Using New Reasoning Technology in Chemical Information Systems

P. N. Judson*

Heather Lea, Bland Hill, Norwood, Harrogate HG3 1TE, U.K.

J. Fox and P. J. Krause

Imperial Cancer Research Fund, P.O. Box 123, Lincoln's Inn Fields, London WC2A 3PX, U.K.

Received September 9, 1995[⊗]

Unreliability of numerical data causes difficulties in computer systems for decision-making, risk assessment, and similar activities. Much human judgment is non-numerical and able to make useful evaluations of alternatives under uncertainty. The Logic of Argumentation (LA) offers a basis for computerized support of decision-making in the absence of numerical data, and it is being used in a project on carcinogenic risk assessment, StAR. There are potential applications of LA in other artificial intelligence systems in chemistry, such as for synthesis planning.

1. INTRODUCTION

Many systems, computerized or manual, for making diverse predictions use numerical methods to assess the likelihood that particular situations will arise—for example, in risk assessment, determining betting odds, weather forecasting, predicting pharmacological or toxicological properties of substances, and synthesis planning. Difficulties arise when there is a lack of good numerical data on which to base calculations, and this paper describes the use of a non-numerical approach, the Logic of Argumentation (LA), in a project to develop a system to assess toxicological hazards.[1]

The paper is not intended to be comprehensive but to illustrate how LA is being applied in this project and might be used in other areas of chemical information.

2. NUMERICAL EVALUATION

A huge amount of work has been published on applications of algebraic methods in the field of quantitative structure−activity relationships (QSAR), particularly applications of the Hansch approach[2] and variations of it. It is generally accepted that these methods work for tightly-defined series of related compounds but are less successful in other cases and that the main factors determining their performance are the choice and reliability of the data they use.

The properties conferred on substances by substructures within them are not constant—they depend on molecular environment and the nature of the medium surrounding a molecule. In the case of molecules that interact with specific biological sites such as the substrate binding sites of enzymes, activity may be sensitive to the relative locations in space of several substructures, and those locations, in turn, depend on molecular flexibility and the energy barriers to adopting particular conformations. This variable behavior of substructures makes it impossible to compute trustworthy numerical values for their contributions to the biological properties of substances other than in sets of closely similar structures (where assuming constancy is a reasonable approximation). By their nature, algebraic methods require fairly precise numbers with which to work, and if they cannot be provided the methods fail.

Some systems base prediction of biological properties on frequentistic probability (e.g., HazardExpert).[3] For example, it might be argued that if two out of ten substances containing an aromatic nitro group are found to be carcinogenic, then the probability that a novel substance containing such a group will be carcinogenic is 0.2. Applying conventional mathematics, reasoning could be extended to argue that if six out of ten substances containing a reactive chloride group are carcinogenic, then the probability that a substance containing both an aromatic nitro group and a reactive chloride group will be carcinogenic is 0.68 ($0.2 + 0.6 - 0.2 \times 0.6$). The problem with this approach is that the mathematical equations relating to frequentistic probability derive from, and apply to, the rules of chance. Biological activity arising from a combination of substructural features is not a chance event, and the rules therefore break down.

In synthesis planning it is necessary to assess the relative merits of alternative synthetic routes either between a precursor and a product or, more generally, from a selection of precursors to a product. The factors that are most critical may depend on the requirements of the user. For example, if raw materials are rare, optimizing the overall percentage yield may be the most important consideration, whereas if raw materials are plentiful and synthesis on a large scale is planned, avoiding hazardous processes or the creation of ecologically harmful waste may be more important. Even if these uncertainties are set aside by assuming that yield is the primary consideration, a problem remains because of the difficulty of predicting the yield from a reaction that has not yet been attempted for the compound of interest.

Hendrickson et al.[4] have pointed out that as a generalization reactions can be optimized to give yields of around 80%. Assuming 80% for each step in a synthesis tree allows likely, optimized, overall yields for different sequences to be calculated and compared. That convergent syntheses are more efficient than nonconvergent ones can be taken into account in choosing likely retrosynthetic disconnections, before formal evaluation takes place.[5] In the WODKA

[⊗] Abstract published in *Advance ACS Abstracts,* April 1, 1996.

S0095-2338(95)00090-4 CCC: $12.00

system[6] a range of factors is taken into account to generate numerical ratings for alternatives in a synthesis plan, including comparison of structural features in starting materials and products, estimates of reactivity, assessment of selectivity, and cost. In LHASA, numerical methods are used in starting-material oriented searches for synthetic routes to compare the synthetic closeness of alternative starting materials to a target product[7] and to decide on the order in which to carry out structural modifications and the best reaction sequences to follow.[8]

LHASA makes use of numerical ratings for transforms (retro-reactions) in a number of its strategies. Originally, transform writers provided an initial, numerical rating for each transform, and wrote statements in the coding of the transform requiring increments or decrements to the rating to take account of specific factors (for example, a writer might specify that the rating of a transform should be decremented by 10 for precursors having a withdrawing group adjacent to the site of reaction, if that were known to inhibit, but not to prevent, the required reaction). When asked to provide an initial rating for a reaction between, say, 0 and 100, taking account of all relevant factors, different chemists were likely to give very different answers. Numerical initial ratings were therefore replaced with a set of non-numerical statements about the different factors contributing to the utility of a reaction.[9]

In the current version of LHASA, the transform writer grades factors such as typical yield, reliability (i.e., consistency of yield for different reactants), reputation (i.e., how often the reaction is reported in the literature), etc. on a scale of bad, poor, fair, good, very good, or excellent. Guidelines are provided to help with the choice of these categories. Using this method to set ratings was found to give acceptable consistency between different writers in a simple trial in which individuals were asked to rate the same transform independently (when grades did differ it was not by more than one degree—e.g., bad versus poor or good versus fair).[10] It was not determined whether this was only because of the separation of the different factors contributing to rating or also because of the use of non-numerical terms.

Although the transform writer is no longer required to provide a numerical initial rating, the LHASA program calculates one from the non-numerical information the writer provides, placing different weightings on the different factors contributing to it, so that a number is available for use in the comparison of synthetic routes. The problems of numerical inconsistency are not fully resolved in LHASA, as transform writers still specify numerical values for increments and decrements to ratings. This is partly for pragmatic reasons—serious inconsistency has not been observed between transform writers, and increments and decrements do not usually have dramatic effects on overall synthesis planning decisions, and so there has been little pressure to change this part of the evaluation process.

The use of a non-numerical representation for initial ratings raises the question of whether evaluation has to be based on numbers or whether alternative forms of computation might be available and appropriate. Moll[11] has commented on the problem of deriving numerical measures for the utility of reactions, and Jorgensen et al.,[12] for example, have discussed mechanistic and heuristic approaches. The synthesis planning system at Sumitomo Chemical Company uses a set of limiting criteria to control the growth of the synthetic tree,[13] such as the number of steps in a sequence, restrictions on the breaking of specific bonds, a requirement that the starting material should be one of a set in a given database, etc.

3. PREDICTIVE REASONING

The commonly-used statistical calculations leading to numerical probabilities grew out of studies of games of chance. There are other kinds of probabilistic estimation, however, such as epistemic probability in which prediction is based on past experience: a chemist is confident that sodium will react vigorously with water because it always has done in the past; asked how favorable the outcome of electrophilic substitution on an activated aromatic ring will be if the aim is to get a high yield of a specified, single product, a chemist is only moderately hopeful, knowing from experience that a mixture of products is likely—depending on other features of the molecule in question, he/she may make a judgment about the approximate likely ratio of the different products. Although the predictions in these examples could be made on the basis of energy calculations, calculations are neither necessary nor normally used in practice.

4. USING THE LOGIC OF ARGUMENTATION

There is no reason to suppose that the human brain spontaneously uses Bayesian probabilities to make judgments. Psychological research points to the heavy use of reasoning.[14] A common strategy for making judgments is to consider the arguments in favor and those against a proposition. LA seeks to formalize this type of reasoning in a mathematical framework.[15] Given an appropriate formalism, and implementation of it, it is possible to incorporate numerical statistical methods as a subset of the logic, so that where adequate numerical data are available, a numerical estimate is generated, but in other cases estimates are based on verbal expressions of probability, such as "certain" or "suspected".[15]

We are using LA in the StAR project[1] to develop a computer system for carcinogenic risk assessment. The following paragraphs illustrate the types of reasoning that can be used. A more detailed description of the use of LA in StAR and of the key elements of the logic can be found elsewhere.[16]

At a simple level, arguments directly in favor of or against a proposition can be constructed and then balanced against each other. So, for example, an argument in support of the proposition that substance X will cause cancer in man might be that it contains a well-researched and understood toxicophore. An argument against the proposition might be that the substance is a nonvolatile solid with physical properties that prevent its absorption through the skin or gut. In LA, arguments for a proposition are independent of those for its antipode: a lack of evidence to support the truth of a proposition is not in itself, in logical terms, a case for its being false—the situation is open; evidence opposing the case for a proposition being true does not automatically support the case for its being false.

Arguments may be about other arguments. An argument in support of the proposition that substance Y will cause cancer in man might be that it caused cancer in male rats in a well-authenticated experiment. An argument weakening

that supporting argument might be that male rats are known to be peculiarly susceptible to certain types of carcinogen and that in these cases cancer is not observed in other species. This argument is not a direct one against the carcinogenicity of Y, but one which undermines the argument in favor of carcinogenicity. Complex trees of interlinked arguments can be constructed and analyzed using LA, making automated decision-making more thorough and effective (applying these concepts with pen and paper can be helpful in human decision-making, too).

The automatic recognition of inter-relationships between arguments allows knowledge-base writers to take best advantage of generalizations. Substances that contain certain toxicophores (e.g., appropriate aromatic nitro compounds) show three biological effects associated with toxicity: they give a positive result in the Ames test, they are mutagenic, and they are carcinogenic. In a database or a simple rule-based system, all three properties could be associated with each such compound or compound type. However, the properties are not unconnected: carcinogenicity arises in consequence of certain mutagenic events; the Ames test is designed as a means of detecting potential mutagenicity. In the StAR knowledge base it is sufficient to write rules that say "mutagens are likely to be carcinogens" and "substances giving positive results in the Ames test are mutagens" and then to associate toxicophores with the capacity to give positive Ames test results. The implications of combining the rules are discovered and reported by the system.

Weighing the pros and cons of arguments implies an ability to compare their relative strengths, which might appear to require resorting to numbers. However, there are levels of certainty which clearly outweigh others, or that differ in strength albeit less clearly. A set of strength descriptors which merely have some well-defined ordinal relationship may well thus suffice in some circumstances. For example, it was suggested above that an argument in *support* of the proposition that substance X would cause cancer in man might be that it contained a well-researched and understood toxicophore. An argument *confirming* β-naphthylamine to cause cancer in man would be that there are well-documented cases of its having done precisely that. So, in LA it can be reasoned that a confirming argument outweighs one which merely gives support. More usefully, confirmation can be defined to over-rule opposition, for example, and disconfirmation to over-rule support. Small sets of terms and definitions for strength of argument that can be reliably manipulated in this way have been proposed.[16,17] In principle they allow a good deal of evaluation to take place in the absence of numerical data.

Automated evaluations are only useful in practice if users can understand how decisions have been reached, and this implies agreement among system developers, knowledge-base creators, domain experts, and lay users on the meanings of the uncertainty terms used. Would everyone agree on the differences in meanings of "improbable", "not probable", and "probably false"? As part of the StAR project, Ayton and Pascoe at City University are conducting psychological experiments to determine the most appropriate uncertainty terms to avoid ambiguities as well as exploring human perception of the relationships between arguments such as those discussed above.[18,19]

5. POTENTIAL FURTHER APPLICATIONS OF LA IN CHEMISTRY

One current aim of the StAR project is to develop a chemical-structure based risk assessment system, and the technology is also to be applied in a non-chemical domain. However, there appear to be applications for the use of LA in synthesis planning, selection of the best examples from database searches, and perhaps even fields such as similarity clustering. The authors invite readers to consider these possibilities.

For example, in the case of synthesis planning, rather than arguing that a synthetic route should be kept on the synthesis tree if it has a numerical rating greater than some cut-off value (or relative to other routes), it would be possible to use arguments such as

- utility of a route is confirmed if the utility of every step in it is confirmed;
- utility of a route is plausible if every step in it is at least plausible;
- utility of a route is opposed if multiple protection/deprotection steps are needed;
- the need for multiple protection/deprotection steps can be circumvented if a single protecting group can be found that is suitable for every step in the sequence;
- etc.

Arguments relating to individual reaction steps might be

- utility of a reaction is confirmed if it has been used successfully to convert the current precursor into the current product;
- utility of a reaction is supported if typical yield is good;
- utility of a reaction is supported if the reaction is reliable;
- utility of a reaction is opposed if it is not reliable;
- utility of a reaction is disconfirmed if it is known to give a product with the wrong stereochemistry for the required target;
- etc.

6. CONCLUSION

LA is being used in the StAR project for a computer system for carcinogenic risk assessment, in order to allow decision-making under uncertainty (i.e., where numerical data is unavailable or unreliable). Research is also proceeding within the project into human perception of uncertainty and the words used to describe it, with a view to developing a standard set of terms.

The capacity of LA to reason both in the absence of and with trustworthy numerical data makes it potentially suitable for a range of computational problems in chemistry. The authors would welcome opportunities to collaborate with other researchers.

REFERENCES AND NOTES

(1) StAR (Standardised Argument Report) is a collaborative project between Imperial Cancer Research Fund, LHASA UK Ltd., Logic Programming Associates Ltd., and City University, supported by the Department of Trade and Industry and the Engineering and Physical Sciences Research Council. For further information, contact the StAR project co-ordinator, Philip Judson, at the address given at the head of this paper.

(2) Hansch, C. A quantitative approach to biochemical structure-activity relationships. *Acc. Chem. Res.* **1969**, *2*, 232−239.

(3) HazardExpert is supplied by CompuDrug Chemistry Ltd, H-1136 Budapest, Fürst Sándor utca 5., Hungary.

(4) Hendrickson, J. B.; Toczko, A. G. Synthesis design logic and the SYNGEN (synthesis generation) program. *Pure Appl. Chem.* **1988**, *60*(11), 1563−1572.
(5) Hendrickson, J. B. Systematic synthesis design. 6. Yield analysis and convergency. *J. Am. Chem. Soc.* **1977**, *99*(16), 5439−5450.
(6) Gasteiger, J.; Ihlenfeldt, W. D.; Röse, P. A collection of computer methods for synthesis design and reaction prediction. *Recl. Trav. Chim. Pays-Bas* **1972**, *111*, 270−290.
(7) Johnson, A. P.; Marshall, C. Starting material oriented retrosynthetic analysis in the LHASA program. 3. Heuristic estimation of synthetic proximity. *J. Chem. Inf. Comput. Sci.* **1992**, *32*(5), 426−429.
(8) Johnson, A. P.; Marshall, C.; Judson, P. N. Starting material oriented retrosynthetic analysis in the LHASA program. 1. General description. *J. Chem. Inf. Comput. Sci.* **1992**, *32*(5), 411−417.
(9) Corey, E. J.; Long, A. K.; Lotto, G. I.; Rubenstein, S. D. Computer-assisted synthetic analysis. Quantitative assessment of transform utilities. *Recl. Trav. Chem. Pays-Bas* **1992**, *111*, 304−309.
(10) Unpublished research at LHASA UK, School of Chemistry, University of Leeds, LS2 9JT, UK.
(11) Moll, R. Context description in synthesis planning. *J. Chem. Inf. Comput. Sci.* **1994**, *34*, 117−119.
(12) Jorgensen, W. L.; Laird, E. R.; Gushurst, A. J.; Fleischer, J. M.; Gothe, S. A.; Helson, H. E.; Paderes, G. D.; Sinclair, S. *Pure Appl. Chem.* **1990**, *62*, 1921.
(13) Takahashi, M.; Dogane, I.; Yoshida, M. Y.; Yamachika, H.; Takabatake, T.; Bersohn, M. The performance of a noninteractive synthesis program. *J. Chem. Inf. Comput. Sci.* **1990**, *30*, 436.
(14) Curley, S. P.; Browne, G. J.; Smith, G. F.; Benson, G. P. Arguments in the practical reasoning underlying constructed probability responses. *J. Behavioural Decision Making* **1995**, *8*, 1−20.
(15) Fox, J.; Krause, P. J.; Ambler, S. J. Arguments, contradictions, and practical reasoning. In Proceedings of ECAI '92; John Wiley and Sons: pp 623−627.
(16) Krause, P.; Fox, J.; Judson, P. Is there a role for qualitative risk assessment? To be published in Proceedings of the 11th Conference on Uncertainty in Artificial Intelligence, Montreal, Quebec, August 1995. Morgan Kaufman: San Mateo, Ca.
(17) Krause, P.; Fox, J.; Judson, P. An argumentation-based approach to risk assessment. *IMA J. Maths. Appl. Business. Ind.* **1993**/**4**, *5*, 249−263.
(18) Ayton, P.; Pascoe, E. Bias in human judgement under uncertainty? *Knowledge Eng. Rev.* **1995**, *10*, 21−41.
(19) Fox, J.; Hardman, D.; Krause, P.; Ayton, P.; Judson, P. Risk assessment and communication: a cognitive engineering approach. In Applications and Innovations in Expert Systems. III; Macintosh, A., Cooper, C., Eds.; Cambridge University Press.

CI9500905