Explicit Collision Simulation of Chemical Reactions in a Graph Based Artifical Chemistry

Gil Benkö[1,2], Christoph Flamm[3], and Peter F. Stadler[2,3]

[1] Graduiertenkolleg Wissensrepräsentation,
[2] Lehrstuhl für Bioinformatik, Institut für Informatik, Universität Leipzig,
Kreuzstraße 7b, D-04103 Leipzig, Germany
{gil,studla}@bioinf.uni-leipzig.de
[3] Institut für Theoretische Chemie,
Universität Wien, Währingerstraße 17, A-1090 Wien, Austria.
{studla,xtof}@tbi.univie.ac.at

Abstract. A Toy Model of an artificial chemistry that treats molecules as graphs was implemented based on a simple Extended Hückel Theory method. Here we describe an extension of the model that models chemical reactions as the result of "collisions". In order to avoid a possible bias arising from prescribed generic reaction mechamisms, the reactions are simulated in a way that treats the formation and breakage of individual chemical bonds as elementary operations.
Keywords: Artificial Chemistry, Energy Conservation, Elementary Reactions

1 Introduction

A chemical reaction might be regarded as a (quite arbitrarily defined) episode in the life of an aggregate of atoms. Viewing a chemical reaction as a clipping from the collection of atoms' walk through its energy landscape, which is eventually defined by quantum mechanics, we get the following picture: In the beginning, the atom or molecules are localized in an energy well from which they cannot escape by vibrations triggered by the thermic energy. Adding thermic energy, by e.g heating the reaction vessel or radiation, enables the molecule to overcome the barriers surrounding the starting well and "hop" into one of the neighboring wells.

In many cases, the newly reached well is shallow and the molecule possesses sufficient thermic energy to escape again. These kind of wells are often called intermediary states. Eventually, the aggregate of atoms will fall into a well deep enough to be stabilized.

Artificial chemistries simulate molecules by means of matrices, strings, Turing machines, graphs or λ calculus [2, 3, 5, 7, 14, 18], for a recent review see [4]. Interesting *algebraic* theories, in particular a theory of chemical organizations, have been developed based on such model. They all lack, however, a crucial ingredient featured prominently in our sketch of a chemical reaction above: they

lack a natural energy function. In earlier work [1], we have implemented an artificial chemistry at an intermediate level of reality and computational complexity. The algorithms used here are derived from computational chemistry but simplified to a very high degree. This level of abstraction was chosen rather than more accurate or more general models for its balance between the computational tractability to guarantee a fast toy chemistry simulation and the fact that important properties of chemistry are still retained. By building on the simple Extended Hückel Theory (EHT) method [11], it provided a wave function and easily energy as a state function. This is required for the "look-and-feel" of chemistry as a constructive system with combinatorial production of new molecules.

Knowledge-based models, in contrast, would be biased due to the intrinsic sampling bias of chemical databases as well as very expensive due to access fees. An explicit description, however, requires a natural definition of mass and, in particular for chemical reactions, energy conservation. This is ensured by representing molecules by graphs and defining energy as a state function. The difference in energy is then the driving force for chemical reactions. A more complete picture of chemical reactions must in addition include the calculation of activation barriers. Here we describe an extension to the Toy Model formalism, inspired by the ideas above, allowing the simulation of chemical reactions in line with the simplicity of the EHT method.

2 The Model

Basis

The Toy Model uses a simplified extended Hückel theory method. Only the connectivity of the molecule is required as input. A basis set consisting of $1s$ for H and $2sp^n$-hybridized Slater-type orbitals for other atoms is used. Valence-shell electron pair repulsion (VSEPR) theory [10] is used to determine the hybridization of an atom.

An overlap matrix and Hamiltonian matrix is set up for a molecule using fixed parameter values that are zero for non-bonded atoms. For bonded atoms, the values are tabulated according to the atoms involved, their hybridization, and their type of interaction. The secular equation relates the matrices to the wave function. By solving the equation we get this function, from which in principle any molecular property might be calculated. In particular, we calculate the spectrum, charges, energy, and, in combination with a spectral embedding, the dipole and solvation energy. The Toy Model comes with parameter values for C, H, N, O, P, and S, and can be easily extended by editing the parameter file. The calculation takes into account σ bonds, π bonds, backbonding and hyperconjugation through indirect sp^n-sp^n and sp^n-p overlaps, banana overlaps in rings, and stronger backbonding in P and S. In line with [8] d orbitals are seen as polarization functions, which we "replace" in our simplified framework by stronger backbonding in P and S.

Extension to chemical reactions

We extended the Toy Model to chemical reactions by decomposing a reaction into small "moves": the formation, breaking, or shifting of bonds, in analogy to the elementary moves of the Dugundji-Ugi model [5, 15]. We took into account both single bond formation and simultaneous formation of two bonds, as there is for example controversy about whether the Diels-Alder reaction (see below) proceeds in a concerted fashion (simultaneous bond formation) or via a diradical (sequential bond formation). The same applies for single and double breaking of bonds. Finally, also the shifting of a bond was included in our simulation. The sequence of moves is determined by a simulation based on a continuous time Monte Carlo method proposed by [9]:

```
loop
    if reaction is monomolecular (one reacting molecule) then
        find neighbors produced by single/double formation, single/double breaking, shifting
    end if
    if reaction is bimolecular (two reacting molecules) then
        find neighbors produced by single/doubleform, shifting
    end if
    if all neighbors are of higher energy then
        stop
    end if
    move to neighbor with lowest energy
    if this neighbor is a split molecule then
        stop
    end if
    find neighbors produced by single/double formation, single/double breaking, shifting
end loop
```

Interaction of orbitals

We now describe how the selection of bonds that may form during a move was restricted. Atomic orbitals were divided into four types in view of possible interactions and thus bond formations: the sp^3 orbital as it features a particularily accessible back lobe (b), the s orbital, the p orbital, and the orbital occupied by a lone electron pair or lone electron (l). The b type is important for its role in the S_N2 reaction (nucleophilic substitution): there a l type orbital interacts with the back lobe of a sp^3 orbital and eventually creates a new bond.

We further define that only the interactions $p-l$, $l-s$, $p-p$, $p-s$ and $b-l$, shown in Fig. 1, are possible between the orbital types. This "sophisticated guess" is based on the importance of those interactions in common organic reactions, as shown in any standard organic chemistry textbook [19] and in Orbital Interaction Theory [6, 13, 16]. In Fig. 1(right column), we show for example the

interactions in the hydroboration, the $E1$ (elimination), the Diels-Alder, the hydrogen shift, and the aforementioned S_N2 reaction.

The possible interactions according to the hybridization of the two interacting atoms are summarized in Tab. 1. Fig. 2 shows some allowed interactions for an example pair of molecules.

The bonds that may break are not restricted to special types. However, we take into account that bond fission can be homolytic or heterolytic. Thus there are three cases: (1) the two fragments each inherit one electron of the bond (radical formation); (2) the first fragment receives both electrons; (3) the second fragment received both electrons (ion formation). The shifting of a bond is simulated in our model by simultaneous fission (unrestricted) and formation (restricted by the allowed interactions above) of a bond.

Table 1. Table of allowed interactions

Atom hybridization and type of orbital 2		Atom hybridization and type of orbital 1: sp^3		sp^2		sp		s
		b	l	p	l	p	l	
sp^3	b		•		•		•	
	l	•		•		•		•
sp^2	p		•	•		•		•
	l	•						•
sp	p		•	•		•		•
	l	•						•
s			•	•	•	•	•	

3 Examples

Chemical reactions were simulated for two sets of reacting molecules: first, ethene and butadiene, and second, the enolate of ethanal (acetaldehyde). We generated all molecular configurations (products) accessible from the reactant(s) within our framework, and whose energies were lower than or within 3 kcal/mol of the energy of the reactant(s).

For ethene and butadiene, both molecules can react by themselves (monomolecular reactions), with another specimen of the same type, or with each other (bimolecular reactions) (Tab. 2). As expected, one of the reactions occuring is the formation of a ring, called the Diels-Alder reaction (DA).

By themselves, the reactants or two ethenes do not form products of lower energy, i.e., they do not react without energy input from outside (first three parts of Tab. 2). The kinetic energy or the energy stored in molecular rotation or vibration may however suffice in dependence of the temperature to attain products of higher energy. Those within 3 kcal/mol are hydride shifts and are artifacts of the model and its parameters which favor sp^3 over sp^2 atoms.

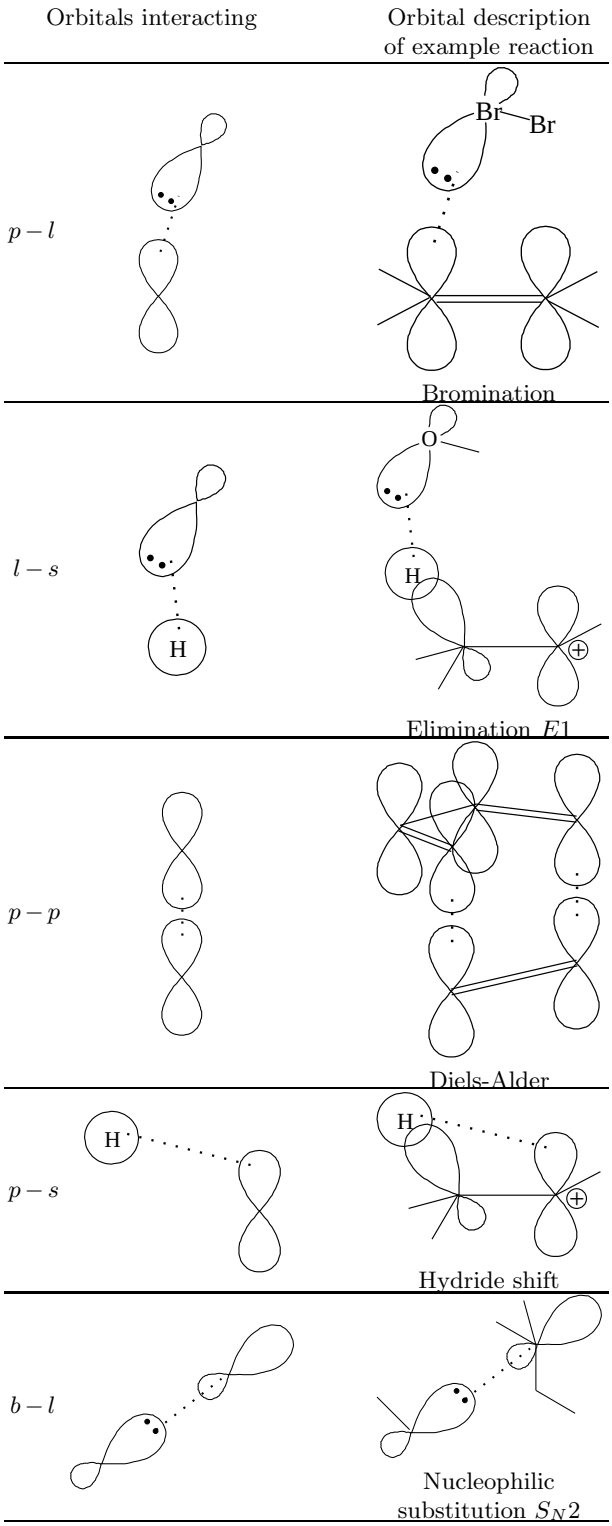

Fig. 1. Allowed interactions. Sketches follow [19].

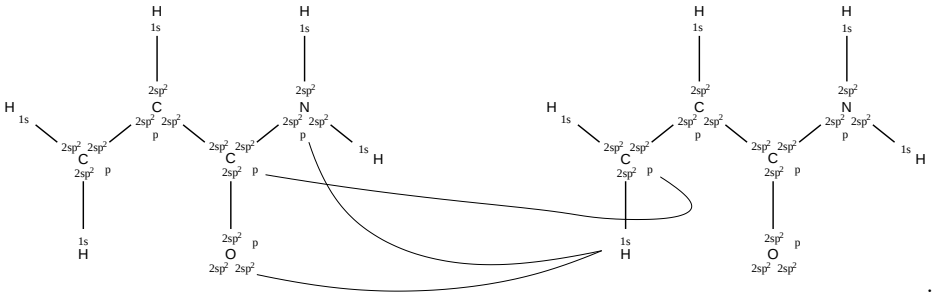

Fig. 2. Selected allowed interactions between two propenamide molecules. The molecules are shown with their atoms surrounded by their valence orbitals. Curved lines indicate some of the allowed interactions: *p*-*p*, *p*-*s*, and *p*-*l* (the sp^2 orbital on O is occupied by a lone pair and thus a *l*-type orbital).

Table 2. Reactions of ethene and butadiene. Molecules are written in the SMILES notation [20], which is very similar to a structural formula. Diels-Alder educts are indicated by (DA). Energies are total atomization energies

	Structural formula	Energy (kcal/mol)
Reactant	C(H)(H)=C(H)(H)	-415.95
Products	none (no reaction)	
Reactant	C(H)(H)=C(H)C(H)=C(H)(H)	-789.62
Products	C(H)(H)(CC(H)=C(H)H)H	-787.86
(of higher	C(H)(H)(C(H)[C-]=C(H)H)[H+]	-787.88
energy)	C(H)(H)([C-]C(H)=C(H)H)[H+]	-787.86
Reactants	C(H)(H)=C(H)(H) + C(H)(H)=C(H)(H)	-831.90
Products	none (no reaction)	
Reactants	C(H)(H)=C(H)C(H)=C(H)(H) + C(H)(H)=C(H)C(H)=C(H)(H)	-1579.24
Products	C1(H)(H)C(H)C(H)C(H)(H)C(C(C(C1(H)H)H)H)(H)H (DA)	-1588.74
	C(H)(H)=C(H)C1(H)C(H)(H)C(C(C(C1(H)H)H)H)(H)H (DA)	-1587.39
Products	C(H)(H)(C(H)C(H)=C(H)H)[C+](C(H)H)C(H)=C(H)H + [H-]	-1577.75
(of higher	C(H)(H)[C+](C(H)=C(H)H)C(H)(H)C(H)C(H)=C(H)H + [H-]	-1577.83
energy)	C(H)(H)[C+](C(H)=C(H)H)C(C(C(=C(H)H)H)H)(H)H + [H-]	-1577.42
	C(H)(H)=[C-]C(H)=C(H)H + [H+]C(H)(H)C(H)C(H)=C(H)H	-1577.49
Reactants	C(H)(H)=C(H)(H) + C(H)(H)=C(H)C(H)=C(H)(H)	-1205.57
Product	C1(H)(H)C(H)(H)C(C(C(C1(H)H)H)H)(H)H (DA)	-1222.33

On the other hand, butadiene can react with either ethene or another butadiene (last two parts of Tab. 2). It forms rings by two simultaneous bond formations bridging twice the gap between two molecules. This Diels-Alder reaction has two possible outcomes for the reaction of butadiene with itself (regioisomers). Again, products of higher energy include artifacts like hydrogen shifts or formation of hydride.

Table 3. Reactions of the enolate of ethanal. Aldol condensation products are indicated by (ac).

	Structural formula	Energy (kcal/mol)
Reactant	[C-](H)(H)C(H)=O	-667.48
Product	[C-](H)(H)([C-]=O)[H+]	-676.65
Reactants	[C-](H)(H)C(H)=O + [C-](H)(H)C(H)=O	-1334.95
Products	[C-](H)(H)[C-]=O + [H+][C-](H)(H)C(H)=O	-1344.14
	[C-](H)(H)(C(H)=O)[C+]([C-](H)H)O + [H-]	-1343.72
	[C-](H)(H)[C+](O)[C-](H)(H)C(H)=O + [H-]	-1343.88
	[C-](H)(H)C(H)(O)[C-](H)(H)C(H)=O (ac)	-1336.94
	[C-2](H)C(H)=O + [H+][C-](H)(H)C(H)=O	-1335.50
Products (of higher energy)	[C-](H)(H)(C(H)=O)C(H)C(H)=O + [H-]	-1332.64

For the enolate of ethanal (Tab. 3), one of the reactions is the expected bond formation between two enolates called the aldol condensation (ac). However, here also the other products of lower energy are artifacts (hydrogen shift or hydride formation).

4 Concluding remarks

The model presented here tries to build a formalism of chemistry with the least possible bias. To this end, we choose a simple molecule representation using graphs which still allows an energy calculation. On the other hand, we use a rather drastic abstraction from reality. While the approach is still based on quantum chemistry, it avoids e.g. the the complications spatial embedding in order to keep the resource requirements tractable even for large scale simulations. Chemical reactions are implemented here not in the forms of prescribed rewrite rules as in [1]. Instead, the model relies on a decomposition into bond formations, fissions, and shifts. The result is an artificial chemical reaction which reproduces some of the experimental results and features energy dependence, reaction specificity, and multiple possible reaction outcomes.

In the present implementation, we only looked at the thermodynamic description of a chemical reaction. In future work, we intend to include the kinetic constraints, in particular by incorporating the reaction barriers calculated using the Klopman-Salem equation [12, 17]. Since our basic model included also the

calculation of solvation energies, we will also take into account solvation effects on chemical reactions.

The simulation results show that our model is indeed capable of producing the chemically expected reaction pathways despite drastic simplifications of this artificial chemistry relative to "real" quantum chemistry. Due to these simplifications, artifacts (relative to reality) are to be expected and indeed to occur. However, these artifacts still conform to the 'look-and-feel' of chemistry, and could be avoided by choosing different parameters — probably at the expense of opening up other reaction pathways that do not occur in reality. We emphasize, that our model is intended as an *artificial chemistry* model, in which not all reactions can occur indiscriminately, but depend on energy, so that multiple reaction outcomes are possible and have different rates that determined within the model.

Acknowledgments

This work was supported by the Graduiertenkolleg Wissensrepräsentation, Universität Leipzig, COST Action D27, and the Bioinformatics Initiative of the DFG.

References

1. Details removed for the purpose of double blind review.
2. R. J. Bagley and J. D. Farmer. Spontaneous emergence of a metabolism. In C. G. Langton, C. Taylor, J. D. Farmer, and S. Rasmussen, editors, *Artificial Life II*, Santa Fe Institute Studies in the Sciences of Complexity, pages 93–141, Redwood City, CA, 1992. Addison-Wesley.
3. W. Banzhaf, P. Dittrich, and B. Eller. Self-organization in a system of binary strings with spatial interactions. *Physica D*, 125:85–104, 1999.
4. P. Dittrich, J. Ziegler, and W. Banzhaf. Artificial chemistries — a review. *Artificial Life*, 7:225–275, 2001.
5. J. Dugundji and I. Ugi. Theory of the *be*- and *r*-matrices. *Top. Curr. Chem.*, 39:19–29, 1973.
6. I. Fleming. *Frontier Orbitals and Organic Chemical Reactions*. John Wiley, 1976.
7. W. Fontana. Algorithmic chemistry. In C. G. Langton, C. Taylor, J. D. Farmer, and S. Rasmussen, editors, *Artificial Life II*, pages 159–210, Redwood City, CA, 1992. Addison-Wesley.
8. D. J. Gilheany. No d orbitals but walsh diagrams and maybe banana bonds: Chemical bonding in phosphines, phosphine oxides and phosphonium ylides. *Chem. Rev.*, 94:1339–1374, 1994.
9. D. T. Gillespie. Exact stochastic simulation of coupled chemical reactions. *J. Phys. Chem.*, 81:2340–2361, 1977.
10. R. J. Gillespie and R. S. Nyholm. Inorganic Stereochemistry. *Quart. Rev. Chem. Soc.*, 11:339–380, 1957.
11. R. Hoffmann. An Extended Hückel Theory. I. Hydrocarbons. *J. Chem. Phys.*, 39(6):1397–1412, 1963.
12. G. Klopman. Chemical reactivity and the concept of charge- and frontier-controlled reactions. *J. Am. Chem. Soc.*, 90:223–243, 1968.

13. G. Klopman. *Chemical Reactivity and Reaction Paths.* Krieger, 1974.
14. J. S. McCaskill and U. Niemann. Graph replacement chemistry for DNA processing. In A. Condon and G. Rozenberg, editors, *DNA Computing*, volume 2054 of *Lecture Notes in Computer Science*, pages 103–116. Springer, Berlin, D, 2000.
15. G. Nowak. Common-sense reasoning cast over D-U model in simulation of chemical reactions. *Int. J. Quantum Chem.*, 84(2):282–289, 2001.
16. A. Rauk. *Orbital Interaction Theory of Organic Chemistry.* Wiley-Interscience, 2000.
17. L. Salem. Intermolecular orbital theory of the interaction between conjugated systems. I. General theory; II. Thermal and photochemical calculations. *J. Am. Chem. Soc.*, 90:543–552 & 553–566, 1968.
18. M. Thürk. *Ein Modell zur Selbstorganisation von Automatenalgorithmen zum Studium molekularer Evolution.* PhD thesis, Universität Jena, Germany, 1993. PhD Thesis.
19. K. P. C. Vollhardt and N. Schore. *Organic Chemistry.* W. H. Freeman, 4th edition, 2002.
20. D. Weininger. SMILES, a chemical language and information system. *J. Chem. Inf. Comput. Sci.*, 28:31–36, 1988.