Vector algebraic determination of molecular symmetry and its application in modeling

Doctoral (Ph.D.) theses

László Gyevi-Nagy

Supervisor:
Dr. Gyula Tasi
associate professor

Doctoral School of Chemistry

Department of Applied and Environmental Chemistry
Faculty of Science and Informatics
University of Szeged

Szeged
2018

1 Introduction

The concept of symmetry plays a crucial role in modern physics, theoretical and physical chemistry. Its importance lies in it fundamentally influencing various properties of a system. Therefore, symmetry may be applied to conveniently explore these properties. Detection of molecular symmetry is essential in both qualitative and quantitative calculations, its application is necessary, for example, during the determination of symmetries of vibrational normal modes and the type of couplings between them or in statistical thermodynamic estimations.

For this reason, many different techniques have been worked out for the detection of static symmetry of molecules. These methods may be classified based on the mathematical concepts they use: algebraic procedures rely mainly on matrix operations, while graph theoretical approaches utilize the automorphisms of the molecular graph to discover symmetry.

However, examining several implementations carrying out symmetry analysis, we discover that these can not detect the highest order point group in every case, preventing the most complete utilization of symmetry during calculations and the correct computation of certain thermochemical values.

Conformational properties play a central role in the appearance of stable molecular structures, therefore molecular conformation has become a dominant concept of modern chemistry and structural biology. As a result, exploration of the conformational space of biological macromolecules with a significant degree of flexibility (mainly proteins) is essential to understand their conformational preferences and interactions governing the formation of stable configurations (i.e. protein folding).

The properties of conformers and their interconversional transition states can be studied with both experimental and theoretical methods. The applied experimental procedures are usually spectroscopic techniques, mostly infrared- and Raman-spectroscopy but sometimes other approaches like diffraction or magnetic resonance methods are also employed. However, the use of linearized model functions to evaluate the experimental data is widespread. Yet, if the error propagation it introduces in not taken into account, the linearization affects the quality of the results.

Theoretical studies rely on molecular dynamics and quantum chemical calculations. Molecular dynamics simulations are based on the solution of Newton's equations of motion – mostly using an empirical, molecular mechanical force field – spanning a trajectory in the phase space of the given molecule.

On the other hand, quantum chemical techniques are more suitable to describe individual structures like stationary points, the reaction paths between them or to calculate single point energies in discrete points of the potential energy surface. Quantum chemical procedures provide valuable results which, in some cases, approximate experimental data in precision, however, due to their much higher computational complexity, are normally not applicable to dynamical systems.

Experimental and theoretical data can also be modeled to speed up and simplify the estimation of relevant quantities. Descriptors of QSPR modeling, derived from the molecular structure, often correlate with various properties, even thermochemical values, providing acceptable description of these quantities.

Conformational enthalpy differences of n-alkanes seem to correlate with the symmetry of the structures, therefore it is worthwhile to carry out a QSPR modeling of these thermochemical quantities using symmetry-based descriptors as variables.

Simple, graph-theoretical descriptors derived form the molecular graph present a promising approach to describe certain properties, furthermore, these parameters may be modified in various ways. Such modified indices are, for example, the electrotopological indices or the three-dimensional Wiener-index including interatomic Euclidean distances, both of these have found widespread application.

The commonly used multilinear model function of QSPR models may also be modified to further improve the quality of the results. A good example is the sometimes applied power series, which has although been put into use in single variable cases exclusively.

2 Aims

The aim of my doctoral study was to work out a method for the detection of static symmetry of molecules that is more efficient than previous procedures, implement this algorithm and model our own high precision quantum chemical results within the framework of a QSPR study.

We carried out symmetry analysis with a purely vector algebraic method. The advantage of this approach is its simplicity compared to the popular matrix-algebraic algorithms. The performance of the newly developed program was demonstrated by comparison to other stand-alone and molecular modeling software capable of point group detection, while its flexibility was proved by integration into two quantum chemical applications.

After that, we wanted to apply molecular symmetry to model quantum chemical results, for which we planned to utilize the results of our own high precision calculations performed on n-pentane conformers. To achieve the greatest precision possible, reference values were determined by a focal point analysis based on coupled cluster calculations. Some small corrections like the diagonal Born–Oppenheimer correction and scalar relativistic correction were also computed.

Our last objective was to give a high precision QSPR model of the calculated thermochemical quantities that also describes the temperature dependence of the modeled values. Since the relationship between the modeled values and the temperature and structure is fundamentally nonlinear, we intended to apply a multivariate power series to explain the enthalpy differences. In the modeling well known symmetry based descriptors and newly defined topographical indices were used. Following a successful modeling, we extended it to other structures like the interconversional transition states on n-pentane and the conformers of n-butane

3 Theoretical background

3.1 Symmetry analysis

Symmetry operations were carried out using vector algebraic operations only. Inversion needs the formation of the additive inverse of atomic position vectors, while the reflection of point $\mathbf{p}$ to the plane defined by its normal vector $\mathbf{n}$ and the origin of the coordinate system may be done using the following equation:

$$\mathbf{p}' = \mathbf{p} - 2\ \mathrm{proj}_{\mathbf{n}}(\mathbf{p}) = \mathbf{p} - 2\frac{\mathbf{p}\cdot\mathbf{n}}{\mathbf{n}\cdot\mathbf{n}}\mathbf{n}$$

Rotations were performed vector algebraically using Rodrigues' rotation formula, neglected by previous methods:

$$\mathbf{p}' = \mathbf{p} + \sin\varphi\ \mathbf{d}\times\mathbf{p} + (1-\cos\varphi)\ \mathbf{d}\times(\mathbf{d}\times\mathbf{p})$$

where φ is the angle of rotation around the axis given by the direction vector $\mathbf{d}$. Improper rotations were carried out as a proper rotation followed by a reflection to the plane orthogonal to the axis of the rotation.

To generate the normal vectors of potential symmetry planes, the difference of the position vectors of every pair of atoms with same atomic numbers were formed, while the direction vectors of possible axes were produced from two (in case of C_2 axes) or three (for higher order axes) atomic position vectors pointing to three vertices of a regular polyhedron.

The determination of the point group was based on the number of various symmetry elements, while other properties could also be discovered with the help of molecular symmetry.

3.2 Focal point analysis

Conformational enthalpy differences of n-pentane $[\Delta H_T^\circ(\alpha\beta)]$ were calculated based on the following equation:

$$\Delta H_T^\circ(\alpha\beta) = \Delta E_{\mathrm{TOT}}(\alpha\beta) + \Delta E_{\mathrm{ZPVE}}(\alpha\beta) + \Delta H_T^{\mathrm{corr}}(\alpha\beta)$$

where $\Delta E_{\mathrm{TOT}}(\alpha\beta)$ is the total electronic energy of conformer $\alpha\beta$ related to the global minimum tt conformer, $\Delta E_{\mathrm{ZPVE}}(\alpha\beta)$ denotes the relative zero point vibrational energy and $\Delta H_T^{\mathrm{corr}}(\alpha\beta)$ stands for the relative thermal correction on temperature T.

Thermal correction values were calculated using statistical thermodynamic equations within the rigid rotor-harmonic oscillator and ideal gas approximations. Zero point vibrational energy contains anharmonic contributions and correction to the hindered internal rotations of the methyl groups and the backbone C–C–C–C torsions.

Further decomposition of the total electronic energy (E_{TOT}) was given by

$$\begin{aligned} E_{\mathrm{TOT}} = & E_{\mathrm{HF}} + \Delta E_{\mathrm{MP2}} + \Delta E_{\mathrm{CCSD}} + \Delta E_{\mathrm{(T)}} + \Delta E_{\mathrm{HO}} \\ & + \Delta E_{\mathrm{DBOC}} + \Delta E_{\mathrm{SR}} \end{aligned}$$

where E_{HF} denotes the Hartree–Fock energy calculated with the cc-pV6Z basis set, ΔE_{MP2}, ΔE_{CCSD} and $\Delta E_{\mathrm{(T)}}$ are the second order Møller–Plesett energy correction and coupled cluster contributions containing single, double [CCSD] and perturbative triple [(T)] excitations extrapolated to the complete basis. The ΔE_{HO} term was also computed at the coupled cluster level containing quadruple and, for the tt and gg conformers, perturbative pentuple excitations. ΔE_{DBOC} stands for the diagonal correction of the Born–Oppenheimer approximation, finally, ΔE_{SR} denotes the scalar relativistic correction.

CCSD(T) calculations were carried out with the MRCC program linked to CFOUR. Relativistic contributions were determined with the MOLPRO program package. Other results were computed with the CFOUR software. Calculations were performed with RHF reference orbitals, correlating all electrons.

3.3 QSPR modeling

The QSPR modeling was performed using symmetry related descriptors (based on symmetry classes, principal moments of inertia and symmetry numbers) and modified topological indices. The graph invariants were altered by parametrization of the weight matrix of the molecular graph (**W**):

$$W_{ij} = \begin{cases} |\psi_{k,i,j,l}|, \text{ if } (k,i),(i,j),(j,l) \in E \\ 0, \text{ otherwise} \end{cases}$$

where E means the edge set of the molecular graph, $\psi_{k,i,j,l}$ denotes the C_k–C_i–C_j–C_l torsional angle. The elements of the distance matrix were calculated as the lengths of the shortest paths between vertices working with signed torsional angles.

4 New scientific results

T1. We introduced a new, purely vector algebraic method for the detection of static symmetry of molecules.[1,3,4,5]

The new procedure, unlike previous algorithms, relies only on simple vector algebraic operations, therefore it is much simpler than existing methods. To increase the effectiveness of our approach, we tried to reduce the number of potential symmetry elements to the minimum.

T2. We implemented our newly defined method for symmetry analysis in Fortran language and showed that it supersedes previous programs both in computation time and precision of the found point group.[1,3,4,5]

The developed code (SYVA) overcomes several stand-alone and quantum chemical software capable of symmetry detection regarding both runtime and precision (see Table 1). Apart from the point group, the program is able to reveal other symmetry related properties like the framework group, the symmetries of normal modes of vibration and the symmetrized geometry as well.

Table 1: Average computation times, the number of unsuccessful symmetry analyses and the precision of various software.

Program	SYVA	SYMMETRY	SYMMETRIZER	SYMMOL	NWCHEM	ORCA	MRCC
Avg. runtime(ms)	8.8	51.7	311.8	9.3	–	–	–
Unsuccessful det.	0	0	0	15	35	52	0
Precision	99.6%	97.5%	93.2%	95.0%	68.5%	96.2%	82.1%

The prove the flexibility of our code, we integrated it into two molecular modeling software (MRCC and GAMESS). The assembled test set and the developed software are freely available.

T3. We suggested benchmark-quality estimates for the conformational enthalpy differences of *n*-pentane based on a focal point analysis.[2]

We showed that satisfying results can be achieved for the enthalpy differences of n-pentane conformers using CCSD(T)/cc-pVTZ reference geometries. The Hartree–Fock energy for every conformer and the second order Møller–Plesset contribution in case of the gg conformer shows rapid convergence in contrast with other correlation corrections. Smaller terms as the diagonal correction of the Born–Oppenheimer approximation and scalar relativistic corrections cancel out, they contribute mainly to the uncertainty of the final values.

The results show that the calculation of the zero-point vibrational energy for *n*-pentane is challenging, both the anharmonic contribution and the correction for hindered internal rotations are significant and about half of the uncertainty of the final values arise from this term.

We proved by comparison to previous studies that the uniform description of the conformers of *n*-pentane is problematic, the individual methods approximate the results of the focal point analysis differently. Based on our results, we suggested new benchmark-quality values for the conformational enthalpy differences of *n*-pentane as $\Delta H_0^\circ(\mathrm{tg}) = 644 \pm 46$ cal mol^{-1}, $\Delta H_0^\circ(\mathrm{gg}) = 1099 \pm 90$ cal mol^{-1} and $\Delta H_0^\circ(\mathrm{gx}) = 2975 \pm 124$ cal mol^{-1} at 0 K and $\Delta H_0^\circ(\mathrm{tg}) = 628 \pm 52$ cal mol^{-1}, $\Delta H_0^\circ(\mathrm{gg}) = 974 \pm 100$ cal mol^{-1} and $\Delta H_0^\circ(\mathrm{gx}) = 2733 \pm 140$ cal mol^{-1} at 298 K.

T4. We determined the conformational enthalpy differences of *n*-pentane by correct reevaluation of experimental data and suggested a method for the correction of the transition disturbing the measurement of the global minimum during the evaluation of temperature dependent conformational enthalpy differences.[2]

The error introduced by the linearization of experimental data was taken into account by appropriate weighting of the transformed values and the original nonlinear function was also utilized. The weighted regression estimates the enthalpy differences with more moderate errors, the uncertainty in case of the gg conformer is even smaller than the deviation from the original value. The best results obtained using the correct nonlinear function are 618 ± 5 cal mol^{-1} for the tg conformer and 953 ± 15 cal mol^{-1} in case of the gg conformer in the $143 - 250$ K temperature range.

A transition that can be assigned to conformer gg influences the relative enthalpies by about 36 cal mol^{-1} based on a former study. Therefore, we performed a correction of this transition and calculated 626 ± 5 cal mol^{-1} and 961 ± 15 cal mol^{-1} for the tg and gg conformers, respectively. This time, the difference of the original and corrected values outside the estimated error for the tg conformer is worth mentioning, only error bars overlap.

T5. We defined new, three dimensional molecular descriptors by modification of existing topological indices and showed that a multivariate power series as model function gives significantly better results in a QSPR modeling, even with the inclusion of quadratic terms.

We created new topographic indices depending on the backbone torsional angles by weighting the molecular graph. The procedure can easily be extended to an arbitrary weighting scheme to define other descriptors. The new predictor variables fulfill the primary requirements like graph invariance, periodicity and continuity.

We studied the applicability of a multivariate power series as model function in QSPR modeling. Such a model has only been applied in univariate cases. We found that even quadratic terms increase the fit and prediction capability of a QSPR model significantly compared to the common multilinear function.

Based on the results, we may conclude that such a model function – even with low degree terms – is much more suitable to describe fundamentally nonlinear relations.

T6. We verified that the newly defined descriptors correlate well with the enthalpy differences related to the global minimum of the appropriate molecules in case of *n*-pentane conformers, their interconversional transition states and the conformers of *n*-butane.

We successfully gave a QSPR model of the relative enthalpies of n-pentane conformers, their interconversional transition states and the conformers of n-butane. The descriptors we defined turned out to be more suitable for the modeling of enthalpy differences than well known symmetry based descriptors, in every case better models could be obtained with topographic indices.

With our model the relative enthalpy related to the global minimum can be simply estimated using only the backbone torsional angles and the temperature in every important critical point of the conformational potential energy surface with a deviation of 40 cal mol^{-1}.

5 Publications, conference lectures

5.1. Publications related to the dissertation

1. **Vector algebra and molecular symmetry: a tribute to Professor Josiah Willard Gibbs**

   G. Tasi, L. Gyevi-Nagy, R. Tobiás, T. Tasi

   Journal of Mathematical Chemistry **2013**, *51*, 2187.

   Independent citations: 0

   IF$_{2013}$: 1.270

2. **Enthalpy Differences of the n-Pentane Conformers**

   J. Csontos, B. Nagy, L. Gyevi-Nagy, M. Kállay, G. Tasi

   Journal of Chemical Theory and Computation, **2016**, *12*, 2679.

   Independent citations: 2

   IF$_{2016}$: 5.245

3. **SYVA: A program to analyze symmetry of molecules based on vector algebra**

   L. Gyevi-Nagy, G. Tasi

   Computer Physics Communications **2017**, *215*, 156.

   Independent citations: 0

   IF$_{2016}$: 3.936

4. **Szimmetria - egy matematikus szemével, 90 éve született a kémiai Nobel-díjas Sir John Anthony Pople**

   L. Gyevi-Nagy, D. Tasi, A. Remete, G. Tasi

   Magyar Kémikusok Lapja **2015**, *70*, 271.

5. **Szimmetria és vektoralgebra**

   L. Gyevi-Nagy, G. Tasi

   Magyar Kémiai Folyóirat - Kémiai Közlemények **2016**, *122*, 19.

5.2. Other publications

1. **Definitive thermochemistry and kinetics of the interconversions among conformers of n-butane and n-pentane**
   R. Tóbiás, A. G. Császár, L. Gyevi-Nagy, and G. Tasi
   Journal of Computational Chemistry **2018**, *39*, 424.
   Independent citations: 0
   IF_{2016}: 3.229

2. **Reaction fronts of the autocatalytic hydrogenase reaction**
   L. Gyevi-Nagy, E. Lantos, T. Gehér-Herczegh, Á. Tóth, C. Bagyinka, D. Horváth
   Journal of Chemical Physics **2018**, *148*, 165103.
   Independent citations: 0
   IF_{2016}: 2.965

3. **Linearizálni vagy nem linearizálni: paraméterek hibaterjedésének analitikus és numerikus számítása**
   D. Barna, L. Gyevi-Nagy, G. Tasi
   Magyar Kémikusok Lapja **2011**, *66*, 383.

4. **A monomolekuláris reakcióhálózatok általános formálkinetikai modellezése**
   R. Tóbiás, L. Gyevi-Nagy, G. Tasi
   Magyar Kémiai Folyóirat - Kémiai Közlemények **2014**, *120*, 102.

5.3. Conference lectures

1. **Miért létezik élet?**
   L. Gyevi-Nagy
   XXXV. Kémiai Előadói Napok, KEN
   Szeged, **2012.** (oral presentation in Hungarian)

2. **Szilánok konformációs flexibilitásának vizsgálata**
   L. Gyevi-Nagy
   XXXI. Országos Tudományos Diákköri Konferencia, OTDK
   Eger, **2013.** (oral presentation in Hungarian)

3. **A hidrogenáz frontreakció modellezése**
   L. Gyevi-Nagy
   XXXIX. Kémiai Előadói Napok, KEN
   Szeged, **2016.** (oral presentation in Hungarian)

3. **A hidrogenáz frontreakció modellezése**
   L. Gyevi-Nagy, E. Lantos, Á. Tóth, D. Horváth
   MTA Reakciókinetikai és Fotokémiai Munkabizottság ülése
   Mátraháza, **2016.** (oral presentation in Hungarian)

Total peer-reviewed publications: 5	out of this, related to the dissertation: 3
Cumulative impact factor: 16.645	out of this, related to the dissertation: 10.451
Total independent citations: 2	out of this, related to the dissertation: 2