Institute for Fragrance Materials), an association whose members represent most of the fragrance industry in the United States, Europe and Japan, and reflects the clinical experience gathered by dermatologists from most countries of the industrialized world.

For the fragrance test set, 33 sensitizers were culled from the IFRA list of compounds of established sensitizing potency; an additional 26 ACD actives were selected from the various lists of standard patch test series of allergens (ICDRG, 1988; Matsuo *et al.*, 1993; NACDG, 1984) and the RIFM Fragrance Raw Materials Monographs reflecting the result of human maximization tests which were also designed to evaluate the potency of allergens (Kligman, 1966). These were published over 25 years in *Food and Chemical Toxicology* (formerly *Food and Cosmetics Toxicology*), with the latest series in 1992 (Ford *et al.*, 1992). Furthermore, clinical test reports of reputable origin were included when they were based on clinical history, physical examination and patch testing (Klecak *et al.*, 1977; Larsen *et al.*, 1992), and data from multicenter ACD investigations (Eiermann *et al.*, 1982; Frosch *et al.*, 1995; Larsen, 1985; Larsen *et al.*, 1996). 15 non-sensitizers (NON) for the test set were selected based on a Medline search for the years 1966–1995, and also from the RIFM Fragrance Raw Materials Monographs.

For the graded test set, entries from the IFRA negative list in the 'prohibited' category were entered *a priori* as strong sensitizers; those in the 'restricted' category as moderate. Furthermore, in order to obtain additional data on potency for ACD compounds other than those on IFRA's negative list and the human maximization tests, it was necessary to include animal data obtained from experimental protocols precluded in human testing. Such complementary data were taken from the GPMT literature (Cronin and Basketter, 1994).

Prerequisite for selection of all compounds used in the validation studies was their listing in the reference

book on perfumery chemicals by S. Arctander (Arctander, 1960).

QSAR model development

QSAR model development relies on the process of learning from experience. It is based on the assumption that a mathematical relationship exists between physicochemical and chemical properties of chemical agents and their action in the living organism. In that process a so-called training set is chosen which involves collection of pertinent clinical data, and the database thus obtained is then analysed for statistically significant parameters by discriminant or regression analysis from a wealth of chemical and physicochemical descriptors. This process yields information on structural elements (independent variables) statistically associated with biological activity (dependent variables), for instance on the interdependence of chemical causes and clinical effects. The output from this process, in the form of an algorithm or a mathematical equation, represents the knowledge gained from experience. QSAR studies applied to physiological phenomena such as the penetration of biological membranes or hypersensitivity reactions, also present the benefit of yielding detailed information on the processes occurring at the cellular and molecular level; those descriptors which survive the scrutiny for statistical significance are also those likely to determine course and end result in the real-life biological process. For validation, the algorithm derived is then applied to the analysis of a new set of compounds, the test set, and correlated with clinical data obtained *in vivo*. The more rigorous the selection of input data in model construction, the higher will be the correlation coefficient in the test set.

QSAR statistics

The statistics used in the models to measure goodness-of-fit are the student's *t*-value for the coefficients and the measures *s*, *r*² and *F*.

The student's *t*-value is the ratio of the coefficient to the error in the coefficient, larger values relating to smaller ranges of error.

The standard error of the equation, *s*, is essentially a standard deviation of the points from the *Y* estimate *v.* *Y* line. For larger datasets (*n* > 50), outliers are defined as points deviating from the line by more than 2.5 standard errors and are generally not included in the regression.

The correlation coefficient, *r*, is derived from the directly calculated *r*², which measures the variance in the data explained by the coefficients of the equation. We have used *r*² exclusively as a better measure of the goodness-of-fit and scatter about the regression line. The unexplained variance, 1-*r*², is a measure of the regression error, which must not exceed that of the original experimental error.

The Fisher distribution, *F*, is a measure of the

overall strength of the equation. Values greater than 10 are considered highly significant.

The *Y*-estimate (*Y*-est) is the model's output, the value derived by the equation for a specific query.

The discriminant model

A statistical discriminant QSAR model based on 72 ACD and NON compounds was developed and tested successfully in diverse studies (Hostýnek *et al.*, 1995; Magee *et al.*, 1994a,b). It is based on a combination of continuous transport/binding descriptors and dichotomous (0/1) descriptors reflecting protein reactivity by simple count of toxophores (haptens and prohaptens). A full description of model development is contained in the primary reference (Magee *et al.*, 1994a). Experience has led to stepwise modification of the model for improved performance, and the one applied here is the latest revised version. The descriptors used in this model are characterized as follows.

Continuous transport/binding descriptors are molar refraction (MR), the polarizable molecular volume modelling London forces in non-specific binding (Charton, 1983), and PL (lipophilic contribution of Log*P*_{octanol:water}; PL + PH = Log*P*) which models lipophilic interactions of compounds with proteins and membranes. Both are easily calculable from contributing fragments (Hansch and Clayton, 1973; Hansch and Dunn, 1972; Hansch and Leo, 1979; Magee, 1991). Of the ordinal descriptors, HBA is a simple count of the electron pairs on N and O that models the concentration potential of hydrogen bond acceptor groups, and HBD is a simple count of the N-H and O-H groups that models the concentration potential of hydrogen bond donor groups. Dichotomous descriptors are used to code for the presence or absence of protein reactive substructures (1/0). The descriptor ICOOR codes for simple aliphatic esters and models esterase degradation; ICONJ codes for conjugated olefins activated by groups that cause Michael addition of SH and NH₂ protein end-groups; IX codes for reactive aliphatic or aromatic halides; IArOH codes for easily metabolized phenols (to reactive quinones); IArNH₂ codes for easily metabolized anilines (to reactive quinones); IOH codes for easily oxidized primary alcohols (to reactive aldehydes); IUNIQ codes for strong nucleophiles and for unusual reactive electrophiles that occasionally require coding. In Table 1, the 1/0 values of the reactivity descriptors are omitted for reasons of space as they are easily recreated by visual inspection of the molecular structures.

In earlier studies involving diverse structures, the discriminating power of the equation fell in the range of 79–88% correct classification, predicting non-allergens somewhat more efficiently than allergens (Magee *et al.*, 1994a). Model output is read from the *Y*-estimate (*Y*-est). Allergens and

Table 1. Compounds, ACD ratings, continuous descriptors and model performance in the fragrance test set

Compound [ratings]	CAS NO.	MR	PL	HBA/HBD	Score	Pass/fail
1,2-Propylene glycol [1a/2b]	57-55-6	19.56	1.62	4/2	0*/1 ^d	0*/0 ^f
1-(4-Methoxyphenyl)-1-pentene-3-one** [1/4]	104-27-8	57.88	4.53	4/0	1/3	1/1
1-Octene-3-yl acetate* [1/3]	2442-10-6	50.26	4.67	4/0	0/1	0/0
2-Amyl-cinnamaldehyde [1/3]	122-40-7	65.30	5.13	2/0	1/3	1/1
2-Hexyl-cinnamaldehyde [1/4]	101-86-0	69.92	5.90	2/0	Bg/3	-/1
2-Methyl-cinnamaldehyde [1/3]	101-39-3	46.82	2.97	2/0	1/4	1/1
2-Phenylethyl alcohol [0/1]	60-12-8	37.45	3.00	2/1	0/1	1/1
2- <i>n</i> -Pentylidene cyclohexanone** [1/4]	25677-40-1	52.77	4.97	2/0	1/3	1/1
3,5,5-Trimethylhexyl acetate [0/1]	58430-94-7	54.16	5.71	4/0	0/1	1/1
3-Methyl-2(3)-nonene nitrile* [1/3]	53153-66-5	49.66	4.41	1/0	0/2	0/1
4-Allyl anisole [1/4]	140-67-0	45.59	3.86	2/0	1/3	1/1
4-Hydroxybenzyl alcohol [1/4]	623-05-2	34.65	2.33	4/2	1/3	1/1
4-Methoxyacetophenone [0/1]	100-06-1	43.30	3.45	4/0	0/1	1/1
4-Methoxybenzaldehyde [1/3]	123-11-5	38.68	2.79	4/0	1/3	1/1
4-Methoxybenzyl alcohol [1/3]	105-13-5	39.27	3.22	4/1	1/3	1/1
4-Methylhydrocinnamaldehyde* [1/3]	99-72-9	46.10	3.87	2/0	1/3	1/1
6-Isopropyl-2-decalol** [1/4]	34131-99-2	59.82	5.79	2/1	0/1	0/0
6-Methylheptadienone* [1/3]	1604-28-0	41.69	3.13	2/0	1/3	1/1
7-Methoxycoumarin** [1/4]	513-59-9	46.63	2.86	6/0	1/3	1/1
Allylheptyne carboxylate* [1/3]	73157-43-4	50.76	4.63	4/0	1/3	1/1
Amylcyclopentenone* [1/3]	25564-22-1	48.15	4.18	2/0	1/3	1/1
Anethole [1/4]	4180-23-8	47.48	3.86	2/0	1/3	1/1
Anisylidene acetone** [1/4]	943-88-4	53.26	3.98	4/0	1/3	1/1
Benzophenone-2 [1/3]	119-61-9	56.57	3.80	2/0	1/4	1/1
Benzyl alcohol [1/4]	100-51-6	32.83	2.56	2/1	1/4	1/1
Benzyl dimethylcarbinol [0/1]	100-51-6	47.69	3.92	2/1	0/1	1/1
Benzyl salicylate [1/2]	118-58-1	63.06	4.23	5/0	B/2	-/1
Benzylidene acetone** [1/4]	122-57-6	46.82	3.32	2/0	1/3	1/1
Carvone [1/3]	2244-16-8	38.91	3.52	2/0	1/3	1/1
Cinnamaldehyde** [1/4]	104-55-2	42.20	2.72	2/0	1/3	1/1
Cinnamic alcohol* [1/3]	104-54-1	42.79	2.97	2/1	1/4	1/1
Citral** [1/4]	5392-40-5	50.93	3.73	2/0	1/3	1/1
Citronellal [1/3]	106-23-0	50.21	4.51	2/0	B/3	-/1
Coumarin [1/2]	91-64-5	40.19	2.20	4/0	1/3	1/1
Cyclamen alcohol** [1/4]	4756-19-8	60.55	5.42	2/1	1/4	1/1
Diethyl maleate** [1/4]	75554-12-3	42.30	3.15	8/0	1/3	1/1
Dihydrocoumarin** [1/4]	119-84-6	39.47	2.15	4/0	1/3	1/1
Dimethyl citraconate** [1/4]	617-54-9	37.68	2.72	8/0	1/3	1/1
Eugenol [1/3]	97-53-0	49.23	3.35	4/1	1/3	1/1
Farnesol [1/2]	4602-84-0	75.34	6.29	2/1	1/4	1/0
Fixolide [0/1]	1506-02-1	81.00	7.50	2/0	0/2	1/1
Folione* [1/3]	111-12-6	40.80	4.22	4/0	1/3	1/1
Geraniol [1/3]	106-24-1	51.52	4.27	2/1	1/4	1/1
Heliotropine [1/3]	120-57-0	38.44	2.33	6/0	1/3	1/1
Hexahydrocoumarin** [1/4]	700-82-3	41.62	3.31	4/0	1/3	1/1
Hexyl salicylate [0/1]	6259-76-3	61.83	5.14	5/0	0/1	1/1
Hydroabietyl alcohol** [1/4]	70775-80-3	90.10	8.87	2/1	B/4	-/1
Hydroxycitronellal* [1/3]	107-75-5	51.31	4.95	4/1	B/3	-/1
Isoamyl salicylate [0/1]	87-20-7	57.21	4.47	5/0	0/1	1/1
Isoeugenol* [1/3]	97-54-1	49.23	3.46	4/1	1/3	1/1
Lilial [0/1]	80-54-6	64.58	5.52	2/0	B/3	-/0
Limonene(dipentene) [0/1]	138-86-3	47.64	4.35	0/0	0/2	1/1
Menthol [1/2]	89-78-1	48.02	5.35	2/1	0/1	0/1
Methyl cinnamate [1/3]	103-26-4	46.87	3.32	4/1	1/3	1/1
Methyl crotonate** [1/4]	623-43-8	27.16	2.19	4/0	1/3	1/1
Methyl eugenol [1/2]	93-15-2	55.91	4.52	4/0	1/3	1/1
Methyl <i>N</i> -methylantranilate [0/1]	85-91-6	45.92	3.45	4/0	0/1	1/1
Methyl salicylate [0/1]	119-36-8	38.73	2.56	5/0	B/1	-/1
Musk ketone [0/1]	81-14-1	81.86	6.21	6/0	0/2	1/1
Musk xylene [0/1]	81-15-2	73.10	4.78	6/0	B/2	-/1
<i>p</i> - <i>tert</i> -Butyl phenol** [1/4]	98-54-4	46.69	3.92	2/1	B/3	-/1
Pentyl cyclopentanone propanone [0/1]	40942-73-2	62.52	5.80	4/0	0/1	1/1
Perillaldehyde* [1/3]	2111-75-3	48.87	3.27	2/0	1/4	1/1
Phenyl salicylate [1/3]	118-55-8	58.44	3.57	5/0	1/3	1/1
Phenylacetaldehyde** [1/4]	122-78-1	36.86	2.67	2/0	1/3	1/1
Phenylethyl salicylate [1/2]	87-22-9	67.68	4.65	5/0	B/2	-/1
Propyl paraben [1/2]	94-13-3	47.97	3.52	6/1	1/3	1/1
3-Propylenephthalide* [1/3]	17369-59-4	49.43	3.28	4/0	1/3	1/1
Pseudomethyl ionone** [1/4]	26651-96-7	70.13	6.17	2/0	1/3	1/1
Thymol [1/3]	89-83-8	46.69	4.05	2/1	B/3	-/1
Trans-2-heptenal** [1/4]	18829-55-5	36.35	3.05	2/0	1/3	1/1
Trans-2-hexenal diethyl acetal** [1/4]	67746-30-9	52.62	5.31	4/0	B/3	-/1
Trans-2-hexenal dimethyl acetal** [1/4]	18318-83-7	43.88	4.47	4/0	1/3	1/1
Versalide [0/1]	88-29-9	81.00	7.45	2/0	0/2	1/1

(a) Input class ACD = 1; NON = 0. (b) input rank 1–4. (c) model class 1 = ACD; 0 = NON. (d) model rank 1–4. (e) model performance: 1 = pass; 0 = fail. (g) indeterminate class. (*) restricted by IFRA. (**) prohibited by IFRA.

non-allergens are scored as Y-est values above and below 0.5, and we consider values falling in the range 0.4–0.6 to be of indeterminate category and beyond the classification ability of the model. The equation for the latest version of the model is given in Table 2.

The rank model

The construction of a rank model based on a different set of 89 compounds has been described elsewhere (Hostýnek *et al.*, 1996). The method of ranking is based on non, weak, moderate and strong responses (1–4); it follows the pattern established by Magnusson and Kligman (1969), but with significant numerical differences. The Magnusson–Kligman scale (1–5) is determined by the number of guinea pigs sensitized in the maximization test with 1 = weak (0–8%), 2 = mild (9–28%), 3 = moderate (29–64%), 4 = strong (65–80%) and extreme (81–100%). Our scale corresponds to 1 = non-allergenic (0%), 2 = weak to mild (1–28%), 3 = moderate (29–64%) and 4 = strong to extreme (65–100%). Thus, the compounds selected to represent the NON group are designated as '1' and are similar to the selection identified as NON in the discriminant model described above. All of the others (ranks 2–4) are allergens of weak-to-mild, moderate and strong-to-extreme ACD inducers. With these differences in mind, the graded responses were correlated by multiple regression against the full set of 22 descriptors studied for our discriminant model (Magee *et al.*, 1994a). The current version of the model is given in Table 3. Despite the difference in procedure and data set, the significant descriptors are found to be identical to those in the discriminant model (Table 2) and only the descriptor coefficients distinguish the two models. Formal differences in the grading model are that IArOH and IArNH₂ were combined to IQUIN, and ICONJ and IUNIQ to IPOS because the coefficients proved to be similar. Note that the sign of HBA is reversed and the dependence on ester degradation, ICOOR, is statistically weak.

Scoring rank model performance

Scoring the rank model is not as straightforward as the 'yes–no' scoring with the discriminant model, due to the significant overlap in the model's score for the respective class levels. The following ranges for classes 1–4 are: 1 = 0.63–1.97; 2 = 1.24–3.10; 3 = 1.81–4.14; 4 = 2.66–4.88. As the equation score Y-est can fall into more than one range, the criterion used is closest proximity to the mid-point value of each range. The mid-point values for each range are: 1 = 1.30; 2 = 2.17; 3 = 2.97; 4 = 3.77. Measured against these benchmarks, there are two levels of concordance possible for performance assessment. Ideally, the Y-est of the equation would accurately reflect the clinical result and fall exactly within one range. On a more realistic level, the experimental variability of $\pm$ one level described by Gad (1988) is applied in scoring here, as no model can perform better than the variance in the data employed.

Classes 1 and 4 are perfectly separated, but the overlap between successive classes is significant. Classes 2 and 3 are well represented by the average Y-est (2.1 and 3.1), but class 1 is overpredicted (av. Y-est = 1.3) while class 4 is underpredicted (av. Y-est = 3.5). The estimated span covers only 2.2–3.0 (73%) of the input scan. The problem, of course, is that dichotomous reactivity descriptors cannot successfully model a range of 1–4. Repairing this shortcoming is the subject of ongoing research towards further improvement of the models.

Aside from these problems, the ranked model performs well within the $\pm$ one level variance already inherent in experimental test data as described by Gad (1988). In an earlier validation study (Hostýnek *et al.*, 1996) of 61 allergens and non-allergens not used in the learning set, we find a concordance of 90% overall and 97% for separation of inactives from strong allergens. Application of the rank model for the rating of ACD response thus appears to be a useful corollary for confirmation of data obtained with the discriminant model.

Table 2. The statistical discriminant model

Descriptor	Coefficient	T-Value
Molar Refraction MR	0.0116	2.49
Lipophilicity Factor PL	–0.161	–3.38
Hydrogen Bond Acceptors HBA	0.0498	2.88
Hydrogen Bond Donor Groups HBD	–0.129	–2.91
Aliphatic Esters ICOOR	–0.284	3.55
Michael Reactants ICONJ	0.570	6.47
Reactive Halides IX	0.160	4.73
Aromatic Phenols IArOH	0.596	6.18
Aromatic Amines IArNH ₂	0.503	4.82
Primary Alcohols IOH	0.207	2.51
Reactive Nucleophiles and Electrophiles IUNIQ	0.512	4.12
Intercept	0.118	

Learning set: ACD (Class 1): n = 36 NON (Class 0): n = 36

t-value: Student's t-value; total sample size n = 70 (two outliers); explained variance r^2 = 0.750; standard error s = 0.275; Fisher statistics F = 15.81.

Table 3. The statistical rank model

Descriptor	Coefficient	T-Value
Molar Refraction MR	0.0294	2.47
Lipophilicity Factor PL	-0.281	-2.55
Hydrogen Bond Acceptors HBA	-0.0960	-2.28
Hydrogen Bond Donors HBD	-0.207	-2.48
Aliphatic Esters ICOOR	-0.409	-1.69
Primary Alcohols IOH	0.547	2.84
Reactive Halides IX	0.519	3.54
Michael Reactants, Reactive Nu & El philes IPOS	1.375	10.70
Aromatic Amines and Alcohols IQUIN	1.390	9.17
Intercept	1.70	

Learning set: NON allergens (Class 1): $n = 29$ WEAK allergens (Class 2): $n = 18$
 MODERATE allergens (Class 3): $n = 15$ STRONG allergens (Class 4): $n = 26$
 t -value: Student's t -value; total sample size $n = 88$; explained variance $r^2 = 0.726$ standard error
 $s = 0.680$; Fisher statistics $F = 23.00$.

RESULTS

Discriminant model

Of the 74 fragrance chemicals in Table 1, the discriminant model places 12 (16%) in the indeterminate region, (Y -est = 0.4–0.6) and thereby they are not predictable by the equation. Of the 62 remaining chemicals, the model correctly predicts 57 (92% concordance). All five incorrectly predicted compounds are false negatives, rated as allergens in the literature. Thus, the model performs better on non-allergens (12/12, 100% specificity) than on allergens (45/50, 90% sensitivity), which is consistent with our previous studies (Magee *et al.*, 1994a,b).

Rank model

To evaluate rank model performance, it must be noted that all non-allergens have to be predicted exactly as such with a score falling within the Y -est brackets of 0.63–1.97 to be correct, and are ranked as "1". By such convention, model specificity is 60% (9/15). For allergens (rank "2–4"), a correct score is the exact level $\pm$ one level for a pass score, in accord with the variance to be expected in experimental data (Gad, 1988). Thereby, model sensitivity is 95% (56/59). Of the 74 fragrance chemicals in the test set (Table 1), overall rank model performance is 88% (65/74). Of the six incorrectly predicted ranks (1–4), four are also incorrectly predicted by the discriminant model. As these four are false negatives, the rank model also performs better in predicting non allergens than allergens. Of the remaining two misses, one is correctly predicted by the discriminant model (ACD) while the other is borderline and not classifiable. Of greatest interest to this study is the fact that 11/12 (92%) of the borderline cases from the discriminant model are correctly assigned by the rank model. Thus the use of the two models is both confirmatory and supplementary.

DISCUSSION

The validation performance achieved by the two QSAR models, tested mainly on clinical evidence of

sensitization in man, demonstrates the importance of skin penetration and skin metabolism as relevant factors in an algorithm expected to predict ACD potential. Other approaches such as DEREK (Ridings *et al.*, 1996) or TOPKAT (Enslein *et al.*, 1996) which are based on a vast body of GPMT data, are eminently qualified for the determination of protein reactivity or allergenic potential of chemical structures of great diversity, but not for estimating the actual risk of sensitization in man.

It is not surprising that the discriminant model performs with high specificity (100%), since classification of the compounds for ACD activity in the test is entirely based on empirical evidence obtained from human exposure under realistic use patterns. Those reactions avoid limit conditions as they are provoked in animal experimentation, such as intracutaneous guinea pig tests or the use of FCA, and account for the barrier function of the skin. Prediction of a compound as an allergen by the model means that substructures were coded as protein reactive, while prediction as a non-allergen indicates that either no substructures were considered to be protein reactive or, as in some cases, the mechanistic factors determining barrier penetration outweigh the substructure descriptors.

Comparative evaluation of performance shows that the two QSAR models are confirmatory and supplement each other to a large extent, as the 'fails' in one overlap with the 'passes' in the other: four of the nine rank 'fails' are correctly predicted by the discriminant model; 10 of the 12 borderline cases in the discriminant model are correctly ranked, and only three of the 62 classified queries fail in both models. Concordance achieved by the two models applied in tandem thus is 93% overall (69/74). This is at least equal to the variance expected for such chemicals from human and animal testing. The present validation exercise demonstrates the advantage of using the two QSAR models in conjunction on a given set of compounds, as they are internally consistent to a large degree with discriminant failures

being identical with three of the rank failures. A number of the 'fails' invite discussion.

1,2-Propylene glycol is reported as a weak/mild allergen (Corazza *et al.*, 1993; Eiermann *et al.*, 1982), presumably by metabolism to the protein reactive lactaldehyde. However, we coded the primary alcohol group as IOH = 0 for lack of direct evidence of oxidative metabolism in human skin. The glycol is too polar to be a good substrate for mixed-function oxydases (Hansch and Leo, 1995) and the very low oral and skin toxicity on several animals ($LD_{50} = 10\text{--}80\text{ g/kg}$) (Sax, 1984) does not suggest any metabolic conversion to a toxic electrophile such as lactaldehyde. Perhaps this is a case where the known mild-moderate skin irritation (Sax, 1984) has been misinterpreted as an ACD response.

1-Octene-3-yl acetate and 3-methyl-(2,3)-nonene nitrile are restricted in perfumery use (classified here as moderate allergens) and 6-isopropyl-2-decalol is prohibited there (classified as a strong/extreme allergen) (Ford, 1991). None contain substructures that we would normally code as protein reactive in these models. However, there is one difference. In the case of 1-octene-3-yl acetate, our current rules of coding do not allow for sequential coding of activation steps such as ester hydrolysis and subsequent oxidation to the obviously ACD active vinyl ketone. Thus, the model fails in these cases even though in some instances we can anticipate probable mechanisms.

The case of 6-isopropyl-2-decalol, prohibited by RIFM (Ford, 1991), is different in that metabolism to other than a ketone of low protein reactivity is not expected. While speculative, it seems that air oxidation of one or more of the tertiary C-H bonds to a reactive hydroperoxide contaminant is a more likely event. Its presence would be readily detected in the commercial grade of the compound.

Menthol is reported to be a weak/mild allergen (Hayakawa *et al.*, 1996; Sharp, 1978; Wilkinson and Beck, 1994) and, like the decalol discussed above, contains none of the protein reactive structures normally coded for these models. Again, metabolic oxidation to the ketone is likely, but as ketones are not normally considered protein reactive haptens, the coding still would be IOH = 0. However, menthol has substructures in common with 6-isopropyl-2-decalol, and the possibility of air oxidation contaminants is of similar credibility. Such scenarios currently lie outside the model and can only be addressed on a case-by-case basis.

Limonene, a hydrocarbon with no protein reactive structures, is correctly predicted as a non-allergen by the discriminant model. However, the rank model overpredicts limonene by one class, based on the Y-test, revealing the present weakness of the model to accurately bracket the lowest and highest ranks. As mentioned under Materials and Methods, in its present state the statistical model overpredicts at the

low end of the scale, and underpredicts at its upper end.

Overall, the successes in the two models clearly outweigh the failures. We have not succeeded in developing models with $n = 72$ and 88 compounds that will cover all conceivable hapten/prohapten structures and at the same time adequately factor in physicochemical parameters. However, each failed prediction represents a far more important piece of information than the successes. It is through insights gained from predictive failures that the models can be further refined and strengthened.

Acknowledgements—The authors thank Dr G. Klecak for in-depth discussion of predictive tests for contact allergy. This work was supported by Avon through the Johns-Hopkins CAAT Program.

REFERENCES

- Arctander S. (1960) *Perfume and Flavor Materials of Natural Origin*. Vol. 1 & 2. S. Arctander, Montclair, NJ.
- Charton M. (1983) Volume and bulk parameters. In *Steric Effects in Drug Design*. Edited by M. Charton and I. Motoc. pp. 57–91. Springer-Verlag, Berlin.
- Corazza M., Virgili A., Mantovani L. and La Malfa V. (1993) Propylene glycol allergy from acyclovir cream with cross-reactivity to hydroxypropyl cellulose in a transdermal estradiol system?. *Contact Dermatitis* **29**, 283–284.
- Cronin M. T. D. and Basketter D. A. (1994) Multivariate QSAR analysis of a skin sensitization database. In *SAR and QSAR in Environmental Research*. pp. 159–179. Gordon and Breach Science Publishers S. A., New York.
- Cronin M. T. D. and Dearden J. C. (1995) QSAR in toxicology: 4. Prediction of non-lethal mammalian toxicological endpoints, and expert systems for toxicity prediction. *Quantitative Structure–Activity Relationships* **14**, 518–523.
- Dooms-Goossens A. (1993) Cosmetics as causes of allergic contact dermatitis. *Cutis* **52**, 316–320.
- Dooms-Goossens A., Kerre S. and Drieghe J. (1992) Cosmetic products and their allergens. *European Journal of Dermatology* **2**, 465–468.
- Eiermann H. J., Larsen W. G., Maibach H. I. and Taylor J. S. (1982) Prospective study of cosmetic reactions. *Journal of the American Academy of Dermatology* **6**, 909–917.
- Enskin K., Gombar V. K., Blake B. W., Maibach H. I., Hostýnek J. J. and Sigman C. C. (1996) New QSTR for the dermal sensitization guinea pig maximization test. *HDI Computational Toxicology News* **19**, 2.
- Fisher A. A. (1995) *Fisher's Contact Dermatitis*. Edited by R. L. Rietschel and J. F. Fowler. Williams & Wilkins, Baltimore, MD.
- Ford R. A. (1991) The toxicology and safety of fragrances. In *Perfumes: Art, Science, Technology*. Edited by P. M. Müller and D. Lamparsky. pp. 441–465. Elsevier, New York.
- Ford R. A., Api A. M. and Letizia C. S. (1992) Monographs on fragrance raw materials. *Food and Chemical Toxicology* **30** (Suppl), 1S–138S.
- Frosch P. J., Pilz B., Andersen K. E., Burrows D., Camarasa J. G., Dooms-Goossens A., Ducombs G., Fuchs T., Hannuksela M., Lachapelle J. M., Lahti A., Maibach H. I., Menné T., Rycroft R. J. G., Shaw S., Wahberg J. E., White I. R. and Wilkinson J. D. (1995) Patch testing with fragrances: results of a multicenter study of the European Environmental and Contact

- Dermatitis Research Group with 48 frequently used constituents of perfumes. *Contact Dermatitis* **33**, 333–342.
- Gad S. C. (1988) A scheme for the prediction and ranking of relative potencies of dermal sensitizers based on data from several systems. *Journal of Applied Toxicology* **8**, 361–368.
- Graham C., Gealy R., Macina O. T., Karol M. H. and Rosenkranz H. S. (1996) QSAR for allergic contact dermatitis. *Quantitative Structure–Activity Relationships* **15**, 224–229.
- Guin J. D. and Berry V. K. (1980) Perfume sensitivity in adult females. *Journal of the American Academy of Dermatology* **3**, 299–302.
- Hansch C. and Clayton J. M. (1973) Lipophilic character and biological activity of drugs II: The parabolic case. *Journal of Pharmaceutical Science* **62**, 1–20.
- Hansch C. and Dunn W. J., III. (1972) Linear relationships between lipophilic character and biological activity of drugs. *Journal of Pharmaceutical Science* **61**, 1–18.
- Hansch C. and Leo A. (1979) *Substituent Constants for Correlation Analysis in Chemistry and Biology*. John Wiley & Sons, New York.
- Hansch C. and Leo A. (1995) *Exploring QSAR—Fundamentals and Applications in Chemistry and Biology*. Vol. 1. Edited by S. R. Heller. American Chemical Society, Washington, DC.
- Hansch C., Leo A. and Hoekman D. (1995) *Exploring QSAR—Hydrophobic, Electronic, and Steric Constants*. Vol. 2. Edited by S. R. Heller. American Chemical Society, Washington, DC.
- Hayakawa R., Yamamura M. and Sugiura M. (1996) Contact dermatitis from l-menthol. *Cosmetics and Toiletries* **111**, 28–29.
- Hostýnek J. J. (1995) Predicting absorption of fragrance chemicals through human skin. *Journal of the Society of Cosmetic Chemists* **46**, 221–229.
- Hostýnek J. J., Lauerma A. I., Magee P. S., Bloom E. and Maibach H. I. (1995) A local lymph-node assay validation study of a structure–activity relationship model for contact allergens. *Archives of Dermatological Research* **287**, 567–571.
- Hostýnek J. J., Magee P. S. and Maibach H. I. (1996) QSAR predictive of contact allergy: scope and limitations. In *Prevention of Contact Dermatitis*. Edited by P. Elsner, J. M. Lachapelle and J. E. Wahlberg. pp. 18–27. Karger, Basel.
- ICDRG (1988) Revised European standard series. *Contact Dermatitis* **19**, 391.
- Klecak G., Geleik H. and Frey J. R. (1977) Screening of fragrance materials for allergenicity in the guinea pig. Comparison of four testing methods. *Journal of the Society of Cosmetic Chemists* **28**, 53–64.
- Kligman A. M. (1966) The identification of contact allergens by human assay. 3. The maximization test: a procedure for screening and rating contact sensitizers. *Journal of Investigative Dermatology* **47**, 393–409.
- Larsen W. G. (1985) Perfume Dermatitis. *Journal of the American Academy of Dermatology* **12**, 1–9.
- Larsen W. G., Jackson E. M., Barker M. O., Bednarz R. M., Engasser P. G., O'Donoghue M. N. and Strauss J. S. (1992) A primer on cosmetics. *Journal of the American Academy of Dermatology* **27**, 469–578.
- Larsen W., Nakayama H., Lindberg M., Fischer T., Elsner P., Burrows D., Jordan W., Shaw S., Wilkinson J., Marks J. J., Sugawara M. and Nethercott J. (1996) Fragrance contact dermatitis: A worldwide multicenter investigation (Part I). *American Journal of Contact Dermatitis* **7**, 77–83.
- Magee P. S. (1991) Complex factors in hydrocarbon/water, soil/water and fish/water partitioning. *Science of the Total Environment* **109/110**, 155–178.
- Magee P. S., Hostýnek J. J. and Maibach H. I. (1994a) A classification model for allergic contact dermatitis. *Quantitative Structure–Activity Relationships* **13**, 22–33.
- Magee P. S., Hostýnek J. J. and Maibach H. I. (1994b) Modeling allergic contact dermatitis. In *In Vitro Skin Toxicology: Irritation, Phototoxicity, Sensitization*. Edited by A. Rougier A. M. Goldberg and H. I. Maibach. pp. 281–291. Mary Ann Liebert, New York.
- Magnusson B. and Kligman A. M. (1969) The identification of contact allergens by animal assay. The guinea pig maximization test. *Journal of Investigative Dermatology* **52**, 568.
- Maibach H. I. (1991) Fragrance hypersensitivity, Part II. Frequency of fragrance allergy and photoallergy. *Cosmetics and Toiletries* **106**, 35–36.
- Matsuo S., Nakayama H., Kasahara N., Shimai N., Matsumura M. and Inoue M. (1993) Cases—course from beginning of therapy to remission. In *Dentistry and Metal Allergy*. Edited by M. Inoue and H. Nakayama. pp. 103–126. Dental Diamond Co., Tokyo.
- NACDG (1984) *Patch Testing in Allergic Contact Dermatitis*. (North American Allergic Contact Dermatitis Group). American Academy of Dermatology, Evanston, IL.
- Nielsen N. H. and Menne T. (1992) Allergic contact sensitization in an unselected Danish population. The Glostrup Allergy Study, Denmark. *Acta Dermato-Venereologica* **72**, 456–460.
- Ridings J. E., Barratt M. D., Cary R., Earnshaw C. G., Eggington C. E., Ellis M. K., Judson P. N., Langowski J. J., Marchant C. A., Payne M. P., Watson W. P. and Yih T. D. (1996) Computer prediction of possible toxic action from chemical structure: an update on the DEREK system. *Toxicology* **106**, 267–279.
- Sax N. I. (1984) *Dangerous Properties of Industrial Chemicals*. 6th Ed. Van Nostrand Reinhold Co., New York.
- Sharp D. W. (1978) The sensitization potential of some perfume ingredients tested using a modified Draize procedure. *Toxicology* **9**, 261–271.
- Wilkinson S. M. and Beck M. H. (1994) Allergic contact dermatitis from menthol in peppermint. *Contact Dermatitis* **30**, 42–43.