Applied Medical Informatics
Original Research - Computer Application Vol. 29, No. 3/2011, pp: 11-18

# Computer Assisted Geometry Optimization for *in silico* Modeling

**Lorentz JÄNTSCHI** [1,2,*]

[1] Technical University of Cluj-Napoca, 28 Memorandumului, 400114 Cluj-Napoca, Romania..
[2] University of Agricultural Sciences and Veterinary Medicine Cluj-Napoca, 3-5 Mănăştur, 400372 Cluj-Napoca, Romania.
E-mail: lorentz.jantschi@gmail.com

* Author to whom correspondence should be addressed; Tel.: +4-0264-401775; Fax: +4-0264-401-768.

**Received: 15 August 2011 /Accepted: 15 September 2011 / Published online: 22 September 2011**

**Abstract**
This paper aimed to present a methodology for the geometry optimization of molecules in order to analyze structure-activity/property relationships. The main objective was to reduce the probability of carrying out misleading analyses while preparing the molecules for modeling. An algorithm that used HyperChem 8.0 software was developed and implemented in a series of PHP programs. A sample of 319 drug-like compounds was used to test the implemented algorithm. The program executed the geometry optimization protocols efficiently and quickly on both small and large sample sizes. The results obtained in the investigated sample of 319 compounds showed that our programs performed well on the investigated compound collection thus increasing the reliability of compound preparation for quantitative structure-activity relationships analysis.

**Keywords:** Computer Assisted/Automated System; Geometry Optimization; *in silico* Modeling; Accuracy.

## Introduction

The molecular structures are optimized at the beginning of any structure-activity/property relationship modeling approach (mathematical representation of the activity/property of molecules in relation with their physical and chemical properties [1]). Energy minimization and geometry optimization are both applied on molecular structures before computing the molecular descriptors.

A series of methods are applied to optimize the geometry of molecules: empirical force field methods (molecular mechanics, a less expensive method in terms of computational speed, able to provide excellent structural parameters), semi-empirical methods (that can solve the Schrödinger equation, with certain approximations, in order to describe the electron properties of atoms and molecules), and ad initio methods (e.g. Hartree-Fock, Post-Hartree Fock, Density Functional Theory).

The characteristics of some molecular mechanics force fields (known to neglect electrons) and their applicability domains are as follows:

- MM+: developed as an extension of MM2 [2] for organic molecules [3].
- AMBER (Assisted Model Building with Energy Refinement): a family of force fields for molecular dynamics (biomolecules) [4-6].
- BIO+: a method similar with the CHARMM (Chemistry at HARvard Molecular Mechanics)

functional form developed for proteins [7].

- OPLS (Optimized Potential for Liquid Simulations): a molecular mechanics method similar to AMBER for chemical calculations developed for proteins and nucleic acids (it reproduces liquid properties such as heat of vaporization and density) [8,9].

The semi-empirical methods (calculations are faster compared to ad ignition methods) are known to neglect core electrons. The characteristics of some semi-empirical methods are as follows:

- CNDO, INDO, NDDO: rarely used today but their methodology constitutes the basis of newer methods [10].
- MINDO, MNDO [11], AM1 [12], PM3 [13,14]: use parameters to fit the experimental heats of formation, dipole moments, ionization potentials and geometries. The Austin Model 1 eliminates the problem of MINDO (tendency to overestimate repulsions between atoms). The PM3 method is the third parametrization of MNDO and it uses parameters derived by "chemical intuition".
- ZINDO [15] and SINDO [16-18]: calculate excited states and predict electronic spectra.
- Semi-empirical extended Hückel [19]: used for single point calculations to produce quantitative and semi-quantitative descriptors.

*Ad initio* methods fully count the electrons and use only theoretical principles to solve the Schrödinger equation associated with the molecular Hamiltonian [20].

Energy minimization and geometry optimization are essential tasks in molecular modeling and drug design. The failure to minimize energy and/or optimize geometry is directly translated into incorrect molecular descriptors and thus into unreliable QSAR/QSPR (quantitative Structure-Activity Relationship / quantitative Structure-Property Relationship) models. The probability of geometry optimization to fail is directly related to the number of compounds in the dataset. This probability increases as the number of compounds in the data set increase. The aim of the research was to develop a methodology able to carry out optimal geometry optimization of biologic active compounds and to implement as well as assess the proposed methodology.

## Material and Method

The PHP (Hypertext Preprocessor) language was used in order to develop the automated geometry optimization application. The application was developed to use the functions implemented in HyperChem (Version 8.0) under Windows operating system.

### *Dataset for Design*

Data previously used to validate a QSAR model were used to design and validate procedures for the proposed algorithm [21]. The sample comprised 319 drug-like molecules belonging to classes of compounds with different structures, which are known as active (BBB+, BBB = blood-brain barrier) or not active (BBB-) compounds in the central nervous system (CNS).

The PubChem (http://pubchem.ncbi.nlm.nih.gov/) database was used to download the *.sdf files of the BBB compounds. The name of the drug-like molecule, the PubChem compound identifier (CID) and active/not active status in the CNS are presented in ***Supplementary material***.

### *Algorithm*

The main objective of the research was to create an automated application able to perform geometry optimization of biological active compounds that are subject of QSAR/QSPR analyses. The 3-step protocol put forward for implementation contains is:

- Turn the 2-dimensional sketch of the molecule into a 3-dimensional structure: [Build – Add H & Model Build].
- Define a model of molecular mechanics (AMBER were defined) and optimize the molecular structure by applying the AM1 method, Polak-Ribiere algorithm.
- Minimize the energy of the molecule.

The protocol is applied to each molecule in the dataset as a first procedure in QSAR/QSPR analyses.

Figure 1 presents the algorithm proposed for the automated optimization of molecular geometry.

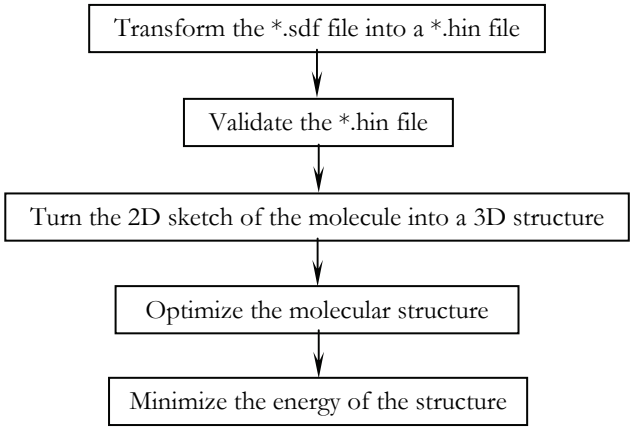



**Figure 1.** Protocol for automated optimization of molecular geometry

The following parameters were used:

1. Apply the molecular mechanics method: AMBER
2. Optimize the molecular geometry: ▪ Scf-convergence = 1e-8; ▪ Max-iterations = 100; ▪ Optim – max – cycles = 1000; ▪ Optim – coverage = 0.01; ▪ Optim algorithm = PolakRibiere [22]; ▪ Query-value optim-converged.
3. Compute partial charges: ▪ Semi-empirical-method = AM1, ▪ Do-single-point.
4. Save the results in Job_Results.txt file: ▪ Value of optim-converged = True/False, ▪ Value current-file-name = name of the file, ▪ Value total-energy = value (kcal/mol), ▪ Value dipole-moment = (Debyes) - Sum(Point-Chg., sp Hybrid, pd Hybrid), ▪ Value scf-atom-energy = Isolated Atomic Energy (kcal/mol), ▪ Value scf-binding-energy = Binding Energy (kcal/mol), ▪ Value scf-core-energy = Core-Core Interaction (kcal/mol), ▪ Value scf-electronic-energy = Electronic Energy (kcal/mol), ▪ Value heat-of-formation = Heat of Formation (kcal/mol)

*Validation.*

The analysis of the developed programs in terms of speed and accuracy was first performed on a small sample and later on the 319 data set of BBB compounds. The program was assessed with Intel® Core™2 Quad CPU, 2.66 GHz, 2 GB RAM using the HyperChem 8.0 for Windows.

A series of parameters were counted and analyzed in order to assess the implemented algorithm: (1) overall time needed for optimization; (2) optimization performance (in terms of accuracy expressed as the ratio between the number of optimized molecules and the total number of molecules in the sample); (3) the numbers of runs needed to optimize all compounds in the sample.

The program was first run for all the molecules in the sample (319 compounds). For non-optimized compounds, the program was run again but this time it only had to perform geometry optimization.

## Results

*Program Description*

A series of steps are applied for obtaining the information of a 3D structure: define a model of molecular mechanics, optimize the molecular geometry until convergence with a specified algorithm, define the semi-empirical method for energy calculation and obtain a series of energetic parameters that characterize the optimum geometry as well as the electrical partial charges (see

Figure 2). The proposed method applied the AMBER mechanical model [4], the Polak-Ribiere optimization algorithm [22] and the AM1 semi-empirical method [12] for energy calculations. The results of optimization and energy minimization were saved in a *.hin file which defined the chemical structure of each model built (a structure with many rows, one row for each atom of molecule (number, name, partial charge, spatial coordination, number and type of bounds)).

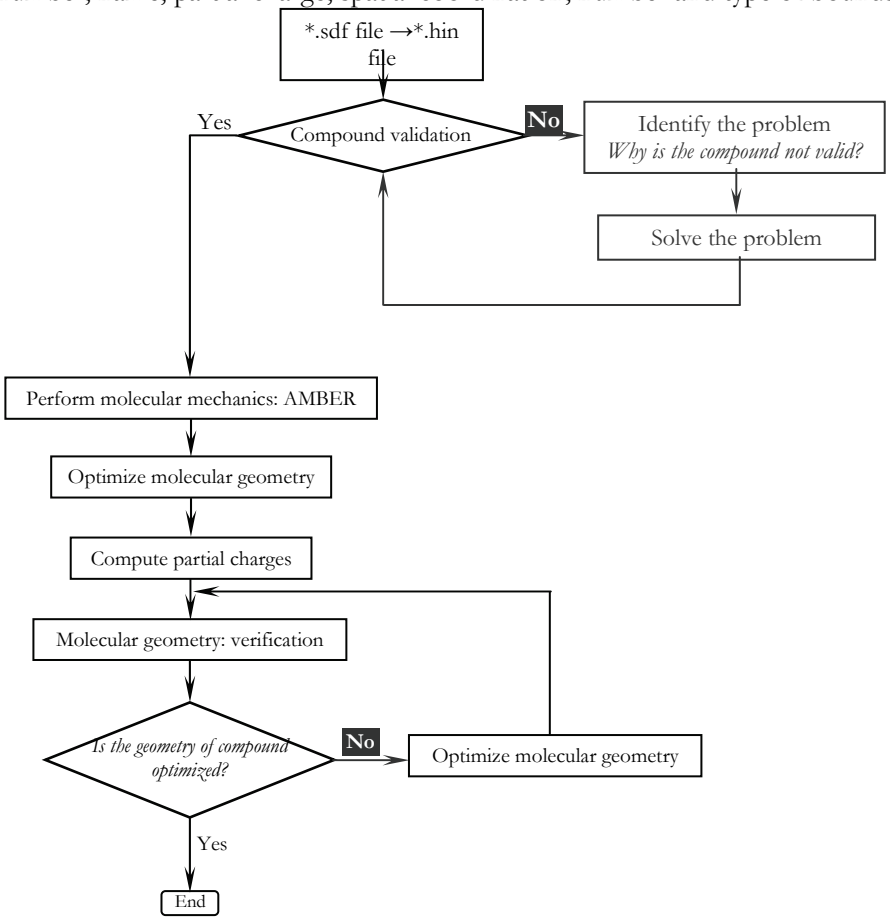



**Figure 2.** Process flow for optimizing the geometry of the compounds

Two changes could be observed on the *.hin file after applying the proposed algorithm for optimization geometry: the atoms and the position of the atoms in the structure (Figure 3).

The *.hin structural files before and after applying the developed and implemented algorithm are presented in Figure 4.

A PrintScreen from the results opened with Microsoft Excel is presented in Figure 5.

The time needed to optimize the geometry depends on computer hardware. The implemented algorithm runs properly with HyperChem 8.0. The program sources are available upon request (a PHP program that generates HyperChem script files and runs the HyperChem application once).

*Analysis of Program Runs*

The analysis of the program was performed while implementing a sample of three molecules drawn randomly from the sample. The program was applied in order to optimize the sample of 319 compounds after release of its final form.

The first compound in the sample was optimized in 2.20 minutes. The average time needed to optimize the geometry was of almost 8 minutes. The overall time needed for the entire sample was of 42.72 hours. The shortest time was needed for molecules with a small number of atoms and a less complex structure.

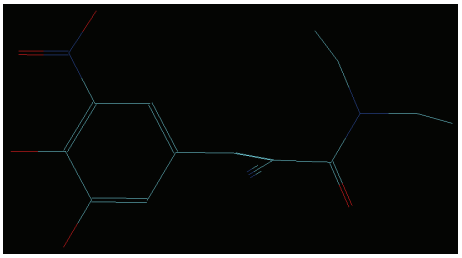
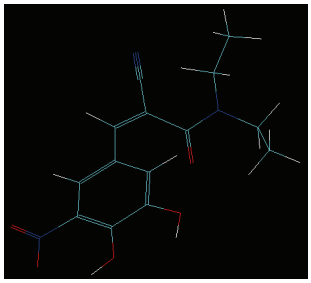
(a)  (b) 

**Figure 3.** Structural refinement of molecule before (a) and after (b) applying the algorithm

(a)

```
forcefield amber94
sys 0 0 1
view 40 0.21229 40 15 1 0 0 0 1 0 0 0 1 -5.6344 -3.0066 -40
seed -1111
mol 1
atom 1 - O ** - -0.56 0.9362 -3.0997 0.2653 2 11 s 13 s
atom 2 - O ** - -0.68 1.9329 -0.0026 0.6466 1 10 s
atom 3 - O ** - -0.68 0.0690 -1.4104 -2.0674 1 12 s
atom 4 - O ** - -0.68 2.4710 -3.3340 -3.0317 1 14 s
atom 5 - N ** - 0.05 -0.9268 -2.0371 1.2756 3 11 s 15 s 16 s
atom 6 - N ** - -0.57 -3.0699 -2.6098 1.0540 2 16 d 17 s
atom 7 - N ** - -0.57 -1.0277 -0.8580 3.3969 2 15 d 19 s
atom 8 - N ** - -0.62 -3.2996 -0.8292 4.2877 2 18 s 19 d
atom 9 - N ** - -0.9 -5.1379 -1.8581 3.1524 1 18 s
atom 10 - C ** - 0.28 0.8306 -0.7087 0.0818 3 2 s 11 s 12 s
atom 11 - C ** - 0.54 0.4981 -1.9183 0.9600 3 1 s 5 s 10 s
atom 12 - C ** - 0.28 1.2289 -1.3146 -1.2508 3 3 s 10 s 13 s
atom 13 - C ** - 0.28 1.7482 -2.6902 -0.8476 3 1 s 12 s 14 s
atom 14 - C ** - 0.28 1.6947 -3.7592 -1.9233 2 4 s 13 s
atom 15 - C ** - 0.11 -1.5653 -1.5465 2.3788 3 5 s 7 d 17 s
atom 16 - C ** - 0.04 -1.8721 -2.6667 0.5100 2 5 s 6 d
atom 17 - C ** - 0.23 -2.8927 -1.9125 2.2225 3 6 s 15 s 18 d
atom 18 - C ** - 0.41 -3.7738 -1.5280 3.2283 3 8 s 9 s 17 d
atom 19 - C ** - 0.47 -1.9756 -0.5418 4.3059 2 7 s 8 d
endmol 1
```

(b)

```
forcefield amber94
sys 0 0 1
view 40 0.15771 55 15 1 0 0 0 1 0 0 0 1 0.30214 1.3994 -56.013
seed 0
mol 1
atom 1 - O OS - -0.2685631 -0.3191915 -3.263214 -0.4359262 2 11 s 13 s
atom 2 - O OH - -0.3055244 2.462053 -1.91374 -1.020778 2 10 s 20 s
atom 3 - O OH - -0.3286266 2.074043 -3.738041 1.932041 2 12 s 21 s
atom 4 - O OH - -0.3174937 -1.474208 -5.542453 0.7886236 2 14 s 22 s
atom 5 - N NT - -0.209632 -0.4680514 -1.287082 0.8758315 3 11 s 15 s 16 s
atom 6 - N NB - -0.1122555 -1.269061 -0.8435638 2.998311 2 16 d 17 s
atom 7 - N NC - -0.2074309 -2.094369 0.1446743 -0.4152072 2 15 d 19 s
atom 8 - N NC - -0.2681285 -3.680485 1.326119 1.050464 2 18 s 19 d
atom 9 - N N2 - -0.3568287 -3.608629 1.116663 3.430436 3 18 s 23 s 24 s
atom 10 - C CT - 0.000673065 1.746398 -2.191056 0.1641861 4 2 s 11 s 12 s 25 s
atom 11 - C CT - 0.1659299 0.2392081 -1.965548 -0.179414 4 1 s 5 s 10 s 26 s
atom 12 - C CT - -0.02909811 1.843217 -3.690682 0.5352615 4 3 s 10 s 13 s 27 s
atom 13 - C CT - -0.05206451 0.4967563 -4.306856 0.1144068 4 1 s 12 s 14 s 28 s
atom 14 - C CT - -0.03038923 -0.2671671 -4.981892 1.253617 4 4 s 13 s 29 s 30 s
atom 15 - C CB - 0.08249438 -1.552392 -0.4065465 0.71006 3 5 s 7 d 17 s
atom 16 - C CK - -0.05454803 -0.3735078 -1.525109 2.275386 3 5 s 6 d 31 s
atom 17 - C CB - -0.2466197 -2.03038 -0.1273508 2.059452 3 6 s 15 s 18 d
atom 18 - C CA - 0.2414911 -3.131856 0.7825465 2.190691 3 8 s 9 s 17 d
atom 19 - C CQ - 0.02605373 -3.129242 0.9753901 -0.1403341 3 7 s 8 d 32 s
atom 20 - H HO - 0.2144616 3.394199 -1.851692 -0.7765953 1 2 s
atom 21 - H HO - 0.2156567 2.16116 -4.667913 2.175689 1 3 s
atom 22 - H HO - 0.2189877 -1.972887 -4.831853 0.3617753 1 4 s
atom 23 - H H - 0.239451 -4.406889 1.693442 3.524526 1 9 s
atom 24 - H H - 0.248053 -3.24482 0.6482923 4.222165 1 9 s
atom 25 - H H1 - 0.1252722 2.104832 -1.547854 1.011427 1 10 s
atom 26 - H H2 - 0.1840025 0.1060296 -1.389873 -1.144839 1 11 s
atom 27 - H H1 - 0.1064256 2.689736 -4.18417 -0.01321531 1 12 s
atom 28 - H H1 - 0.129646 0.6359202 -5.035758 -0.7324823 1 13 s
atom 29 - H H1 - 0.1001421 0.3203558 -5.84969 1.655588 1 14 s
atom 30 - H H1 - 0.08138472 -0.4673436 -4.241935 2.074612 1 14 s
atom 31 - H H2 - 0.2171755 0.4002238 -2.198637 2.674911 1 16 s
atom 32 - H H2 - 0.1899019 -3.607626 1.444208 -1.028106 1 19 s
endmol 1
```

**Figure 4.** Content of a *.hin file before and after applying the algorithm

Microsoft Excel - results-optimization.xls

File Edit View Insert Format Tools Data Window StatsDirect Help

A1 optim-converged

| | A | B | C | D | E | F | G | H | I |
|---|---|---|---|---|---|---|---|---|---|
| 1 | optim-converged | current-file-name | total-energy | dipole-moment | scf-atom-energy | scf-binding-energy | scf-core-energy | scf-electronic-energy | heating formation |
| 2 | TRUE | 131_191.hin | -87051.97 | 2.616018 | -83774.8 | -3277.171 | 458296.2 | -545348.2 | -87.70886 |
| 3 | FALSE | 132_51263.hin | -122879.8 | 3.952207 | -116793.7 | -6086.073 | 997701.8 | -1120582 | 26.55836 |
| 4 | TRUE | 133_2099.hin | -82336.04 | 4.345716 | -78055.3 | -4280.736 | 510573.9 | -592909.9 | 73.78911 |
| 5 | TRUE | 134_16231.hin | -69651.55 | 6.323501 | -67366.21 | -2285.342 | 286214 | -355865.5 | 36.36287 |
| 6 | FALSE | 135_60795.hin | -122814.2 | 2.667873 | -116925.7 | -5888.429 | 815940.2 | -938754.3 | -35.10727 |
| 7 | TRUE | 136_2344.hin | -82012.22 | 0.6501277 | -77036.29 | -4975.933 | 577559.3 | -659571.5 | 87.86626 |
| 8 | FALSE | 137_2369.hin | -89295.8 | 2.939988 | -84308.25 | -4987.556 | 544658.1 | -633954 | -108.9007 |
| 9 | TRUE | 138_2405.hin | -97344.4 | 2.237763 | -92123.44 | -5220.968 | 571194.4 | -668538.8 | -178.54 |
| 10 | TRUE | 139_2435.hin | -67611.72 | 2.101606 | -64738.91 | -2872.805 | 320807.5 | -388419.2 | 119.7452 |
| 11 | FALSE | 140_31101.hin | -176120.1 | 4.527873 | -167579.8 | -8540.328 | 1680861 | -1856981 | -98.23335 |
| 12 | TRUE | 141_2487.hin | -90706.32 | 1.755899 | -85377.08 | -5329.239 | 691597.9 | -782304.3 | 2.527497 |
| 13 | TRUE | 142_2707.hin | -47513.1 | 2.239162 | -46689.62 | -823.4765 | 102907.9 | -150421 | -119.3025 |
| 14 | TRUE | 143_2712.hin | -81903.99 | 4.777861 | -78133.48 | -3770.507 | 472394.3 | -554298.2 | 120.7101 |

**Figure 5.** Structural refinement of molecule before (a) and after (b) applying the algorithm

A number of 199 compounds out of 319 were optimized after the first run of the program. Thus, a 62% (95%CI [57% – 68%]) accuracy was obtained after the first run.

The statistical characteristics in terms of number of atoms in optimized and non-optimized compounds after the first run are presented in Table 1.

**Table 1.** Statistical characteristics for optimized and non-optimized compounds after the first run: number of atoms

| Parameter | Optimized (n = 199) | | Non-optimized (n = 119) | |
|---|---|---|---|---|
| | Total no. of atoms | No. of atoms other than H | Total no. of atoms | No. of atoms other than H |
| Mean [95%CI] | 38 [37 – 40] | 21 [20 – 21] | 50 [47 – 52] | 26 [25 - 27] |
| StdDev | 12.53 | 6.01 | 11.93 | 5.71 |
| Median | 39 | 21 | 49 | 25 |
| Mode | 32 | 21 | 48 | 25 |
| Min | 10 | 7 | 17 | 14 |
| Max | 72 | 36 | 87 | 44 |

Mean = arithmetic mean, 95%CI = 95% confidence interval; StdDev = standard deviation

The mean of the total number of atoms in optimized compounds proved to be statistically significant smaller compared to the number of atoms in not-optimized compounds when first run was analyzed ($t = -7.76$, $p = 1.19 \cdot 10\text{-}13$, significance level of 5%). The mean of the number of atoms other than H in optimized compounds proved to be statistically significant smaller compared to not-optimized compounds in first run ($t = -8.17$, $p = 7.58 \cdot 10\text{-}15$, significance level of 5%).

The second run aimed to optimize the not-optimized geometry after the first run. A number of 6.34 hours was needed with an average of 0.05 hours per compound. All compound proved to be optimized after the second run (resulting an accuracy of 100%).

## Discussion

The proposed algorithm was implemented for the HyperChem 8.0 software. The implemented PHP programs led to the efficient optimization of geometry on the investigated sample of compounds. The users can select the parameter of energy minimization and geometry optimization as well as different protocols. The programs were initially tested on a small sample size. The program is currently being developed to include other QSAR/QSPR calculations.

Computer based geometry optimization was successfully developed, implemented and assessed. These files represent the input data in molecular modeling and are used to calculate the molecular descriptors for each compound.

Structure-activity/property approaches use molecular structures by neglecting the contribution of hydrogen atoms as structural information related to the active activity/property. The following are explanations for not using hydrogen atoms:

- Biological activities are determined in vivo in an aqueous medium where partial dissociation processes take place. Thus, a form without hydrogen atoms is presented in this experiment.
- Hydrogen atoms have a single bound. Since hydrogen atoms have one bound, they do not contribute to molecular complexity (they do not create chains and braches being only terminal elements) and they do not provide information able to characterize compound structure, unlike the information supplied by all the other atoms.
- Since hydrogen atoms are not taken into consideration, the amount of calculations required reduces significantly if hydrogen atoms are deleted (e.g. the amount of calculations for the $C_nH_{2n+2}$ alkane is reduced by 1/3 if the calculations are proportional to molecule size).

The following strategies could be used in order to verify that the program works properly:

- Display the optimized molecules (Figure 3): ▪ The file before optimization did not contain hydrogen atoms while molecules after optimization contain all atoms including hydrogen atoms. & ▪ The position of the atoms in the structure is quite different after applying the proposed algorithm.

The analysis of the files showed in Figure 4 revealed the following:

- AMBER was applied to the molecules stored in PubChem.
- The structure of the compound was correct if only one molecule was present. Our program is not able to optimize the geometry of any structures that have more than 1 molecule.
- The number of atoms before applying the methodology was smaller compared to the number of atoms after using the program. This difference was due to the absence of hydrogen atoms from the structures downloaded from PubChem (see Figure 3).

Some of the numbers that characterize each atom were not changed by applying the proposed methodology (the last five numbers of each atom) while all the other values changed as the result of energy minimization and geometry optimization. The program was developed to include in the report file a series of parameters that characterize the optimized compound (see Figure 5).

Two main advantages were identified while running the program:

- The time needed to prepare the molecule for structure-activity/property analyses was reduced. All the processes implemented in the algorithm were performed automatically; there was no coffee break during the process and the program will never get tired regardless of the number of compounds processed. Time was reduced by almost ½ when the automated program was used.
- The possibility of any mistakes / false analyses was minimized. Energy minimization and/or

geometry optimization are always accurate since the processes are performed by the program. The program stops in two situations: the analysis is complete (all molecules in the set were analyzed) or a power cut occurs in the laboratory. The power shortage can be prevented by using UPS devices.

Some structures could not be optimized by applying the parameter used for optimization. The program ran for our sample requires a maximum of 1000 cycles and an optimal coverage of 0.01. The program stops while processing a compound if the maximum number of cycles required is reached or if optimal coverage is obtained. This problem could be solved by changing the parameters of the optimization algorithm. This solution may be used if the structures are similar in terms of number of compounds and/or complexity. Another solution is to apply the program again on the structures that were not optimized. This strategy was more suitable for our sample given the heterogeneity of compounds.

A series of *in silico* molecular modeling methods use automated processes to prepare compounds but none could be identified for geometry optimization using the HyperChem software. Pencheva et al. [23] proposed an automated molecular mechanics optimization tool for in silico screening. The development and implementation of automated methods applied to in silico screening represent a research problem to which a series of researchers have already found several solutions: conformation generator and rigid docking [24]; web server ligand-based screening [25]; 3D ligand similarity search [26], etc.

In 2009, Lagorce et al. proposed a DG-AMMOS source program, which allows the generation of 3D conformation of small molecules using distance geometry and their energy minimization via automated molecular mechanics optimization [27]. Helgee et al. developed and implemented a method that optimizes a compound based on the interpretation of quantitative structure-activity relationship models [28]. Miteva et al. proposed a web tool for generating the conformation ensembles of small molecules starting from 1D, 2D or 3D descriptions of the compounds [29]. We were not able to identify any program similar with our proposed algorithm. We are currently working on implementing other facilities to our program in terms of QSAR descriptors.

## Conclusions

An efficient geometry optimization algorithm was developed and implemented to be used with HyperChem 8.0 software in order to prepare molecules for QSAR/QSPR analyses. Fully automatic PHP programs can perform geometry optimization according to the defined characteristics specified by the user. The program executes efficiently and quickly the implemented geometry optimization protocols on both small and large sample sizes. The results obtained in the investigated sample of 319 compounds showed that our program performed well on the investigated compound collection thus increasing the reliability of the compounds included in the QSAR analysis.

## Conflict of Interest

The author declares that he has no conflict of interest.

## Acknowledgements

The study was supported by POSDRU/89/1.5/S/62371 through a fellowship for L. Jäntschi.

## References

1. Hansch C, Fujuta T. p-σ-π Analysis. A Method for the Correlation of Biological Activity and Chemical Structure. Am Chem Soc 1964;86(8):1616-1630.
2. Allinger NL. Conformational analysis. 130. MM2. A hydrocarbon force field utilizing V1 and V2 torsional terms. J Am Chem Soc 1977;99:8127-8134.
3. Hocquet A, Langgard M. An Evaluation of the MM+ Force Field. J Mol Model 1998;4:94-112.

4. Weiner SJ, Kollman PA, Case DA, Singh UC, Chio C, Alagona G, et al. A new force field for molecular mechanical simulation of nucleic acids and proteins. J Am Chem Soc 1984;106:765-784.
5. Weiner SJ, Kollman PA, Nguyer DT, Case DA. An all atom force field for simulations of proteins and nucleic acids. I Comput Chem 1986;7:230-252.
6. Cornell WD, Cieplak P, Bayly CI, Gould IR, Merz KM Jr, Ferguson DM, Spellmeyer DC, Fox T, Caldwell JW, Kollman PA. A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules. J Am Chem Soc 1995;117:5179-5197.
7. Brooks BR, Bruccoleri RE, Olafson BD, States DJ, Swaminathan S, Karplus M. Charmm: a program for macromolecular energy, minimization, and dynamics calculations. J Comp Chem 1983;4(2):187-217.
8. Jorgensen WL, Tirado-Rives J. The OPLS Force Field for Proteins. Energy Minimizations for Crystals of Cyclic Peptides and Crambin. J Am Chem Soc 1988;110:1657-1666.
9. Pranata J., Wierschke S.G., Jorgensen W.L. OPLS potential functions for nucleotide bases. Relative association constants of hydrogen-bonded base pairs in chloroform. J Am Chem Soc 1991;113:2810-2819.
10. Cramer CJ. Essential of Computational Chemistry. Whiley, Chichester, 2002, pp. 126-131.
11. Dewar MJS, Thiel W. Ground staters of molecules. 38. The MNDO method. Approximations and parameters. J Am Chem Soc 1977;99(15):4899-4907.
12. Dewar MJS, Zoebisch EG, Healy EF, Stewart JJP. AM1: A New General Purpose Quantum Mechanical Molecular Model. J Am Chem Soc 1985;107:3902-3909.
13. Stewart JJPJ. Optimization of parameters for semiempirical methods I. Method. Comput Chem 1989;10:209-220.
14. Stewart JJP. Optimization of parameters for semiempirical methods. II. Applications. J Comput Chem 1989;10:221-264.
15. Zerner M. Reviews in Computational Chemistry, Volume 2, Eds. Lipkowitz KB, Boyd DB, VCH, New York, 313, (1991).
16. Nanda DN, Jug K. SINDO1. A semiempirical SCF MO method for molecular binding energy and geometry I. Approximations and parametrization. Theor Chim Acta 1980;57(2):95-106.
17. Jug K, Nanda DN. SINDO1 II. Application to ground states of molecules containing carbon, nitrogen and oxygen atoms. Theor Chim Acta 1980;57(2):107-130.
18. Jug K, Nanda DN. SINDO1 III. Application to ground states of molecules containing fluorine, boron, beryllium and lithium atoms. Theor Chim Acta 1980;57(2):131-144.
19. Hoffmann R. An Extended Hückel Theory. I. Hydrocarbons. J Chem Phys 1963;39:1397-1412.
20. Joachim Sauer. Molecular models in ab initio studies of solids and surfaces: from ionic crystals and semiconductors to catalysts. Chem Rev 1989;89(1):199-255.
21. Kortagere S, Chekmarev D, Welsh WJ, Ekins S. New Prredictive Models for Blood-Brain Barrier Permeability of Drug-like Molecules. Pharm Res 2008;25(8):1836-1845.
22. Polak B, Ribiere G. Note surla convergence des méthodes de directions conjuguées. Rev Fr Imform Rech Oper 1969;16:35-43.
23. Pencheva T, Lagorce D, Pajeva I, Villoutreix BO, Miteva MA. AMMOS: Automated Molecular Mechanics Optimization tool for in silico Screening. BMC Bioinformatics 2008;9:438.
24. Sauton N, Lagorce D, Villoutreix BO, Miteva MA. MS-DOCK: accurate multiple conformation generator and rigid docking protocol for multi-step virtual ligand screening. BMC Bioinformatics 2008;9:184.
25. Sperandio O, Petitjean M, Tuffery P. wwLigCSRre: a 3D ligand-based server for hit identification and optimization. Nucleic Acids Res 2009;37(Web Server issue):W504-9.
26. Quintus F, Sperandio O, Grynberg J, Petitjean M, Tuffery P. Ligand scaffold hopping combining 3D maximal substructure search and molecular similarity. BMC Bioinformatics 2009;10:245.
27. Lagorce D, Pencheva T, Villoutreix BO, Miteva MA. DG-AMMOS: a new tool to generate 3d conformation of small molecules using distance geometry and automated molecular mechanics optimization for in silico screening. BMC Chem Biol 2009;9:6.
28. Helgee EA, Carlsson L, Boyer S. A method for automated molecular optimization applied to Ames mutagenicity data. J Chem Inf Model 2009;49(11):2559-63.
29. Miteva MA, Guyon F, Tufféry P. Frog2: Efficient 3D conformation ensemble generator for small compounds. Nucleic Acids Res 2010;38(Web Server issue):W622-7.

# Computer Assisted Geometry Optimization for *in silico* Modeling

**Lorentz JÄNTSCHI** [1,2,*]

[1] Technical University of Cluj-Napoca, 28 Memorandumului, 400114 Cluj-Napoca, Romania..
[2] University of Agricultural Sciences and Veterinary Medicine Cluj-Napoca, 3-5 Mănăştur, 400372 Cluj-Napoca, Romania.
E-mail: lorentz.jantschi@gmail.com

* Author to whom correspondence should be addressed; Tel.: +4-0264-401775; Fax: +4-0264-401-768.

**Received: 15 August 2011 /Accepted: 15 September 2011 / Published online: 22 September 2011**

**Table 1.** The name of the drug-like molecule, the PubChem compound identifier (CID) and active/not active status in the CNS (central nervous system)

| Name | CID | CNSA | Name | CID | CNSA | Name | CID | CNSA |
|---|---|---|---|---|---|---|---|---|
| Adenosine | 191 | 1 | Amprenavir | 2177 | 0 | Isradipine | 3784 | 0 |
| Alfentanil | 51263 | 1 | Amrinone | 3698 | 0 | Ketotifen | 3827 | 0 |
| Alosetron | 2099 | 1 | Anastrozole | 2187 | 0 | Lamivudine | 3877 | 0 |
| Amiloride | 16231 | 1 | Anthralin | 2202 | 0 | Lansoprazole | 3883 | 0 |
| Aripiprazole | 60795 | 1 | Argatroban | 92722 | 0 | Latanoprost | 5311221 | 0 |
| Benztropine | 2344 | 1 | Azathioprine | 2265 | 0 | Leflunomide | 3899 | 0 |
| Betaxolol | 2369 | 1 | Aztreonam | 5362041 | 0 | Letrozole | 3902 | 0 |
| Bisoprolol | 2405 | 1 | Baclofen | 2284 | 0 | Levamisole | 26879 | 0 |
| Brimonidine | 2435 | 1 | Balsalazide | 5362070 | 0 | Lindane | 727 | 0 |
| Bromocriptine | 31101 | 1 | Beclometasone | 20469 | 0 | Linezolid | 3929 | 0 |
| Butorphanol | 2487 | 1 | Benazepril | 2311 | 0 | Lisinopril | 5362119 | 0 |
| Chloralhydrate | 2707 | 1 | Bepridil | 2351 | 0 | Lodoxamide | 44564 | 0 |
| Chlordiazepoxide | 2712 | 1 | Brinzolamide | 68844 | 0 | Loracarbef | 3956 | 0 |
| Chlorpheniramine | 2725 | 1 | Budesonide | 63006 | 0 | Losartan | 3961 | 0 |
| Chlorzoxazone | 2733 | 1 | Bumetanide | 2471 | 0 | Lovastatin | 53232 | 0 |
| Citalopram | 2771 | 1 | Bupivacaine | 2474 | 0 | Mechlorethamine | 4033 | 0 |
| Clemastine | 2781 | 1 | Calcitriol | 6398761 | 0 | Medroxyprogesterone | 10631 | 0 |
| Clonazepam | 2802 | 1 | Candesartan | 2541 | 0 | Melphalan | 4053 | 0 |
| Clorazepate | 2809 | 1 | Capsaicine | 2548 | 0 | Mercaptopurine | 667490 | 0 |
| Clozapine | 2818 | 1 | Captopril | 2550 | 0 | Meropenem | 64778 | 0 |
| Cyclobenzaprine | 2895 | 1 | Cefaclor | 2609 | 0 | Mesalamine | 4075 | 0 |
| Cyproheptadine | 2913 | 1 | Cefadroxil | 2610 | 0 | Metaproterenol | 4086 | 0 |
| Dezocine | 40841 | 1 | Cefazolin | 33255 | 0 | Metformin | 4091 | 0 |
| Dipivefrin | 3105 | 1 | Cefdinir | 6399011 | 0 | Methimazole | 1349907 | 0 |
| Dolasetron | 3148 | 1 | Cefditoren | 6437877 | 0 | Methylergonovine | 8226 | 0 |
| Doxazosin | 3157 | 1 | Cefepime | 2622 | 0 | Metoclopramide | 4168 | 0 |
| Doxepin | 667477 | 1 | Cefixime | 54362 | 0 | Metolazone | 4170 | 0 |
| Dronabinol | 2978 | 1 | Cefmetazole | 2626 | 0 | Metyrosine | 3125 | 0 |
| Droperidol | 3168 | 1 | Cefonicid | 43592 | 0 | Mexiletine | 4178 | 0 |
| Emedastine | 3219 | 1 | Cefoperazone | 135784 | 0 | Miglitol | 441314 | 0 |
| Entacapone | 5281081 | 1 | Cefotaxime | 2632 | 0 | Milrinone | 4197 | 0 |
| Esmolol | 59768 | 1 | Cefoxitin | 37194 | 0 | Minoxidil | 4201 | 0 |
| Estazolam | 3261 | 1 | Cefpodoxime | 6335986 | 0 | Moexipril | 91270 | 0 |
| Fexofenadine | 3348 | 1 | Ceftazidime | 157706 | 0 | Moricizine | 34633 | 0 |

CNSA=active(1) or inactive(0) in Central nervous system; CID=compound ID

**Table 1.** (continued)

| Name | CID | CNSA | Name | CID | CNSA | Name | CID | CNSA |
|---|---|---|---|---|---|---|---|---|
| Fluoxetine | 3386 | 1 | Ceftibuten | 5282242 | 0 | Moxifloxacin | 4259 | 0 |
| Flurazepam | 3393 | 1 | Ceftizoxime | 2655 | 0 | MycophenolicAcid | 446541 | 0 |
| Fluvoxamine | 5324346 | 1 | Ceftriaxone | 5479530 | 0 | Nabumetone | 4409 | 0 |
| Formoterol | 3410 | 1 | Cefuroxime | 2659 | 0 | Naloxone | 4425 | 0 |
| Fosphenytoin | 56339 | 1 | Celecoxib | 2662 | 0 | Naphazoline | 4436 | 0 |
| Galantamine | 3449 | 1 | Cephalexin | 27447 | 0 | Naproxen | 1302 | 0 |
| Granisetron | 3510 | 1 | Chlorpropamide | 2727 | 0 | Nateglinide | 4443 | 0 |
| Hydrocodone | 411697 | 1 | Chlorthalidone | 2732 | 0 | Nedocromil | 50294 | 0 |
| Hydromorphone | 3648 | 1 | Cholecalciferol | 6221 | 0 | Nicardipine | 4474 | 0 |
| Isotretinoin | 5538 | 1 | Cholestyramine | 3086319 | 0 | Nifedipine | 4485 | 0 |
| Labetalol | 3869 | 1 | Ciclopirox | 2749 | 0 | Nimodipine | 4497 | 0 |
| Levobunolol | 39468 | 1 | Cidofovir | 60613 | 0 | Nisoldipine | 4499 | 0 |
| Levocabastine | 54385 | 1 | Cladribine | 1546 | 0 | Nitazoxanide | 41684 | 0 |
| Maprotiline | 4011 | 1 | Clindamycin | 29029 | 0 | Nitrofurantoin | 4509 | 0 |
| Meperidide | 3034126 | 1 | Clopidogrel | 2806 | 0 | Nitroglycerin | 4510 | 0 |
| Metaxalone | 15459 | 1 | Clotrimazole | 2812 | 0 | Nizatidine | 4513 | 0 |
| Methadone | 4095 | 1 | Colchicine | 2833 | 0 | Norgestrel | 13109 | 0 |
| Methocarbamol | 4107 | 1 | Cromolyn | 2882 | 0 | Ofloxacin | 4583 | 0 |
| Methoxamine | 6082 | 1 | Cyclopentolate | 2905 | 0 | Olopatadine | 60865 | 0 |
| Methyldopa | 4138 | 1 | Cyclophosphamide | 2907 | 0 | Olsalazine | 6816262 | 0 |
| Molindone | 23897 | 1 | Cytarabine | 596 | 0 | Oseltamivir | 65028 | 0 |
| Nalbuphine | 4419 | 1 | Dantrolene | 2952 | 0 | Oxaprozin | 4614 | 0 |
| Naratriptan | 4440 | 1 | Dapsone | 2955 | 0 | Oxybutynin | 4634 | 0 |
| Nefazodone | 4449 | 1 | Delavirdinemesylate | 5625 | 0 | Pantoprazole | 4679 | 0 |
| Nortriptyline | 4543 | 1 | Dexamethasone | 5743 | 0 | Pemirolast | 57697 | 0 |
| Ondansetron | 4595 | 1 | Dexpanthenol | 4678 | 0 | Penbutolol | 37464 | 0 |
| Orphenadrine | 4601 | 1 | Diazoxide | 3019 | 0 | Penciclovir | 4725 | 0 |
| Oxcarbazepine | 34312 | 1 | Dibucaine | 3025 | 0 | Pentamidine | 4735 | 0 |
| Oxycodone | 4635 | 1 | Dicloxacillin | 3041 | 0 | Pentoxifylline | 4740 | 0 |
| Oxymorphone | 4639 | 1 | Digoxin | 15478 | 0 | Perindopril | 107807 | 0 |
| Paroxetine | 4691 | 1 | Diltiazem | 3076 | 0 | Pindolol | 4828 | 0 |
| Phenelzine | 3675 | 1 | Dinoprostone | 9691 | 0 | Pioglitazone | 4829 | 0 |
| Phenylephrine | 6041 | 1 | Disopyramide | 3114 | 0 | Pramoxine | 4886 | 0 |
| Pirbuterol | 4845 | 1 | Dofetilide | 71329 | 0 | Procainamide | 4913 | 0 |
| Pramipexole | 4885 | 1 | Dorzolamide | 3154 | 0 | Procarbazine | 4915 | 0 |
| Prazosin | 4893 | 1 | Econazole | 33745 | 0 | Propafenone | 4932 | 0 |
| Procyclidine | 4919 | 1 | Edrophoniumchloride | 3202 | 0 | Propylthiouracil | 657298 | 0 |
| Propoxyphene | 10100 | 1 | Ephedrine | 5032 | 0 | Pyridoxine | 1054 | 0 |
| Pseudoephedrine | 7028 | 1 | Eplerenone | 443872 | 0 | Quinapril | 54892 | 0 |
| Quazepam | 4999 | 1 | Epoprostenol | 5280427 | 0 | Quinidine | 1065 | 0 |
| Quetiapine | 5002 | 1 | Eprosartan | 60879 | 0 | Ramipril | 5038 | 0 |
| Rizatriptan | 5078 | 1 | Estramustine | 18140 | 0 | Rivastigmine | 77991 | 0 |
| Scopolamine | 5184 | 1 | Etidronicacid | 3305 | 0 | Rofecoxib | 5090 | 0 |
| Secobarbital | 5193 | 1 | Etodolac | 3308 | 0 | Rosiglitazone | 77999 | 0 |
| Sertraline | 5203 | 1 | Famciclovir | 3324 | 0 | Sildenafil | 5212 | 0 |
| Sibutramine | 5210 | 1 | Famotidine | 3325 | 0 | Simvastatin | 54454 | 0 |
| Sufentanil | 41693 | 1 | Fenoldopam | 3341 | 0 | Streptozocin | 5299 | 0 |
| Sumatriptan | 5358 | 1 | Fenoprofen | 3342 | 0 | Sulfacetamide | 5320 | 0 |
| Thiethylperazine | 5440 | 1 | Flavoxate | 3354 | 0 | Sulfasalazine | 5353980 | 0 |
| Thiothixene | 5454 | 1 | Flecainide | 3356 | 0 | Sulfinpyrazone | 5342 | 0 |
| Tiagabine | 5466 | 1 | Floxuridine | 3363 | 0 | Sulindac | 5352 | 0 |
| Timolol | 5478 | 1 | Flunisolide | 82153 | 0 | Tamsulosin | 129211 | 0 |
| Tolazoline | 5504 | 1 | Fluoxymesterone | 6446 | 0 | Tazarotene | 5381 | 0 |
| Tramadol | 5523 | 1 | Flurbiprofen | 3394 | 0 | Terazosin | 5401 | 0 |
| Trazodone | 5533 | 1 | Flutamide | 3397 | 0 | Terbutaline | 5403 | 0 |
| Trimethobenzamide | 5577 | 1 | Fluvastatin | 446155 | 0 | Theophylline | 2153 | 0 |
| Venlafaxine | 5656 | 1 | Fosfomycin | 3417 | 0 | Thiamine | 1130 | 0 |
| Zaleplon | 5719 | 1 | Furosemide | 3440 | 0 | Ticlopidine | 5472 | 0 |

CNSA=active(1) or inactive(0) in Central nervous system; CID=compound ID

**Table 1.** (continued)

| Name | CID | CNSA | Name | CID | CNSA | Name | CID | CNSA |
|---|---|---|---|---|---|---|---|---|
| Ziprasidone | 60854 | 1 | Ganciclovir | 3454 | 0 | Tocainide | 38945 | 0 |
| Zolpidem | 5732 | 1 | Gatifloxacin | 5379 | 0 | Tolazamide | 5503 | 0 |
| Zolmitriptan | 5731 | 1 | Gemcitabine | 60750 | 0 | Tolbutamide | 5505 | 0 |
| Acarbose | 41774 | 0 | Gemfibrozil | 3463 | 0 | Tolmetin | 5509 | 0 |
| Acetazolamide | 1986 | 0 | Glimepiride | 3476 | 0 | Torasemide | 41781 | 0 |
| Acetylcysteine | 581 | 0 | Glipizide | 3478 | 0 | Trandolapril | 5484727 | 0 |
| Acyclovir | 2022 | 0 | Glyburide | 3488 | 0 | Triamcinolone | 31307 | 0 |
| Adefovir | 60172 | 0 | Hydralazine | 3637 | 0 | Triamterene | 5546 | 0 |
| Allopurinol | 2094 | 0 | Ibutilide | 60753 | 0 | Valacyclovir | 5647 | 0 |
| Alprostadil | 214 | 0 | Idarubicin | 42890 | 0 | Voriconazole | 5231054 | 0 |
| Altretamine | 2123 | 0 | Ifosfamide | 3690 | 0 | Warfarin | 6691 | 0 |
| Aminoglutethimide | 2145 | 0 | Imiquimod | 57469 | 0 | Zileuton | 60490 | 0 |
| Amlodipine | 2162 | 0 | Indapamide | 3702 | 0 | ZoledronicAcid | 68740 | 0 |
| Amoxicillin | 2171 | 0 | Isoetharine | 3762 | 0 | | | |
| Ampicillin | 2174 | 0 | Isosorbidedinitrate | 170113 | 0 | | | |

CNSA=active(1) or inactive(0) in Central nervous system; CID=compound ID