ELSEVIER

Journal of Molecular Catalysis A: Chemical 119 (1997) 405–414

JOURNAL OF MOLECULAR CATALYSIS A: CHEMICAL

Interactive elucidation (without programming) of reaction mechanisms in heterogeneous catalysis

Raúl E. Valdés-Perez [a,*], Andrew V. Zeigarnik [b]

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

[b] *Laboratory of Chemical Kinetics and Catalysis, Lomonosov Academy of Fine Chemical Technology, pr. Vernadskogo 86, Moscow 117571, Russia*

Received 4 June 1996; accepted 9 December 1996

Abstract

MECHEM is a computer program for the interactive elucidation of reaction mechanisms. A recent application to a model catalytic system (ethane hydrogenolysis to produce methane over a transition metal catalyst) turned up a simple, plausible and seemingly novel mechanism. However, that application required some programming on the part of the user and also required subdividing the reaction into two stages in order to handle the reaction complexity. Recent advances in MECHEM now enable straightforward handling of such complex reactions without any programming nor division into stages. The capability is illustrated on the previous ethane hydrogenolysis example.

Keywords: Ethane hydrogenolysis; Reaction mechanisms; Computer programs; MECHEM

1. Introduction

A recent article [1] described the application of MECHEM, a computer program for the interactive elucidation of reaction mechanisms, to a model catalytic system: ethane hydrogenolysis to produce methane over a transition metal catalyst. Use of the program helped turn up a simple, plausible mechanism that seemingly had been overlooked and which involved transfers of hydrogen from a carbon atom of an adsorbate to another adsorbate's carbon atom, rather than the successive dehydrogenation and then hydrogenation steps that have been hypothesized by most researchers.

From the viewpoint of developing tools for the field of catalysis, the earlier application presented two drawbacks. Firstly, use of MECHEM on a given reaction involved some programming to express constraints on plausible mechanisms; the program makes use of these constraints during its search for simple mechanisms. Secondly, because of the inherent limitations of the approach, which are due to the combinatorial complexity, the problem had to be solved in two stages: (1) starting materials → intermediates and (2) intermediates → products. At each stage, partial pathways were

* Corresponding author.

PII S1381-1169(96)00504-3

found with MECHEM and then combined to form overall pathways. Both drawbacks, especially the first, represented a significant hindrance to the program's potential as a discovery tool that could be used autonomously by chemists without the aid of a programmer.

The aim of this article is to report advances in MECHEM that have removed the two cited drawbacks, at least for ethane hydrogenolysis and other reaction mechanisms of similar complexity. This article will show that the simplest mechanisms for ethane hydrogenolysis can be reached by an interactive use of MECHEM that involves no programming nor division of the problem into simpler sub-problems, but instead involves only expressing new constraints via a convenient interface and running the program several times.

2. Background on the program

The MECHEM program has developed over the last seven years. It consists of about ten thousand lines of Common Lisp, which is a language whose facilities for list processing support very well the type of symbolic programming that the program's task requires.

MECHEM represents an application of heuristic search principles, which have led to the working hypothesis that some problems of scientific reasoning involve guided search through a large space of possibilities. Model building in science is a very promising outlet for this working hypothesis. In line with this overall research direction, MECHEM deals with one type of model building: the elucidation of reaction pathways or mechanisms on the basis of experimental evidence and other constraints.

MECHEM is not an autonomous program, it depends on the talent and art of the user who is able to judiciously select what local constraints can reasonably be supposed about a reaction. The program's role is to *draw out the consequences of these constraints* by reporting the simplest reaction mechanisms that are consistent with them. At first, the simplest mechanisms are implausible, so the user articulates his objections, enters them into the program via the interface and runs the program again. This interactive process repeats until either the user is satisfied with the program's output, or the problem becomes too complex to handle. For reasons of sheer combinatorics, not all reactions are within MECHEM's capability, but many are e.g. most of the typical reactions of heterogeneous catalysis.

Compared with other projects in computer-aided chemistry [2], MECHEM's characteristic features are these:

- Its only 'hard-wired' assumptions are that steps must be materially balanced and that every step has at most two reactants and at most two products. *Everything else* is an optional constraint or a re-settable parameter.
- It reports the simplest mechanisms i.e. having the fewest number of species (intermediates) and fewest overall steps. Mechanisms are sought in order of increasing total number of species: mechanisms having fewer species are generated first. With the number of species held constant, mechanisms having fewer reaction steps are generated earlier. That is, the number of steps is incremented faster than the number of species. These two constraints on mechanism generation constitute simplicity criteria that guide the search.
- It makes no use of species or reaction libraries. That is, it does not require specifying a complete list of possible species, reactions, or reaction transforms. Instead, it does everything from scratch. Therefore, the program can potentially address any branch of chemistry, although our research has focused on catalysis (Ref. [3] reports an application to the partial oxidation of methane using some representative constraints).
- It makes use of mostly qualitative constraints that are typically used by chemists to reason about catalytic reaction pathways. In MECHEM, these constraints are drawn out via interactive use of the program and serve,

in turn, as a replicable record of one's reasoning.

The program incorporates a number of combinatorial algorithms that have taken years to refine because of the many subtle opportunities for early pruning of the program's search space, while maintaining correctness (i.e. an exhaustive search). All chemical reactions are potentially within the program's scope, the only question is whether on a particular problem the combinatorics will overwhelm the program, so early pruning through dedicated algorithm design and experimentation is critical. There have been a number of published reports on aspects of these algorithms [4–8], one of which [4] is now obsolete in MECHEM due to recent advances which have not yet been published. These advances have been motivated by our recent work on a considerably more complex catalytic reaction from organometallic catalysis. Further discussion of this application or of algorithms is outside the scope of this article.

3. Future work

The next steps in this project are to provide a window-based graphical interface (but which maintains the interaction style of the demonstration in the appendix) and to incorporate constraints based explicitly on energetics. In parallel, we will add further constraint types such as 'prohibit CO insertion between the elements E1 and E2' or 'no step can decrease the oxidation state of atom A'. The user can always reject *specific* steps or species that MECHEM finds, so that the effect of energetics and other reasoning can be approximated by means of specific prohibitions.

4. A detailed demonstration on ethane hydrogenolysis

An earlier article [1] reported a reaction mechanism, found by interactive use of MECHEM, for ethane hydrogenolysis that was simple, plausible and which seemingly had been overlooked by workers in the field. This new mechanism involved crucial hydrogen transfer steps between carbonaceous adsorbates, rather than the successive dehydrogenations, CC bond scission and hydrogenations that are featured in the accepted mechanism. The main aim of that article was to give a sharp demonstration of the potential utility of MECHEM.

The (abbreviated) demonstration in the appendix to this paper serves to illustrate what constraints are needed, using the existing version of MECHEM, in order to (almost) reach the conclusion that the *accepted* ethane-hydrogenolysis mechanism is the unique simplest [1]. The newer mechanism based on hydrogen transfer that MECHEM found earlier will not appear because of the explicit rejection of all hydrogen transfer steps in the form of a constraint.

From this exercise, our conclusion is that MECHEM can be useful on complex problems of current interest without any need for programming by the user, who instead merely needs to formulate constraints via an interactive interface.

Appendix A

This demonstration begins after MECHEM is loaded, which takes a few seconds and the program is entered. User input is shown below in bold face, whereas commentary is shown in italics and in square brackets. The interactive process that led to the constraints is not, for lack of space, reflected here, just a final, illustrative list of constraints.

MECHEM is an interactive discovery program that helps ensure that no simple multistep reaction pathways are overlooked. MECHEM is

[1] The program currently disregards any step of the form $A + B \rightarrow A + B'$. Such steps are quite reasonable, however, so we plan to revise MECHEM accordingly.

widely applicable, it builds on the user's current knowledge of a reaction and it provides a convenient interface for articulating that knowledge flexibly to the program.

This interface is organized as a hierarchical collection of contexts. For example, if you want to specify the starting materials, you should enter the 'starting materials' context as shown below. To talk about mechanisms that MECHEM finds, enter the 'mechanisms' context when it appears. The same holds for 'steps', 'conjectured species' and so on.

Type the commands in square brackets [ ] to enter or quit contexts.

Starting Materials [sm]
Products and Intermediates [p]
Species catalogue [sc]
About MECHEM [ab]
Quit [q]

? **sm**

Add one starting material or a parenthesized list [a]
List the starting materials [l]
Delete a starting material [d]
Load from a file [f]
Show the catalogued species that MECHEM recognizes [sh]
Quit dealing with starting materials [q]

? **a**

starting material(s): (H2 ETHANE M MM) [*M and MM are notation for single and dual free sites on a catalyst.*]

[*the previous context menu is repeated.*]

? **q**

Working on (H2 + ETHANE + M + MM) go to ()

Starting Materials [sm]
Products and Intermediates [p]
Constraints [c]
Parameters and Switches [ps]
Run MECHEM [r]
Species catalogue [sc]
Combinatorics hints [co]
About MECHEM [ab]
Quit [q]

? **p**

Note: no distinction is made in this context between products and intermediates

Add one product or a parenthesized list [a]
List the products [l]
Delete a product [d]
Load from a file [f]
Show the catalogued species that MECHEM recognizes [sh]
Quit dealing with products [q]

? **a**

product(s): (C2-H2-M2 C2-H4-M2 METHANE) [*The structures of intermediates have already been entered.*]

[*the previous context menu is repeated.*]

? **q**

Working on (H2 + ETHANE + M + MM) go to (C2-H2-M2 + C2-H4-M2 + METHANE)

Starting Materials [sm]
Products and Intermediates [p]
Constraints [c]
Parameters and Switches [ps]
Run MECHEM [r]
Species catalogue [sc]
Combinatorics hints [co]
About MECHEM [ab]
Quit [q]

? **c**

Add a constraint [a]
List the current constraints [l]
Quit dealing with constraints [q]

? **a**

Atoms [a]
Formulas [f]
Topology [to]
Structures [str]
Reactants [r]
Products [p]
Steps [st]
Sequences of Steps [ss]
Full Pathways [fp]
Quit adding constraints [q]

[*the user now selects any of these options, thus descending within the constraint hierarchy to a new menu of options, e.g., constraints that pertain to pairs of reactants, to the molecular structure of intermediates, and so on. The constraint hierarchy cannot be shown here for lack of space.*]

? **q**

Add a constraint [a]

List the current constraints [l]
Delete a constraint [d]
Delete all constraints [all]
Quit dealing with constraints [q]

? **l**

1. Overall stoichiometry is: (H2 ETHANE) → (METHANE METHANE).
2. No multiple occurrences of any reactant pair on left-hand side of pathway.
3. H2 can appear at most one time on the left-hand side of the pathway.
4. ETHANE can appear at most one time on the left-hand side of the pathway.
5. C2-H2-M2 is a necessary precursor of METHANE.
6. C2-H4-M2 is a necessary precursor of C2-H2-M2.
7. No gas-phase + adsorbate products involving (M) sites.
8. H2 cannot appear on right-hand side of pathway.
9. ETHANE cannot appear on right-hand side of pathway.
10. No gas-phase + adsorbate reactants involving (M) sites.
11. The element M must be involved in all steps.
12. A C atom cannot bridge M and M atoms.
13. Distinct instances of the links CH CH are unbreakable in C2-H2-M2.
14. Distinct instances of the links CH CH are unbreakable in C2-H4-M2.
15. Distinct instances of the links CH CH are unbreakable in ETHANE.
16. Every conjectured species has at least one M atom.
17. Every conjectured species has at most two M atoms.
18. Every conjectured species has at most two C atoms.

[*the previous context menu is repeated.*]

? **q**

Working on (H2 + ETHANE + M + MM) go to (C2-H2-M2 + C2-H4-M2 + METHANE)

Starting Materials [sm]
Products and Intermediates [p]
Constraints [c]
Parameters and Switches [ps]
Run MECHEM [r]
Species catalogue [sc]
Combinatorics hints [co]
About MECHEM [ab]
Quit [q]

? **ps**

List the current parameter settings [l]
Change the current parameter settings [ch]

Restore the default parameter settings [re]
Quit dealing with parameters [q]

? **ch**

1. Start with simplicity parameters: 0 conjectured species and 0 steps
2. Will keep searching by considering more complex mechanisms
3. Consider at most 6 conjectured species
4. At most 3 changes to the topology of molecules per step
5. Exchange reactions (AB + CD → AC + BD) are OK
6. Hydrogen transfer steps like (ads1-H + ads2 → ads1 + ads2-H) are OK
7. Conjectured species need not be at full valence
8. A starting material can be produced from the other starting materials
9. Isomers are allowed
10. Allowed valences are: C(4) H(1) M(1)

[*the values of parameters 5 and 6 are next inverted, thus prohibiting exchange reactions and hydrogen-transfer steps.*]

? **q**

Working on (H2 + ETHANE + M + MM) go to (C2-H2-M2 + C2-H4-M2 + METHANE)

Starting Materials [sm]
Products and Intermediates [p]
Constraints [c]
Parameters and Switches [ps]
Run MECHEM [r]
Species catalogue [sc]
Combinatorics hints [co]
About MECHEM [ab]
Quit [q]

? **r**

Conjecturing 0 species ...
    With 2 steps
Conjecturing 1 species ...
    With 2 steps
Conjecturing 2 species ...
    With 3 steps
Conjecturing 3 species ...
    With 3 steps
    With 4 steps
Conjecturing 4 species ...
    With 4 steps

With 5 steps
With 6 steps
Conjecturing 5 species ...
With 4 steps
With 5 steps
With 6 steps
With 7 steps
Conjecturing 6 species ...
With 5 steps
With 6 steps
With 7 steps
With 8 steps
With 9 steps

Found five mechanisms!
(after about fifty-six minutes) [*on a Silicon Graphics Indigo.*]

Starting Materials [sm]
Products and Intermediates [p]
Constraints [c]
Parameters and Switches [ps]
Run MECHEM [r]
Mechanisms [m]
Steps [st]
Conjectured species [cs]
Species catalogue [sc]
Combinatorics hints [co]
About MECHEM [ab]
Quit [q]

[*By entering the "mechanisms" menu, the user proceeds to group these mechanisms according to their array of stoichiometric numbers.*]

[*Conjectured species are shown within pound signs #·#. They are currently listed as molecular formulas; we are working on printing that reflects their structures. Note that some stoichiometric numbers (embedded in the arrows) turn out to be negative, in order to account for the net stoichiometry ETHANE + H2 → 2 METHANE. This means that the step's net degree of advancement is toward the left.*]

(One pathway similar to this one)
1. H2 + MM –(1)–> 2(#H-M#)
2. ETHANE + MM –(1)–> #H-M# + #C2-H5-M#
3. MM + #C2-H5-M# –(1)–> C2-H4-M2 + #H-M#
4. M + C2-H4-M2 –(1)–> #H-M# + #C2-H3-M2#
5. M + #C2-H3-M2# <–(-1)– C2-H2-M2 + #H-M#

6. C2-H2-M2 -(1)-> 2(#C-H-M#)
7. #C2-H3-M2# + #C-H-M# -(2)-> C2-H2-M2 + #C-H2-M#
8. #H-M# + #C-H2-M# -(2)-> M + #C-H3-M#
9. #H-M# + #C-H3-M# -(2)-> MM + METHANE

[*Step 7 above is not a hydrogen-transfer step, it involves breaking a CC bond and forming another.*]

(One pathway similar to this one)

1. H2 + MM -(1)-> 2(#H-M#)
2. ETHANE + MM -(1)-> #H-M# + #C2-H5-M#
3. MM + #C2-H5-M# -(1)-> C2-H4-M2 + #H-M#
4. M + C2-H4-M2 <-(-1)- #H-M# + #C2-H3-M2#
5. M + #C2-H3-M2# -(1)-> C2-H2-M2 + #H-M#
6. C2-H2-M2 -(1)-> 2(#C-H-M#)
7. C2-H4-M2 + #C-H-M# -(2)-> #C2-H3-M2# + #C-H2-M#
8. #H-M# + #C-H2-M# -(2)-> M + #C-H3-M#
9. #H-M# + #C-H3-M# -(2)-> MM + METHANE

[*Step 7 above is not a hydrogen-transfer step, it involves breaking a CC bond and forming another.*]

(Three pathways similar to this one)

1. H2 + MM -(1)-> 2(#H-M#)
2. ETHANE + MM -(1)-> #H-M# + #C2-H5-M#
3. MM + #C2-H5-M# -(1)-> C2-H4-M2 + #H-M#
4. C2-H4-M2 -(1)-> #H-M# + #C2-H3-M#
5. MM + #C2-H3-M# -(1)-> C2-H2-M2 + #H-M#
6. C2-H2-M2 -(1)-> 2(#C-H-M#)
7. #H-M# + #C-H-M# -(2)-> M + #C-H2-M#
8. #H-M# + #C-H2-M# -(2)-> M + #C-H3-M#
9. #H-M# + #C-H3-M# -(2)-> MM + METHANE

[*The mechanisms are then manipulated further, although this interaction is omitted for lack of space. MECHEM stores a log of the entire interaction.*]

References

[1] R.E. Valdes-Perez, Catal. Lett. 28(1) (1994) 79–87.
[2] I. Ugi, J. Bauer, K. Bley, A. Dengler, A. Dietz, E. Fontain, B. Gruber, R. Herges, M. Knauer, K. Reitsam and N. Stein, Angew. Chem. Int. Ed. Engl. 32 (1993) 201–227.
[3] R.E. Valdes-Perez, in: M. Bhasin and D. Slocum (Eds.), Methane and Alkane Conversion Chemistry (Plenum Press, New York, 1995).
[4] R.E. Valdes-Perez, J. Comput. Chem. 15(11) (1994) 1266–1277.
[5] R.E. Valdes-Perez, J. Chem. Inf. Comput. Sci. 34(4) (1994) 976–983.
[6] R.E. Valdes-Perez, J. Comput. Chem. 14(12) (1993) 1454–1459.
[7] R.E. Valdes-Perez, J. Comput. Chem. 13(9) (1992) 1079–1088.
[8] R.E. Valdes-Perez, J. Chem. Inf. Comput. Sci. 31(4) (1991) 554–556.