

# Molecules from the Minkowski space

## An approach for 3D structure prediction

Ö. Farkas<sup>1</sup> and G. Imre<sup>1,2</sup>

(1) Department of Organic Chemistry, Eötvös Lóránd University, 1/A Pázmány Péter St, Budapest H-1117, Hungary

(2) Department of Automation and Applied Informatics, Budapest University of Technology and Economics, Goldman György sq. 3., Budapest, H-1521, Hungary

E-mail: farkas@organ.chem.elte.hu, imreg@organ.chem.elte.hu

### INITIAL CONDITIONS - input

The purpose of the present algorithm is the automatic, human interaction free generation of complete three dimensional (3D) atomic coordinates for small or medium sized molecules. The process requires the same level of structural information than a chemical formula can represent.

The minimal set of given structural information must contain the participating atoms and bonds (and bond orders) between them.

Although, it is required to determine the hybridization state (relative spatial layout of the ligands) for atoms and the desired length and type (single/double/triple or aromatic) for bonds, it happens at an automated preprocessing stage. The structure generation will try to keep up with these requirements.

It is possible to specify chirality information for chiral atoms and stereo information for double bonds.

### ABSTRACT

In the field of theoretical chemistry it is usual to have only a partial set of structural information about compounds, like the connectivity or the formula. Individual studies can easily be performed using "human interfaces" for building input structures. However, automatic, "batch" processes cannot be applied on a large number of molecules if they imply human intervention. Studies, like QSAR, pharmacophore analysis, reaction prediction might need full, complete 3D information for the compounds of interest. The wide-spread tools used for structure determination (force-fields or quantum chemical methods) even needs a complete set of initial 3D structure.

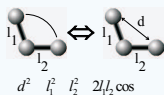
Our approach intends on generating globally valid set of 3D coordinates for small and medium sized molecules, based on local structural criteria. Over against iterative, backtrack based structure predicting algorithms, our method is capable of satisfying partially inconsistent requirements. Such situations are incidental to examining structures holding polycyclic, rigid details.

Goals mentioned above can be achieved using coordinates interpreted in a Minkowski metric. Our coordinate assignment process is divided into the following parts: (I) Automatic generation of distance criteria based on chemically relevant local properties, such as bond stretches, bond angles, dihedral angles, etc. (II) Multi dimensional coordinate assignment which fulfills all of the criteria. (III) Geometry optimization using a force field extended to multi-dimensional Minkowski metric. The optimization eliminates the over-3D components and yields the 3D coordinates.

### INTERATOMIC DISTANCES

Some heuristic used when estimating inner coordinates:

- Torsion angles in chains are set to 180
- Torsion and bond angles in aromatic rings are set according to the ring size



We transform the input data into internal coordinate requirements which leads to a subset of the interatomic distances. (As an extra criterion avoids the proximity of non-bonded atoms by specifying a minimal distance between them if necessary.)

Some of the internal coordinates can be estimated directly (bond lengths are available, bond angles can be associated to hybridization states) or through a selection (dihedral), while most of them remain unknown in this phase.

The next placement step will use the predicted interatomic distance set.

### THE SOUGHT CONFORMER

The ideal coordinate assignment would satisfy all of the distances generated in the previous step. In most of the cases, they can contain contradictions (because of ring closure or tension, for example).

The assignment tries to make a compromise with them by minimizing the total energy originating from the deviation between the actual and desired internal coordinates.

By our approach the assignment is done in two parts: at first, we calculate the coordinates in a multi dimensional Minkowski space, where all of the distance requirements can be satisfied, then a geometry optimization collapse it into 3 (real) dimensions.

The placement algorithm orders the atoms, then assign coordinates for the first two of them, producing 1D system.

All of the atoms will be placed one-by-one by evaluating the distance requirements with triangulation. When triangulation fails then a refinement step will add two extra dimension to the coordinates (with +1 and -1 metric). Two singular directions in the subspace of added dimensions will be designated.

By moving the atom under placement to one of them and the other placed atoms to the direction of other one, the mismatched distances corrected while inner distances in the previously placed subset are not affected.

### MINKOWSKI PLACEMENT

The square of distance (metrid) between two points determined by the Minkowski metric:

$$d^2 = \begin{pmatrix} 1 & 0 & 0 & \dots & 0 \\ a & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 & \dots \\ m_1 & 0 & 0 & \dots & 0 \\ 0 & m_2 & \dots & \vdots & m_1 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & m_2 & \dots & 1, 1 \end{pmatrix}$$

The special nature of this metric is the presence of singular directions indicated by the negative metric components: these directions are placed on a multi-dimensional cone surface. The distance (metrid) along these directions is zero, consequently differing coordinate values do not necessarily induce non-zero distance.

Multi dimension Minkowski placement can be done with a direct algorithm executing placement and refinement steps for each atom.

### OPTIMIZATION TO 3D

A simple force field can be constructed based on the distance criteria found:

The energy components derived from the distances using spring model:

$$E_s = \frac{1}{2} \sum_i (D_i - D_{required,i})^2$$

The energy components associated with higher dimensions:

$$E_h = \frac{1}{2} \sum_j (x_j)^2$$

In an analog way to classical force fields, it is possible to define an energy value for the structure with Minkowski coordinates. The energy components can be calculated using local sets of atom-atom distances - which are also available in the Minkowski space.

By adding extra energy components along over-3D directions, which slightly pulls the structure to the origo, using geometry optimization the structure can be forced to 3D.

The modified force field converges to the original one during the elimination of extra coordinates. The optimization process, therefore, leads to a valid 3D structure by means of the original force field.

### ACKNOWLEDGEMENTS



Department of Organic Chemistry,  
Eötvös Lóránd University  
Budapest, Hungary



Department of Automation and Applied Informatics,  
Budapest University of Technology and Economics  
Budapest, Hungary



ChemAxon Ltd  
<http://www.chemaxon.com>

### REFERENCES

Gasteiger, J.; Rudolph, C.; Sadowski, J. *Tetrahedron Comp. Method.* **1990**, *3*, 537-547

Sadowski, J.; Gasteiger, J. *Chem. Reviews* **1993**, *93*, 2567-2581

Stephen L. Mayo; Barry D. Olafson; William A. Goddard III. *J. Phys. Chem.* **1990**, *94*, 8897-8909