Using Argumentation for Absolute Reasoning about the Potential Toxicity of Chemicals

Philip N. Judson,* Carol A. Marchant, and Jonathan D. Vessey
LHASA Limited, Department of Chemistry, University of Leeds, Leeds LS2 9JT, United Kingdom

Received June 10, 2002

The application of a new argumentation model is illustrated by reference to DEREK for Windows, a knowledge-based expert system for the prediction of the toxicity of chemicals. Examples demonstrate various aspects of the model such as the undercutting of arguments, the resolution of multiple arguments about the same proposition, and the propagation of arguments along a chain of reasoning.

INTRODUCTION

This paper describes a practical use of the model for argumentation described in a preceding paper,[1] in an application to predict the potential toxicity of chemicals, DEREK for Windows.[2] The introduction of the argumentation model into DEREK for Windows has resulted in a number of benefits, including the association of levels of belief with predictions and the ability to reason about the effects of the physicochemical and known toxicological properties of a chemical. Extrapolation of a prediction for one toxicological endpoint to a second related endpoint to fill data gaps and the expression of confidence in an absence of toxicological activity are also possible.

In this application, use is made only of "absolute reasoning", by which is meant reasoning about the level of belief in an isolated event or chain of events, as distinct from "relative reasoning", by which is meant reasoning about the ranking of likelihoods of two or more alternative outcomes of events or chains of events. The use of relative reasoning is discussed in another paper in this series.[3]

DEREK FOR WINDOWS

DEREK for Windows is a knowledge-based expert system for toxicity prediction which has evolved from the DEREK program and the StAR project. The DEREK program was originally conceived at Schering Agrochemicals Limited[4] and jointly developed over a number of years by LHASA Limited and the LHASA group at Harvard University.[5,6] The StAR project was established to develop a PC-based computer system to aid in chemical carcinogenicity risk assessment, which has become DEREK for Windows. The project was a collaboration between LHASA Limited, Imperial Cancer Research Fund, Logic Programming Associates and City University.[7] This paper refers to DEREK for Windows version 5.0.1.

The DEREK for Windows knowledge base is composed of alerts, examples, and rules which may each contribute to the toxicity predictions made by the system.

Alerts. Each alert in the knowledge base describes the relationship between a structural feature, or toxicophore, and the toxicological endpoint with which it is associated. One or more patterns similar to Markush-type structures are used to describe the toxicophore itself. Figure 1 shows an example of an alert as it appears in the editor which is used to build and maintain the knowledge base. Alert 401 describes the relationship between carboxylic acid halides and skin sensitization. Its toxicophore is a simple one and can be represented by a single pattern.

When a query structure is processed, the system compares it with the toxicophores in the knowledge base and the finding of toxicophores results in the activation or firing of the corresponding alerts. Figure 2 shows the result of processing decanoyl chloride for the skin sensitization endpoint. A match with the toxicophore of alert 401 results in activation of this alert and a prediction of skin sensitization. The location of the toxicophore in the structure is highlighted.

Examples. Alerts in the DEREK for Windows knowledge base are supported by a summary of the evidence which forms the basis of the alert, a list of references from which this evidence was derived, and a selection of example compounds and their associated toxicological data. This information can be viewed via the Alert Description window which is directly accessible from the results display. When processing, DEREK for Windows also checks for the presence of an exact structure match between the query structure and the structures of example compounds. The existence of a match is reported on the results display in a similar way to that for alerts, as shown in Figure 3 for nonanoyl chloride. This compound has been reported to give a positive response in the murine local lymph node assay and is included in the knowledge base as a supporting example for alert 401.[8]

Rules. It can be seen from Figures 2 and 3 that the prediction of skin sensitization is associated with a level of belief such as 'plausible' or 'probable' which is provided by use of the argumentation model described in the preceding paper in this series.[1] Arguments used by the model are described by rules in the DEREK for Windows knowledge base which have the following general structure:

If [Grounds] is [Threshold] then [Proposition] is [Force] (1)

For example, a simple rule might have the following form:

If [Skin sensitization alert] is [certain] then [Skin sensitization] is [plausible] (2)

* Corresponding author phone: +44 (0) 1943 880241; e-mail: judson@dircon.co.uk.

10.1021/ci020272g CCC: $25.00
Published on Web 08/23/2003

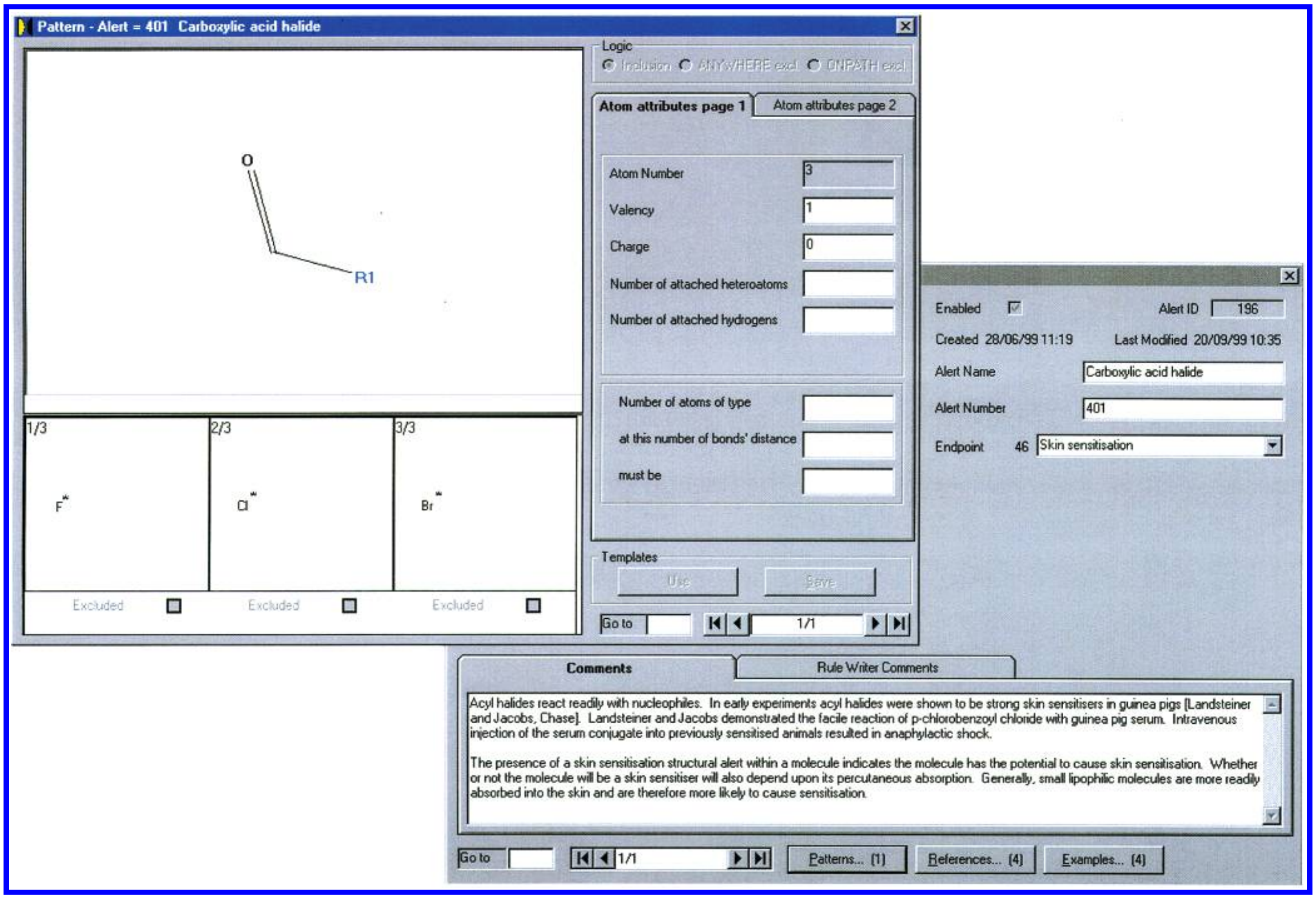

Figure 1. Alert 401, which describes the skin sensitization of carboxylic acid halides, as it appears in the DEREK for Windows editor.

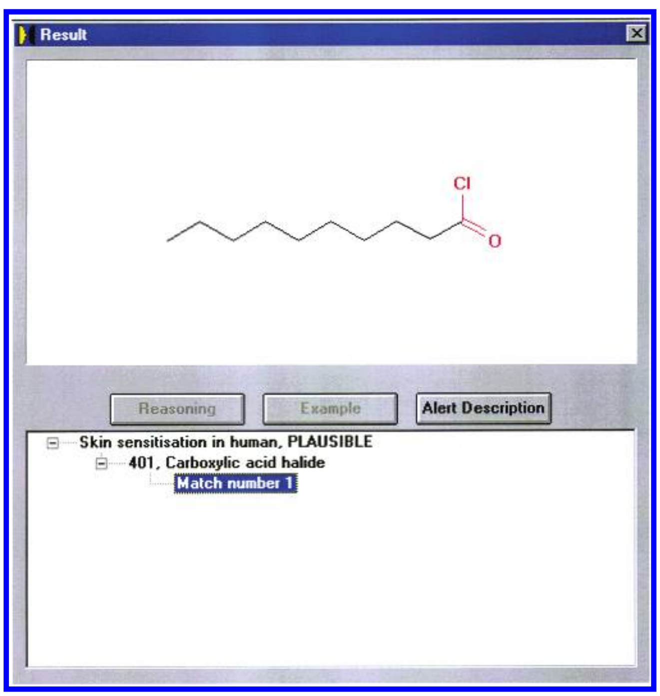

Figure 2. Skin sensitization prediction for decanoyl chloride in humans in DEREK for Windows.

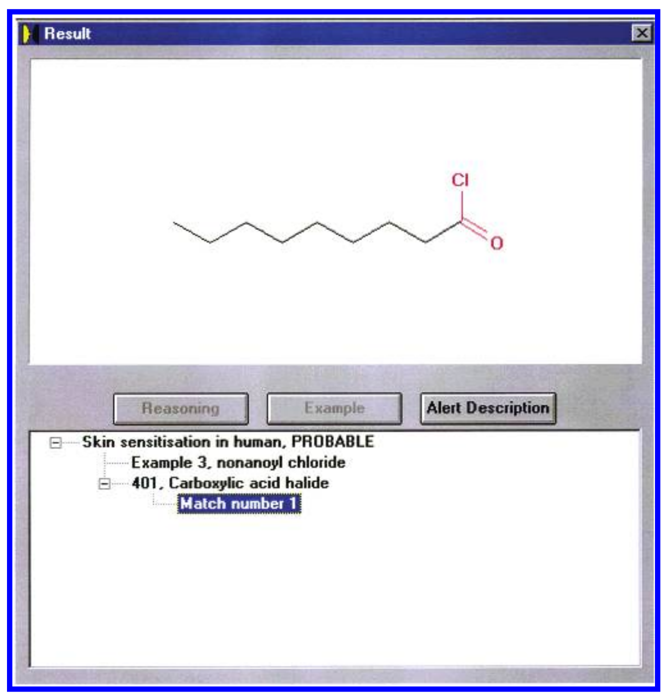

Figure 3. Skin sensitization prediction for nonanoyl chloride in humans in DEREK for Windows.

In this case, the rule describes the argument relating to the proposition that a query structure will cause skin sensitization on the grounds that it activates an alert for skin sensitization. In the event that a query structure activates one or more alerts for skin sensitization then the threshold for the argument is met, leading to the conclusion that it is 'plausible' that the compound will be a skin sensitizer.

The facts upon which DEREK for Windows can base its reasoning (i.e. those things about which rules can be written which DEREK for Windows can understand) currently

Table 1. Uncertainty Terms Used in DEREK for Windows and Their Definitions

certain	There is proof that the proposition is true.
probable	There is at least one strong argument that the proposition is true and there are no arguments against it.
plausible	On balance the weight of evidence supports the proposition.
equivocal	There is an equal weight of evidence for and against the proposition.
doubted	On balance the weight of evidence opposes the proposition.
improbable	There is at least one strong argument that the proposition is false, and there are no arguments that it is true.
impossible	There is proof that the proposition is false.
contradicted	There is proof both that the proposition is true and that it is false.
open	There is no evidence that supports or opposes the proposition.
undefined	The grounds for a decision have not yet been considered or have not been provided.

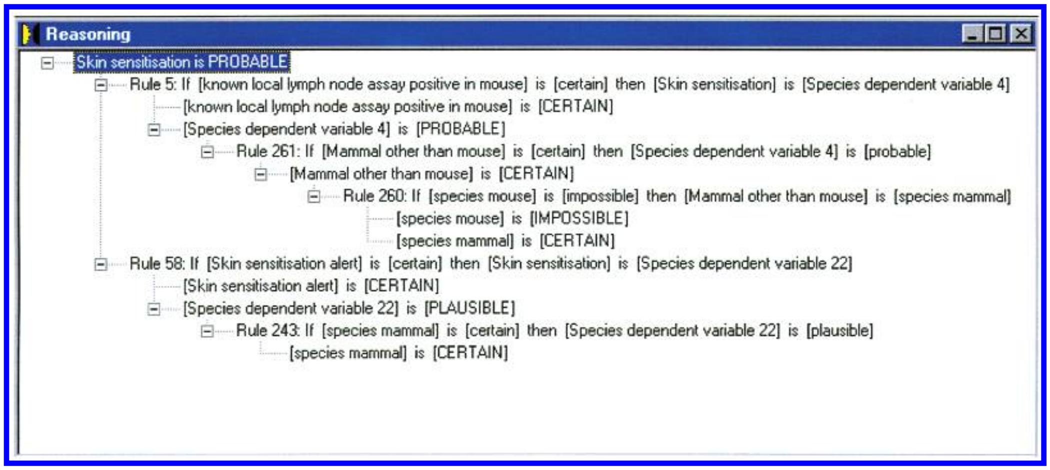

Figure 4. Reasoning display for the skin sensitization prediction of nonanoyl chloride in humans in DEREK for Windows.

include individual alerts, toxicological endpoints, sets of alerts linked by a common endpoint, species, physicochemical properties, and toxicological data.

It is necessary to define the levels used in the argumentation model in such a way that they can both be processed consistently with respect to one another and communicated to the user. The words used in DEREK for Windows to describe these levels (the uncertainty terms) and their definitions are listed in Table 1. They were chosen and defined in the light of psychological experiments which found them to be acceptable and to be interpreted consistently enough for practical use[9] and are similar to those suggested at various times by Fox and co-workers.[10] It is important to note, however, that the definitions of these terms are not the same as those in normal usage and that the subjects who took part in the psychological experiments were all individuals whose first language was UK English.

These values of force include three polarizing forces[1] in each of the sets For (‘certain’, ‘probable’, ‘plausible’) and Against (‘impossible’, ‘improbable’, ‘doubted’). The sets, For and Against, are described more fully below.

Undercutting Arguments. Figure 2 shows the prediction of skin sensitization for decanoyl chloride in humans. Predictions in other species can also be made by selection of the query species of choice from the DEREK for Windows processing constraints menu. To support this functionality, many of the rules in the DEREK for Windows knowledge base need to take account of species specificity. Thus, for example, the simple rule (2) given above describing the relationship between the presence of a skin sensitization alert within a structure and the occurrence of skin sensitization is in practice replaced by the following group of rules (3)–(5).

Table 2. Priority and Complementarity Lists for Forces of Arguments

for	against
contradicted	contradicted
certain	impossible
probable	improbable
plausible	doubted
equivocal	equivocal
open	open
undefined	undefined

If [Skin sensitization alert] is [certain] then
[Skin sensitization] is [Variable name] (3)

If [species mammal] is [certain] then
[Variable name] is [plausible] (4)

If [species bacterium] is [certain] then
[Variable name] is [impossible] (5)

Any name could be used for the variable, “Variable name”. In practice, it is called “Species dependent variable 22”.

As a consequence of this group of rules, while the level of belief in the skin sensitization prediction of decanoyl chloride in humans is ‘plausible’, it becomes ‘impossible’ if a bacterial query species such as *Salmonella typhimurium* is selected instead.

Combining Multiple Arguments for or against a Proposition. The uncertainty terms used in DEREK for Windows are ranked as shown in Table 2.

Where multiple arguments either for or against a proposition occur then the overall conclusion is taken to be the individual conclusion with the highest priority in the list. Returning to the DEREK for Windows prediction for nonanoyl chloride shown in Figure 3, for example, the overall

```
int   Resolve[10][10] =
{
   /*          C     PB    PL    E     D     IB    IS    CD    O     U    */
   /*C*/   {   C,    C,    C,    C,    C,    C,    CD,   C,    C,    U    },
  /*PB*/   {   C,    PB    PB,   PB,   PL,   E,    IS,   CD,   PB,   U    },
  /*PL*/   {   C,    PB,   PL,   PL,   E,    D,    IS,   CD,   PL,   U    },
   /*E*/   {   C,    PB,   PL,   E,    D,    IB,   IS,   CD,   E,    U    },
   /*D*/   {   C,    PL,   E,    D,    D,    IB,   IS,   CD,   D,    U    },
  /*IB*/   {   C,    E,    D,    IB,   IB,   IB,   IS,   CD,   IB,   U    },
  /*IS*/   {   CD,   IS,   IS,   IS,   IS,   IS,   IS,   IS,   IS,   U    },
  /*CD*/   {   C,    CD,   CD,   CD,   CD,   CD,   IS,   CD,   CD,   U    },
   /*O*/   {   C,    PB,   PL,   E,    D,    IB,   IS,   CD,   O,    U    },
   /*U*/   {   U,    U,    U,    U,    U,    U,    U,    U,    U,    U    }
};
Key:
C = certain, PB = probable, PL = plausible, E = equivocal, D = doubted, IB = improbable,
IS = impossible, CD = contradicted, O = open, U = undefined.
```

Figure 5. Matrix for prioritizing and resolving forces as implemented in C++.

Table 3. Classification of Terms as Exclusive or Inclusive

exclusive	inclusive
contradicted	impossible
equivocal	improbable
open	doubted
undefined	plausible
	probable
	certain

conclusion is that skin sensitization is 'probable'. The reasoning behind this conclusion can be viewed by selecting the Reasoning button from the results display, as shown in Figure 4.

There are two arguments contributing to the skin sensitization prediction. The first of these relates to the fact that the structure is known to give a positive response in the local lymph node assay in the mouse and it results in a conclusion of 'probable'. The second relates to the presence of a skin sensitization alert within the structure and results in a conclusion of 'plausible'. The overall conclusion for skin sensitization is then 'probable', the individual conclusion with the higher priority according to the list in Table 2. Any number of arguments for a proposition can be combined in this way, and a comparable procedure can likewise be used for combining multiple arguments against a proposition.

Although the 'contradicted' term heads the priority lists in both the 'for' and 'against' cases, in practice the situation where the conclusion of a single argument is 'contradicted' will be rare—perhaps always avoidable. As a result, 'certain' and 'impossible' will normally dominate when combining multiple arguments for or against a proposition.

Exclusivity. Whether a term is exclusive or inclusive has implications for the way it is handled in comparisons in reasoning algorithms.[1] In DEREK for Windows, terms are classed as exclusive or inclusive as shown in Table 3.

Resolving Arguments for and against a Proposition. Arguments for and against the same proposition are resolved using the matrix shown in Table 4. The resolutions conform to Implementation 1 in the preceding paper in this series.[1] In the actual implementation in C++, Table 4 is replaced by the matrix illustrated in Figure 5 which is also used to resolve the ordering of forces expressed in Table 2. Where there are multiple arguments for and/or against a proposition, the arguments with the highest priorities for and against the proposition are determined first and then resolved. The operations must be done in this order for results to be correct and consistent.

From the matrix in Table 4, it can be seen that 'certain' or 'impossible' dominates if it is one of a pair of terms in a resolution, except when 'certain' and 'impossible' oppose each other (the classical case of contradiction), there is independent contradiction of the rare kind resulting from a single argument, or one of the terms is undefined (in which case 'undefined' dominates as a fail-safe measure). Apart from 'certain' and 'impossible', terms of equivalent weight

Table 4. Matrix for Resolving Forces for and against the Same Proposition

	contradicted	certain	probable	plausible	equivocal	open	undefined
contradicted	contradicted	contradicted	contradicted	contradicted	contradicted	contradicted	undefined
impossible	contradicted	contradicted	impossible	impossible	impossible	impossible	undefined
improbable	contradicted	certain	equivocal	doubted	improbable	improbable	undefined
doubted	contradicted	certain	plausible	equivocal	doubted	doubted	undefined
equivocal	contradicted	certain	probable	plausible	equivocal	equivocal	undefined
open	contradicted	certain	probable	plausible	equivocal	open	undefined
undefined	undefined	undefined	undefined	undefined	undefined	undefined	undefined

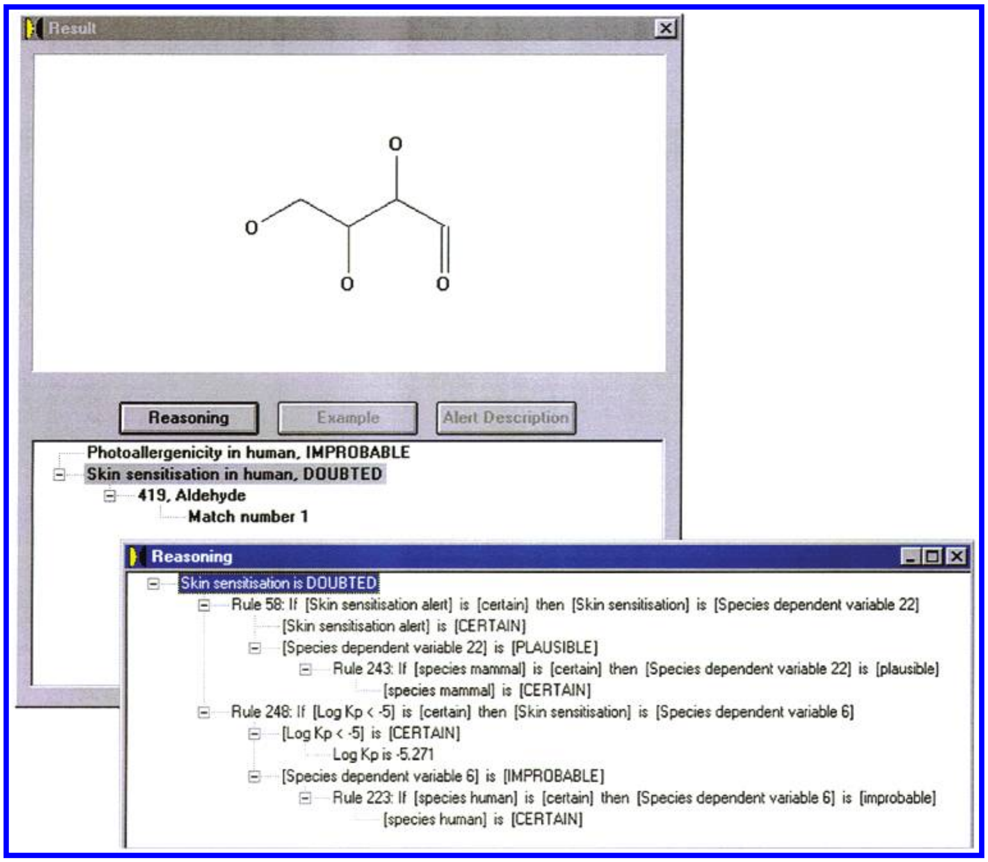

Figure 6. Result and reasoning displays for the skin sensitization prediction of 2,3,4-trihydroxybutanal in humans in DEREK for Windows.

for and against are resolved to ‘equivocal’. The resolutions of ‘probable’ with ‘doubted’ to give ‘plausible’, and of ‘improbable’ with ‘plausible’ to give ‘doubted’, are consistent with the definitions of the terms given in Table 1.

The practical application of the matrix in Table 4 is illustrated in Figure 6 which shows the skin sensitization prediction for 2,3,4-trihydroxybutanal. The level of belief associated with this prediction is ‘doubted’ and it can be seen from the reasoning window that this conclusion is the result of two arguments. The first of these relates to the presence of a skin sensitization alert within the structure and results in a conclusion of ‘plausible’. The second relates to the fact that the logarithm of the coefficient of skin permeability measured in cm/h (log Kp) for this compound is estimated to be less than −5 and results in a conclusion of ‘improbable’. Resolving these arguments for and against the proposition that 2,3,4-trihydroxybutanal may cause skin sensitization using the matrix in Table 4 results in an overall conclusion of ‘doubted’ as shown on the result display.

The basis of the argument making use of the log K_p value is that compounds with low skin permeability are considered unlikely to cause skin sensitization.[11] The log K_p value used by DEREK for Windows is estimated from the Potts and Guy equation[12] using log P values calculated by a ClogP plug-in.[13]

The results display in Figure 6 also includes the DEREK for Windows prediction for the related endpoint of photo-allergenicity which is given an ‘improbable’ level of belief. This outcome is the consequence of a single argument, similar to the one described above for skin sensitization, relating the log K_p value for the compound to its potential for photoallergenicity. Unlike the case for skin sensitization, however, there is no second argument relating to the presence of a photoallergenicity alert since none is present in the structure.

The prediction for the photoallergenicity of 2,3,4-trihydroxybutanal illustrates two benefits of incorporating the reasoning model into DEREK for Windows. First, it makes it possible for the prediction to take account of the physicochemical properties of the compound. Second, the prediction expresses an expectation that 2,3,4-trihydroxybutanal will not cause photoallergenicity. Such predictions were not possible in earlier versions of the DEREK for Windows system, which lacked the reasoning model and based their predictions solely on the presence of alerts. It is not possible to conclude inactivity on the basis only of the absence of any alerts for a given toxicological endpoint, since the set of alerts for that endpoint in the knowledge base may be incomplete, but merely to conclude that there is no evidence either for or against activity (corresponding to the ‘open’ state in the reasoning model).

Propagating Arguments along a Chain. All forces in DEREK for Windows belong to complementary pairs as listed in Table 2.

So far the examples used to illustrate the application of the reasoning model in DEREK for Windows have been limited to predictions for skin sensitization and the related endpoint of photoallergenicity, but the DEREK for Windows knowledge base incorporates coverage for a range of endpoints including carcinogenicity. One of the limiting factors in developing a knowledge base for carcinogenicity is the comparatively small number of compounds which have undergone adequate testing in a carcinogenicity bioassay. In DEREK for Windows, this may result in insufficient

Table 5. Examples of the Types of Evidence Associated with Particular Uncertainty Terms in DEREK for Windows

term	evidence
certain	The query structure is present in the knowledge base as a supporting example which includes positive in vivo evidence of toxicity for the endpoint in question in the selected query species.
probable	The query structure is present in the knowledge base as a supporting example which includes positive in vivo evidence of toxicity for the endpoint in question in a species related to the selected query species.
plausible	The query structure activates an alert for the endpoint in question.
	The query structure is present in the knowledge base as a supporting example which includes positive in vitro evidence of toxicity for the endpoint in question in the selected query species or a species related to it.
equivocal	The query structure activates an alert for the endpoint in question but there is limited evidence either for the alert or its relevance to the selected query species.
	The query structure is present in the knowledge base as a supporting example which includes borderline in vivo evidence of toxicity for the endpoint in question in the selected query species or a species related to it.
doubted	The query structure is present in the knowledge base as a supporting example which includes negative in vitro evidence of toxicity for the endpoint in question in the selected query species or a species related to it.
improbable	The query structure activates an alert for the endpoint in question but there is significant evidence to indicate that the endpoint is not relevant to the selected query species.
	The query structure has physicochemical properties which indicate that significant absorption is unlikely.
	The query structure is present in the knowledge base as a supporting example which includes negative in vivo evidence of toxicity for the endpoint in question in the selected query species or a species related to it.
impossible	The physiological apparatus required for the endpoint in question to occur is known not to be present in the selected query species.
open	The query structure activates an alert for the endpoint in question but the alert is considered irrelevant to the selected query species.
	The query structure is present in the knowledge base as a supporting example which includes toxicity data for the endpoint in question which is not considered relevant to the selected query species.

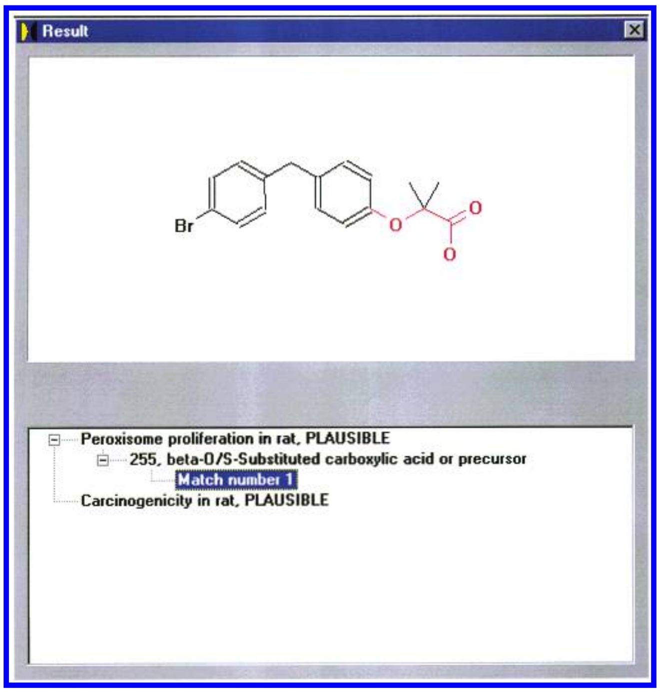

Figure 7. Peroxisome proliferation and carcinogenicity predictions in the rat for 2-[4-[(4-bromophenyl)methyl]phenoxy]-2-methylpropanoic acid in DEREK for Windows.

bioassay data being available to support implementation of a new carcinogenicity alert. One means of circumventing this limitation is to make use of the more abundant data derived from short-term tests known to be mechanistically related to carcinogenicity. Thus, for example, it is widely accepted that there is a relationship between peroxisome proliferation and the observation of tumors of the liver in certain rodent species.[14] In the DEREK for Windows knowledge base this relationship is expressed in terms of a rule with the following format:

$$\text{If [Peroxisome proliferation] is [certain] then [Carcinogenicity] is [probable]} \quad (6)$$

Figure 7 shows the result of processing a compound containing an alert for peroxisome proliferation in the rat. As a result of the presence of this alert, the level of belief associated with the prediction of peroxisome proliferation is ‘plausible’. This in turn leads to a ‘plausible’ prediction of carcinogenicity on the basis of the above rule. Note that the conclusion from the argument is ‘plausible’ and not ‘probable’ because the ‘plausible’ prediction for peroxisome proliferation does not meet the ‘certain’ threshold for the rule. Thus, the weakness in certainty about peroxisome proliferation is propagated along the chain of argument to the proposition of carcinogenicity.

Processing the same compound with the human query species results in an ‘improbable’ level of belief for peroxisome proliferation since humans are known to be much less susceptible to this effect.[14] In this case, rule 6 does not lead to any conclusion about carcinogenicity since it only describes the relationship between the *presence* of peroxisome proliferation and carcinogenicity. A separate rule would be required if any conclusion about carcinogenicity was to be drawn from the *absence* of peroxisome proliferation. In practice, such a rule would be inappropriate since it is not possible to conclude anything about the carcinogenic potential of a compound from an absence of peroxisome proliferator activity. The conclusion should remain ‘open’ because there are many other alternative mechanisms by which carcinogenicity might arise.

Assigning Levels of Uncertainty for Grounds at the Beginning of Chains of Reasoning. Definitions of the uncertainty terms used for the reasoning model in DEREK for Windows are given in Table 1. Some of these definitions refer to “strong” evidence but do not state what constitutes such evidence. This allows flexibility so that definitions can be made appropriate to the domain in which the model is being used. It does not affect the reasoning process or the validity of the model itself.

The predictions already described illustrate some of the types of evidence in the toxicology domain which are considered appropriate for association with some of the uncertainty terms. These are summarized in Table 5 together with further examples. It is expected that the range of evidence and its association with particular uncertainty terms will be expanded and refined as the use of the reasoning model in DEREK for Windows develops.

EXPERIMENTAL SECTION

Figures 1−4, 6, and 7 were taken from the program LPS version 5.0.1[2] running DEREK for Windows on a PC with a 500 MHz Intel Pentium III processor and 128 Mb RAM running Windows NT version 4 with service pack 6. LPS version 5.0.1 was written and compiled with Microsoft Visual Basic and Visual C++ version 6 with service pack 5; Figure 5 is taken from a section of such code.

CONCLUSION

The use of argumentation for absolute reasoning about the potential toxicity of chemicals has been illustrated by reference to the DEREK for Windows system. Introduction of the reasoning model into this system has resulted in a number of benefits over earlier versions of the software. These include the association of levels of belief with predictions and the ability to reason about the effects of the physicochemical and known toxicological properties of a chemical. Extrapolation of a prediction for one toxicological endpoint to a second related endpoint also becomes possible, so that advantage can be taken of general toxicological principles to fill gaps in available data. Since DEREK for Windows predictions no longer rely solely on the presence of alerts, confidence in an absence of toxicological activity can also be expressed in some cases.

REFERENCES AND NOTES

(1) Judson, P. N.; Vessey, J. D. A comprehensive approach to argumentation. *J. Chem. Inf. Comput. Sci.* **2003**, *43*, 1356−1363.

(2) The DEREK for Windows program is available from LHASA Limited, Department of Chemistry, University of Leeds, Leeds LS2 9JT, United Kingdom. DEREK is a trade mark of LHASA Limited and Windows is a trade mark of Microsoft Corporation.

(3) Button, W. G.; Judson, P. N.; Long, A.; Vessey, J. D. Using argumentation for absolute and relative reasoning about potential metabolic sequences. *J. Chem. Inf. Comput. Sci.* **2003**, *43*, 1371−1377.

(4) Judson, P. N. The use of expert systems for detecting potential toxicity. A paper presented at the Autumn meeting of the British Toxicology Society, 20th−22nd September 1989.

(5) Sanderson, D. M.; Earnshaw, C. G. Computer prediction of possible toxic action from chemical structure; the DEREK system. *Hum. Exp. Toxicol.* **1991**, *10*, 261−273.

(6) 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.; Yih, T. D. Computer prediction of possible toxic action from chemical structure: an update on the DEREK system. *Toxicology* **1996**, *106*, 267−279.

(7) Greene, N.; Judson, P. N.; Langowski, J. J.; Marchant, C. A. Knowledge-base expert systems for toxicity and metabolism prediction: DEREK, StAR and METEOR. *SAR QSAR Environ. Res.* **1999**, *10*, 299−314.

(8) Ashby, J.; Basketter, D. A.; Paton, D.; Kimber, I. Structure activity relationships in skin sensitisation using the murine local lymph node assay. *Toxicology* **1995**, *103*, 177−194.

(9) Unpublished work by Ayton, P.; Hardman, D.; Pascoe, E. at City University, London.

(10) Krause, P. J.; Fox, J.; Judson, P. N.; Patel, M. Qualitative risk assessment fulfils a need. In *Applications of Uncertainty Formalisms*; Hunter, A., Parsons, S., Eds.; Springer-Verlag: Berlin, 1998; pp 138−156.

(11) Basketter, D. A.; Scholes, E. W.; Chamberlain, M.; Barratt, M. D. An alternative strategy to the use of guinea pigs for the identification of skin sensitisation hazard. *Food Chem. Toxicol.* **1995**, *33*, 1051−1056.

(12) Potts, R. O.; Guy, R. H. Predicting skin permeability. *Pharm. Res.* **1992**, *9*, 663−669.

(13) The ClogP plug-in is produced by BioByte Corporation and is also available from LHASA Limited, Department of Chemistry, University of Leeds, Leeds LS2 9JT, United Kingdom.

(14) Corton, J. C.; Lapinskas, P. J.; Gonzalez, F. J. Central role of PPARα in the mechanism of action of hepatocarcinogenic peroxisome proliferators. *Mutat. Res.* **2000**, *448*, 139−151.

CI020272G