

# Proposed Methodological Improvement in the Elucidation of Chemical Reaction Mechanisms Based on Chemist–Computer Interaction

Andrew V. Zeigarnik,<sup>†</sup> Raúl E. Valdés-Pérez,\* and Brian S. White

Computer Science Department, Carnegie Mellon University, Pittsburgh, PA 15213; \*valdes@cs.cmu.edu

## Introduction

A recent book by Roald Hoffmann, entitled *The Same and Not the Same*, is a popular exposition of the achievements, methodology, and reasoning in chemistry (1). Hoffmann's Chapter 29, entitled "Mechanism", describes a reaction—the vacuum ultraviolet photolysis of ethane—studied experimentally by Okabe and McNesby and reported in 1961 (2). This reaction serves as a vehicle for Hoffmann's exposition of the study of mechanisms of chemical reactions, which he characterizes on page 144 as a

"...textbook case for the application of the scientific method. You have an observation. You form several alternative hypotheses explaining that observation, and you proceed to eliminate the hypotheses (through experiment or theory, but mainly through experiment), one by one, until you are left with one and that one must be right."

Hoffmann's choice of this study seems justified because of the simplicity of the chemistry and the crispness of the experimental results obtained in the 1961 article. These qualities, combined with the book's methodological commentary, make this reaction suitable for our own purposes as well: to improve on the method accurately characterized by Hoffmann's book. Thus, we use ethane photolysis as a vehicle for our arguments, but without claiming any definitive contribution to the mechanism of the reaction, because the actual mechanism is much more complicated—as becomes clear from the subsequent literature on this reaction (3–5).

Hoffmann discusses three hypotheses for the mechanism of vacuum ultraviolet photolysis of ethane:

1. The one-step mechanism  $\text{CH}_3\text{CH}_3 \rightarrow \text{H}_2 + \text{CH}_2=\text{CH}_2$ .
2. Abstraction of two hydrogen atoms from the same carbon atom to form  $\text{H}_2$ , followed by the isomerization of ethylidene into ethylene:  $\text{CH}_3\text{CH}_3 \rightarrow \text{H}_2 + \text{CH}_3-\text{CH}$  and  $\text{CH}_3-\text{CH} \rightarrow \text{CH}_2=\text{CH}_2$ .
3. A radical-chain reaction in which the initiation step is  $\text{CH}_3\text{CH}_3 \rightarrow \text{H} + \text{CH}_3\text{CH}_2$  and further steps are radical abstraction and possibly rearrangement of ethyl radical into ethylene.

Okabe and McNesby carried out one isotopic labeling experiment in which the starting material was  $\text{CD}_3\text{CH}_3$ , and another that used a mixture of  $\text{CH}_3\text{CH}_3$  and  $\text{CD}_3\text{CD}_3$  (2). According to Hoffmann, the experimental results were unambiguous because in both cases the amount of HD formed is negligible, so hypotheses 1 and 3 above can be eliminated, and

"...therefore mechanism (2) is proven.

"Or is it? Now we come to the workings of the scientific method and the role of human psychology. Of course mechanism (2) is not proven. You only falsify or disprove hypotheses, eliminate mechanisms—you do not prove them. What I am expounding here is one modern view of the philosophy of science, that associated primarily with the name of Karl Popper. Popperians would say that you could grade theories according to how easily they may be falsified; a theory incapable of being falsified or of being tested is not a good theory. You might as well throw it away.

"Let me restate, in colloquial language, what one might say from Popper's point of view about this beautiful experiment of Okabe and McNesby: We have, in the weakness of our minds, written down three and only three hypotheses for how ethane might fragment under ultraviolet irradiation. And in the strength and beauty of our hands and our minds, we have constructed experiments to eliminate two such hypotheses. That does not prove the third one at all. There may be a fourth or a fifth one we just were not clever enough to devise.

"Now, everyone knows that. I know that, the people who did this experiment know that. But these are people who are doing experiments and interpreting them. It is in the nature of people not to want to write wishy-washy conclusions in papers, such as: 'I have disproven A and B. I hope it's C, but maybe it's something else.' No, people want to say, "I have proven C." Scientists want to do something positive."

This lengthy quotation, in which we have underlined five words for emphasis and future reference, summarizes concisely and accurately the methodological status quo in the determination of reaction mechanisms from experimental evidence (cf. 6 for a fuller account).<sup>1</sup>

The remainder of this paper will show that, for ethane photolysis and many other reactions, (i) a team of chemist (people) and computer is clever enough to devise conveniently all simple mechanisms consistent with given evidence; (ii) these simple mechanisms are surprisingly numerous; and (iii) nevertheless, a scientist–computer team can make positive chemical contributions that are appropriately circumspect without being wishy-washy.

## Chemist–Computer Collaboration

The computer program MECHEM is intended largely as an aid for the elucidation of reaction mechanisms. It is generally usable on any reaction, although so far its application has focused on catalysis (heterogeneous and homogeneous). Given the set of reaction starting materials, any observed products and intermediates, and user-defined constraints, the

<sup>†</sup>Permanent address: Laboratory of Chemical Kinetics and Catalysis, Lomonosov Academy of Fine Chemical Technology, Moscow, 11751 Russia; azeigarn@mitht.rssi.ru.

program searches comprehensively for all simplest reaction mechanisms, that is, containing the fewest number of conjectured species (i.e., any species that is not a declared starting material, intermediate, or product) and steps. Every declared intermediate and product is required to be present in any mechanism output by the program.

The user can override this default search behavior and search for next-simplest mechanisms, for example by (i) asking the program to generate mechanisms containing new steps or new species (i.e., that are absent from the mechanisms that were already found), and (ii) imposing more constraints, with which previously generated (simplest) mechanisms are inconsistent. In either case, the program will necessarily explore more complex mechanisms that have either more conjectured species or more steps. Thus, we get a phenomenon that usually surprises first-time observers: adding more constraints often leads to the program's finding a larger number of simplest mechanisms whenever the current batch of simplest are entirely ruled out by the new constraints. The reason for the increase is that allowing more conjectured species is analogous to allowing more free parameters in a theory, which usually enlarges the number of theories consistent with given evidence.

MECHEM is not a stand-alone program; rather, it is an interactive aid that can reach credible results only by teaming with a knowledgeable chemist. The chemist formulates the problem and articulates plausibility constraints that express the available experimental evidence and any prior background assumptions that are desired. These constraints serve to discard implausible branches of the comprehensive search. Currently, MECHEM incorporates a rich array of approximately 100 different types of constraints. MECHEM's main goal is to find all simplest mechanisms that are consistent with constraints formulated by the user. There are a number of user conveniences, such as an ability to inspect what partially formed pathways the program is considering at the moment, as well as new auxiliary tools that can aid in the design of isotopic labeling experiments.

The genesis of the program was the realization that the "space" of reaction mechanism hypotheses consistent with experimental evidence could be searched comprehensively by an algorithm that (i) avoided duplicate generation via a canonical ordering of mechanisms, (ii) generated the simpler (fewer species and elementary steps) mechanisms first, and (iii) made use of the available evidence and heuristic assumptions to constrain the generator (7, 8). Points i and iii were also crucial in the development of the early Dendral program for molecular structure elucidation (9).

Although the combinatorial space of mechanisms can become very large, the practical availability of diverse types of experimental evidence, the ability to respect and build on intuitions articulated by the chemist, and the continual algorithmic improvements and faster computers have made this approach practical for significant classes of chemical reactions.

### How MECHEM Works

Here we sketch the basic methods that underlie MECHEM. The technical details of various algorithms can be found in specialized journals (8, 10-13).

The basic approach in MECHEM is to comprehensively search the possible elementary reactions and pathways in a "first principles" spirit. The principle involved is that an elementary step involves a small (user-adjustable) number of changes in the bonding of the reactants. Since in mechanism elucidation the reaction starting materials are known (unlike the case of synthesis), MECHEM builds elementary steps of the form "known reactants  $\rightarrow$  X + Y" and then solves for all possible structures of the unknowns X and Y by using graph algorithms and assuming, say, at most three or four total changes (cleavage or formation) to the connectivity of all the molecular graphs, including X and Y. One could contrast this "logical" approach to generating elementary steps with an alternative empirical approach that, say, generates only steps that follow specific reaction schemata such as migratory insertion, reductive elimination, radical recombination, dissociative adsorption, and so on.

Thus, it is clear how the program generates initial elementary steps from the starting materials. After X and Y become specific species, then the program can consider all possible second steps in a similar manner. The "space" of possible pathways is simply the set of possible lists of such elementary steps. The only unchangeable assumption in MECHEM is that all elementary steps have at most two reactants and at most two products; there are no further built-in assumptions about either elementary steps or pathways.

The program organizes its search in stages of simplicity, by first trying to find mechanisms that involve no machine-generated species, then one such species, then two, three, and so on. Also, for a fixed number of species, it looks first for mechanisms having fewer steps. Thus, a preference for parsimony is inherent in the search process.

The relative absence of built-in assumptions also means that the program has no factual knowledge of any specific chemistry. Thus, one might expect that its output mechanisms will not be very credible. For example, in a catalytic oxidation reaction the program might ignore the catalyst and propose mechanisms that directly oxidize the starting materials. It is the chemist/user's responsibility to drive the process by supplying the right assumptions, based on experimental evidence and background knowledge, so that the final output mechanisms are the simplest plausible ones. The program is designed so that these assumptions (constraints) constrain mechanism generation as early as possible, so that the program will not waste time exploring partial pathways that already violate a constraint. Such efficiencies are crucial for a practical program, given the huge combinatorial spaces that are involved.

The internal program, including a command-line interface, is written entirely in Common Lisp and the graphical user interface is written in Tcl/Tk. Thus, a Lisp interpreter and compiler are needed to run MECHEM.

The final choice of best mechanism, if such a choice is needed in the face of several plausible alternatives, is left to the user (researcher, student, or professor).

### Comprehensive Generation of Hypotheses

We used MECHEM on a slightly more complicated task than the one described by Hoffmann: in addition to ethylene and dihydrogen, we declared methane as a by-product of the

reaction in agreement with the original paper by Okabe and McNesby (2).

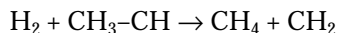
Our first run involved only constraints 1–4 from the list below. Constraints 1 and 2 express the observations from isotopic labeling experiments (the methods for predicting [14] and explaining the results of an isotopic labeling experiment are recent developments in MECHEM). Constraint 3 serves to simplify the search; if this constraint is removed or the maximum number of carbon atoms is higher, more mechanisms are found. Constraint 4 represents one way to articulate our intuition that every reaction intermediate should be consumed by some step.

#### Ethane Photolysis Constraints

1. Given starting materials  $\text{CH}_3\text{-CH}_3$  and  $\text{CD}_3\text{-CD}_3$ , the reaction does not yield HD.
2. Given a starting material  $\text{CD}_3\text{-CH}_3$ , the reaction does not yield HD.
3. Reactants involve jointly at most two carbon atoms.
4. Every conjectured species must appear on left- and right-hand sides of some (different) steps in the pathway.
5. Reject mechanisms whose only intermediate species is  $\text{CH}_3\text{-CH}$ .
6. Reject the step:  $\text{H}_2 + \text{CH}_3\text{-CH} \rightarrow \text{CH}_4 + \text{CH}_2$ .

After 11 seconds, the program found a unique simplest mechanism which contains the cited Okabe–McNesby mechanism (hypothesis 2 from the Introduction). Since our aim is to show that—quoting Hoffmann—the “weakness of our minds” can be overcome by chemist–computer collaboration, next we searched for more interesting alternative

mechanisms. Hence, we introduced constraint 5, which requires at least one additional intermediate to be formed. The program found some mechanisms that contain the dubious step:



so we explicitly rejected this step as constraint 6 in the above list. Finally, after several minutes the program reported ten mechanisms (Fig. 1 shows the state of the interaction at this point), of which five contain the two steps from the Okabe–McNesby mechanism cited above. The five remaining mechanisms are as follows:

#### Mechanism 1

1.  $\text{CH}_3\text{-CH}_3 \rightarrow \text{H}_2 + \text{CH}_3\text{-CH}$
2.  $\text{CH}_3\text{-CH}_3 \rightarrow \text{CH}_4 + \text{CH}_2$
3.  $2\text{CH}_2 \rightarrow \text{CH}_2=\text{CH}_2$
4.  $\text{CH}_3\text{-CH} \rightarrow 2\text{CH}_2$

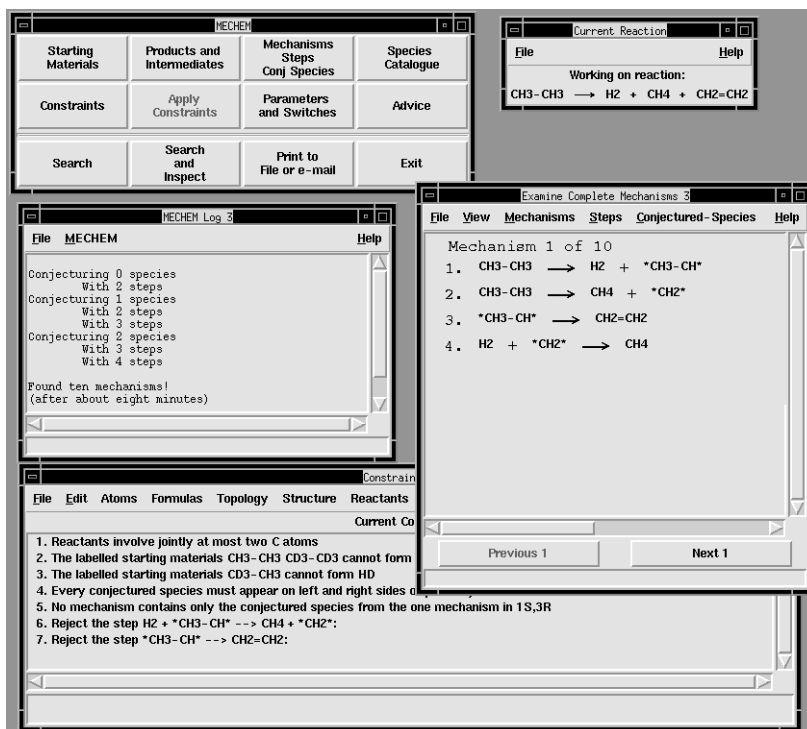
#### Mechanism 2

1.  $\text{CH}_3\text{-CH}_3 \rightarrow \text{H}_2 + \text{CH}_3\text{-CH}$
2.  $\text{CH}_3\text{-CH}_3 + \text{H}_2 \rightarrow 2\text{CH}_4$
3.  $\text{CH}_3\text{-CH} \rightarrow \text{H}_2 + \text{CH}=\text{CH}$
4.  $\text{H}_2 + \text{CH}=\text{CH} \rightarrow \text{CH}_2=\text{CH}_2$

#### Mechanism 3

1.  $\text{CH}_3\text{-CH}_3 \rightarrow \text{H}_2 + \text{CH}_3\text{-CH}$
2.  $\text{CH}_3\text{-CH}_3 + \text{H}_2 \rightarrow 2\text{CH}_4$
3.  $\text{CH}_3\text{-CH} \rightarrow 2\text{CH}_2$
4.  $2\text{CH}_2 \rightarrow \text{CH}_2=\text{CH}_2$

Figure 1. User/MECHEM interaction in ethane photolysis.



**Mechanism 4**

1.  $\text{CH}_3\text{-CH}_3 \rightarrow \text{H}_2 + \text{CH}_3\text{-CH}$
2.  $\text{CH}_3\text{-CH} \rightarrow 2\text{CH}_2$
3.  $\text{H}_2 + \text{CH}_2 \rightarrow \text{CH}_4$
4.  $2\text{CH}_2 \rightarrow \text{CH}_2=\text{CH}_2$

**Mechanism 5**

1.  $\text{CH}_3\text{-CH}_3 \rightarrow 2\text{CH}_3$
2.  $\text{CH}_3 \rightarrow \text{H}_2 + \text{CH}$
3.  $\text{CH}_3\text{-CH}_3 + \text{H}_2 \rightarrow 2\text{CH}_4$
4.  $\text{CH}_3 + \text{CH} \rightarrow \text{CH}_2=\text{CH}_2$

These mechanisms are more complicated than the simplified Okabe–McNesby mechanism, but they do account for the formation of  $\text{CH}_4$ . Mechanisms 1, 3, and 4 suggest the formation of two methylene species ( $\text{CH}_2$ ) from ethylidene, which seems doubtful. In mechanism 2, acetylene is a precursor of ethylene, although ethylene is generally believed to be the source for acetylene in this system. Mechanism 5 implies the successive formation of methyl radical and methyldiene. Neither process is improbable (4), but both seem to be less important than competitive ones. Direct formation of methane from ethane (step 3 of mechanism 5) also seems doubtful.

The above mechanisms are a comprehensive list of the simplest mechanisms that account for  $\text{H}_2$ , ethylene, and methane formation and are consistent with the explicit constraints. Other constraints could be tried by someone with direct expertise in alkane photochemistry and knowledge of the (somewhat contradictory) experimental results from the literature, but—although interesting and desirable—this would not detract from the aims of this paper.

**Positive but Circumspect Contributions**

As explained colorfully by Hoffmann above, chemists (and scientific modelers generally) wish to make positive contributions without including wishy-washy provisos concerning all the mechanisms or models that might be lurking but were not considered or even conceived of by the author.

We propose a different way to contribute positively, which presupposes a scientist–program collaboration: the user selects the appropriate problem formulation and constraints, based on experimental evidence and other background knowledge, while the computer is charged with carrying out a comprehensive search—in order of simplicity—within the ensuing “hypothesis space”.

Specifically, for many of the simpler chemical reactions, the mechanistic conclusions of an experimental chemistry paper can be of the form: Given (i) a problem formulation (starting materials, observed products or intermediates, and possibly a catalyst) and (ii) a set of assumptions (constraints), all the simpler plausible mechanisms are as follows: .... Of course, an author could perform experiments to rule out some of these mechanisms, thus generating new constraints, but at some point the conclusion—within the limits of current resources and experimental technique—is that some number of mechanisms remain plausible.

We have recently made exactly this type of positive contribution for two rather complex homogeneous organometallic reactions (15, 16). In one case (15), our positive contribution was to list the several dozen plausible simpler mechanisms that lacked precedents in the literature. In the newer case (16), we were able to reduce the set of 41 simpler hypotheses to a handful, based on kinetic isotope effect experiments.

**Educational Role of Interactive Mechanism Elucidation**

A student/user can drive the process of interactive mechanism elucidation (with MECHEM) in two ways: (i) by formulating the plausible assumptions, and (ii) by deciding on the next steps when the computer delivers a set of multiple, equally simple mechanisms.

To formulate plausible assumptions (constraints), the user must know or hypothesize, for example, the elementary steps that are unlikely to occur and why; the role of the catalyst in the reaction, if any; what kinds of species are plausible or not plausible; what type of mechanism (e.g., Langmuir–Hinshelwood or Eley–Rideal) is unlikely to be observed; and so on. To accomplish this interaction with the computer, the students have to consolidate much of their knowledge of physical, organic, inorganic, and other chemistry, and their knowledge of basic heuristic chemical principles such as orbital symmetry and other rules.

Choosing a next step when confronted with multiple plausible mechanisms requires other student skills, such as proposing how to obtain discriminating evidence (whether experimental or theoretical) to discriminate among the alternatives. For example, from the set of hypothetical mechanisms, the student may propose discriminating them by an isotopic labeling or kinetic isotope effect experiment, quantum chemistry computations, chemical kinetics, spectroscopy, and so on. The task is challenging because the student must once again appeal to very different sources of knowledge. Occasionally students may conclude that the methods known to them are insufficient or inefficient, and then they should be called on to explain why. If the student proves capable of proposing and critiquing methods known from elsewhere in the curriculum, then he or she has acquired a basic skill of every mechanistic chemist.

Besides the knowledge and skills that the student is led to acquire or to exercise, he or she begins to appreciate that chemistry is not a hodgepodge of separate courses and techniques, but instead is a coherent discipline where different ideas are closely related.

**Conclusion**

We have proposed a methodological improvement in the elucidation of chemical reaction mechanisms. Although this article supports our proposal with only a very simple reaction, our main work in more specialized areas of homogeneous and heterogeneous catalysis (15–18) can be adduced to claim generality.

A collaborative team consisting of chemist and computer program (e.g., MECHEM) is clever enough to overcome the weakness of our minds by making positive—but appropriately

circumspect—chemical contributions of this form: Given the available experimental evidence and reliable assumptions based on background knowledge, the following is a *comprehensive* list of the simpler, plausible, reaction mechanism hypotheses.

Our methods and software are a promising approach to enhance the teaching of chemical intuition in mechanistic chemistry as expounded, for example, by Scudder in this *Journal* (19). In particular, MECHEM's capabilities for predicting and explaining isotopic labeling patterns can be used to teach these valuable skills of experiment design and interpretation, or to aid lecturers in the preparation of interesting classroom exercises or exam questions.

We invite inquiries about the availability of MECHEM (20) for use in either education or research. Proposals for collaboration are especially welcome.

### Acknowledgments

This work was supported in part by the grant IRI-9421656 from the National Science Foundation and by the NSF Division of International Programs.

### Note

1. This ethane photolysis example also appears as Problem 188 in the ConcepTests Web page (<http://www.chem.wisc.edu/~concept>) at the University of Wisconsin Chemistry Department. These conceptual problems are intended for use in the "peer instruction" pedagogical style developed originally for introductory Physics (as Project Galileo) by Eric Mazur of Harvard.

### Literature Cited

- Hoffmann, R. *The Same and Not the Same*, Columbia University Press: New York, 1995.
- Okabe, H; McNesby, J. *J. Chem. Phys.* **1961**, *34*, 668–669.
- Hampson, R. Jr.; McNesby, J.; Akimoto, H.; Tanaka, I. *J. Chem. Phys.* **1964**, *40*, 1099–1105
- Lias, S. *J. Chem. Phys.* **1970**, *52*, 1841–1851.
- Roquette, B. *J. Phys. Chem.* **1970**, *74*, 1204–1207.
- Hoffmann, R.; Minkin, V. I.; Carpenter, B. K. *Bull. Soc. Chim. Fr.* **1996**, *133*, 117–130.
- Valdés-Pérez, R. E. *Artificial Intelligence* **1994**, *65*, 247–280.
- Valdés-Pérez, R. E. *Artificial Intelligence* **1995**, *74*, 191–201.
- Lindsay, R.; Buchanan, B.; Feigenbaum, E.; Lederberg, J. *Applications of Artificial Intelligence for Organic Chemistry: The Dendral Project*, McGraw Hill: New York, 1980.
- Valdés-Pérez, R. E. *J. Comput. Chem.* **1992**, *13*, 1079–1088.
- Valdés-Pérez, R. E. *J. Comput. Chem.* **1993**, *14*, 1454–1459.
- Valdés-Pérez, R. E. *J. Chem. Inf. Comput. Sci.* **1994**, *34*, 976–983.
- Valdés-Pérez, R. E. *J. Comput. Chem.* **1994**, *15*, 1266–1277.
- Zeigarnik, A. V.; Valdés-Pérez, R. E. *J. Comput. Chem.* **1998**, *19*, 741–753.
- Zeigarnik, A. V.; Valdés-Pérez, R. E.; Temkin, O. N.; Bruk, L. G.; Shalgunov, S. I. *Organometallics* **1997**, *16*, 3114–3127.
- Bruk, L. G.; Gorodskii, S. N.; Zeigarnik, A. V.; Valdés-Pérez, R. E.; Temkin, O. N. *J. Mol. Catal. A: Chemical* **1998**, *130*, 29–40.
- Valdés-Pérez, R. E. *Catal. Lett.* **1994**, *28*, 79–87.
- Valdés-Pérez, R. E.; Zeigarnik, A. V. *J. Mol. Catal. A: Chemical* **1997**, *119*, 405–414.
- Scudder, P. H. *J. Chem. Educ.* **1997**, *74*, 777–782.
- Valdés-Pérez, R. E. MECHEM Demos; <http://www.cs.cmu.edu/~sci-disc/mechem.html> (accessed Dec 1999).