

Dendral and Meta-Dendral - The Myth and the Reality

Abstract

At a time when research in Artificial Intelligence (AI) was concerned with general mechanisms for reasoning, the applied Heuristic Dendral project focussed on the need to represent and manipulate knowledge. Heuristic Dendral provided an empirical demonstration of an approach to problem solving that uses detailed domain-specific knowledge to guide a relatively weak inference system. Further, Dendral showed how complex real-world knowledge could be represented in terms of discrete pattern→action rules that are processed by a rule-interpreter. This focus on knowledge became a model for much subsequent work in applied AI, and lead indirectly to the entire Knowledge-Based Systems industry of the '80s.

Further work within the Dendral project has shown that search methods can be used as a basis for programs that attempt theory formation and learning tasks. Numerous technical problems relating to the computer-compatible representation of chemical structure, particularly stereochemical structure, have also been solved in the course of the Dendral project. Algorithms have been implemented that can identify all possible structures consistent with known chemical constraints; systems have been demonstrated for the prediction and interpretation of various types of chemical spectral data.

Despite its successes, and its importance as a model for applied AI, the Dendral project never succeeded in its original aim of providing an effective problem solving tool that could be utilized routinely by organic chemists, nor did the theory-formation programs ever derive new knowledge. Like so many other AI programs, Dendral has succeeded more in the minds of the believers than in the practice of the laboratory.

Contents

1. The Dendral Project
2. Origins
3. Heuristic Dendral
4. Diversification
5. Dendral
6. MetaDendral
7. Lessons from the Dendral Project

7. Lessons from the Dendral Project

The achievements of AI and in particular of Expert Systems are mainly presented through secondary literature – reviews, books, product presentations and so forth. These presentations are normally designed to popularise some technique, or to encourage the wider use of particular approaches to problem solving, or to sell a product. They use data on previous AI systems to provide supporting evidence for their claims. The Dendral project has figured prominently in such reports – and, for the most part, is presented in a totally inappropriate way.

Even in those Dendral/MetaDendral programs that did utilize rules for spectral interpretation and prediction, the AI-related code for creating and using rules represented only part (often only a small part) of the total system code. Most of the code of these systems was involved in the user interface, in structure and substructure editors, in specialized file system, in graph-matcher and canonicalization routines, and so forth. It is likely that the rule creation/interpretation code will only represent a small part of the total code needed for problem solving in most complex domains. Systems as elaborate as the MetaDendral programs simply cannot be built from standard Expert System shells.

The Dendral programs of greatest practical value to chemists are the basic structure generators Congen and Genoa, and the Stereo module. Even these programs are of relatively limited use – they simply don't address a problem that arises with sufficient frequency as to really justify the use of computer aids. The chemical structural problems that do arise frequently (and that could utilise computer aids) are of the form "*What is the nature of this pollutant?*" – in these problems a rough classification of the compound (or mixture) as polynuclear aromatic, organo-phosphorous, chlorinated biphenyl etc is required, the data available to the classification procedure concern the origin of the sample and the results of a few quick physico-chemical tests. These problems are insufficiently constrained for use of structure generators like Congen, and in any case there is no interest in the specific structural form of the pollutant.

Complete structure elucidation problems, such as can be handled by Congen, are relatively infrequent. Most of the time of the typical individual graduate student in a natural products laboratory is spent on the isolation and re-identification of previously known compounds; the three or four new compounds encountered each year represent the main results of that individual's work, and the elucidation of their structure is the only intellectually stimulating

part of the entire exercise. The natural products researcher is unlikely to hand over the only "fun" part of the work to a computer program, particularly as in most cases biochemical constraints on skeletal form are such as to make the structural problem readily soluble with no need to exhaustively consider hundreds of potential candidates. Further, structure elucidation ultimately requires proof – proof either through X-ray crystallography or by means of one of the newer skeleton-tracing nmr techniques. Identifications based on Dendral-like procedures of candidate generation and ranking do not constitute proof of correct structure elucidation and so do not constitute complete solutions to a structure elucidation problem.

Although available for many years through the networked Sumex-Aim facility, Congen was never widely used (when an attempt was made to discover its usage, only about eight applications citing its use could be traced). The Genoa/Stereo program was made available through the company Molecular Design[40], but it represented only a minor and not very successful product line added to the repertoire of an already existing company. Claims of widespread use of the Dendral programs are simply without foundation. In any case, the programs Congen, Genoa, and Stereo have essentially no AI content and even if they had been widely used such usage would not constitute evidence for the successful application of AI techniques. (If Dendral is "one of the most successful AI programs ever" then god help the unsuccessful AI programs.)

Other misrepresentations of Dendral indicate how reports of the actual achievements of programs require much more careful phrasing. For example, original reports have compared the performance of graduate students and of the Heuristic Dendral program at the task of identifying a structure solely from composition and mass spectral data[10], and the correct results obtained from the structure generation algorithms of StrGen have been compared with the less successful attempts of post-doctoral students assigned similar structure-generation problems[21]. However these are not realistic tasks. A chemist normally interprets spectra in the context of considerable additional information concerning the origin and isolation of a compound. Structure generation problems, as actually encountered by chemists, involve small numbers of distinct multi-atom fragments and involve quite different conceptual problems from those associated with the generation of structures from large numbers of identical atoms. The lack of realism of these tasks is not stressed in the original reports – because these reports presume substantial shared background knowledge. Those familiar with the domain understand that the tasks reported represent only a small part of the total

that's his definition of AI
There's plenty of expertise in

process of structure elucidation, and that the examples chosen have been selected to give an easily comprehended measure of the scope of problems that can be tackled by the programs. However, those who do not share this presumed background knowledge tend to take these measures of performance quite out of context – and so the Dendral programs come to be reported as having greater expertise than post-doctoral scientists.

The MetaDendral systems had little potential for real application. In principle, the Planner system could be used to help identify structures that had some novel pattern of substituents on a standard skeleton. However, Planner was limited by its requirement that the fragmentation processes of a skeleton would not be substantially altered by the presence of the substituents – while this requirement was met in the case of the simply substituted estrogens, it is not a requirement that would generally be satisfied. In any case, such structural problems are much more readily solved through the use of other spectral techniques. The RuleGen/RuleMod analysis of mass spectral data could not generate rules that would be of significant value in structure elucidation. The spectral \leftrightarrow structural correlations that these programs could identify were for compounds in well defined chemical classes and so were irrelevant for general structure elucidation problems. Further because the selection method picked those processes for which there was general evidence among the available reference data on example compounds, the rules were usually of low discriminatory power. These rules could not really be used as the basis of ranking candidates even when one did get a structure elucidation problem involving compounds in the class for which the rules had been developed.

Claims that "Dendral surpasses all humans at its task and, as a consequence, has caused a redefinition of the roles of humans and machines in chemical research" are made in ignorance. Such claims attempt to establish a myth – a myth that may in the short term help to sell some techniques or advance some careers. Such myths only obscure the real achievements of the Dendral project.

The Dendral project has real achievements. Many are esoteric. Dendral led to advances in combinatorial mathematics, particularly graph labelling techniques[20,21,25]. Lederberg's Vertex Graph method for describing molecular structures, combined with methods for embedding complex subgraphs within other graphs led first to systems that could help enumerate isomers and then to practical structure generation systems. Although other methods are now used to establish unique identifiers for chemical structures[26], Lederberg's scheme for classifying structures is still of some current research interest.

Brown, Carhart, and Nourse invented and refined many algorithms for the representation of canonical forms for structures, for subgraph matching, for analysing symmetry, and for handling stereochemistry. Few of these algorithms have been made as accessible to other researchers as one might have wished; however, these algorithms continue to be used and developed by specialists in companies such as Molecular Design[40].

Possibly, the real achievement of Dendral was to help change the way people thought about computers and problem solving. The success of Heuristic Dendral, and its stepchild Mycin, encouraged people to look beyond numerical computation and data processing tasks. Dendral and Mycin showed how complex real-world classification and diagnostic problems could be tackled provided that adequate domain-specific knowledge was given to a program that incorporated even a relatively simple inference procedure.

References

1. R.K. Lindsay, B.G. Buchanan, E.A. Feigenbaum, and J. Lederberg, *Applications of artificial intelligence for organic chemistry: The DENDRAL project*, McGraw-Hill, New York, 1980.
2. N.A.B. Gray, *Computer assisted structure elucidation*, Wiley-Interscience, New York, 1986.
3. E.J. Corey, A.K. Long, and S.D. Rubenstein, Computer-Assisted Analysis in Organic Synthesis, *Science*, 228 (1985) 408-418.
4. A.P. Johnson, Computer Aids to Synthesis Planning, *Chemistry in Britain*, 1985 (January) 59-67.
5. W.T. Wipke, G.I. Ouchi, and S. Krishnan, Simulation and Evaluation of Chemical Synthesis – SECS: An Application of Artificial Intelligence Techniques, *J. Artificial Intelligence*, 11 (1978) 173-193.
6. J. Lederberg, G.L. Sutherland, B.G. Buchanan, E.A. Feigenbaum, A.V. Robertson, A.M. Duffield, and C. Djerassi., Applications of artificial intelligence for chemical inference. I. The number of possible organic compounds: acyclic structures containing C, H, O, and N., *J. Am. Chem. Soc.*, 91 (1969) 2973-2976.
7. H. Budzikiewicz, C. Djerassi, and D.H. Williams, *Structure elucidation of natural products by Mass Spectrometry*, Holden-Day, San Francisco, 1964.
8. B.G. Buchanan, G.L. Sutherland, and E.A. Feigenbaum, Heuristic Dendral: A Program for Generating Explanatory Hypotheses in Organic Chemistry, in B. Meltzer and D. Michie (Editors), *Machine Intelligence*, 4, Edinburgh University Press, Edinburgh, 1969, pp. 209-254.

9. A.M. Duffield, A.V. Robertson, C. Djerassi, B.G. Buchanan, G.L. Sutherland, E.A. Feigenbaum, and J. Lederberg, Applications of artificial intelligence for chemical inference. II. Interpretation of low resolution mass spectra of ketones. *J. Am. Chem. Soc.*, 91 (1969) 2977-2981.
10. B.G. Buchanan, G.L. Sutherland, and E.A. Feigenbaum, Rediscovery some Problems of AI in the Context of Organic Chemistry", in B. Meltzer and D. Michie (Editors), *Machine Intelligence*, 5, Edinburgh University Press, Edinburgh, 1969, pp. 253-280.
11. G. Schroll, A.M. Duffield, C. Djerassi, B.G. Buchanan, G.L. Sutherland, E.A. Feigenbaum, and J. Lederberg, Applications of artificial intelligence for chemical inference. III. Aliphatic ethers diagnosed by their low resolution mass spectra and NMR data, *J. Am. Chem. Soc.*, 91 (1969) 7440-7445.
12. A. Buchs, A.M. Duffield, G. Schroll, C. Djerassi, A.B. Delfino, B.G. Buchanan, G.L. Sutherland, E.A. Feigenbaum, and J. Lederberg, Applications of artificial intelligence for chemical inference. IV. Saturated amines diagnosed by their low resolution mass spectra and NMR data, *J. Am. Chem. Soc.*, 92 (1970) 6831-6838.
13. D.A. Waterman, Machine learning of heuristics, Ph.D. thesis, Department of Computer Science, Stanford University, 1968.
14. A. Newell and H.A. Simon, *Human Problem Solving*, Prentice-Hall, Englewood Cliffs, 1972.
15. A. Buchs, A.B. Delfino, A.M. Duffield, C. Djerassi, B.G. Buchanan, E.A. Feigenbaum, and J. Lederberg, Applications of artificial intelligence for chemical inference. VI. Approach to a general method of interpreting low resolution mass spectra with a computer, *Helvetica Chimica Acta*, 53 (1970) 1394.
16. B.G. Buchanan, E.A. Feigenbaum, and J. Lederberg, A heuristic programming study of theory formation in science, in *Proceedings of the 2nd International Joint Conference on Artificial Intelligence*, September 1971, W. Kaufmann Inc, Los Altos, pp.40-50.
17. B.G. Buchanan and E.H. Shortliffe (Editors), *Rule Based Expert Systems: The Mycin Experiments of the Stanford Heuristic Programming Project*, Addison-Wesley, 1984.
18. R.E. Carhart, D.H. Smith, H. Brown, and N.S. Sridharan, Applications of artificial intelligence for chemical inference. XVI. Computer generation of vertex graphs and ring systems, *J. Chem. Inf. Comp. Sci.*, 15 (1975) 124-130.
19. H. Brown, L. Masinter, and L. Hjelmeland, Constructive graph labeling using double cosets, *Discrete Mathematics*, 7 (1974) 1-30.
20. H. Brown and L. Masinter, An algorithm for the construction of the graphs of organic molecules, *Discrete Mathematics*, 8 (1974), 227-244.
21. L. Masinter, N.S. Sridharan, R.E. Carhart, and D.H. Smith, Applications of artificial intelligence for chemical inference. XII. Exhaustive generation of cyclic and acyclic isomers, *J. Am. Chem. Soc.*, 96 (1974) 7702-7714.

22. L. Masinter, N.S. Sridharan, R.E. Carhart, and D.H. Smith, Applications of artificial intelligence for chemical inference. XIII. Labeling of objects having symmetry, *J. Am. Chem. Soc.*, 96 (1974) 7714-7723.
23. D.H. Smith, Applications of artificial intelligence for chemical inference. XV. Constructive graph labeling applied to chemical problems. Chlorinated hydrocarbons. *Analytical Chemistry*, 47 (1975) 1176-1179.
24. D.H. Smith, Applications of artificial intelligence for chemical inference. XVII. The scope of structural isomerism, *J. Chem. Inf. Comp. Sci.*, 15 (1975) 203-XXX.
25. H. Brown, Molecular Structure Elucidation. 3., *SIAM J. Appl. Math.*, 32 (1977) 534-551.
26. H.L. Morgan, The generation of a unique machine descriptor for chemical structure – A technique developed at Chemical Abstracts Service, *J. Chem. Doc.*, 5 (1965) 107-113.
27. R.E. Carhart, D.H. Smith, H. Brown, and C. Djerassi, Applications of artificial intelligence for chemical inference. XVII. An approach to computer-assisted elucidation of molecular structure, *J. Am. Chem. Soc.*, 97 (1975) 5755-5762.
28. A. Lavanchy, T. Varkony, D.H. Smith, W.C. White, R.E. Carhart, B.G. Buchanan, and C. Djerassi, Rule-based mass spectrum prediction and ranking. Applications to structure elucidation of novel marine sterols, *Org. Mass. Spectrom.*, 15 (1980) 355-366.
29. N.A.B. Gray, R.E. Carhart, A. Lavanchy, D.H. Smith, T. Varkony, B.G. Buchanan, W.C. White, and L. Creary, Computerized mass spectrum prediction and ranking, *Anal. Chem.*, 52 (1980) 1095-1102.
30. N.A.B. Gray, C.W. Crandell, J.G. Nourse, D.H. Smith, M.L. Dageford, and C. Djerassi, Applications of artificial intelligence for chemical inference. XXXIV. Computer-assisted structural interpretation of C-13 spectral data, *J. Org. Chem.*, 46 (1981) 703-715.
31. Y. Kudo and S.I. Sasaki, Principle for exhaustive enumeration of unique structures consistent with structural information, *J. Chem. Inf. Comp. Sci.*, 16 (1976) 43-49.
32. J.G. Nourse, D.H. Smith, R.E. Carhart, and C. Djerassi, Computer assisted elucidation of molecular structure with stereochemistry, *J. Am. Chem. Soc.*, 102 (1980) 6289-6295.
33. R.E. Carhart, D.H. Smith, N.A.B. Gray, J.G. Nourse, and C. Djerassi, GENOA: A computer program for structure elucidation utilizing overlapping and alternative substructures, *J. Org. Chem.*, 46 (1981) 1708-1718.
34. D.H. Smith, B.G. Buchanan, R.S. Engelmores, A.M. Duffield, A. Yeo, E.A. Feigenbaum, J. Lederberg, and C. Djerassi, Applications of artificial intelligence for chemical inference. VIII. An approach to the computer interpretation of the high resolution mass spectra of complex molecules. Structure elucidation of estrogenic steroids, *J. Am. Chem. Soc.*, 94 (1972) 5962-5973.

35. D.H. Smith, B.G. Buchanan, W.C. White, E.A. Feigenbaum, C. Djerassi, and J. Lederberg, Applications of artificial intelligence for chemical inference. X. Intsum: a data interpretation program as applied to the collected mass spectra of estrogenic steroids, *Tetrahedron*, 29 (1973) 3117-3134.
36. B.G. Buchanan, D.H. Smith, W.C. White, R. Gritter, E.A. Feigenbaum, J. Lederberg, and C. Djerassi, Applications of artificial intelligence for chemical inference. XXII. Automatic rule formation in mass spectrometry by means of the MetaDendral program, *J. Am. Chem. Soc.*, 96 (1976) 6168-6178.
37. T.M. Mitchell and G.M. Schwenzer, Applications of artificial intelligence for chemical inference. XXV. A computer program for automated empirical ¹³C nmr rule formation, *Org. Magn. Reson.*, 11 (1978) 378-384.
38. J.E. Gurst and A.K. Schrock, Mass spectra of Androstane-7,17-diones, *J. Org. Chem.*, 45 (1980) 4062.
39. N.A.B. Gray, Applications of artificial intelligence for organic chemistry. Analysis of C-13 spectra, *J. Artificial Intelligence*, 22 (1984) 1-21.
40. Molecular Design Ltd, 2132 Farallon Drive, San Leandro, CA94577.