Modeling the kinetics of conformational interconversions using first-order chemical reaction networks

Theses of the PhD dissertation

ROLAND JÓZSEF TÓBIÁS

Msc in Chemistry

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

2017

1. Introduction

Molecular conformation has become one of the central concepts of modern chemistry and structural biology. The conformational properties of bioactive molecules, due to the diverse valence states of their heavy atoms, play a crucial role in the appearance and maintenance of life. The stable spacial structures of molecules are formed during a large number of kinetic processes, governed by inter- and intramolecular interactions. To explore the conformational space and the conformational preferences of a molecule, including the dominant secondary interactions, one needs to rely both on experimental and theoretical methods.

Among the experimental methods capable of determining stable conformations, most prominent are the diffraction and spectroscopic techniques. The theoretical strategies, producing stationary molecular geometries, are of quantum chemical or classical physical origin. Quantum-chemical procedures are able to approximate the observed structures of molecules with considerable accuracy. In contrast, the classical physical alternatives yield only semi-quantitative estimates for the three dimensional molecular geometries.

In chemical and biochemical structural research enumerating all the possible conformational states of a molecule and understanding the interconversions of its conformers are equally important topics. This kind of scientific interest culminates in protein and nucleic-acid folding, both of which are rather complex processes and none of which have been modeled with detailed kinetic schemes.

Similarly to the conformational equilibria of small molecules, the folding of an arbitrary bioactive macromolecule is driven by an interconversion reaction network, containing several interconversions of the form $A_1 \rightarrow A_2$, where A_1 and A_2 are two molecular conformers. The interconversion networks are described by systems of first-order linear differential equations, whose solutions correspond to the temporal distributions of the conformer populations. Since the concentration profiles of the conformers can be treated without numerical difficulties, even a couple of thousand of species can be taken into account in such networks, facilitating the establishment of an in-depth kinetic folding model.

The principal objectives of the present study are as follows: (*i*) perform a simple algebraic characterization of first-order reaction networks (FCRN), and (*ii*) provide benchmark-quality estimaes for the interconversion rate coefficients of *n*-butane and *n*-pentane. While the former goal introduces a novel approach into formal kinetics, the latter helps to understand the flexibility and the dispersion behavior of small organic molecules. The improved understanding is expected to have direct application in the analysis of *n*-alkyl side chains of larger molecular systems.

2. Theoretical background

2.1. Chemical reaction networks

The ordered quadruple $\langle \mathbf{D}, \mathbf{G}, \mathbf{A}, \mathbf{k} \rangle$ is a *chemical reaction network* (CRN) if (*i*) $\mathbf{D} = \{d_{ij}\}$ and $\mathbf{G} = \{g_{ij}\}$ are the left and right stoichiometric matrices of the considered reaction system, respectively, and (*ii*) $\mathbf{A} = \{A_j\}$ and $\mathbf{k} = \{k_i\}$ are vectors of chemical species and rate coefficients, respectively. If only *first-order reactions* take place in this CRN, *i.e.*, in each rows of the matrix $\mathbf{D}$ exactly one entry equals to one and all the others elements are zero, then $\langle \mathbf{D}, \mathbf{G}, \mathbf{A}, \mathbf{k} \rangle$ is a *first-order chemical reaction network* (FCRN). For the concentration vector $\mathbf{c} = \mathbf{c}(t)$ of a FCRN the following system of linear differential equations holds:

$$\dot{\mathbf{c}} = \mathbf{F}\mathbf{c}\,, \tag{1}$$

where $\dot{\mathbf{c}}$ is the time derivative of the vector $\mathbf{c}$, and

$$\mathbf{F} = \mathbf{S}^{\mathrm{T}} \mathrm{diag}(k_i) \mathbf{D} \tag{2}$$

is the coefficient matrix of the FCRN with the stoichiometric matrix defined as $\mathbf{S} = \mathbf{G} - \mathbf{D}$. It is important to note that Eq. (1) can be solved in the case of a known *initial value vector* $\mathbf{c}(t=0) = \mathbf{c}_0$. If the reverse of every reaction is also of first-order in $\langle \mathbf{D}, \mathbf{G}, \mathbf{A}, \mathbf{k} \rangle$, then the processes in $\langle \mathbf{D}, \mathbf{G}, \mathbf{A}, \mathbf{k} \rangle$ are *isomerization reactions*, and $\langle \mathbf{D}, \mathbf{G}, \mathbf{A}, \mathbf{k} \rangle$ is an *isomerization reaction network* (IRN).

2.2. Estimation of rate coefficients using quantum-chemical methods

The interconversions of molecular conformers – within the Born–Oppenheinmer approximation – are motions on a so-called *conformational potential energy surface* (CPES). Having found the minima and the first-order saddle points of the CPES, the rate coefficient of the interconversion $\mathcal{R} = \mathcal{S}_1 \rightarrow \mathcal{S}_2$ through the transition state $\mathcal{S}_{12}$, denoted by $k_{\mathcal{R}}(T)$, can be computed with the help of the Eyring–Polányi equation:

$$k_{\mathcal{R}}(T) = \frac{k_{\mathrm{B}} T}{h} \exp\left\{ -\frac{\Delta^{\ddagger} G_{\mathcal{R}}(T)}{N_{\mathrm{A}} k_{\mathrm{B}} T} \right\}, \tag{3}$$

where N_{A}, k_{B}, and h are the Avogadro, the Boltzmann, and the Plack constants, respectively, while $\Delta^{\ddagger} G_{\mathcal{R}}(T)$ is the *activation free energy*, at temperature *T*, of the reaction $\mathcal{R}$, interpreted as the difference between the total free energies of the transition state $\mathcal{S}_{12}$ and the reactant $\mathcal{S}_1$. It must be emphasized that Eq. (3) can only be employed within the realm of transition state theory and the ideal gas approximation.

Activation free energies, $\Delta^{\ddagger}G_{\mathcal{R}}(T)$, are computed within the framework of the focal-point analysis (FPA) approach, using state-of-art quantum-chemical methods, as follows:

$$\Delta^{\ddagger}G_{\mathcal{R}}(T) = \Delta^{\ddagger}G_{\mathcal{R}}(0) + \delta^{\ddagger}G_{\mathcal{R}}(T), \tag{4}$$

where $\Delta^{\ddagger}G_{\mathcal{R}}(0)$ and $\delta^{\ddagger}G_{\mathcal{R}}(T)$ are the *activation free energy at 0 K* and the *thermal contribution to the activation free energy*, respectively. The quantity $\Delta^{\ddagger}G_{\mathcal{R}}(0)$ is estimated as

$$\Delta^{\ddagger}G_{\mathcal{R}}(0) = \Delta^{\ddagger}E_{\mathcal{R}} + \Delta^{\ddagger}E_{\mathrm{ZPE,h},\mathcal{R}} + \delta^{\ddagger}E_{\mathrm{ZPE,a},\mathcal{R}}, \tag{5}$$

where $\Delta^{\ddagger}E_{\mathrm{ZPE,h},\mathcal{R}}$ and $\delta^{\ddagger}E_{\mathrm{ZPE,a},\mathcal{R}}$ are the harmonic and anharmonic contributions to *the zero-point energy* (ZPE) of the reaction $\mathcal{R}$, respectively, and $\Delta^{\ddagger}E_{\mathcal{R}}$ is the activation energy extrapolated to the *complete basis set limit*. $\Delta^{\ddagger}E_{\mathcal{R}}$ is decomposed in this study as

$$\Delta^{\ddagger}E_{\mathcal{R}} = \Delta^{\ddagger}E_{\mathrm{HF},\mathcal{R}} + \delta^{\ddagger}E_{\mathrm{MP2(fc)},\mathcal{R}} + \delta^{\ddagger}E_{\mathrm{CCSD(fc)},\mathcal{R}} + \delta^{\ddagger}E_{\mathrm{CCSD(T)(fc)},\mathcal{R}} + \delta^{\ddagger}E_{\mathrm{CV},\mathcal{R}}, \tag{6}$$

where

(*i*) $\Delta^{\ddagger}E_{\mathrm{HF},\mathcal{R}}$ is the *activation Hartree-Fock* (HF) *energy*,

(*ii*) $\delta^{\ddagger}E_{\mathrm{MP2(fc)},\mathcal{R}}$ is the *second-order Møller–Plesset* (MP2) *perturbation energy*,

(*iii*) $\delta^{\ddagger}E_{\mathrm{CCSD(fc)},\mathcal{R}}$ is the *coupled-cluster singles and doubles* (CCSD) energy increment,

(*iv*) $\delta^{\ddagger}E_{\mathrm{CCSD(T)(fc)},\mathcal{R}}$ is the energy increment derived from the *coupled-cluster singles, doubles, and perturbative triples* (CCSD(T)) method,

(*v*) 'fc' denotes the use of the frozen core approximation,

(*vi*) $\delta^{\ddagger}E_{\mathrm{CV},\mathcal{R}}$ is a term including the core-core and core-valance interations, and

(*vii*) the individual terms are computed at carfully selected *reference geometries.*

In the case of molecules with just a few heavy atoms the reference structures are usually optimized with the CCSD(T) method. Nevertheless, optimizations at this level would require extreme computational resources for larger molecular systems; thus, the CCSD(T) procedure should be substituted by a density functional theory (DFT) technique suitable for providing geometries near their CCSD(T) counterparts for „ordinary" molecules.

3. Methodological details

The research presented in the underlying dissertation, beyond simple algebraic means and the FPA scheme, a significant amount of programming was performed using the Fortran 90 language. The quantum-chemical computations were carried out with the Gaussian 09 Rev. E.01 and Molpro 2012.1 softwares. Molpro was employed for compute the CCSD(T) energies, while Gaussian was applied to all the other electronic structure computations. In all electron-correlation treatments restricted Hartree–Fock orbitals were utilized.

4. Summary

4.1. The Luther–Rost representation (LRR) of the equation $\dot{\mathbf{c}} = \mathbf{Fc}$ [1]

In the first part of the present PhD research we dealt with a simple and sophisticated strategy for solving the equation $\dot{\mathbf{c}} = \mathbf{Fc}$. This technique is called the Luther–Rost representation, as it was proposed by Luther and Rost for handling similar problems.The LRR method is coupled with the Leverrier algorithm concerning the calculation of the characteristic polynomial coefficients for the matrix **F**.

It is important to point out that the combination of these two procedures may be applied conveniently to solve linear systems of kinetic differential equations both in education and in research. As revealed by a later study of Kyurkiev and Markov, the protocol recommended in our study, that avoids the cumbersome application of Jordan chains of **F**, also appeared to be useful for symbolic-numeric calculations.

The utility of this pair of methods is demonstrated on four relevant kinetic models, two of which, specifically the quadrangle and pentangle reactions, induce systems of linear differential equations whose solution functions had not been available in the literature. On the ground of the worked examples and the Abel–Ruffini theorem, it is anticipated that the universal algebraic solutions to the kinetic problems of the *K*-angle reaction systems can be expressed only in the case of $K \leq 5$, where K is the number of species in the reaction network under study.

4.2. New results in the qualitative theory of first-order reaction networks [2]

After the constructive investigations related to solving the equation $\dot{\mathbf{c}} = \mathbf{Fc}$, we focused on the unexplored structural characteristics of FCRNs. After introducing the decomposability of CRNs, we proved that the first-order chemical reaction networks of nonnegative **G** matrices can be decomposed iff their **F** matrices are block diagonalizable.

Furthermore, we have also showed that each reaction of the FCRNs with nonnegative integer right stoichiometric coefficients has to meet the following so-called mass incompatibility relations: (*i*) the reactant must have a mass not smaller than those of the products, (*ii*) if the masses of the reactant and an individual product are identical, the right stoichiometric coefficient of this product is exactly one, and (*iii*) provided that the masses of the reactant and one of the products are the same, further products are not created. These mass incompatibility conditions also imply that if there are not only isomerization reactions in a

FCRN, the coefficient matrix of the network inspected can be transformed, with a suitable permutation matrix, to block-diagonal form.

It is also derived that for all conservative FCRNs there is a marker network, which (*i*) is linearly conjugate to the original network, (*ii*) contains only isomerization reactions, and (*iii*) entirely describes the temporal behavior of the network marked. The existence of marker networks supports that the statements related to IRNs can be translated to arbitrary conservative FCRNs.

In the end, a method built upon the principle of successive eigenvalue elimination [284] is presented. This method helps to decide whether the eigenvalue problem of a specific $\hat{\mathbf{F}}_{ii}$ diagonal block in Frobenius form of the matrix $\mathbf{F}$ can be solved algebraically. As a consequence, a sufficient condition for the exact expressibility of the eigenvalues connected to the blocks $\hat{\mathbf{F}}_{ii}$ has been obtained: if $\hat{\mathbf{F}}_{ii}$ has at most four nonzero eigenvalues, then they can be localized with the well-known root formulae of algebraic equations.

4.3. Fortran programs for simulating the kinetics of first-order reaction networks

At the next stage of our research two new Fortran programs (`fcrn_lrr` and `fcrn`) have been developed, which are discussed together with some representative kinetic examples. Both programs determine the concentrations of the chemical species included in FCRNs at separate sampling times, but these codes display significant deviations in their operating conditions and the algorithms implemented.

Regarding the procedures selected, the `fcrn_lrr` program utilizes the LRR strategy, the Leverrier method, and the root formulae of polynomial equations. In contrast, the `fcrn` code seeks the eigenvalues of the matrix $\mathbf{F}$ with the most elaborate numerical tools available (Biloti–Matioli–Yuan and Bini–Gemignani–Tisseur method), and yields the concentrations of the species via the spectral decomposition of the coefficient matrix. As to the operating conditions, the following differences must be mentioned: (*i*) the `fcrn_lrr` program demands (*a*) an $\mathbf{F}$ matrix whose eigenvalues can be obtained in closed form, and (*b*) stable Krylov- and Vandermonde matrices, and (*ii*) the `fcrn` code – according to our experiments is robust up to $K \leq 300$ components – can be applied to those FCRNs with diagonalizable $\mathbf{F}$ matrices. (This constraint is typically enforced for the common first-order kinetic models.) While it may occur that building a basis from Krylov and Vandermonde matrices can be stabilized via an appropriate scaling procedure, this question requires additional examination.

4.4. Interconversion kinetics of *n*-butane and *n*-pentane [3]

Finally, so-called interconversion parameters (activation energies, activation free energies at 0 and 298 K, and interconversion rate coefficients at 298 K) of *n*-butane and *n*-pentane are estimated via the FPA approach. Within this project a FPA model (FPA_{na}) is introduced, applying electronic structure theory techniques as inexpensive as possible, and providing definitive approximations with reliable uncertainties for the aforementioned physico-chemical quantities. The best estimates of the FPA_{na} model are collected into **Table 4.4**, where $\{t, g^{\pm}\}$ and $\{tt, tg^{\pm}, g^{\pm}g^{\pm}, g^{\pm}x^{\mp}\}$ are two sets containing the unique conformers of *n*-butane and *n*-pentane, respectively.

$\mathcal{R}$	$\Delta^{\ddagger}E_{\mathcal{R}}$ / cal mol^{-1}	$\Delta^{\ddagger}G_{\mathcal{R}}(0)$ / cal mol^{-1}	$\Delta^{\ddagger}G_{\mathcal{R}}(298\ \mathrm{K})$ / cal mol^{-1}	$k_{\mathcal{R}}(298\ \mathrm{K})$ / s^{-1}
$t \rightarrow g^{\pm}$	3331(30)	3347(95)	3901(158)	$103(7)\times10^{7}$
$g^{\pm} \rightarrow t$	2738(50)	2652(90)	3216(148)	$33(2)\times10^{8}$
$g^{\pm} \rightarrow g^{\pm}$	4840(47)	4916(91)	5499(151)	$69(4)\times10^{6}$
$tt \rightarrow tg^{\pm}$	3124(28)	3115(96)	3327(146)	$27(2)\times10^{8}$
$tg^{\pm} \rightarrow tt$	2524(42)	2412(91)	3047(153)	$44(3)\times10^{8}$
$tg^{\pm} \rightarrow g^{\pm}g^{\pm}$	2930(36)	2932(106)	3716(190)	$14(1)\times10^{8}$
$g^{\pm}g^{\pm} \rightarrow tg^{\pm}$	2597(38)	2404(77)	2551(112)	$101(5)\times10^{8}$
$g^{\pm}g^{\pm} \rightarrow g^{\pm}x^{\mp}$	6162(55)	6110(143)	6186(179)	$22(2)\times10^{6}$
$g^{\pm}x^{\mp} \rightarrow g^{\pm}g^{\pm}$	4258(33)	4236(77)	4723(143)	$26(1)\times10^{7}$
$tg^{\pm} \rightarrow g^{\pm}x^{\mp}$	2784(30)	2809(110)	3635(200)	$16(1)\times10^{8}$
$g^{\pm}x^{\mp} \rightarrow tg^{\pm}$	544(9)	404(17)	1004(86)	$137(5)\times10^{9}$
$tg^{\pm} \rightarrow tg^{\pm}$	423(6)	212(38)	682(118)	$24(1)\times10^{10}$
$g^{\pm}x^{\mp} \rightarrow g^{\pm}x^{\mp}$	4763(56)	4829(101)	5497(170)	$69(5)\times10^{6}$

Table 4.4 FPA_{na} estimates for the gase-phase interconversion parameters related to the unique conformers of *n*-butane and *n*-pentane. The uncertainties of the estimated values are given in parentheses.

In the FPA_{na} protocol the following computations are included: (*i*) reference geometries are optimized at the DSD-PBEP86-D2/cc-pVTZ level of DFT theory, (*ii*) single-point energies are determined at these structures using the CCSD(T)(fc)/cc-pV*X*Z, MP2(full)/cc-pCV*X*Z, MP2(fc)/cc-pV5Z, and MP2(fc)/cc-pV*X*Z ($X = 2,3,4$) electron correlation methods, (*iii*) harmonic vibrational analyses are executed at the DSD-PBEP86-D2/cc-pV*X*Z (*X* = 2,3) levels, and (*iv*) HDCPT2 computations are performed with the MP2(fc)/6-31G* procedure. Knowing that our FPA_{na} model results in high-accuracy estimates for the interconversion parameters of *n*-butane and *n*-pentane, we believe that the same protocol will work well for modeling the same quantities of longer *n*-alkanes, as well.

5. Publications related to the dissertation

[1] Simple algebraic solutions to the kinetic problems of triangle, quadrangle, and pentangle reactions

R. Tóbiás and G. Tasi,

Journal of Mathematical Chemistry, **2016**, *54*, 85.

IF_{2016} = 1.308

Independent citations: 4

[2] First-order chemical reaction networks I: theoretical considerations

R. Tóbiás, L. L. Stachó, and G. Tasi,

Journal of Mathematical Chemistry, **2016**, *54*, 1863.

IF_{2016} = 1.308

Independent citations: 1

[3] 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, **2017** (*in press*).

IF_{2016} = 3.229

Independent citations: 0

6. Other publications

[1] Vector algebra and molecular symmetry: a tribute to Professor Josiah Willard Gibbs

G. Tasi, L. Gyevi-Nagy, R. Tóbiás, and T. S. Tasi,

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

IF_{2013} = 1.270

Independent citations: 0

[2] Cycle bases to the rescue

R. Tóbiás, T. Furtenbacher, and A. G. Császár,

Journal of Quantitative Spectroscopy and Radiative Transfer, **2017**, *203*, 557.

IF_{2016} = 2.419

Independent citations: 0

[3] Critical evaluation of measured rotational-vibrational transitions of four sulfur isotopologues of $S^{16}O_2$

R. Tóbiás, T. Furtenbacher, A. G. Császár, O. V. Naumenko, J. Tennyson, P. Kumar, and B. Poirier,

Journal of Quantitative Spectroscopy and Radiative Transfer, **2018** (*under revision*).

7. Conference lectures and posters

[1] Exotic isomerization reactions: solution to the kinetic problem of a general reaction scheme?

R. Tóbiás and G. Tasi,
XXXV. Chemistry Presentations for Young Scientists, KEN
Szeged, **2012** (lecture)

[2] Experimental rotation-vibration transitions and energy levels for sulfur dioxide

T. Furtenbacher, R. Tóbiás, A. G. Császár, B. Poirier, J. Tennyson, V.-M. Hornemann, O. V. Naumenko, I. A. Vasilenko, A. Z. Fazliev
The 25th Colloquium on High-Resolution Molecular Spectroscopy, HRMS 2017
Helsinki, Finnland, **2017** (poster)

Total peer-reviewed publications: 5	out of this, related to the dissertation: 3
Cumulative impact factor: 9.534	out of this, related to the dissertation: 5.845
Total independent citations: 5	out of this, related to the dissertation: 5