CHAPTER NINETEEN

Structural Concepts in the Prediction of the Toxicity of Therapeutical Agents

HERBERT S. ROSENKRANZ
Department of Biomedical Sciences
Florida Atlantic University
Boca Raton, Florida

Contents

1 Introduction, 828
  1.1 Development of Database, 828
  1.2 Model Building, 829
  1.3 Model Characterization, 831
  1.4 Model Validation, 834
  1.5 Applications and Mechanistic Studies, 836
2 Conclusions, 844
3 Acknowledgments, 844

Burger's Medicinal Chemistry and Drug Discovery
Sixth Edition, Volume 1: Drug Discovery
Edited by Donald J. Abraham
ISBN 0-471-27090-3 © 2003 John Wiley & Sons, Inc.

1 INTRODUCTION

The increased acceptance and availability of various structure-activity relationship (SAR) approaches in health hazard identification (1, 2) is accompanied by many opportunities and some pitfalls. The latter are derived from the availability of various computer-based SAR platforms whose basis and performance characteristics are not transparent to the user. Such programs, in the hand of the non-expert, may be misused. SAR models and associated technologies on the other hand, while not crystal balls, provide the expert toxicologist with meaningful information regarding the putative toxicological profile of candidate agents. They can guide in the design of agents with decreased or without unwanted side effects and yet retain or even enhance therapeutic effectiveness. Finally, the SAR technology can provide insights into the mechanism whereby a chemical exerts its toxic effects and thereby provide a better understanding of the risk that the agent poses to humans.

However, to achieve these aims, it is essential that the performance characteristics and the basis of the SAR model be known. This involves several critical steps in the SAR model development. These are listed below, each of which will be amplified:

1. Development of database
2. Model building
3. Model characterization
4. 'Model validation
5. Model application to individual agents and for mechanistic evaluations

Although the present review focuses on the MULTICASE SAR methodology (3–5), the concepts discussed herein apply to all generally available SAR techniques used to study toxicological phenomena. Basically SAR approaches that have been used fall into two categories: those that are based on statistical automated algorithms not dependent on prior expert judgment and those that are *a priori* rule-based requiring prior expert knowledge (Table 19.1). However, as will be stressed herein, even the approach not requiring prior expert input is very much dependent on human expertise at various stages of the model development and interpretation process.

Table 19.1 Some SAR Approaches Used in Toxicology

Designation	Approach[a]	References
MULTICASE	I	3–5
TOPKAT	I	6
COMPACT	I	7
DEREK	II	8–10
ONCOLOGIC	II	11, 12
Hazard Expert	II	13
PROGOL	I	14
Structural Alerts	II	15

[a]Approach I indicates statistical automated algorithms not dependent on prior expert judgment. Approach II indicates a rule-based technique that requires prior expert judgment.

Reviews and assessments of the various SAR methodologies used to analyze toxicological phenomena are available (16–21).

1.1 Development of Database

Most experimental data compilations of toxicological effects both in the public domain as well as in proprietary databases were not developed for SAR purposes. Thus, with respect to some toxicological phenomena, the database may be rich with certain chemical classes such as chloroarene and lacking in data relating to others, e.g., aminoarenes. Yet, unlike the SAR models developed for drug discovery, SAR models of toxicological phenomena must be able to handle databases composed of non-congeneric chemicals. Additionally, as a consequence of how toxicological data are generated, there may be a paucity of data altogether. Yet, for optimal SAR models of toxicological phenomena, the "learning set" should include at least 300 non-congeneric chemicals (3, 22).

Accordingly, the human expert may suggest, that for certain purposes, the results of certain assays be pooled, e.g., rat and mouse

carcinogenicity or the results of the Salmonella and E. coli WP uvrB mutagenicity assays (23, 24). Obviously, such data pooling must be based on a sound scientific basis as well as data that show extensive concordance between the experimental results of the systems to be pooled, i.e., that a substantial number of chemicals must give identical results in the two systems, thereby indicating that results obtained with one system can be amalgamated with those obtained in the other (25). For example, when the same chemicals were tested for their ability to induce sister chromated exchanges and chromosomal aberrations in cultured Chinese hamster ovary (CHO) cells, they showed divergent results (26). Hence, the results of the two assays cannot be amalgamated into a single database to develop an SAR model of cytogenetic effects. Similarly, even using the same indicator system, results cannot be merged if different criteria are used to interpret the significance of the results. That situation prevails with respect to the induction of mutations at the thymidine kinase locus of mouse lymphoma cells vis-a-vis the criteria used by the U.S. National Toxicology Program versus those employed by the U.S. Environmental Protection Agency's GeneTox Program. In fact, each data set gives rise to a distinct SAR model (27–29).

On the other hand, the consensus database of potential developmental toxicity in humans, based on experimental results in animals, observations in exposed humans, and expert judgment, yields a coherent SAR model of developmental risks to humans (30). That model is distinct from SAR models of developmental toxicity to individual rodent species (31).

1.2 Model Building

Once a "learning set" (i.e., database) satisfying preset criteria for acceptance (3) has been developed, the model building phase can begin. In general, this is a straightforward process that is specific for the SAR method employed.

Here, I will exemplify the various stages with the MULTICASE SAR system (3–5, 32). Thus. once the structures of the chemicals and an indication of their potency (i.e., either active, marginally active, and inactive, or a continuous scale of potencies) are entered, the program identifies the chemical substructures significantly associated with the toxicological phenomenon under investigation (Table 19.2). Each of these structural determinants ("toxicophore") is associated with a base potency and a probability of activity (see Fig. 19.1). The latter is derived from the distribution of active and inactive molecules that contain the toxicophore. The program also identifies the chemicals that give rise to the toxicophore (Table 19.3 and Fig. 19.7). This enables the human expert (see below) to ascertain whether the structures of the chemicals giving rise to the toxicophores are germane to the test chemical whose toxicity is predicted.

In addition to the toxicophores, the program also identifies modulators for specific toxicophores (Table 19.4). These are substructures or physicochemical parameters that determine whether the specific toxicity inherent in the toxicophore will be expressed or whether it is augmented further.

Thus, when faced with a chemical of unknown activity, the program uses the presence or absence of toxicophores and of modulators to predict its toxicity (Figs. 19.1–19.3). Thus, the presence of the toxicophore OH—C= (a phenol) endows a chemical with an 87.5% probability of being a contact allergen and a potency of 51 (moderate activity, see Table 19.3). That basal activity is modulated by $-25.8 \times$ electronegativity (see Table 19.3). For the example in Fig. 19.1, this results in a further increase in potency. The total potency of 55 units corresponds to a moderately strong activity (Table 19.3). A chemical with that toxicophore may also contain a structural modulator that augments the basal activity further (Fig. 19.2). On the other hand, the chemical may contain a modulator which completely abolishes a chemical's potential to be an allergen (Fig. 19.3). Additionally, the MULTICASE SAR program will identify substructures that are absent from the learning set and therefore may introduce an element of uncertainty in the prediction, i.e., the "unknown" substructure could represent either potential toxicophore or a modulator that alters a rec-

Table 19.2 Major Toxicophores Associated with Allergic Contact Dermatitis in Humans

Fragment							N*		Inactives*	Marginals*	Actives*	Toxicophore No.
1	2	3	4	5	6	7	9	10				
N	—CH_2—						50		1	0	49	1
N	—CH_3—						26		1	0	25	2
Cl	—CH_2—						8		0	0	8	3
OH	—c=						86		10	0	76	4
SH	—CH_2—						19		0	1	18	5
NH	—CH_2—						26		1	1	24	6
NH_2	—CH_2—						7		0	0	7	7
cH	=cH	—c	=cH—			⟨3—NH_2⟩	18		0	0	18	8
CH″	—CO	—CH=					15		0	0	15	9
O	—CO	—C	=CH_2—				21		1	5	15	10
NH	—c	=cH—					17		0	1	16	11
CO	—CH	=CH_2—					9		0	0	9	12
N	=C—						24		1	0	23	13
CO	—N—						14		0	0	14	14
OH	—CO	—c=					8		1	0	7	15
cH	=c	—c	=c	—cH=			25		1	0	24	16
cH	=cH	—cH	=cH	—cH	=c⟩—		17		4	0	13	17
S	—c.=						10		0	0	10	18
CO	—CH_2	—CH_2	—c=				5		5	0	0	19

The database and derivation of the SAR model have been described (33).

*N indicates the number of chemicals in the database that contain that toxicophore. "Inactives," "marginals," and "actives" indicate the distribution of that toxicophore among activity groups.

Toxicophore No. 4 is shown embedded in chemicals in Figs. 19.1–19.3 and No. 5 is shown in Fig. 19.5.

C indicates a carbon atom shared by two rings; ⟨3—NH_2⟩ indicates an amino group attached to the third non-hydrogen atom from the left. In toxicophore No. 17, the last carbon to the right is shown as unsubstituted. This means that it can be substituted with any atom except a hydrogen. On the other hand, in toxicophore No. 8, the penultimate carbon is shown unsubstituted; it can only be substituted by an amino group (i.e., ⟨3—NH_2⟩). However, the last carbon of that toxicophore is shown with an attached hydrogen. It cannot be substituted by any other atom.

Table 19.3 Derivation of Toxicophore: The 19 Molecules Containing Fragment SH—CH_2

Chemicals	Potency"
2,3-Dimercapto-1-propanol	55
2-Mercaptoethanesulfonicacid	55
2-Mercaptoethyl methyl sulfone	45
2-Mercaptoethyl urea	35
2-Methoxyethyl mercaptoacetate	35
N-(1,1-dimethylolethyl) mercaptoacetamide	35
N,N-dimethyl mercaptoacetarnide	35
N-(2-mercaptoethyl) acetamide	25
N-(2-mercaptoethyl) pyrolidone	55
N-(mercaptoacetyl) urea	35
N-(mercaptoacetyl) glycine	35
N-(mercaptoethyl) morpholine	35
N-methyl mercaptoacetamide	35
N-trimethylolmethyl mercaptoacetamide	45
Cysteine	45
Mercaptoacetamide	55
Mercaptoacethydrazide	45
Mercaptoacetic acid	45
Thioglycerol	55

The program identifies the chemicals that are responsible for toxicophore No. 5 of Table 19.1 (see also Fig. 19.7). The toxicophore is shown embedded in **a** molecule in Fig. 19.5.

"The allergenic potencies were defined based on the percent responders in the human maximization test as follows (33): 10, Non-sensitizer; 25, "marginal" (4–7% responders); 39, "weak" (8–23% responders); 49, "moderate": (24–55% responders); 59, "strong" (56–83% responders); 69, "extreme" (84–100% responders).

ognized toxicophore or a noninformative structure unrelated to toxicity (Fig. 19.4).

It should be stressed that not every experimental data set gives rise to a coherent SAR model. Failure to construct a model may be caused by the fact that the experimental data are invalid or that they do not reflect a specific toxicological phenomenon. Additionally, the phenomenon under investigation may be so complex or be the result of so many different mechanisms that the experimental database is not sufficiently large to describe it. With this in mind, it should be stressed that the predictivity of the SAR model will be a reflection of the complexity of the phenomenon, the size of the database (i.e., the number of chemicals for which experimental data are available), and the ratio of actives/inactives in the data set (3, 22).

In view of the above considerations, once an SAR model has been developed, it requires extensive validation and characterization.

1.3 Model Characterization

As mentioned above, the nature of the SAR model that is derived is a reflection of the complexity of the toxicological phenomenon that it describes, as well as of the size of the learning and the extent to which it includes chemical classes and/or substructures that are representative of the chemical species to which it will be applied. Thus, the chemical substructures present among therapeutics are much greater and diverse than, for example, those used or generated in the chemical or agricultural industries. This means that SAR models used to examine pharmacologically active substance must contain a greater variety of chemical substructures. This may well translate into a requirement for a larger experimental data set (i.e., one containing an increased number of chemicals).

In evaluating the SAR model, it is of importance to determine the relationship between its predictivity and the size of the database to determine whether the model is oytimal. This can be ascertained by first determining the model's predictivity (see below), and then systematically decreasing the size of the database by random deletion of chemicals to determine the predictive parameters of the model derived from the reduced data set. Doing this iteratively will allow a determination of the relationship between database size and concordance between predicted and experimentally derived results (22). If the relationship, including the value for the SAR model derived from the total database is linear, then the model will not be optimally predictive and consideration should be given to obtaining additional experimental data and deriving a further model. On the other hand, if the relationship including the data for the SAR model derived from the total database is no longer linear, the size of the data set may be satisfactory. Incremental data may not yield a correspondingly significant increase in the model's performance. Thus, the predictivity of the SAR model of mutagenicity in ***Salmonella*** improves linearity until a database size of 350 chemicals is reached, and then it plateaus (22).

Table 19.4 List of MODULATORS Related to Toxicophore OH —c ═

Fragment									Constant = 51.0 QSAR	Toxicophore No.
1	2	3	4	5	6	7	8			
2D	[N—]⟨——6.5A——⟩[NH—]								16.7	1
CO	—CH_2	—CH_2—							−53.3	2
cH	═c	)—c	═cH—					⟨3—CO⟩	−38.8	3
OH	—c	═c	—CO	—c═					25.2	4
cH	═c	—cH	═cH	—c═				⟨2—OH⟩	8.3	5
CH_2	—CH	—c	═cH	—cH	═cH—			⟨3—cH═⟩	13.4	6
cH	═cH	—cH	═c	—cH	═cH—			⟨4—OH⟩	−10.3	7
cH	═cH	—cH	═cH	—c	═cH—			⟨5—OH⟩	−10.3	8
OH	—c	═c	—O	—CO	—CH_2	—CH—			−20.3	9
OH	—c	═cH	—cH	═cH	—cH	═cH—			−20.6	10
OH	—c	═c	—cH	═c	—CH_2	—CH_2	—CH_3		−22.4	11
(HOMO + LUMO)/2									−25.8	12

Modulators associated with toxicophore No. 4 of Table 19.2. Each of the modulators augments or decreases the activity inherent in the toxicophore (i.e., 51.0 units; see Figs. 19.1–19.3). (HOMO + LUMO)/2 describes the electronegativity of the molecule. That value is multiplied by −25.8. Modulator No. 5 and No. 2 are shown embedded in chemicals in Figs. 19.2 and 19.3, respectively. Modulator No. 1 describes a 2D distance descriptor of 6.5 Å between two atoms. For interpretation of the structures see legend to Table 19.2.

The molecule contains the Toxicophore (nr.occ.= 2):

OH -c"

*** 76 out of the known 86 molecules (88%) containing such a Toxicophore are Contact Allergen8 with an average activity of 49. (conf.level=100%)

*** QSAR Contribution : Constant is 51.04

** The following Modulator is also preeent:

Electronegativity = -0.15 ; Its contribution is 3.83

** Total projected QSAR activity 54.87

*** The probability that this molecule is a Contact Allergen is 87.5% **

** The projected Allergic Potency is 54.9 CASE units **

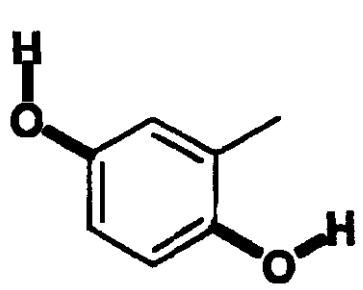


Figure 19.1. Prediction of the contact allergenicity of 2-methyl-1,4-benzenediol. The prediction is based on the presence of the toxicophore (shown in bold). The potency is modulated further by the electronegativity (see Table 19.4). A potency of 55 units indicates a moderately strong allergen (see Table 19.3).

Another concern relates to the effect of the ratio of active to inactive chemicals in the data set. Some SAR models are most predictive when that ratio is unity (3, 22). Hence, for a model that will be widely used for hazard identification and risk assessment purposes, it would be of importance to determine whether its performance is optimal. Thus, if the number of inactives exceeds the number of actives, the number of inactives can be decreased by randomly removing the appropriate number of inactives and determining the performance of the resulting SAR model. The random deletion of inactives and the model derivation should be repeated several times to ascertain that a robust model has been derived. We found that because the nature of the toxicophores is determined primarily by the actives and because the "quality" of the toxicophores is a function of the size of the database (22, 34, 35), it follows that if the number of actives exceeds the number of inactives that removal of actives to achieve a ratio of unity is not the optimal solution. Rather, we have found that supplementing the database with randomly selected chemicals from a "pool" of normal physiological chemicals (amino acids, sugar, lipids, purines, pyrimidines, etc., but excluding hormones, prostaglandins, and vitamins), assuming these chemicals to be inactive, is a viable alternative (36, 37). This is based on the recognition that the biological and/or toxicological phenomena being modeled occur in a milieu that is rich in these physiological chemicals.

Finally, the "informational content" of an SAR model determines its coverage. Thus, if a test molecule contains a substructure un-

```
The molecule contains the Toxicophore  (nr.occ.= 2):

                OH  -c"

     +** 76 out of the known  86 molecules ( 88%) containing such a
Toxicophore are contact allergiee with an average activity of  49.
(conf.level=100%)

     *** QSAR Contribution :                    Constant is    51.04

     ** The following Modulators are also present:

     ( 1) cH =c  -cH =cH -c  =            c2-OH >   Activating   8.33

         Electronegativity   = -0.17 ;     Its contribution is    4.29
                                                                 ------
      ** Total projected QSAR activity                           63.65

     *** The probability that this molecule is a Contact Allergen is
87.5%  **

      ** The projected Allergic Potency is 63.7 CASE units **
```

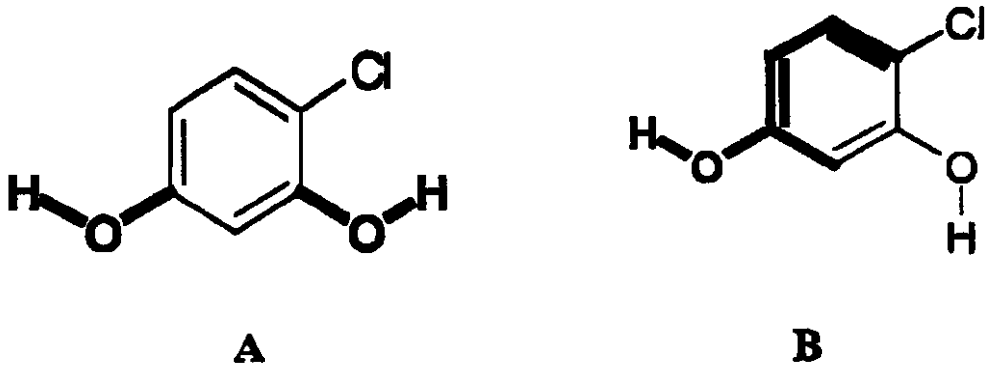


Figure 19.2. Prediction of the contact allergenicity of 4-chloro-1,3-benzenediol. In addition to the probability of activity and the basal potency derived from the toxicophore (shown in bold in **A),** the chemical also contains an activating modulator (shown in bold in B), which further augments the potency. **A** potency of **64** units indicates a very strong potency (Table **19.3).**

known to the model, this introduces a measure of uncertainty into the SAR prediction. In the MULTICASE SAR program, such an "unknown" moiety is flagged (Fig. 19.4). We have found that a satisfactory approach to determining informational content is to challenge an SAR model with a panel of 10,000 chemicals representative of the "universe of chemicals" and determining the frequency with which the SAR predictions are accompanied by a "warning" of the presence of "unknown" substructures. An enumeration of the frequency with which the individual unknown moieties are present will allow a determination of their importance and thereby identifies chemicals that should be tested and the results included in the model to improve the predictive performance. This is based on the observation that the greater the informational content (i.e., the fewer warnings of "unknown" moieties), the greater the model's predictivity (22, 34, 35).

1.4 Model Validation

In its application to toxicology, SAR can serve two functions: (1)to predict a specific toxico-

```
 The molecule contains the Toxicophore   (nr.occ.= 1):

                OH  -c"

      *** 76 out of the known  86 molecules ( 88%) containing such a
Toxicophore are  Contact Allergen8 with an average activity of 49.
(conf.level=100%)

      *** QSAR Contribution :                        Constant is    51.04

       ** The following Modulator8 are also present:

     ( 1) CO -CH2-CH2-                              Inactivating -53.33

        Electronegativity   = -0.10 ;     Its contribution is     2.51
                                                                  ------
       ** Total projected QSAR activity                            0.22

       ** The molecule contains the following DEACTIVATING Fragment:

              CO  -CH2
                    \
                     CH2 -c"

      *** The probability that this molecule is a Contact Allergen is 63.6%  **
```

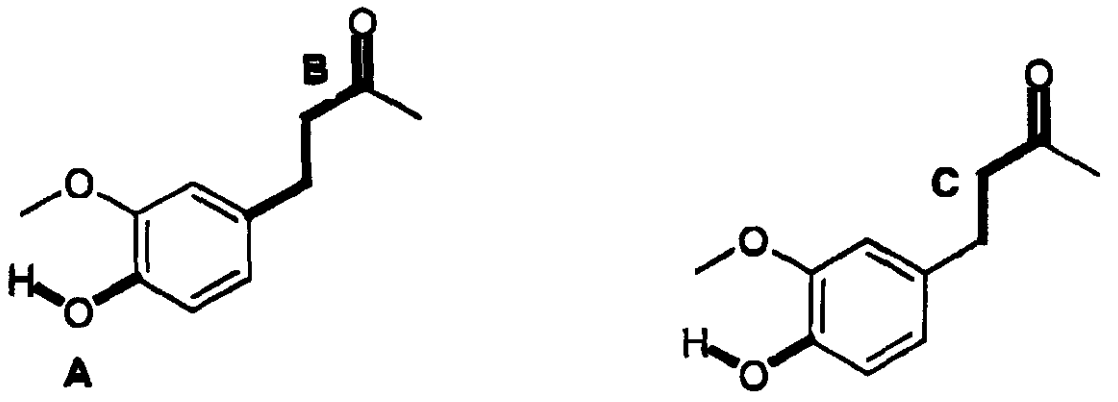


Figure 19.3. Prediction of the lack of contact allergenicity of zingerone. Whereas the presence of the toxicophore (A) is associated with a probability of activity and a potency, the presence of the inactivating modulator (B) abolishes the potency. Moreover, the presence of a deactivating moiety (C), which is present in five chemicals in the database that are devoid of allergenicity (Table 19.2, No. 19), further decreases the likelihood that the zigerone is a contact allergen.

logical effect based on the identification of substructures significantly associated with that activity and (2) to gain insight into the mechanistic basis of that effect.

To be useful in its predictive mode, the performance of a model does not need to be perfect, but it must be known. The predictivity of an SAR model is defined by the concordance between the predictions of chemicals external to the SAR model and the experimentally determined toxicities. The predictivity is governed by the sensitivity (number of correct positive predictions/total number of positive chemicals) and the specificity (number of correct negative predictionsltotal number of negative chemicals) (22). Moreover, because the basic function of SAR applied to toxicological phenomena is the prevention, reduction, or elimination of harmful chemicals from the home, the environment, and the workplace, risk averse prediction models are preferred. That is achieved by the development of SAR models that yield a low frequency of false negative predictions, i.e., high specificity. Obviously, ideally the model should have high sensitivity as well as high specificity **(38).**

```
0*** WARNING *** The following functionalities are UNKNOWN to me:
      *** O  -C. =C. -
      *** CO -O  -C. -
       ** The molecule does not contain any known Biophore   **
            it is therefore presumed to be INACTIVE
```

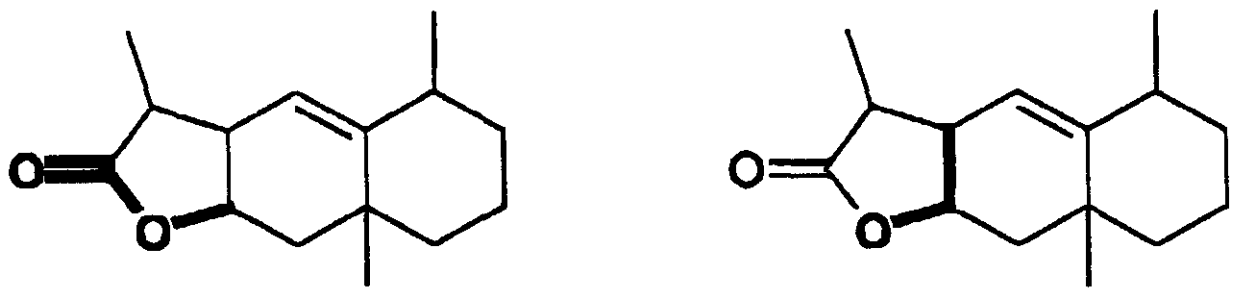


Figure 19.4. Prediction of the lack of contact allergenicity of of dehydroalantolactone. The chemical contains no toxicophore; therefore, it is presumed to be inactive. However, it contains two structures (shown in bold) that are "unknown" to the model. That introduces an element of uncertainity in the prediction.

The simplest way to determine predictivity parameters is to remove initially from the data set a random representative sample (e.g., 5%) to be used as a "tester set," to develop the SAR model on the remaining chemicals (i.e., 95%), and then challenge the model with the "tester set" and ascertain the predictivity. However, as has been demonstrated on a number of occasions, the predictivity of an SAR model is determined by the size of the database (22), and as in most instances, the size of the available data set is not optimal, therefore, further decreasing the size of the learning set by sequestering the "tester set" is not optimal.

To overcome this limitation, a cross-validation approach has been used (39). In that procedure, a portion of the database (e.g., 5%) is randomly selected and removed, and a model is developed from the remaining 95%. That model is challenged with the "tester set" (5%). That procedure is repeated 20 times, and the cumulative predictivity is determined. The final SAR model includes the complete database (i.e., 100%). Because the predictive performance is a function of the size of the database, the performance of the final model will be better than that based on 95% of the data. When, however, the learning set consists of less than 150 chemicals, a more tedious procedure may be required, wherein one to two chemicals (i.e., n-1 or n-2) are removed at a time to serve as the "tester set" and the process is repeated n or n/2 times.

1.5 Applications and Mechanistic Studies

As has been mentioned earlier (Table 19.1), SAR methodologies can be divided into two general non-mutually exclusive approaches: (1) hypothesis driven and (2) knowledge based. The former is rule driven, wherein specific properties or chemical substructures are looked for, e.g., mutagens are electrophiles and hence one would look for electrophilic or proelectrophilic moieties. This approach assumes that mutations are caused solely by covalent binding of electrophiles to DNA. Agents that induce mutations by a nonelectrophilic (i.e., non-DNA damaging) mechanism will not be detected. Thus, agents that mutagenize purely as a result of intercalation between DNA base pairs (e.g., acridine orange, ethidium bromide) will not be identified. Such rules are based on prior knowledge and/or in-

```
 The molecule contains the Biophore
                SH  -CH2
  *** 18 out of the known 19 molecule8 ( 95%) containing such a
toxicophore are Contact Allergens with an average activity of 42.
(conf.level=100%)

     *** QSAR Contribution :                    Constant is    52.50
     ** The following Modulators are also present:
     (2D) [CO -] e-- 5.2A --> [SH -]            Inactivating   -7.6
         Electronegativity   =  0.10 ;     Ite contribution is   -0.67
                                                               -----
      ** Total projected QSAR activity                         44.23
     *** The probability that this molecule is a Contact
           Allergen is 95.0%  **
      ** The projected Allergic Cont activity is 44.3 CASE units **
```

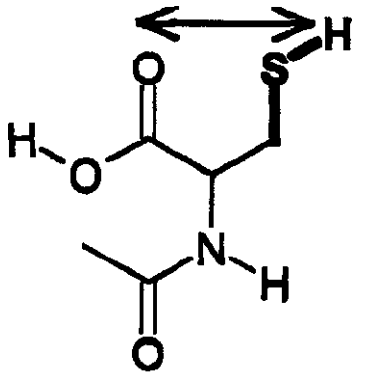


Figure 19.5. Prediction of the contact allergenicity of *N*-acetyl-L-cysteine. The prediction is based on the presence of the toxicophore (shown in bold), which is present in 19 chemicals in the database (18 allergens and 1 marginal allergen; see Table 19.3). The arrow indicates the 5.2 Å distance described by the inactivating modulator.

tuition and do not necessarily require adherence to strict statistical criteria.

The approach illustrated herein, exemplified by MULTICASE (3), is knowledge based. The input consists of the structures and toxicological activities of the chemicals in the learning set. The program then identifies structural descriptors (toxicophores) that are significantly associated with activity (see Table 19.3). The human expert participates in setting criteria for the inclusion of experimental results in the database (3) as well as in examining the plausibility of the final model based on exact knowledge of the toxicological phenomenon under investigation. The human expert again also determines the acceptability of individual predictions (see below).

Once an SAR model has been developed and validated, it can be applied in a number of fashions. SAR methodologies, such as MULTICASE **(3–5),** which document predictions (Table 19.2), are obviously preferable to those

```
 The molecule contains the Toxicophore

          S   -C
                \\
                  N
                /
             C"

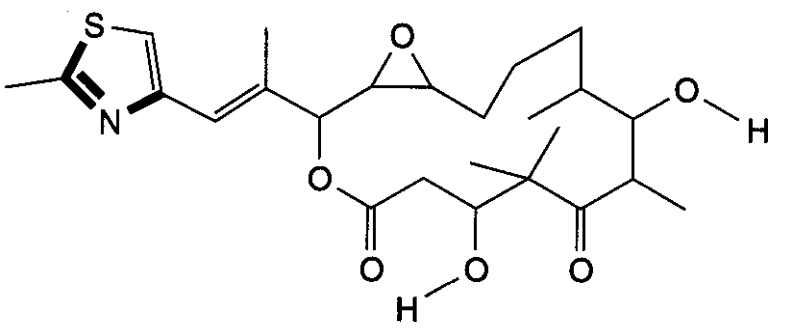
*** 5 out of the known 5 molecules (100%) containing such Biophore
    are Mouse carcinogens with an average activity of 62.
    (conf.level=97%)
 ** This Biophore exists in a significantly different environment
    than in the data baee (i.e. 5.45);  It may not be relevant
*** QSAR Contribution :                    Constant is    64.00
 ** Total projected QSAR activity                         64.00
*** The probability that this molecule is a Mouse carcinogen is 05.7%
 ** The projected Mouse carcinogenic potency is 64.0 CASE units **
```

Figure 19.6. Prediction of the carcinogenicity in mice of epitholone A. The structure of epitholone A (toxicophore shown in bold) is given in Fig. 19.7.

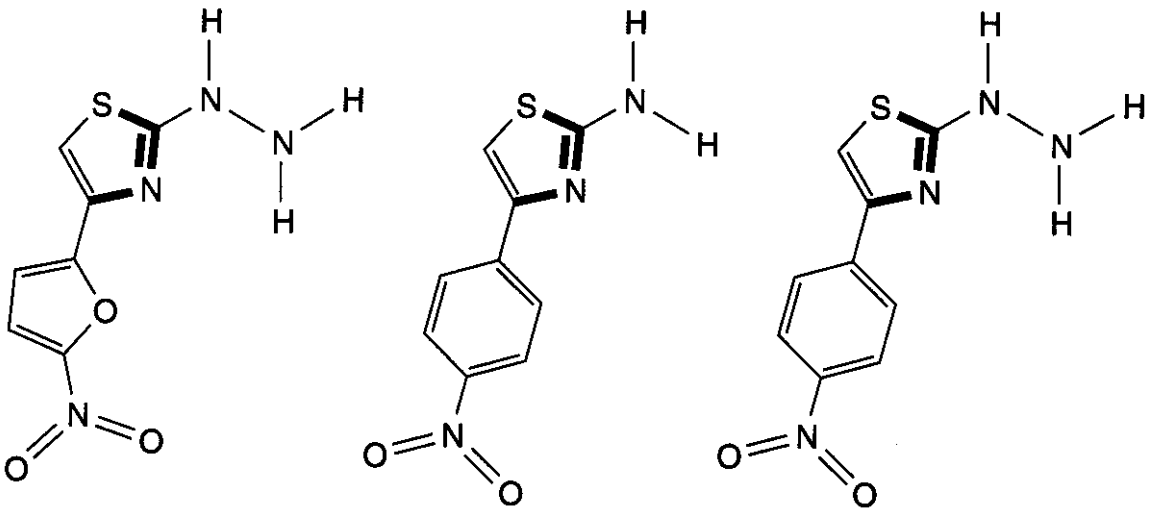
that operate like a "black box." The latter simply provides a likelihood that a test chemical is active or inactive. When, however, the SAR prediction is accompanied by documentation of the basis of that forecast, the human expert can determine whether it is justified and whether it applies to the specific test chemical.

Thus, the mucolytic agent *N*-acetyl-L-cysteine is predicted to have a potential to induce allergic contact dermatitis by virtue of the biophore $SH—CH_2$ (Fig. 19.5). Moreover, examination of the chemicals that contribute to that toxicophore reveals that indeed they all have the substructure in an environment that is similar to the one found in N-acetyl-L-cysteine (Table 19.3). On the other hand, the tubulin polymerization perturber (and potential antineoplastic agent) epitholone A (Fig. 19.6) is predicted to be a mouse carcinogen by virtue of the toxicophore units shown in bold. That toxicophore is present in five molecules in the learning set. The presence of that toxicophore is associated with an 89% probability of carcinogenicity and a potency of 63 units, which corresponds to a TD_{50} value of 0.039 mmol/kg per day (40). However, the program flags the toxicophore because its environment in epitholone A is significantly different from that of the molecules in the learning set (Fig. 19.6). In fact, examination of the structures of the molecules that contribute to the biophore (Fig. 19.7) indicates that indeed the molecules are quite different from that of epitholone A, and hence, the prediction of carcinogenicity can be disregarded (however, also see below).

Moreover the molecules that contributed to this toxicophore (Fig. 19.7), even though they contain the S—C=N—C= moiety (Fig. 19.6), also contain functionalities (i.e., "structural alerts") that are associated with carcinogenicity/genotoxicity such as nitro, amino, and hydrazino groups. In fact, these could be responsible for the murine carcinogenicity of these chemicals. Obviously, these latter functionalities are absent in epitholone A.



Epitholone A



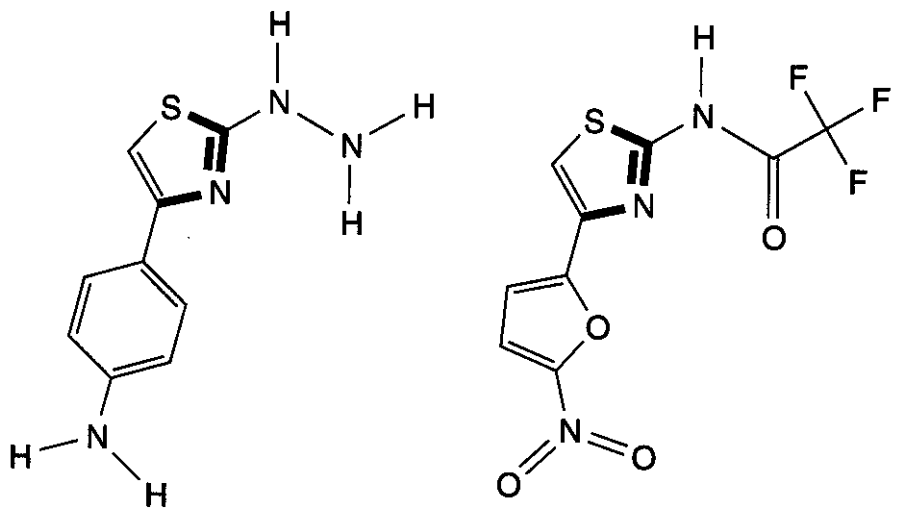


Figure 19.7. Structures of epitholone A and of chemicals that contain the toxicophore. The toxicophore (Fig. 19.6) is shown in bold.

Table 19.5 SAR Predictions Related to the Potential Carcinogenicity of Epitholone A

SAR Model	Prediction	References
Mutagenicity: Salmonella	Negative	22, 47
Error-prone DNA repair	Negative	48
Unscheduled DNA synthesis	Negative	49
Mouse MTD	Positive	50
Rat $LD_{\text{„}}$	Positive	SAR model based on RTECS
Cell toxicity	Positive	51
Inhibition GJIC	Negative	52

A positive response indicates a potential for maximum tolerated dose of less than 0.9 mmol/kg; an LD_{50} value of less than 7.2 mmol/kg or a toxicity (IC_{50}) for cultured BALB/3T3 cells of less than 1 μM.

GJIC, gap junctional intercellular communication; RTECS, Registry of Toxic Effects of Chemical Substances.

Table 19.6 Predicted Toxicological Profile of N-Acetylcysteine

	Multicase	
SAR Model	Probability (%)	Potency (units)
Structure alerts	0	0
Salmonella mutagenicity	0	0
SOS chromotest	0	0
umu/SOS repair	0	0
Carcinogenicity: rodents-NTP	0	0
Carcinogenicity: mice-NTP	0	0
Carcinogenicity: rats-NTP	0	0
Carcinogenicity: rodent-CPDB	0	0
Carcinogenicity: mice-CPDB	0	0
Carcinogenicity: rats-CPDB	0	0
Inhibition gap junction intercell comm	0	0
Binding to Ah receptor	0	0
Mutations in mouse lymphoma (NTP)	0	0
Mutations in mouse lymphoma (GenTox)	0	0
Sister chromatic exchanges in *vitro*	0	0
Chromosomal aberrations in vitro	0	0
Unscheduled DNA synthesis in vitro	0	0
Cell transformation	0	0
Drosophila somatic mutations	0	0
Sister chromatic exchanges in vivo	0	0
Induction of micronuclei in vivo	0	0
Yeast malsegregation	0	0
Inhibition of tubulin polymerization	0	0
Sensory irritation	89	72
Eye irritation	72	52
Respiratory hypersensitivity	0	0
Allergic contact dermatitis	95	44
Rat lethality (LD50)	0	0
Mouse MTD	0	0
Rat MTD	0	0
Cellular toxicity (3T3)	0	0
Cellular toxicity (HeLa)	0	0
Nephrotoxicity: male rats ($\alpha 2\mu$globulin)	0	0
Inhibition human cyt. P4502D	0	0
Developmental toxicity: hamster	0	0
Developmental toxicity: human	0	0
Aquatic toxicity (minnows)	0	0
Water solubility: 3.88	log P (Octanol: water): −1.79	
Electronegativity: 0.10		

NTP and CPDB refer to the U.S. National Toxicology Program carcinogenicity assays (45) and to the Carcinogenic Potency Data Bases (46), respectively.

Based on all of these considerations, the "human expert" would overrule the prediction of rodent carcinogenicity. Additionally, in overriding the computer-based prediction, cognisance was also taken of the understanding that the vast majority of recognized human carcinogens are genotoxicants, i.e., "genotoxic carcinogens" (41–44). Epitholone A, on the other hand, was not predicted to be genotoxic (i.e., a DNA-damaging agent), evidenced by its lack of potential to induce mutations in Salmonella, error-prone DNA repair, or unscheduled DNA synthesis in rat hepatocytes (Table 19.5). Thus, even if the potential for murine carcinogenicity were accepted, in view of the fact that the vast majority of rec-

```
The molecule contains the Biophore   (nr.occ.= 1):

                NH  -CH
                     \
                       C."

*** 38 out of the known 41 molecules (93%) containing such a Biophore
        are perturbers of Tubulin Polymerization
*** QSAR Contribution :                          Constant is     85.89
                                                                 ------
 ** Total projected QSAR activity                                85.89
 ** The probability that this molecule inhibits Tubulin
     Polymerization is 93% **
 ** The projected Tubulin Polymerization Inhibitory activity is 85.9
     CASE units **
```

Figure 19.8. Prediction of the ability of colchicine to inhibit tubulin polymerization. The structure of colchicine is shown in Fig. 19.9. The biophore is shown in bold (a) in Fig. 19.9.

ognized human carcinogens are mutagens/genotoxicants or are hormones and epitholone A is neither, it would not represent a human risk.

If, based on the above, it were accepted that epitholone A is not genotoxic, and if the human expert examining the documentation wished not to override the prediction of carcinogenicity in mice based on the differences in chemical environments between epitholone A and the molecules responsible for the toxicophore (Figs. 19.6 and 19.7), he could examine mechanisms of non-genotoxic carcinogenicity, even though its relevance to human may not be applicable. One of the mechanisms of non-genotoxic carcinogenicity is inhibition of intercellular communication (53). Epitholone A does not possess such a potential (Table 19.5). Another mechanism for non-genotoxic rodent carcinogenesis may involve systemic or cell toxicity followed by mitogenesis (54–56). This may occur as a consequence of including the maximum tolerated dose (MTD) in the cancer bioassay protocol. When this is done, up to 50% of chemicals tested are found to be rodent carcinogens (54). Obviously, this MTD situation rarely, if ever, applies to humans. Still, epitholone A has the potential for inducing cellular as well as systemic toxicity (Table 19.5), which may explain its potential carcinogenicity in mice, were we to discount the difference in chemical environment.

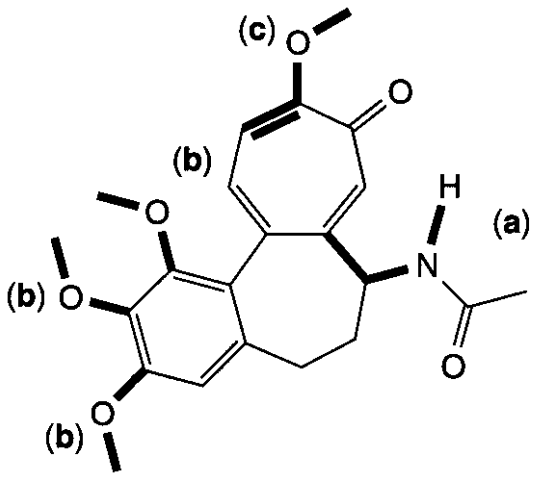


Figure 19.9. Structure of colchicine. The biophore A (bold, see Fig. 19.8) is responsible for the therapeutic effectiveness. Toxicophore B (see Fig. 19.10; shown in bold) is responsible for the induction of sister chromatid exchanges (SCE) in *vivo*. Removal of toxicophore B or its replacement be isopropoxy groups abolishes the induction of SCEs without affecting the therapeutic potential.

Obviously, the availability of a number of characterized and validated SAR models allows the development of a putative toxicologi-

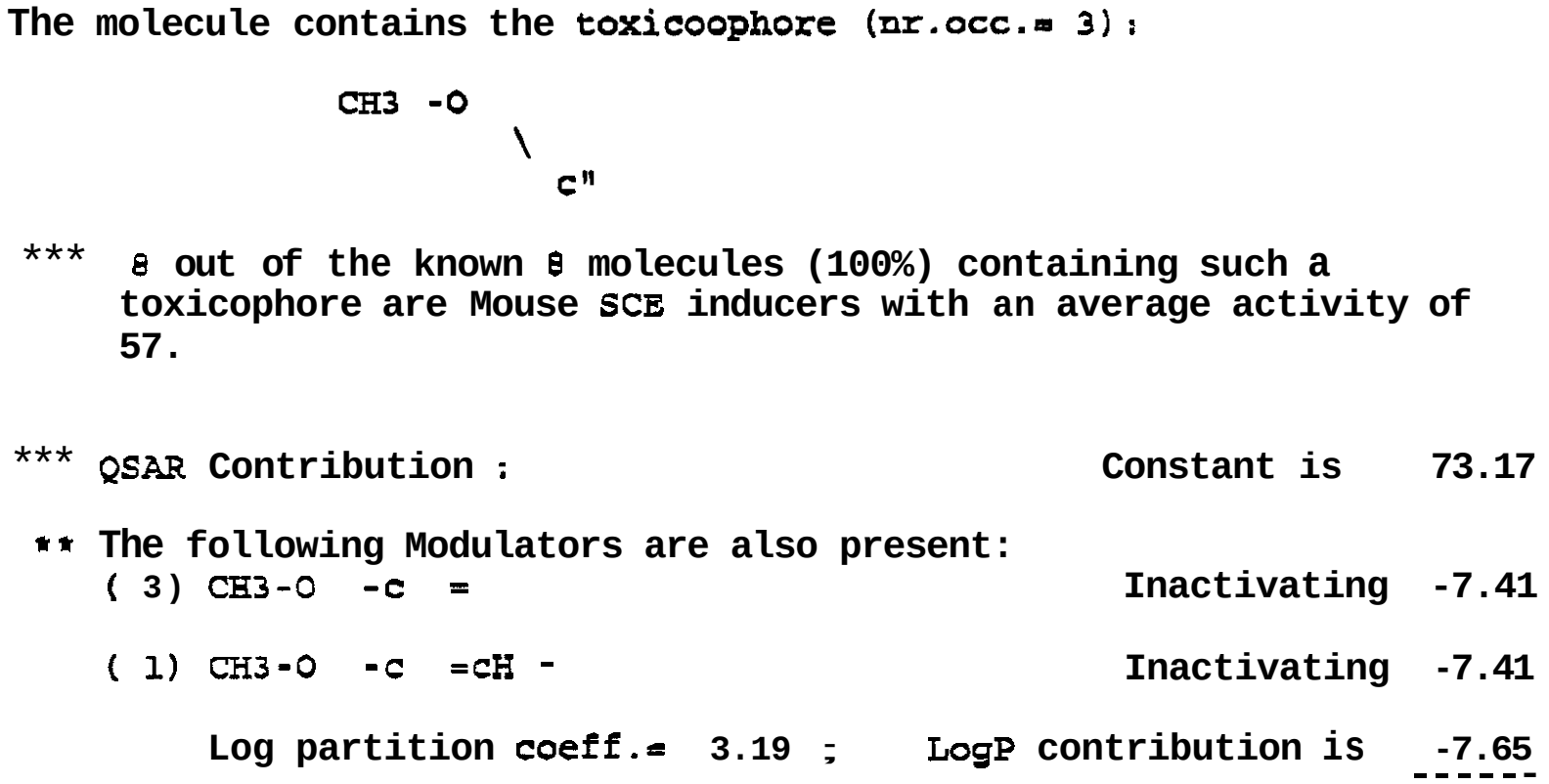


Figure **19.10.** The potential of colchicine to induce sister chromatic exchanges in ***vivo.*** The structure of colchicine and of the toxicophore B is given in Fig. 19.9. One of the inactivating modulators (c) is also shown in bold in Fig. 19.9.

cal profile (Table 19.6). This can be used as a guideline in the product developmental phase to select lead compounds least likely to induce unwanted side effects. However, the SAR approach can also be used to optimize beneficial effects and decrease or eliminate unwanted toxic effects.

Thus, let us examine colchicine (CH), an anti-inflammatory agent that has been in use for several centuries for the treatment of gout. The anti-inflammatory potential of CH is understood to derive from its ability to inhibit tubulin polymerization (iTP) (57). That is also the basis of the anticancer activity of paclitaxel (Taxol) (58–60). The structural basis of that activity derives from the presence in CH of the NH—CH—C.= moiety (Figs. 19.8 and 19.9), which endows the molecule with a 93% probability of activity. However, colchicine also has the potential for inducing sister chromatid exchanges (SCEs) *in vivo* (Fig. 19.10). This SCE-inducing ability may endow it with genotoxic and developmental toxicity potentials. However, the potential for inducing SCEs *in vivo* is associated with the methoxy moiety (Figs. 19.9 and 19.10). Removal of that moiety or replacing it with an isopropoxy group abolishes the SCE-inducing ability of CH without affecting its potential for iTP (i.e., the basis of its anti-inflammatory action).

Finally, SAR approaches can also be used to provide a basis for making intelligent risk assessments. Thus, it has been shown that the similarity in biophores/toxicophores present in different SAR models of toxicological phenomena provides a measure of mechanistic similarity (3). The SAR models of mutagenicity in *Salmonella* and of error-prone DNA repair (SOS Chromotest) show significant overlaps (Table 19.7). This is not unexpected because DNA is the target of both phenomena, and the tester strain used for the *Salmonella* mutagenicity assays contains a plasmid that codes for error-prone DNA repair (61). In fact there is a substantial (though not complete) overlap among chemicals that cause the two

Table 19.7 Structural Commonalities among SAR Models

SAR Models	Percent
Salmonella mutagenicity and SOS chromotest	57
Salmonella mutagenicity and iGJIC	10
Salmonella mutagenicity and iTP	9
Salmonella mutagenicity and Mnt	53
Mnt and iTP	71

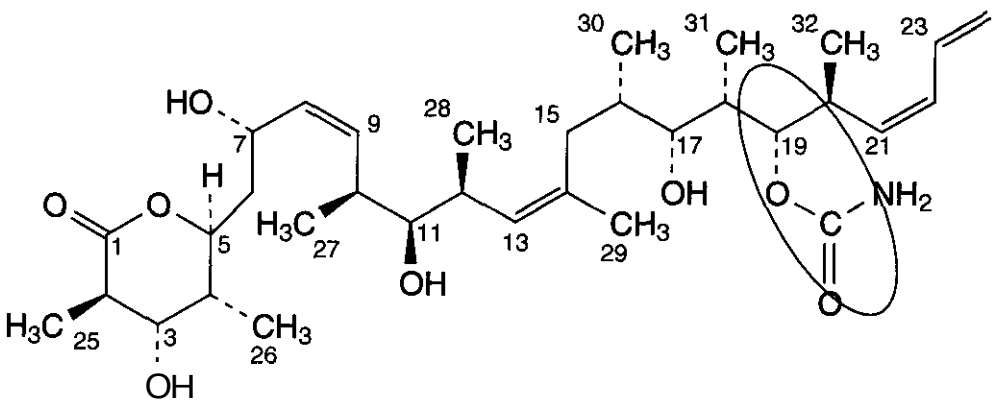
iGJIC, inhibition of gap functional intercellular communications; iTP, inhibition of tubulin polymerization; Mnt, induction of bone marrow micronuclei *in vivo*.

phenomena (48, 62). On the other hand, there is little overlap between *Salmonella* mutagenicity and inhibition of gap junctional intercellular communication (Table 19.7), which is considered the epigenetic (non-genotoxic) phenomenon *par excellence* (53). Nor do the SAR models for *Salmonella* mutagenicity and inhibition of tubulin polymerization overlap significantly (Table 19.7), which is further support for the fact that genotoxicity and inhibition of tubulin polymerization can be dissociated (see above).

With respect to the *in vivo* induction of micronuclei (Mnt), a different situation prevails. There is considerable overlap between the toxicophores associated with Mnt and those with the induction of mutation is *Salmonella* (Table 19.7). This is not surprising, because the induction of Mnt is known to involve a genotoxic mechanism (63, 64). Indeed, when attempting to identify potential genotoxic carcinogens, when a chemical is found to induce mutations in *Salmonella,* this result is frequently followed by a Mnt test to determine whether the chemical is genotoxic *in vivo* as well (43, 65) and thus represent a risk to humans.

However, it was found that there is also substantial overlap between Mnt and iTP, the latter being a non-genotoxic phenomenon (Table 19.7) (66). This finding suggests that Mnt can occur by genotoxic as well as non-genotoxic mechanisms. Thus, a positive Mnt response by a chemical that does not induce mutations in *Salmonella* does not necessarily represent a carcinogenic risk to humans.

Discodermolide (Fig. 19.11) is a promising antineoplastic agent, which like paclitaxel, inhibits tubulin polymerization (67), but being considerably more water-soluble than paclitaxel, discodermolide may present certain therapeutic advantages while also being effective against paclitaxel-resistant cells (67). Neither discodermolide nor paclitaxel are mutagenic in *Salmonella* (and in fact neither is predicted to be a rodent carcinogen). However, both of these agents have a potential (determined by SAR) to induce Mnt *in vivo*. In fact, for paclitaxel that potential has been determined experimentally. This has led to the suggestion that paclitaxel, because of its ability to induce Mnt, presented a carcinogenic risk (68). However, based on the above findings (Table 19.7), it can be assumed that the ability of discodermolide and of paclitaxel to induce Mnt is independent of genotoxicity, and in fact, derives from iTP. Thus, it does not represent an unreasonable risk to humans who are treated with those antineoplastic agents. In fact, the biophores in discodermolide responsible for the induction of Mnt and iTP are identical (Fig. 19.11).



Discodermolide

Figure 19.11. Structure of discodermolide. The circled biophore is responsible for the inhibition of tubulin polymerization.

2 CONCLUSIONS

SAR methodologies, in their present state, coupled with human expertise, can be used to determine and to understand the potential toxicity of therapeutic agents. In fact, this approach can be used to engineer molecules devoid of the moieties associated with these unwanted side effects. It must be understood, however, that while SAR techniques can be used to accelerate the identification and development of safe therapeutic agents, it is to be used as an adjunct to experimental determinations.

3 ACKNOWLEDGMENTS

The support of the Vira Heinz Endowment is gratefully acknowledged.

REFERENCES

1. National Research Council, Science and Judgment in Risk Assessment, National Academy Press, Washington, DC, **1994.**
2. I. D. McKinney, A. Richard, C. Waller, M. C. Newman, and F. Gerberich, Toxicol. Sci., **56, 8–17 (2000).**
3. H. S. Rosenkranz, A. R. Cunningham, Y. P. Zhang, H. G. Claycamp, O. T. Macina, N. B. Sussman, S. G. Grant, and G. Klopman, SAR QSAR Environ. Res., **10,277–298** (1999).
4. G. Klopman and H. S. Rosenkranz, *Mutat.* Res., **305**, 33–46 **(1994).**
5. G. Klopman and H. S. Rosenkranz, Toxicol. *Lett.*, **79,145–155 (1995).**
6. **K.** Einslein, **V.** K. Gombar, and B. W. Blake, *Mutat.* Res., **305**, 47–61 (1994).
7. D. F. **V.** Lewis, C. Ioannides, and D. V. Parke, Environ. Health Perspect., **104, 1011–1016 (1996).**
8. D. M. Sanderson and C. G. Earnshaw, Human Exp. Toxicol., **10,261–273 (1991).**
9. N. Greene, *J.* Chem. *Inf.* Comput. Sci., **37**, 148–150 (1996).
10. J. E. Ridings, M. D. Barratt, R. Cary, C. G. Earnshaw, C. E. Eggington, M. K. Ellis, P. N. Judson, J. J. Langowski, C. A. Marchant, M. P. Payne, W. P. Watson, and T. D. Yih, Toxicology, **106,** 267–279 (1996).
11. **Y.** T. Woo, D. Y. Lai, M. F. Argus, and J. C. Arcos, Toxicol. *Lett.*, **79,219–228 (1995).**
12. **Y.** T. Woo, D. Y. Lai, M. F. Argus, and J. C. Arcos, Environ. Carcin. Ecotoxicol. Rev. C., **16, 101–102 (1998).**
13. F. Darvas, **A.** Papp, A. Allerdyce, E. Benfenati, G. Fini, M. Tichy, N. Sobb, and A. Citti, A M Spring Symposium on Predictive Toxicology of Chemicals: Experiences and Impact of AI Tools, Technical Report **SS-99-01,** AAAI Press, Mento Park, CA, **1999.**
14. R. D. King, S. H. Muggleton, A. Srinivasan, and M. J. E. Sternberg, Proc. Natl. *Acad.* Sci., **93, 438–442 (1996).**
15. J. Ashby, Environ. Mutagen., **7, 919–921 (1985).**
16. A. M. Richard, *Mutat.* Res., **305**, 73–77 **(1994).**
17. A. M. Richard, Knowledge Engineer. Rev., **14, 307–317 (1999).**
18. R. D. Combes and P. Judson, Pestic. Sci., **45, 179–194 (1995).**
19. A. M. Richard, Toxicol. **Lett.102–103, 611–616 (1998).**
20. C. Helma, E. Gottmann, and S. Kramer, Stat. Methods. Med. Res., **9, 1 3 0 (2000).**
21. A. M. Richard and R. Benigni, SAR QSAR *Environ.* Res., 13, **1–19 (2002).**
22. M. Liu, N. Sussman, G. Klopman, and H. S. Rosenkranz, *Mutat.* Res., **358**, 63–72 **(1996).**
23. D. M. Maron and B. N. Ames, *Mutat.* Res., **113, 173–215 (1983).**
24. D. Brusick, **V.** F. Simmon, H. S. Rosenkranz, V. Ray, and R. S. Stafford, *Mutat.* Res., **76,** 169–**190 (1980).**
25. J. Pet-Edwards, **H.** S. Rosenkranz, V. Chankong, and Y. Y. Haimes, *Mutat.* Res., **153, 167–185 (1985).**
26. H. S. Rosenkranz, F. K. Ennever, M. Dimayuga, and G. Klopman, Environ. Mol. Mutagen., **16, 149–177 (1990).**
27. A. D. Mitchell, A. E. Auletta, D. Clive, P. E. Kirby, M. M. Moore, and B. C. Myhr, *Mutat.* Res., **394,177–303 (1997).**
28. B. Henry, S. G. Grant, G. Klopman, and H. S. Rosenkranz, *Mutat.* Res., **397,313–335 (1998).**
29. S. G. Grant, Y. P. Zhang, G. Klopman, and H. S. Rosenkranz, *Mutat.* Res., **465,201–229 (2000).**
30. M. Ghanooni, D. R. Mattison, Y. P. Zhang, O. T. Macina, H. S. Rosenkranz, and G. Klopman, *Am.* J. Obstet. Gynecol., **176, 799–806 (1997).**
31. J. Gómez, O. T. Macina, D. R. Mattison, Y. P. Zhang, G. Klopman, and H. S. Rosenkranz, *Ter*atology, **60, 190–205 (1999).**
32. H. S. Rosenkranz, Y. P. Zhang, and G. Klopman, *Altern.* Anim. Test., **26, 779–809 (1998).**

33. C. Graham, R. Gealy, O. T. Macina, M. H. Karol, and H. S. Rosenkranz, *Quant. Struct. Activ. Relat.*, 15, 224–229 (1996).
34. N. Takihi, Y. P. Zhang, G. Klopman, and H. S. Rosenkranz, *Mutagenesis*, 8,257–264 (1993).
35. N. Takihi, Y. P. Zhang, G. Klopman, and H. S. Rosenkranz, *Qual. Assur. Good Pract. Regul. Law*, 2,255–264 (1993).
36. H. S. Rosenkranz and A. R. Cunningham, *Mutat. Res.*, 476, 133–137 (2001).
37. H. S. Rosenkranz and A. R. Cunningham, *SAR QSAR Environ. Res.*, 12,267–274 (2001).
38. V. Chankong, Y. Y. Haimes, H. S. Rosenkranz, and J. Pet-Edwards, *Mutat. Res.*, 153, 135–166 (1985).
39. Y. P. Zhang, N. Sussman, G. Klopman, and H. S. Rosenkranz, *Quant. Struct. Activ. Relat.*, 16, 290–295 (1997).
40. A. R. Cunningham, H. S. Rosenkranz, Y. P. Zhang, and G. Klopman, *Mutat. Res.*, 398, 1–17 (1998).
41. F. K. Ennever, T. J. Noonan, and H. S. Rosenkranz, *Mutagenesis*, 2, 73–78 (1987).
42. H. Bartsch and C. Malaveille, *Cell Biol. Toxicol.*, 5, 115–127 (1989).
43. J. Ashby and R. S. Morrod, *Nature*, 352, 185–186 (1991).
44. M. D. Shelby, *Mutat. Res.*, 204, 3–15 (1988).
45. J. Ashby and R. W. Tennant, *Mutat. Res.*, 257, 229–306 (1991).
46. L. S. Gold, N. B. Manley, T. H. Slone, G. B. Garfinkel, L. Rohrbach, and B. N. Ames, *Environ. Health Perspect.*, 100, 65–135 (1993).
47. H. S. Rosenkranz and G. Klopman, *Mutat. Res.*, 228, 51–80 (1990).
48. V. Mersch-Sundermann, G. Klopman, and H. S. Rosenkranz, *Mutat. Res.*, 340, 81–91 (1996).
49. Y. P. Zhang, A. van Praagh, G. Klopman, and H. S. Rosenkranz, *Mutagenesis*, 9, 141–149 (1994).
50. H. S. Rosenkranz and G. Klopman, *Environ. Mol. Mutagen.*, 21, 193–206 (1993).
51. H. S. Rosenkranz, E. J. Matthews, and G. Klopman, *Altern. Anim. Test.*, 20, 549–562 (1992).
52. M. Rosenkranz, H. S. Rosenkranz, and G. Klopman, *Mutat. Res.*, 381, 171–188 (1997).
53. J. E. Trosko and C. C. Chang in R. W. Hoerger and F. D. Hoerger, Eds., *Banbury Report* 31: *Carcinogen Risk Assessment: New Directions in Qualitative and Quantitative Aspects*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1988, pp. 139–174.
54. B. N. Ames and L. S. Gold, *Proc. Natl. h a d . Sci. USA*, 87, 7772–7776 (1990).
55. S. M. Cohen and L. B. Ellwein, *Science*, 249, 1007–1011 (1990).
56. S. Preston-Martin, M. C. Pike, R. K. Ross, P. A. Jones, and B. E. Henderson, *Cancer Res.*, 50, 7415–7421 (1990).
57. E. ter Haar, H. S. Rosenkranz, E. Hamel, and B. W. Day, *Bioorg. Med. Chem.*, 4, 1659–1671 (1996).
58. E. Hamel, *Med. Res. Rev.*, 16,207–231 (1996).
59. P. B. Schiff and S. B. Horwitz, *Proc. Natl. h a d . Sci. USA*, 77,1561–1565 (1980).
60. P. B. Schiff, J. Fant, and S. B. Horwitz, *Nature (Lond.)*, 277,665–667 (1979).
61. J. McCann, N. E. Spingarn, J. Kobori, and B. N. Ames, *Proc. Natl. Acad. Sci. USA*, 72, 979–983 (1975).
62. V. Mersch-Sundermann, U. Schneider, G. Klopman, and H. S. Rosenkranz, *Mutagenesis*, 9, 205–224. (1994).
63. J. A. Heddle, M. C. Cimino, M. Hayashi, F. Romagna, M. D. Shelby, J. D. Tucker, Ph. Vanparys, and J. T. MacGregor, *Environ. Mol. Mutagen.*, 18,277–291 (1991).
64. K. H. Mavournin, D. H. Blakey, M. C. Cimino, M. F. Salamone, and J. A. Heddle, *Mutat. Res.*, 239, 29–80 (1990).
65. H. Tinwell and J. Ashby, *Environ. Health Perspect.*, 102, 758–762 (1994).
66. E. ter Haar, B. W. Day, and H. S. Rosenkranz, *Mutat. Res.*, 350, 331337 (1996).
67. E. ter Haar, R. J. Kowalski, E. Hamel, C. M. Lin, R. E. Longley, S. P. Gunasekera, H. S. Rosenkranz, and B. W. Day, *Biochemistry*, 35, 243–250 (1996).
68. H. Tinwell and J. Ashby, *Carcinogenesis*, 15, 1499–1501 (1994).