ELSEVIER

Artificial Intelligence 74 (1995) 191-201

Artificial Intelligence

Research Note

Machine discovery in chemistry: new results

Raúl E. Valdés-Pérez *

Computer Science Department and Center for Light Microscope Imaging and Biotechnology, Carnegie Mellon University, Pittsburgh, PA 15213, USA

Received June 1994; revised September 1994

Abstract

Earlier we proposed an idea for conjecturing unseen entities in science, and described its application within MECHEM to the chemistry task of inferring the mechanism of a chemical reaction based on experimental evidence. However, the program was a prototype, and lacked several capabilities that rendered it incompetent on current science.

We now describe extensions that enable reasoning about the molecular structural transformations that are the focus of modern chemistry. We also report successful applications of MECHEM to chemical problems of current interest, and point out subsequent machine discovery work that the MECHEM project has strongly influenced. These new results demonstrate the efficacy and generality of the original idea for machine discovery, and vindicate the research strategy of emphasizing specific task competence and deferring concerns with generality.

1. Introduction

A previous paper [1] in this journal proposed an idea for conjecturing unseen entities in science, and described its application within the MECHEM program to the chemistry task of inferring the mechanism of a chemical reaction based on experimental evidence [26]. These earlier reports were not yet convincing demonstrations of MECHEM's scientific competence for two reasons: first, the program reasoned about molecular formulas, but lacked the crucial ability to reason about molecular structures, and second, no specific evidence of incorporation within the practice of chemistry was yet available.

The aim of this note is to describe the extension of MECHEM to reason about the structural transformations that chemically-reacting molecules undergo [23,25], and to report concrete applications of the program to chemical problems of current interest [19,

* E-mail: valdes@cs.cmu.edu.

[1] Submitted in August 1991, hence reflecting the state at that time.

SSDI 0004-3702(94)00073-5

28]. We submit these new results as convincing evidence for the efficacy of the original idea for conjecturing hidden entities. We also sketch the probable future evolution of the program within chemical practice, and lastly point out some connections to subsequent machine discovery work that MECHEM has influenced.

2. Development of MECHEM

The original motivation for MECHEM was to select some challenging discovery task from current scientific practice and to demonstrate its automation by a logical analysis of hypothesis generation and of the relation between evidence and hypothesis [20]. We began by searching within chemistry for circumstantial reasons, and then settled on the task of elucidating the multi-step character (or *mechanism*) of a chemical reaction on the basis of experimental evidence (the reader may wish to glance ahead to Fig. 4 for an idea of what a reaction mechanism looks like, but should ignore for now the angle-bracket notation in the reaction steps). Elucidating reaction mechanisms has been a nearly universal task of experimental chemists since it was first proposed by van 't Hoff in the late 1800s that many chemical reactions do not occur as a single act, but instead involve a number of consecutive or parallel steps. The fact that the mechanism elucidation task is of a long scientific tradition adds an extra interest to its successful automation, since it enhances the credibility of the overall machine discovery enterprise.

The task of mechanism elucidation had been first addressed within AI by the Ph.D. Thesis of V.W. Soo at Rutgers [15, 16], although his work focused on enzymatic reactions. One important drawback to that work was its dependence on a small catalogue of candidate mechanisms, which were discriminated by applying known experiment-analysis rules from enzymology. Outside of AI, chemists and engineers have also addressed task automation. Typically, chemical engineering work has assumed complete knowledge of all reaction intermediates and products, an untenable assumption in practice. Virtually all the chemistry work has followed a schema based on searching a space of chemical-reaction operators (e.g., [31]); we will not contrast that work here, but this difference is discussed in the chemical papers cited throughout this note.

A main obstacle to formulating mechanistic hypotheses competently, here and in many other scientific tasks, is finding some means to conjecture unseen entities, e.g., unseen reaction intermediates and products. Typically, one knows the starting materials of a reaction and has identified some of the products or even intermediates, but others remain undetected because of practical limitations of experimental technique.

The obstacle of conjecturing unseen entities is overcome in MECHEM by a simple and seemingly naïve method: conjecture "wild cards" such as X, Y, Z, etc., use these wild cards together with the seen entities to formulate hypotheses, and then use the domain laws of a science to constrain these variables sufficiently (within the context of a specific hypothesis) to entail a small set of possible identities for the variables. For example, using the conservation constraint of reaction balance, the unknown X in the following single-step hypothesis

$$CH_3 + MCH_2OOH \rightarrow X + CH_3OH$$

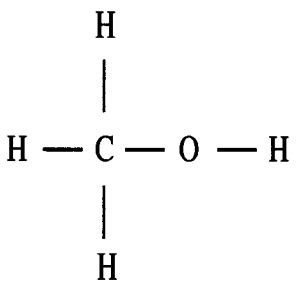

Fig. 1. The molecular structure of methanol.

is inferred to consist of 1 *M*, 1 carbon, 2 hydrogen, and 1 oxygen atoms. In more complex cases involving multiple steps and unknowns, and possibly more than one unknown per step, a generalized linear equation solver is used to infer the chemical composition of the wild cards.

The previous idea was combined with a number of other methods in order to implement a working prototype. For example, a constrained-generation algorithm [21,22] generated hypotheses non-redundantly under a bias for simplicity (fewer reaction steps and conjectured entities), which was needed if a systematic search was to have any hope against the powerful combinatorial increase in the hypothesis space with increments in steps and conjectured entities. MECHEM's canonical generator of mechanisms evokes the earlier DENDRAL [12] generator CONGEN [11], although the issue of simplicity did not arise there in the same way, since CONGEN was not required to conjecture unseen entities (also, CONGEN generated structures, not mechanisms). Returning to the MECHEM prototype, there were also a number of other program components that tested hypotheses against various given experimental evidence, such as overall stoichiometry.

The above was the state of MECHEM as described in an earlier paper in this journal [26]. Enough machinery was in place to enable the program to systematically find, from historical data, the simplest reaction pathways[2] for urea synthesis in biochemistry, whose discovery in 1932 by Hans Krebs [7] had been modeled by Kulkarni and Simon in their KEKADA program [10].

However, despite the novelty and promise of MECHEM, as demonstrated on the urea pathway, the program could not reason *structurally* about chemical substances and reactions. That is, the graph-like nature of chemical molecules was ignored, since substances were represented as molecular formulas (simple vectors). For example, methanol (whose molecular structure is depicted in Fig. 1) was represented as consisting of 1 carbon, 4 hydrogens, and 1 oxygen, which obscured the known structural connectivity among these atoms. Although Krebs's (and MECHEM's) discovery of the urea pathway did not involve much structural reasoning (historically speaking [6]), the graph-like structure of molecules is nevertheless at the heart of modern chemistry.

[2] For the present purposes, a reaction "pathway" and "mechanism" will be used synonymously. They have somewhat different connotations in chemical practice.

This gap in MECHEM was problematic for two reasons. Firstly, it suggested that the program's basic hypothesis-formation method of conjecturing hidden entities was naïve and not really up to the task of current science, which usually involves reasoning more complex than the simple balancing of accounts. Secondly, the program would have no impact on practicing experimentalists, since it would report mechanistic hypotheses that were obviously implausible for structural reasons that the program did not know about, hence an experimentalist would quickly lose interest (we speak from experience). Our goal for MECHEM had become to complete the transition from theory to practical impact, hence the problem.

3. Recent advances

One aspect of the gap was relatively easy to close: we designed a heuristic graph algorithm to test whether a given, single reaction step *reactants* → *products* was structurally plausible [23]. The criterion of plausibility is that at most a small number N of bonds could be broken or formed[3] during the conversion of the reactants into products; N is adjustable, but is set to 3 by default, which covers almost all elementary chemical reactions. The algorithm gains critical efficiency by amortizing the cost of matching structural fragments through the use of tables that store all the fragments derivable from a given molecule. With this new algorithm, if the reactants and products of a step are known, then MECHEM never reports that step if it is structurally implausible.

However, this graph-algorithmic test could not be applied to any step that contained a wild card, even if a formula has already been inferred for it, since the structural information is missing from wild cards. There seemed little chance of overcoming this last problem, hence it appeared that MECHEM would remain an AI research program that fails to have an impact outside of AI. Finally, we realized that it must be possible to *infer* the molecular structure of any wild cards, given their already-inferred formulas, and given the overall multi-step context in which they appeared. For example, Fig. 2 shows a nine-step mechanism in which there appear the six wild cards $\mathcal{U}$, $\mathcal{V}$, $\mathcal{W}$, $\mathcal{X}$, $\mathcal{Y}$, and $\mathcal{Z}$ (M is not a wild card, but a catalyst reaction site). Later, Fig. 4 shows the same mechanism but with the wild cards replaced by the molecular structures that are inferred for them.

In short, the same general idea for conjecturing hidden entities could work for structures as well as formulas, although the algorithm in the newer case would be considerably more complex.[4] The details of this algorithm are given elsewhere [25]; here we merely state its feasibility and remark that the algorithm relies on a case-by-case breakdown of the various schematic ways that structures and formulas can appear within a single reaction step, together with the previous assumption of at most N bond changes per step.

[3] Actually, it is not precisely the number of bonds, but the number of changes to connectivity. That is, currently MECHEM treats breaking a double bond the same as breaking a single bond.

[4] We conjecture that if steric, or three-dimensional reasoning becomes necessary to reach competence on certain application areas of MECHEM, it should similarly be feasible to infer the possible 3D configurations of molecular structures based on the multi-step context of their formation.

1.	$H_2 + M_2 \rightarrow 2\mathcal{U}$
2.	$M_2 +$ ethane $\rightarrow \mathcal{U} + \mathcal{V}$
3.	$M_2 + \mathcal{V} \rightarrow CH_2M\text{-}CH_2M + \mathcal{U}$
4.	$M + CH_2M\text{-}CH_2M \rightarrow \mathcal{U} + \mathcal{W}$
5.	$2\mathcal{W} \rightarrow CH_2M\text{-}CH_2M +$ CHM-CHM
6.	CHM-CHM $\rightarrow 2\mathcal{X}$
7.	$\mathcal{U} + \mathcal{X} \rightarrow M + \mathcal{Y}$
8.	$2\mathcal{Y} \rightarrow \mathcal{X} + \mathcal{Z}$
9.	$\mathcal{U} + \mathcal{Z} \rightarrow M_2 +$ methane

Fig. 2. An example of six conjectured wild cards.

Together, these two new algorithms ensured that MECHEM would no longer report reaction mechanisms that contained obviously wrong steps. A final problem was the combinatorial increase in run time due to increments in the number of wild cards that were conjectured. Under MECHEM's systematic search regimen, the program could not handle any problem for which more than four unseen entities had to be conjectured. Hence, the program was limited to relatively easy problems, thus narrowing its scope and potential impact. This final problem was solved by the invention of a divide-and-conquer heuristic [27] that partitions the given products and intermediates into two or more sets according to various chemical criteria. This heuristic, together with others of more modest power reported in the cited paper, enlarged tremendously the class of practical chemical problems to which MECHEM could be applied. We now describe some members of that class.

4. Applications

As already mentioned, one of our goals is to insert MECHEM within chemical practice as a highly-capable creative aid for experimentalists. Since this project has not developed in collaboration with a chemist, much less an experimentalist concerned with

1. The overall stoichiometry is 1 (ethane) + 1 (H_2) → 2 (methane).
2. A catalyst reaction site (modeled as M) forms one bond.
3. Two catalyst reaction sites are modeled as M_2 (this notation does not by itself imply a bond across the two sites).
4. $CH_2M\text{-}CH_2M$ and CHM-CHM are required intermediates (these do form "bridges" over two reaction sites on the catalyst).
5. H_2 is not a product of a step (except as the reverse of initial dissociation).
6. Every reaction intermediate is adsorbed on the catalyst, i.e., every intermediate contains M.
7. No species contains three carbons nor spans three catalyst reaction sites.
8. There is a maximum of three bond changes (cleavage or formation) per step.
9. $CH_2M\text{-}CH_2M$ is a precursor (not necessarily single-step) of CHM-CHM.

Fig. 3. Constraints on ethane hydrogenolysis reaction.

1. $H_2 + M_2 \rightarrow 2\langle HM\rangle$
2. $M_2 + \text{ethane} \rightarrow \langle HM\rangle + \langle CH_3\text{-}CH_2M\rangle$
3. $M_2 + \langle CH_3\text{-}CH_2M\rangle \rightarrow CH_2M\text{-}CH_2M + \langle HM\rangle$
4. $M + CH_2M\text{-}CH_2M \rightarrow \langle HM\rangle + \langle CH_2M\text{-}CHM\rangle$
5. $2\langle CH_2M\text{-}CHM\rangle \rightarrow CH_2M\text{-}CH_2M + CHM\text{-}CHM$
6. $CHM\text{-}CHM \rightarrow 2\langle CHM\rangle$
7. $\langle HM\rangle + \langle CHM\rangle \rightarrow M + \langle CH_2M\rangle$
8. $2\langle CH_2M\rangle \rightarrow \langle CHM\rangle + \langle CH_3M\rangle$
9. $\langle HM\rangle + \langle CH_3M\rangle \rightarrow M_2 + \text{methane}$

Fig. 4. MECHEM's reaction mechanism for catalytic hydrogenolysis of ethane.

reaction mechanisms, we faced (and still face) the problem of how, as an outsider, to convince that community of the value of the new technique. We have encountered all the familiar hurdles that have been reported by others in similar circumstances of crossing disciplinary lines.

Our recourse has been to vigorously seek out chemistry experts by various means. During a visit to William Jorgensen's laboratory at Yale, he suggested a focus on industrial chemistry as the most promising application area, since chemists (whether academic or not) who study industrially-significant reactions spend the most time gathering the experimental evidence that MECHEM needs in order to credibly elucidate a reaction mechanism. We then chose to focus on catalytic industrial chemistry as an applications area for these reasons:

(1) catalytic pathways are cyclic and complicated, hence difficult for the unaided human mind to deal with;
(2) the molecules are typically small, which alleviates the combinatorial problems occasioned by large molecules; and
(3) the reactions are of keen economic importance.

Understanding reaction mechanisms is of practical significance, because such knowledge can suggest changes to reaction conditions that will enhance the yield of a chemical process.

One of our attempts resulted in the suggestion by a local experimentalist colleague (Edmond Ko) of applying MECHEM to a relatively simple reaction (catalyzed ethane hydrogenolysis) whose reaction mechanism was considered well understood; Professor Ko also helped with formulating the input constraints. At first, MECHEM could not handle this reaction, because it involved conjecturing at least five species (ultimately, six were needed) and the program had never terminated successfully on any problem needing more than four such conjectures. Being frustrated by this setback, and by the suggestion that MECHEM would never do anything practical if it couldn't handle this relatively simple practical problem, we devised the cited divide-and-conquer heuristic [27] that made this reaction, and many others, suddenly within the scope of the program.

Given the input constraints shown in Fig. 3, MECHEM found several simplest mechanisms for ethane hydrogenolysis: one mechanism matched exactly an earlier proposal [3], which was contained in another slightly extended mechanism given recently [13]. The latter authors reviewed the state of knowledge about this reaction as follows:

> Several different mechanisms have been proposed in the literature for ethane hydrogenolysis over supported metals. Although there is disagreement on the detailed nature of individual steps, there is general agreement on the nature and the sequence of these steps. Qualitatively, general aspects of the mechanism can be summarized as follows. Hydrogen is adsorbed on the surface in an atomic form. Ethane is adsorbed dissociatively, undergoing cleavage of a C-H bond. Further dehydrogenation of the C_2H_x species occurs, accompanied by the creation of additional bonds between the C_2H_x species and the metal surface. The C-C bond breaks and CH_y species are produced. Hydrogenation of the CH_y species takes place, followed by the desorption of methane.

Surprisingly, one of MECHEM's mechanisms differed significantly from the others. Fig. 4 shows this mechanism, in which the intermediates conjectured by MECHEM appear within angle brackets to distinguish them from the given substances. We reported this mechanism, together with an analysis of how well it explains various qualitative evidences reported in the chemical literature, to *Catalysis Letters* as a successful example of a human–computer collaboration [28]. In turn, the latter is proposed within that paper as a fruitful new technique in studies of catalytic reaction mechanisms.

We have also applied MECHEM to more complex reactions involving a dozen or more steps, such as the potentially lucrative catalytic conversion of alkanes (e.g., natural gas, or methane) by partial oxidation [4]. We have not yet found any specific mechanism of clear chemical interest, for lack of close involvement with experimentalists who are active in that type of chemistry. However, we have illustrated MECHEM's capabilities on representative input constraints at the 1994 Spring National Meeting of the American Chemical Society, during a symposium on alkane conversion [19].

5. Interactive use of MECHEM

From a chemist's viewpoint, MECHEM is an interactive program. First, the experimentalist/user states the starting materials, what species have been observed experimentally, and other constraints on the reaction. The program then reports the simplest mechanisms, which typically prompt the user to object to various aspects of these, which may lead to rejecting all of the proposed mechanisms. The user then articulates his objections by formulating new constraints, and the program is re-run by adding these constraints to the earlier ones. This interactive process continues until the user is satisfied that none of the reported mechanisms is objectionable, or until the problem becomes too complex for the current program to handle. This interaction lasted two or three cycles on the above ethane reaction, resulting finally in the constraints shown in Fig. 3.

In general, MECHEM will report several plausible reaction mechanisms. Knowledge of these (whether generated by human or machine) can prompt the chemist to try a different catalyst, design an experiment, carry out a kinetic analysis, and so on. However, such steps are outside the current scope of this research, and are left to the expert or to other techniques.

6. MECHEM's task as constraint satisfaction

It is accurate to view MECHEM as carrying out a heuristic breadth-first tree search with complicated node generators and node evaluators. It is also fruitful to view it as a complex constraint-satisfaction program (Simon [14] analyzes the relation between the heuristic search and constraint satisfaction problem-solving metaphors). A major project effort has been to identify new chemical constraints—arising from background theory or as typical experimental evidence—and to design algorithms that test whether a given, partially-built reaction mechanism is consistent with a given constraint.

A possible approach to this and other constraint-satisfaction problems of scientific inference is to draw on parallel work in constraint satisfaction, e.g., constraint logic programming (CLP) [5]. We have previously experimented with the CLP language Prolog III [1] in the context of MECHEM's pathway generator [8], but returned to programming from scratch in Lisp because of the complicated algorithmic nature of the further constraints that were needed to make MECHEM into a competent program. With current constraint-satisfaction tools, it seems awkward to implement the constraint that every individual reaction be realizable in at most N bond changes, not to mention the preliminary step of inferring the molecular structure of wild cards. Another problem with typical CLP-based search engines is that they carry out a depth-first search, whereas in scientific model-building applications a breadth-first search is preferable, since the shallower nodes correspond to simpler models.

Nevertheless, a CLP or generic constraint-satisfaction approach may be promising when addressing scientific tasks that involve mostly simple constraints, or when building demonstration prototypes for tasks that are potentially quite complicated.

7. The future

If there were a theory to predict reliably the exact course of a complex chemical reaction based on starting materials and initial conditions, then the experimentalist's task of elucidating reaction mechanisms on the basis of evidence would be superseded. However, no such theory is yet available, and the study of reaction mechanisms continues to be dominated by experimentation. From our readings of papers in catalytic chemistry, we know of no computer tools that can assist experimenters in devising reaction mechanism hypotheses.

Up to now, MECHEM has been applied only a few times to evidences gathered by others, for reasons already discussed. We will attempt to make chemists aware of the benefits that can accrue from making use of the program early in mechanistic studies, so that the program's outputs can help guide experimental decision-making. We believe that the program can increase the speed with which mechanistic conclusions are reached, as well as improve their accuracy. As evidence for this belief, we recall that the program has already turned up new simple hypotheses on the first reaction (ethane hydrogenolysis) of current interest to which it was applied. Given our observations of the density of plausible hypotheses of equal simplicity, as reflected in MECHEM's outputs, we can predict frequent such occurrences. If these predictions hold true, they will raise the

question of whether complicated reaction mechanisms can be elucidated at all reliably without computerized hypothesis-generation tools.

Our immediate plans are to implement a coarse-grain parallel version of MECHEM, perhaps using PVM [17], but keeping Lisp as the main implementation language. MECHEM's tree search is "embarrassingly parallel" so that a distributed version should be possible with relatively minor changes. In addition, this author is currently serving both as programmer and as "user interface", and the latter role will need to change before MECHEM becomes available for autonomous use by chemists.

8. Subsequent research in scientific discovery

The work on MECHEM has strongly influenced several subsequent results, which one may interpret as evidence for the generality of the ideas. Perhaps more significantly, these results vindicate the research strategy of selecting a specific problem in science and providing an automation of it, while postponing much consideration of generality until a successful automation is near at hand.

During a sabbatical visit by J. Żytkow to Carnegie Mellon, the two of us together with H.A. Simon sought generalizations among several discovery systems, including MECHEM, that had been developed separately. The result was the new concept of search in matrix spaces [30], which expresses the idea that the top-level search space of many scientific model-building tasks can be fruitfully viewed as a matrix-algebraic equation whose unknown entries are filled in subject to a variety of domain constraints.

We have since used this concept [18] to re-design (with some advantages) the GELL-MANN program [2] that postulates quark models in particle physics. Also, concurrently with the work on matrix-space search, we used some of the representational and algorithmic techniques in MECHEM to re-design [24] the BR-3 program [9], which finds phenomenological quantum properties in particle physics. The latter effort resulted in the PAULI program, which carries out simplicity-guided search with a linear optimization in its inner loop. The work on PAULI has led a novel theorem in particle physics [29] which space prevents discussion of here.

9. Conclusion

The original idea for conjecturing hidden entities in chemistry [26] was seemingly naïve in the sense that the modern focus on molecular structure was not accommodated. This note reports the extension of the idea to conjecturing molecular structures, which are in essence topological graphs, hence are more complex objects than the molecular formulas which constituted the representational scope of MECHEM previously. We also report the first convincing evidence that MECHEM has reached competence on a significant class of chemical reactions of current interest. Finally, MECHEM has strongly influenced some of our subsequent research in machine discovery, which is evidence for the general applicability of the ideas underlying the program, and is a vindication of

the research strategy of deferring considerations of generality in favor of specific task competence.

Acknowledgments

This work was supported in part by the W.M. Keck Center for advanced training in computational biology at Carnegie Mellon University, the University of Pittsburgh, and the Pittsburgh Supercomputing Center, and in part by a Science and Technology Center grant from the National Science Foundation, #BIR-8920118.

References

[1] A. Colmerauer, An introduction to Prolog III, *Commun. ACM* **33** (7) (1990) 69–90.
[2] P. Fischer and J.M. Zytkow, Discovering quarks and hidden structure, in: Z. Ras, M. Zemankova, and M. Emrich, eds., *Proceedings Fifth International Symposium on Methodologies for Intelligent Systems* (North-Holland, Amsterdam, 1990) 362–370.
[3] B. Gudkov, L. Guczi, and P. Tetenyi, Kinetics and mechanism of ethane hydrogenolysis on silica-supported platinum and platinum-iron catalysis, *J. Catalysis* **74** (1982) 207–215.
[4] J. Haggin, Direct conversion of methane to fuels, chemicals still intensely sought, *Chem. Engng. News* **70** (17) (1992) 33–35.
[5] P.V. Hentenryck, H. Simonis, and M. Dincbas, Constraint satisfaction using constraint logic programming, *Artif. Intell.* **58** (1992) 113–159.
[6] F.L. Holmes, Hans Krebs and the discovery of the ornithine cycle, in: *Proceedings 63rd Annual Meeting of the Federation of American Societies for Experimental Biology* **39** (1980). Symposium on Aspects of the History of Biochemistry.
[7] F.L. Holmes, *Hans Krebs* (Oxford University Press, New York, 1991).
[8] J. Jourdan and R.E. Valdés-Pérez, Constraint logic programming applied to hypothetical reasoning in chemistry, in: *Logic Programming: Proceedings 1990 North American Conference*, Cambridge, MA (1990) 154–172,
[9] S. Kocabas, Conflict resolution as discovery in particle physics, *Mach. Learn.* **6** (3) (1991) 277–309.
[10] D. Kulkarni and H. Simon, The processes of scientific discovery: the strategy of experimentation, *Cognitive Sci.* **12** (1988) 139–175.
[11] J. Lederberg, Topological mapping of organic molecules, *Proc. Nat. Acad. Sci.* **53** (1) (1965) 134–139.
[12] R. Lindsay, B. Buchanan, E. Feigenbaum and J. Lederberg, DENDRAL: a case study of the first expert system for scientific hypothesis formation, *Artif. Intell.* **61** (2) (1993) 209–261.
[13] D. Rudd and J. Dumesic, Catalyst synthesis by analogy, *Catalysis today* **10** (1991) 147–165.
[14] H.A. Simon, Search and reasoning in problem solving, *Artif. Intell.* **21** (1983) 7–29.
[15] V. Soo, C. Kulikowski, D. Garfinkel and L. Garfinkel, Theory formation in postulating kinetic mechanisms: reasoning with constraints, Technical Report CBM-TR-150, Department of Computer Science, Rutgers University, New Brunswick, NJ (1987).
[16] V. Soo, C. Kulikowski, D. Garfinkel and L. Garfinkel, Theory formation in postulating enzyme kinetic mechanisms: Reasoning with constraints, *Comput. Biomedical Res.* **21** (1988) 381–403.
[17] V. Sunderam, PVM: a framework for parallel distributed computing, *Concurrency: practice and experience* **2** (4) (1990) 315–339.
[18] R.E. Valdés-Pérez, Algebraic representation in scientific model building: discovery of quark structure in physics, submitted for publication.
[19] R.E. Valdés-Pérez, Computer-aided elucidation of reaction mechanisms: application to the partial oxidation of methane, in: *Proceedings 1994 American Chemical Society Symposium on Methane and Alkane Conversion Chemistry* (in press).

[20] R.E. Valdés-Pérez, Machine discovery of chemical reaction pathways, Ph.D. Thesis, Technical Report CMU-CS-90-191, School of Computer Science, Carnegie Mellon University, Pittsburgh, PA (1990).
[21] R.E. Valdés-Pérez, A canonical representation of multistep reactions, *J. Chem. Inf. Comput. Sci.* **31** (4) (1991) 554–556.
[22] R.E. Valdés-Pérez, Algorithm to generate reaction pathways for computer-assisted elucidation, *J. Comput. Chem.* **13** (9) (1992) 1079–1088.
[23] R.E. Valdés-Pérez, Algorithm to test the structural plausibility of a proposed elementary reaction, *J. Comput. Chem.* **14** (12) (1993) 1454–1459.
[24] R.E. Valdés-Pérez, Algebraic reasoning about reactions: discovery of conserved properties in particle physics, *Mach. Learn.* **17** (1) (1994) 47–68.
[25] R.E. Valdés-Pérez, Algorithm to infer the structures of molecular formulas within a reaction pathway, *J. Comput. Chem.* **15** (11) (1994) 1266–1277.
[26] R.E. Valdés-Pérez, Conjecturing hidden entities via simplicity and conservation laws: machine discovery in chemistry, *Artif. Intell.* **65** (2) (1994) 247–280.
[27] R.E. Valdés-Pérez, Heuristics for systematic elucidation of reaction pathways, *J. Chem. Inf. Comput. Sci.* **34** (4) (1994) 976–983.
[28] R.E. Valdés-Pérez, Human/computer interactive elucidation of reaction mechanisms: application to catalyzed hydrogenolysis of ethane, *Catalysis Lett.* **28** (1) (1994) 79–87.
[29] R.E. Valdés-Pérez and M. Erdmann, Systematic induction and parsimony of phenomenological conservation laws, *Comput. Phys. Commun.* **83** (1994) 171–180.
[30] R.E. Valdés-Pérez, J.M. Zytkow and H.A. Simon, Scientific model-building as search in matrix spaces, in: *Proceedings AAAI-93*, Washington, DC (1993) 472–478.
[31] W. Wipke, G. Ouchi and S. Krishnan, Simulation and evaluation of chemical synthesis – SECS: an application of artificial intelligence techniques, *Artif. Intell.* **11** (1978) 173–193.