

Towards the Simulation of Reaction Networks in Astrochemistry

Pierre Philippe, David Weiss Solís, Tom Lenaerts and Hugues Bersini

IRIDIA - Université Libre de Bruxelles

CP 194/6

50, av. Franklin Roosevelt

1050 Bruxelles - Belgium

pphilipp@ulb.ac.be

Abstract

The aim of this paper is to apply artificial chemistry (AC) to the modelling of astrochemistry. More generally, we have attempted to construct a computational model of 'a' minimal real chemistry in an object oriented (OO) approach (UML, Java). To this end we map astrochemistry into the ACs theoretical framework thanks to the developed OO model. The OO approach has permitted us to separate the actual and logical representation of chemical structure. To perform this mapping we identify and simplify features from astrochemistry to tackle the issues of, representation and canonization, reaction rules, and meta-dynamics for a suitable level of abstraction. In the search for a chemical criterium for existence relations between objects we have used chemical thermodynamics. The result is an abstract, simplified model of a real chemical dynamical reaction network which conceptually captures real chemistry in general.

Introduction

Astrochemistry is the study of the chemicals found in outer space, usually in molecular gas clouds, and of their formation, interaction and destruction. For our purposes, the chemistry of interstellar molecule-bearing nebulae is the prime candidate in the search for a real-world archetype of a strongly constructive chemical network. The model subject was motivated by the following observations: primarily, the system behaves as a mixture of very diluted ideal gases, giving the possibility to restrict the molecules to a finite and computationally tractable set of chemical objects while still considering large variations in space and time. Secondly, the low size and diversity of interstellar molecules intrinsically eases the combinatorial explosion and the canonization problem. Thirdly, if we focus on a large enough portion of the cloud, the system behaves as a closed system which allows for simple thermodynamic treatment. Fourthly, the rate equations which in biological kinetics (and bioprocess engineering) give rise to non-polynomial non-linearities (Nicolis and Prigogine, 1989) are reduced to constants (Duley and Williams, 1984). On the other side computational approaches have been identified as the sole way to understand these inaccessible cosmic objects (Le Boulrot, 1997). As a

modelling framework we have chosen artificial chemistries (ACs) for their synthetic approach towards chemistry.

Artificial chemistries, in their most general form, have been defined as man-made systems which capture different features from real chemical systems (Dittrich et al., 2001). The ambitious goal put forward for these systems is to construct theoretical models which allow us to improve our understanding of the origin of evolutionary systems. To achieve this goal, real (bio)chemical systems were formally abstracted into a triplet $\langle S, R, A \rangle$, with S referring to a collection of molecules, R referring to a collection of reaction rules and A identifying the reaction vessel or reactor. The intricate interaction between these components is used to study their complex dynamics in terms of organizational and functional properties. The focus of these systems is clearly on the results produced by A .

Although ACs provide interesting tools to formally study the kinetics and meta-dynamics of chemical systems, there is often a large gap with chemical reality. For instance, molecules captured in S are regularly represented as numbers or strings. Such a representation ignores the real complexity of molecular structures and their structure/function relationship. Moreover, the uniqueness of the representation is extremely relevant when the synthetic algorithm can construct different isomorphic forms of the same molecule (Schubert and Ugi, 1978). Furthermore, the reactions in R are often depicted as linguistic operations on molecules (Fontana, 1996). Yet in real chemistry this is not enough. There are additional energy considerations, apart from the structural ones that need to be taken into account, which determine the possible molecular structure. One of the main criticisms towards ACs is that they are models where one cannot quantify a distance between model and modelled system (Kauffman, 1995). This criticism however is not restricted to ACs but extends to biochemical or bioprocess models where parametric estimation is delicate. Hence, careful considerations are in order when drawing conclusions obtained from AC for real chemistry. In order to close the gap and to make well-founded observations and predictions, AC should move closer to real chemistry.

Copyrighted Material

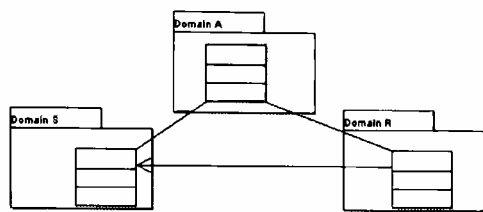


Figure 1: UML package diagram.

A solution can be found in that context. ACs do not stand on their own in their synthetic approach towards chemistry. In chemistry itself a rich research area called *computer chemistry* exists which has been addressing this topic already for a couple of decades (not to be mistaken with the area of computational chemistry) (Ugi et al., 1993). In this context both formal and semi-formal algorithms (which focus on *S* and *R*) have been constructed for the synthetic design of molecular structures (Dugundji et al., 1976; Ugi et al., 1993; Faulon and Sault, 2001).

These last two disciplines are closely related in their methodology but differ significantly in their goals. In order to integrate knowledge from both one requires a realistic case-study. The goal set for our research is double. First, we want to study the relation between reaction network topology and dynamics in a real chemical system. Second, we want to improve the dynamic perception of computer chemistry and pave the way towards full integration of AC and computer chemistry. To achieve this distant goal we address in this paper the issue of mapping the *S*, *R*, and *A* components into astrochemistry and making the necessary simplifying assumptions to effectively apply AC to astrochemistry, in a simple yet incremental approach.

In the following section we discuss the structure of our AC. Due to its roots both as a knowledge representation device in artificial intelligence research and as a simulation tool in the natural sciences (see SIMULA), object-oriented (OO) modelling is ideally suited for this task. Moreover, both in computer chemistry and AC this relevance has already been acknowledged (Bauerschmidt and Gasteiger, 1997; Bersini, 2000; Bersini, 1999). The resulting model allows us to reason about the logical structure of reactions and molecules in astrochemistry. Hence, in the next section, we develop a simple mathematical model to construct the actual algorithmic solution. In a modest attempt to close the loop, this procedure amounts to a subset of the original (too) complex problem that gave rise to ACs.

Object-Oriented Model

In this section we briefly discuss the different components of the triplet $\langle S, R, A \rangle$. Fig. 1 shows the package diagram for a general AC. The object oriented approach permits

to separate and develop independently the three parts of the model because each part can be encapsulated, i.e. each component's internal workings are invisible to the other components. Each component can be developed independently and each can rely on a particular level of abstraction, or algorithmic solution to test and improve the performance of the whole.

The component *S* is in charge of the molecular representation and canonization. The component *R* is responsible for defining the reactions, determining which reactions can take place, and the actual manipulation of molecules. The *A* component will take account of the kinetics and meta-dynamics of the system. Each part in the diagram can be specified in more detail.

Molecular Representation *S* It is customary for chemists to apply graphs (Lewis structures) to represent molecules and to add extra features to these structures to represent issues like stereo-chemistry. For instance, molecules can be represented, in their simplest form, by connection tables or matrices (Dugundji et al., 1976). These representations are not enough to allow for a complete expression of most molecular particularities. Therefore, extensions to these base representations or other representations have been suggested (Bauerschmidt and Gasteiger, 1997). These extensions were often the result of requirements posed on the structure by their users. Also a certain amount of flexibility is required from the actual molecular model since depending on the chemical problem, different levels of molecular detail are required. OO models allow for such a flexibility.

The novelty of the current OO model lies in the fact that we make a distinction between the actual representation of the molecule, for instance graph or matrix, and the logical components which constitute a molecule, for instance atoms and bonds. By making this distinction an OO chemistry can be constructed independent of the computational representation. This achievement is crucial since such a separation allows for a reasoning about a 'chemical world' without being bound to computational or representational issues. For current experiments we opted for an undirected graph representation which contains the principal chemical reactive elements in gas phase in its nodes.

Through the algorithm *A*, new and existing molecular types are constructed. Each of these molecules is represented by a directed graph. In such systems generated redundant structures should be identified or removed to ensure correct meta-dynamical results. Therefore, an algorithm has been designed which transforms constitutional similar graphs into a particular base-form. The algorithm is a simplification of general algorithms for canonical labelling and consists of three steps:

- Each atom in a molecule is assigned an arbitrary base value which is determined by its atomic number, its unbound electrons and the energy of the bonds.

-Afterward each atom is validated, i.e. it assigned a ranking using the base value. The higher the ranking value the higher the precedence of the atom in the molecule.

- Finally, the complete molecule is re-numerated according to the validated values.

As a result of this process, molecules can be differentiated in the system. At this point only structurally similar molecules can be detected, stereochemistry is not taken into account yet.

Class	Sub class	Reaction template	rate k
1	1.a	$A + \psi \rightarrow B^+ + e^-$	$10^{-17} s^{-1}$
	1.b	$A + \psi \rightarrow B^+ + C + e^-$	$10^{-19} s^{-1}$
2	2.a	$A^+ + B \rightarrow C^+ + D$	$10^{-9} \frac{cm^3}{s}$
3	3.a	$A + B \rightarrow C + D$	$10^{-11} \frac{cm^3}{s}$
4	4.a	$A^+ + e^- \rightarrow B$	$10^{-15} s^{-1}$
	4.b	$A^+ + e^- \rightarrow B + C$	$10^{-10} s^{-1}$
5	5.a	$A + h\nu \rightarrow B + C$	$10^{-11} s^{-1}$
	5.b	$H_2 + h\nu \rightarrow 2H$	$10^{-14} s^{-1}$
6	6.a	$2H \xrightarrow{grains} H_2$	$10^{-17} \frac{cm^3}{s}$

Table 1: Examples of classes of astrochemical reactions and their reaction rates (ψ = cosmic ray energy, $h\nu$ = electromagnetic radiation energy (photon) and e^- = electrons). Note that hydrogen reactions have special behavior within a given class of reaction.

Reactions R Traditionally each concrete reaction needs to be specified for the reaction generator to work. Yet general classes of reactions can be identified which capture the architecture of a reaction independent of the actual molecular elements. OO models allow for such a conceptualization. In Table 1 the different classes and subclasses are listed.

The functionality of a formal reaction is visualized in Fig. 2.

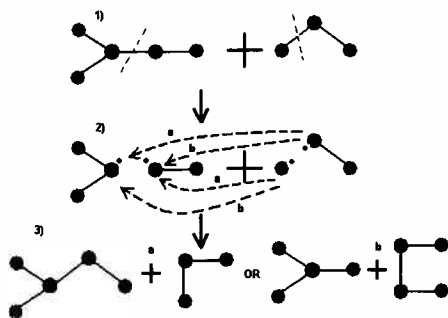


Figure 2: Reaction as conditional graph rewriting.

Due to our choice of undirected graphs for the molecular representation, a reaction corresponds to a graph rewriting

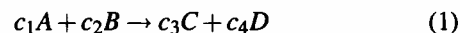
rule (cfr. (Benkő et al., 2003)), although in our model this graph rewriting is conditional. No new atoms are introduced. Hence, as in real chemistry there is a conservation of mass (electrons are considered as non-massive particles). Formally, a reaction consists of two phases:

-A certain amount of bonds of the reactants are broken according to some chemically plausible criteria.

-The products are formed by trying all possible combinations of half-reactants.

As a consequence of this reaction scheme a combinatorial explosion in the number of possible combinations is possible. Especially in the case of larger molecules, this mechanism will lead to an inefficient method. The fact that hydrogen is in large excess (99 percent in molecular form), (cfr. mass action law) and acts as a chain terminator, already limits the combinatorial explosion. Note, also that the largest molecules in astrochemistry have only 13 atoms, composed of only a few elements. These astrochemical observations were one of the motivations to first investigate the applicability of AC.

Yet, using the described reaction mechanism, the number of different molecules still increases rapidly. Only 120 different molecules have been observed. The reaction graph (the number of reactions that have been identified as plausible) exceeds 4000 different reactions on a small set of atoms (including ionic and radical reactions). Astrochemical conditions allow us to define a general reaction template:



A , B , C and D are chemical species (ions radicals or neutrals) and the c_i s are the stoichiometric coefficients; additionally:

- $c_i \in \{0, 1\}$, and

- $\sum_{LHS} c_i - \sum_{RHS} c_i = \Delta N$.

That is only monomolecular and bimolecular reactions are allowed because the probability of a three body collision is nil in such extreme dilution conditions (Herbst, 2001). In theory most reactions can be decomposed in elementary reactions of the form of Eq. (1), representing a collision event, in interstellar chemistry, however, elementary reactions and reactions are confounded. Moreover, the classes of reactions derived from the general template reaction (1) and their reaction rates were adopted from the literature (Duley and Williams, 1984). The different reaction classes are summarized in Table 1.

Reactor Algorithm A Given both the molecules and their unique representation, and the reactions in combination with the suggested methods to reduce the combinatorial explosion, we can now address the reaction engine A.

We assume a mixture of molecular types present at certain densities. On this mixture A is executed producing the reaction network dynamics. The algorithm consists of the

following steps:

Copyrighted Material

- for every molecular species present in the sample, the algorithm tests if it can react on its own or if it can react with any of the other species present.
 - if a reaction is thermodynamically possible,
 - then examine whether the specific reaction already exists
 - * if the reaction has already been added to the reaction list,
 - * then reactant and product concentrations are updated
 - * else a new reaction object is created and the products are generated.
 - if the products already exist
 - then the concentrations of reactants and products are updated,
 - else they are added to the list of species and the concentration of reactants are updated.

Although we already introduced some simplifications, the combinatorial explosion resulting from the examination of all possible species combinations can still be huge (Bersini, 2000; Bersini, 1999) if we consider the products as candidates to participate as reactants in further reactions. Mechanisms are required which can determine whether or not a particular reaction can take place. In the next section we for the first time introduce thermodynamics in a meaningful way in astrochemical reactions networks to define a criterion for the automated selection of relevant reactions. The mapping of the *R* and *A* components is done in some detail but the *S* component will not be discussed as there are no formal universal chemical rules to describe it.

Astrochemical Model

Reaction rules *R* The graph rewriting mechanism acts as a generator of diversity and needs an objective, binary criterium that can discriminate between possible and impossible products. We argue that thermodynamics is a candidate to filter the combinatorial explosion faced by the *R* component. In interstellar chemistry, thermodynamics is reputed useless because despite the timescale involved, dilution is such that equilibrium is never attained. Therefore the whole time evolution of the species has to be solved to explain present observations. Furthermore, the set of molecules is given by astronomical observations and the only thing left to determine are the rate constants, this is referred in the literature as solving the problem kinetically (Herbst, 2001). We, in turn, are interested in the more general problem of automatically generating the species themselves from a minimal set of initial molecules or ideally, atoms, thus considering the constructivity of the system explicitly. We do not seek to use a thermodynamic criterium as a predictor of equilibrium concentrations, but rather as an energy constraint to differentiate probable from improbable molecules. In the remaining part of this section we will go in more detail into this

approach applying the principles of thermodynamics to the system at hand along with suitable approximations.

Because we consider a closed thermodynamic system at constant temperature and pressure, spontaneous processes (permitted reactions), are selected on the basis of the sign of the change in the Gibbs free energy function *G*:

$$\Delta G = \Delta H - T\Delta S \quad (2)$$

where ΔH and ΔS are the enthalpy and the entropy change associated with a reaction.

Because we are dealing with an ideal gas, the work of expansion or contraction ($\kappa\Delta N$) must be taken into account:

$$\Delta H = \Delta U + \kappa\Delta N \quad (3)$$

where κ is a constant depending on pressure and ΔN is the change in the number of molecules. In order to apply the suggested thermodynamic criterium, one must satisfy the hypotheses made, particularly the assumption that it is a closed system. We thus include as part of the single reaction system, electrons, cosmic rays and photons as non-massive entities which role is energetic. This leads to the definition of 4 classes of cosmic *events* as potential energy contributions:

- ψ cosmic ray energy,
- $h\nu$ electromagnetic radiation energy,
- E_i first ionization potential energy (we consider only singly ionized molecules or atoms), and
- E_{dp} energy transferred to an interstellar dust particle.

Some simplifying assumptions concerning ΔG can be made. Since the temperature is so low (10-50 K), the entropy term is negligible compared to the internal energy term which is independent of temperature. Also, the internal energy contributions from rotation, translation, and vibration are essentially negligible when compared to the potential energy stored in the chemical bonds E_b .

U is then approximated by the total potential energy E_p , contained in or transferred to a compound :

$$U \approx E_p = E_b + E_i + E_{dp} + \psi + h\nu \quad (4)$$

Note that the possibility of cosmic events essential in order to have a constructive chemistry (accounting for E_b only would favor atoms over molecules). This equation lists only microscopic quantities and should be related to the macroscopic applicability domain of thermodynamics. The correct approach would consist in using statistical mechanics to determine the value of ΔG for a single reaction event, but for the first simulations, a rough approximation consists in defining a free energy variation per reaction event ΔG_r . In this approximation the standard conditions are not molar concentrations, but instead single particles that participate in the reaction:

$$\Delta G_r = \sum_{RHS} E_p - \sum_{LHS} E_p + \kappa\Delta N \quad (5)$$

We can apply the criterium of Tab. 2 in astrochemistry with-

$\Delta G_r < 0$	Exergonic reaction	permitted
$\Delta G_r > 0$	Edergonic reaction	not permitted
$\Delta G_r = 0$	Equilibrium	irrelevant

Table 2: Selection rules, ΔG_r is the free energy change associated with a reaction.

out the fear that a thermodynamically favored product will be kinetically unfavored or vice-versa (see Fig. 3).

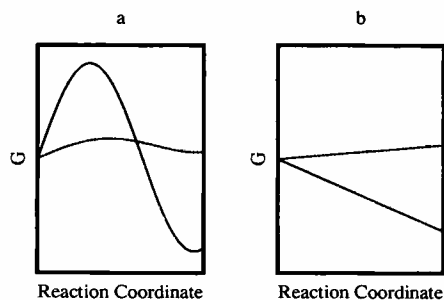


Figure 3: Schematically, starting with one educt on the left of the reaction coordinate axis two structurally possible products are considered a) in general, a thermodynamically unfavored reaction (the reaction leading to the product with the highest G value) can be kinetically favored: energy is available from the environment to pass through maximums along the reaction coordinate, the favored product is thus the one with the lowest maximum (dotted line path), b) at low temperature and pressure as in astrochemistry, if the system is correctly defined as a closed system, there is no possibility of energy exchange with the environment and the only reaction that takes place is the one resulting in a lower G value than the educt (solid line path).

Eq. (4) permits to define conditional reactions. If the bond potential energy change plus the contribution from $\kappa\Delta N$ is sufficient to account for a negative free energy change and if the reactants match reaction class 2 or 3, of Tab. 1 the reaction will exist unconditionally, otherwise, it is tested whether it can occur in the case of a cosmic event for all of the other reaction classes. Once the permitted reactions have been found, further selection is done kinetically, the reaction that can occur is the one that proceeds faster amongst the possible reactions. This does not mean that reactants are replaced by products in the vessel. This latter possibility will be managed by the reactor algorithm A that uses the rates of a reaction to determine if it was effectively reactive.

Reactor Dynamics A The value and thus the sign of ΔG in a macroscopic mixture will vary as a function of the compo-

sition and can be permitted at one time during the simulation and forbidden at another time. Given the macroscopic relation:

$$\Delta G = RT \ln \left(\frac{Q}{K_e} \right) \quad (6)$$

a reaction quotient:

$$Q = \frac{C^{c_3} D^{c_4}}{A^{c_1} B^{c_2}}$$

and a ΔG value computed from Eq. (5) the equilibrium constant K_e , that describes the fixed points of the reaction can be computed. Once it is known, any subsequent value of ΔG can be computed from a particular reaction quotient, which must be provided by the A component that has the information about the time evolution of the abundances.

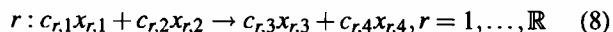
This emphasizes again the importance of the synthetic approach of ACs where the whole problem solution is only accessible through the interaction of all of the components of the tuple.

Introducing the rate coefficient k , the reactor algorithm A uses the following standard differential chemical relations to describe a single reaction:

$$-\frac{dA}{dt} = -\frac{dB}{dt} = \frac{dC}{dt} = \frac{dD}{dt} = kA^{c_1} B^{c_2} \quad (7)$$

Tab. 1 lists the values of k for all reaction classes. It is important to remark that these rates do not have an arrhenius-like temperature dependence ($E_{act} \gg k_b T$) (Herbst, 2001) and are treated as constants.

To count the contributions from all reactions, reaction 1 must be recast as:



where the indices identify the position in the reaction of a compound for every reaction r , if each compound is also labelled in arbitrary order by a superscript and dropping the $c_{r,j}$ s then,

$$\frac{dx^i}{dt} = \sum_{\substack{r \in \mathbb{R} \\ x^i \in RHS}} k_r x_{r,1} x_{r,2} - \sum_{\substack{r \in \mathbb{R} \\ x^i \in LHS}} k_r x_{r,3} x_{r,4}, i = 1, \dots, \mathbb{S} \quad (9)$$

with $x_{r,}$ being x^i 's partner in the reaction's LHS, describes the mathematical problem that has to be solved by explicit simulation of the molecules; or, for comparison, by standard ODE numerical integration routines. The numbers \mathbb{S} and \mathbb{R} describe the size of the list of species and reactions which grow over time. The whole problem amounts to the search of two object spaces that are dependent on the whole trajectory of the dynamic system.

Discussion

Our first goal was to build a realistic model and a new framework for AC developers. In fulfilling the goal set for this paper of making a computational model of a minimal actual chemistry, we have succeeded in constructing an OO model that captures chemistry in general in a realistic way. We have identified the *S*, *R* and *A* components in astrochemistry and have made the necessary abstractions to simplify astrochemistry in order to incorporate it in our AC OO framework. Along the way we have made two contributions: recognizing the importance of a separation between actual and logical representation of chemical structure, and using thermodynamics in astrochemistry in a meaningful way. We have also recognized computer chemistry as a source of knowledge for artificial chemists. At the same time as the conception of a theoretical model, we have developed a set of tools useful to achieve a complete AC. These tools are easily pluggable because they are based on the OO model, created for all type of AC. But the full integration of tools are still in progress and we will present the final software in a future work.

Moreover the objective for astrophysicists would be to predict results (new molecules and reactions) with the new approach of this model: indeed the study of dynamics and meta-dynamics interest greatly the searchers in this specific field. The need of computational model (difficulties to observe species, huge *ab initio* calculations) combined with this new approach go through a necessary simplification. It is a source of new ideas when explications of observations are out of reach (Le Bourlot, 1997). Our future work will focus on validation of the model against astrochemical data and results will be presented in our next paper.

Acknowledgements

The authors would like to acknowledge Jacques Reisse for his insight about the possibility of modelling interstellar chemistry with ACs.

References

- Bauerschmidt, S. and Gasteiger, J. (1997). Overcoming the limitations of a connection table description: A universal representation of chemical species. *Journal of chemical information and computer science*, 37:705–714.
- Benkö, G., Flamm, C., and Stadler, P. (2003). Artificial chemistry based on graph rewriting. In Banzhaf, W., Christaller, T., Dittrich, P., Kim, J. T., and Ziegler, J., editors, *ECAL*, volume 2801 of *Lecture Notes in Computer Science*, pages 10–19. Springer.
- Bersini, H. (1999). Design Patterns for an Object-Oriented Computational Chemistry. In Floreano, D., Nicoud, J.-D., and Mondada, F., editors, *ECAL*, volume 1674 of *Lecture Notes in Computer Science*, pages 389–398. Springer.
- Bersini, H. (2000). Reaction Mechanisms in the OO Chemistry. In Mark A. Bedau, John S. McCaskill, N. H. P. and Rasmussen, S., editors, *Proceedings of the 7th International Conference on Artificial Life*, Portland, Oregon. MIT Press.
- Dittrich, P., Ziegler, J., and Banzhaf, W. (2001). Artificial Chemistries - A Review. *Artificial Life*, 7(3):225–275.
- Dugundji, J., Gillespie, P., Marquarding, D., Ugi, I., and Ramirez, F. (1976). *Metric Spaces and Graphs Representing the Logical Structure of Chemistry*, chapter 6, pages 108–174. Academic Press.
- Duley, W. and Williams, D. (1984). *Interstellar Chemistry*. Academic Press, 24-28 Oval Road, London NW1 7DX.
- Faulon, J.-L. and Sault, A. (2001). Stochastic Generator of Chemical Structure: 3. Reaction Network Generation. *Journal of chemical information and computer science*, 41:894–908.
- Fontana, W. (1996). *The future of science has begun: Approaches to Artificial Life and Artificial Intelligence*, volume 4, Digest On Organization, pages 23–40. Fondazione Carlo Erba.
- Herbst, E. (2001). The chemistry of interstellar space. *Chem. Soc. Rev.*, (30):168–176.
- Kauffman, S. (1995). *At Home in the Universe: The Search for Laws of Self-Organization and complexity*, chapter 3, pages 47–69. Oxford University Press, 198 Madison Avenue, New York, New York 10016.
- Le Bourlot, J. (1997). *Images de la physique*, chapter La modelisation des nuages moleculaires interstellaires, pages 22–30. Département Sciences physiques et mathématiques du CNRS.
- Nicolis, G. and Prigogine, I. (1989). *Exploring Complexity, An Introduction*. W. H. Freeman and Company, New York, fifth edition.
- Schubert, W. and Ugi, I. (1978). Constitutional symmetry and unique descriptors of molecules. *American Chemical Society*, 100(1):225–275.
- Ugi, I., Bauer, J., Bley, K., Dengler, A., Dietz, A., E. Fontain, Gruber, B., Herges, R., Knauer, M., Reitsam, K., and Stein, N. (1993). Computer-Assisted Solution of Chemical problems – The Historical Development and the Present State of the Art of a New Discipline of Chemistry. *Angew. Chem. Intl. Ed.*, (32):201–227.