A Mathematical Model for Analytical Fitting of Amino Acid Diamide Conformational Potential Energy Surfaces

John Justine S. Villar, Anita Rágyanszki, David Setiadi, Béla Viskolcz, Imre G. Csizmadia, and Adrian Roy L. Valdez

***Abstract*—The use of mathematical functions to model the topology of conformational potential energy surfaces (PES) is an alternative to more computer-intensive electronic structure calculations, but the choice and complexity of mathematical functions are crucial in achieving more accurate results.**

This paper presents an improved model to model the topology of three amino acid diamide PESs, through a linear combination of a Fourier series and a mixture of Gaussian functions. Results yield a significantly small error, with an average RMSE of 2.9786 kJ·mol^{-1} for all fits, which suggest that these functions may accurately represent the topology of the PESs, with minimal error.

This study lays a preliminary assessment for mathematical representation of amino acid PES, with less number of parameters. This may also be used to assess the conformational stability of peptides, in relation to its component amino acids.

***Index Terms*—Potential energy surfaces, conformational analysis, mathematical modeling, numerical analysis, peptide structure prediction.**

I. Introduction

The protein folding problem is considered as one of the most difficult problems in biology, and it is still not fully understood despite the advancements from the available experimental and computational methods. Attacking this problem computationally also entails a large amount of data which deemed too complex even with the advent of advanced computers, while empirical methods demand stringent experimental protocols that is usually expensive and tedious to perform.

One aspect of dissecting the problem is to look onto the energy profile of a protein of interest, as the potential energy of a foldamer allows us to determine the relative stability of each possible conformation conformation. In principle, finding the stable foldamers of the protein requires an efficient sampling of the entire conformational space of the protein, in which there is an associated *potential energy surface* (PES), and the global minimum of the surface may correspond to the energy of native fold.

The potential energy surfaces are important in analyzing conformations because these are usually used in visualizing the relationship between the energy profile and geometry of the molecules of interest, as well as in understanding how structure prediction methods locate and characterize its preferred conformations. However, the time and space complexity of empirically producing PESs (through electronic structure calculations) increases exponentially with the number of atoms in the system.

One dimensional trigonometric fit to simple internal rotation energies has been proposed by Pople *et al* in the early 1970s [1], [2]. For potential energy surfaces (2D) and hypersurfaces (nD, $n \geq 3$), functions of at least two (and at most n) variables are necessary, and these were achieved for relatively simple surfaces [3]-[5]. More recently, the problem has been reinvestigated to see how does the mathematical complexity of Fourier series relates with the complexity of the topology of the PEC or PES [6].

The computational investigation of the PESs of the amino acids in terms of backbone dihedral angles ϕ and ψ spans more than twenty years. Several studies include alanine [7]-[10], valine [11], and glycine [12]. However, a previous work attempted to model the energy landscape of the amino acid diamides mathematically [13], where it was initially tested only with three amino acid diamides, i.e., glycine, alanine and valine, and using a combination of two forms of Fourier series and a mixture of Gaussian functions.

This paper presents an improvement of a fitting procedure presented in [13], in which a combination of a form of Fourier series and a mixture of Gaussian functions was used, yielding a reduced number of terms for both functions, which implies a significant dimension reduction and less computational time. Results also show that this process gives a significant improvement in fit quality, with respect to R^2 and RMSE values, compared to fitting Fourier series to peptide conformational PES.

II. Analytical Fitting of Potential Energy Surfaces

A simpler mathematical representation of the conformational potential energy surface may be used to decipher problems related to peptide folding. Consequently, fitting mathematical functions for computed grid points will lead to a mathematical representation of such a

Manuscript received July 25, 2016; revised December 29, 2016. This work is supported by grants from Philippine Council for Industry, Energy and Emerging Technologies Research and Development through Engineering Research and Development for Technology HRDP and "Bridging the Human Resource Competency Gaps in Support of the National R&D Agenda" Program.

John Justine S. Villar and Adrian Roy L. Valdez are with the Scientific Computing Laboratory, Department of Computer Science, University of the Philippines, Diliman, 1101 Quezon City, Philippines (e-mail: john_justine.villar@upd.edu.ph, alvaldez@dcs.upd.edu.ph).

Anita Ragyánszki and Béla Viskolcz are with the Institute of Chemistry, Faculty of Materials Science and Engineering, University of Miskolc, H-3515 Miskolc-Egyetemváros, Hungary (e-mail: anita8602@yahoo.com, bela.viskolcz@uni-miskolc.hu).

David Setiadi and Imre G. Csizmadia are with the Department of Chemistry, University of Toronto, M5S 3H6 Toronto, Ontario, Canada (e-mail: davidsetiadi@yahoo.com, icsizmad@rogers.edu).

conformational problem. An example of a conformational PES is shown in Fig. 1.

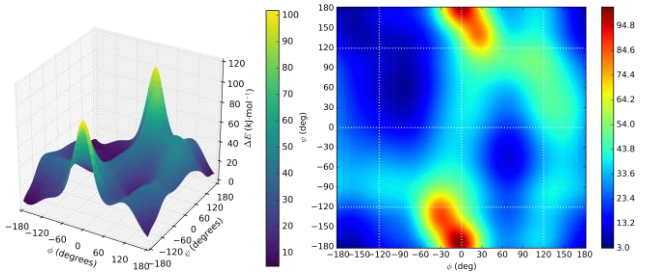

Fig. 1. Ramachandran potential energy surface of N-Ac-Ala-N-Methylamide ($HCONH\text{-}CHCH_3\text{-}CONH_2$), with associated heat map.

Due to the computational time needed to produce a smooth potential energy surface, there are studies that used different set of functions to illustrate its topology, such as trigonometric [14], power series [15] or Gaussian [14]. Power series can be used successfully to fit PES in the reaction subspace [15] while in the conformational subspace, trigonometric functions are favored [14].

Note that a conformational PES shows periodicity in general, but most of the amino acids are asymmetrical (glycine being the exception). This yields dissymmetry in the corresponding PES.

Considering this constraint, the fitting procedure was developed on the following two premises [13]:

- use global functions that cover a large region with "fairly good" accuracy, and
- use functions that perform well in a local region.

In this paper, the Fourier series were chosen as the global function, as it captures the general periodic nature of the PES, and improving the fit by using a mixture of Gaussian functions to represent the local dissymmetries.

Specifically, a linear combination of two functions, in which later referred to as $E_{f+g}(\phi,\psi)$ were used to fit the amino acid diamide potential energy surfaces:

- mixture of Fourier series of the form

$$E_f(\phi,\psi) = k + \sum_{i=1}^{N_f} (w_i \cos iw\phi + x_i \cos iw\psi + y_i \sin iw\phi + z_i \sin iw\psi) \quad (1)$$

where $\phi,\psi \in [-\pi,\pi]$, N_f be the multiplicity (i.e., number of the terms), k is the constant term in the series, ω is the conversion factor from degrees to radians, i.e., $\omega = 2\pi/360$, and $\{w_i\},\{x_i\},\{y_i\},\{z_i\} \in \mathbb{R}$ with $i = 1,\ldots,N_f$; and

- mixture of spherical Gaussian functions, which is fitted at the recognizable peaks of the input grid, of the form

$$E_g(\phi,\psi) = \sum_{j=1}^{N_g} A_j \exp\left[-\frac{c_{\phi j}(\phi-\bar{\phi}_j)^2}{2\sigma_{\phi j}^2} - \frac{c_{\psi j}(\psi-\bar{\psi}_j)^2}{2\sigma_{\psi j}^2}\right] \quad (2)$$

where $\phi,\psi \in [-\pi,\pi]$, N_g be the multiplicity, $\{A_j\}$ be the amplitude, $\{(\bar{\phi}_i,\bar{\psi}_i)\}$ and $\{\sigma_{\phi j}^2,\sigma_{\psi j}^2\}$ be the center and eccentricity of the ellipsoids, respectively, all with $j = 1,\ldots,N_g$. Also, $\exp(\cdot)$ is the exponential function. Thus,

$$E_{f+g}(\phi,\psi) = E_f(\phi,\psi) + E_g(\phi,\psi), \phi,\psi \in [-\pi,\pi] \quad (3)$$

III. Methods

This paper considers three amino acid diamides, such as glycine, alanine and valine, with corresponding chemical structures shown in Fig. 2.

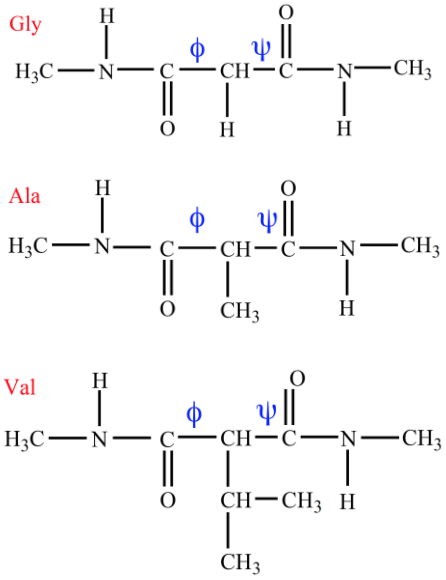

Fig. 2. Chemical structures of three amino acid diamides.

For each of these molecules, the energy corresponding to the rotations of two peptide backbone dihedral angles ϕ and ψ were computed over the interval $[-180°, 180°]^2$ with step size of $15°$. These values were generated from electronic structure calculations under the B3LYP hybrid functional and 6-31G(d) basis set (in gas phase) implementation on Gaussian09 [16] software package, with the input Z-matrices are constructed following [17]. For every amino acid considered, the geometrical parameters were fully relaxed except for the constrained torsion angles ϕ and ψ.

The relative energy, defined as the energy difference from the grid point energy minimum, were also subsequently calculated. This will be used in generating a 625-point grid surface, which an input for the fitting procedure.

The paper also used a nonlinear least square fitting process, through the Levenberg-Marquardt algorithm [18], [19], and is implemented in Python programming language.

IV. Results and Discussion

In this section, the resulting potential energy surfaces generated mathematically for the three amino acid diamides are presented, as well as discussion on the relevance of the mixture of Gaussian functions in improving the goodness-of-fit.

A. Ramachandran-Type Potential Energy Surfaces

Fig. 3 shows the heat maps of the Ramachandran-type PES, with the number of parameters used is detailed in Table I. Here, the number of terms of the Fourier series N_f is fixed at 5, after checking the relative RMSE values obtained. The number of terms in the Gaussian mixture is obtained through inspection of the minima points of the grid *G*.

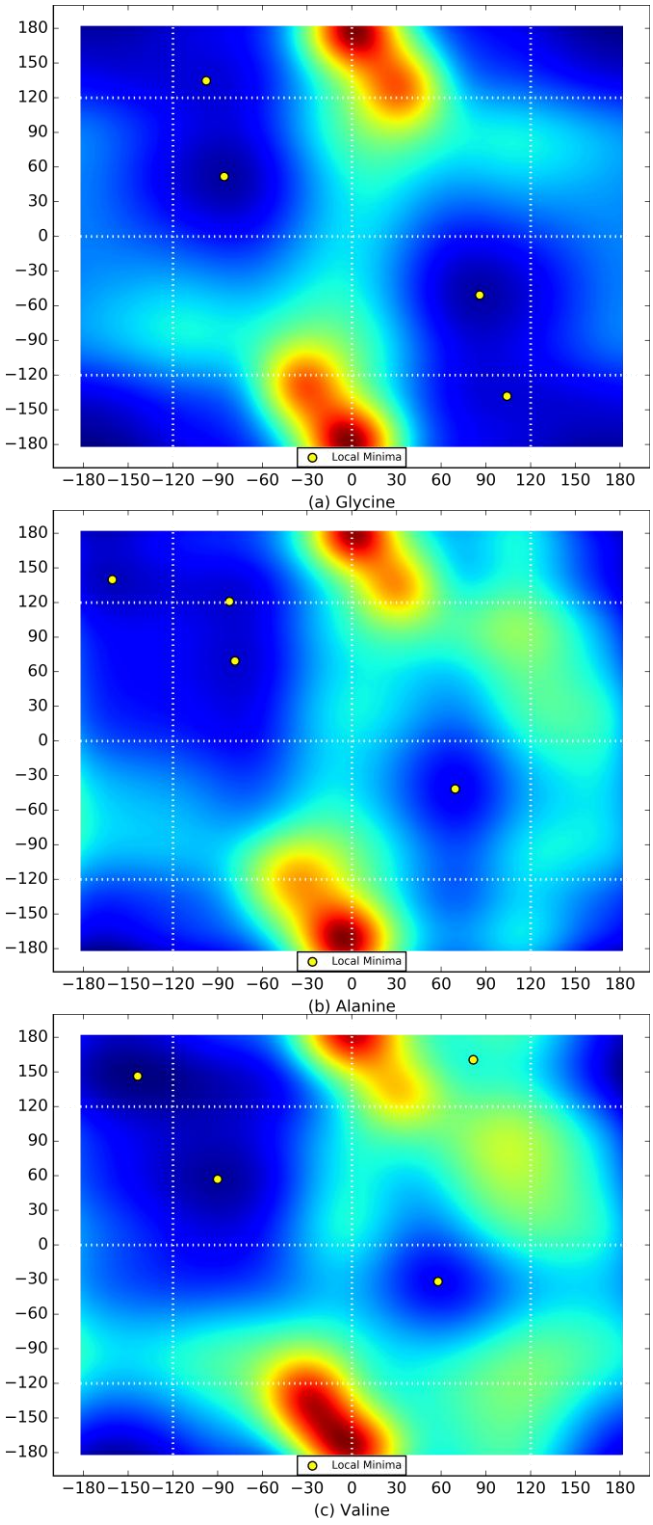

Fig. 3. Resulting PES Heat Maps. Yellow dots correspond to the position of the local minima.

From the preliminary analyses, it has been found out that if we set N_f=5, the relative change in the RMSE on the three amino acid diamide PESs change only by at most 0.0001 even with increased N_f. Also, requiring more number of terms for the Fourier series yields a more ragged surface.

Furthermore, checking the grid maxima positions in the PES for the Gaussian mixture helped in increasing the accuracy of the fit, as only prominent peaks are only used in the fitting procedure.

TABLE I: SUMMARY OF MATHEMATICAL FIT RESULTS

Amino Acid	N_f	N_g	Parameters	RMSE (kJ mol^{-1})
Gly	5	8	77	2.6343
Ala	5	9	84	2.8644
Val	5	9	84	3.4370

The overall fit procedure yields an average RMSE of 2.9786 kJ·mol^{-1} with the glycine PES yielded the lowest RMSE of 2.6343 kJ·mol^{-1}. Note that the results obtained in this study is comparable to results in [13], which yielded an average RMSE of 5.733 kJ·mol^{-1}.

Note that [13] used a more complex function to fit PES, which is given by

$$E_{f+g}(\phi,\psi) = E_f(\phi,\psi) + E_n(\phi,\psi) + E_g(\phi,\psi) \quad (4)$$

where

$$E_n(\phi,\psi) = \sum_{j=1}^{N_n} [a_i \cos(iw\phi + iw\psi) b_i \cos(iw\phi - iw\psi) + c_i \cos(iw\phi + iw\psi) d_i \sin(iw\phi - iw\psi) + e_i \sin(iw\phi + iw\psi) f_i \cos(iw\phi - iw\psi)] \quad (5)$$

The large error obtained in [13] may be due to Equation (5) that may have contributed to the overfitting of the PES grid *G*. Both papers used the same grid points for analytical fitting.

Aiming for further dimension reduction in mathematical fitting, the choice of this study to use only one Fourier series, instead of two, is that the (mathematical) interactions between ϕ and ψ axes can be captured by the Gaussian mixture (Equation (2)), which also belongs to the family of exponential functions.

B. Position of Local Minima

Table II details the location of the local minima for the amino acid diamides considered in the study. The conformation labels are given in Fig. 4. The energy values are reference to input grid minimum energy values.

The heat maps in Fig. 3 show that it had captured all the prominent minima of the PES, with its (ϕ,ψ) positions comparable with the result in [13], but with less number of parameters used. In fact, all of the fit PES in the previous study used 118 parameters, while this study used at most 84 parameters only.

Furthermore, electronic structure calculations to locate the minima positions were performed for the three amino acids, through a single-point optimization in Gaussian09. Results show that all the detected fit minima are also obtained in the electronic structure calculations.

TABLE II: LOCAL MINIMA POSITION OF THE FIT PES

Amino Acid	Conformation	ϕ	ψ	$E_{f+g}(\phi,\psi)$ (kJ mol^{-1})
Gly	γ_L	-85.68	51.79	2.8775
	γ_D	85.84	-50.72	3.1369
	ε_L	-97.68	134.13	6.8727
	ε_D	104.14	-138.18	6.9149
Ala	β_L	-160.71	139.82	6.6793
	γ_L	-78.43	69.31	8.3289
	ε_L	-82.17	120.96	8.3411
	γ_D	69.36	-41.51	12.3002
Val	β_L	-152.24	143.61	1.4271
	γ_L	-88.73	54.06	6.5325
	γ_D	58.21	-30.87	15.2100
	ε_L	86.88	152.62	43.6705

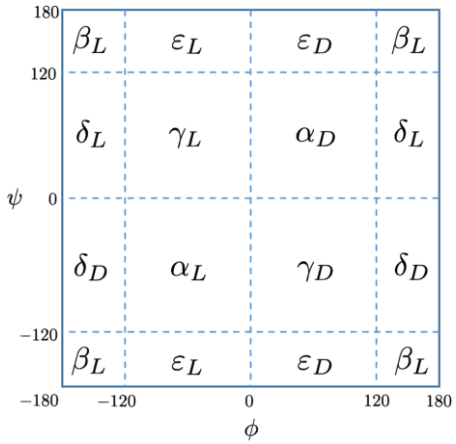

Fig. 4. Schematic Topology of Conformational PES of an amino acid. Labels correspond to the approximate location of the conformers.

C. Use of Gaussian Mixtures in Improving Accuracy

Table III summarizes the number of parameters, and corresponding RMSE values used in the different fit procedures.

TABLE III: RMSE VALUES OF THE TWO FIT FUNCTIONS

Amino Acid	First Fit (N_f)		Second Fit (N_{f+g})	
	Parameters	RMSE	Parameters	RMSE
Gly	21	11.9973	77	2.6343
Ala	21	12.5792	84	2.8644
Val	21	13.0977	84	3.4370

The fitting using only the Fourier series E_f yields an average RMSE of 12.5581 kJ mol^{-1}, while using E_{f+g} returned an average RMSE of 2.9786 kJ mol^{-1}. Although E_{f+g} uses around four times the number of parameters as E_f, this has lessened the RMSE by around 9.5795 kJ mol^{-1}. Furthermore, the average RMSE obtained in this study is within the acceptable value of 3.5 kJ mol^{-1}.

With the significant differences in RMSE between the first and second fit procedures for the three amino acid PES, this shows the relevance of further using mixture of Gaussian functions to improve fit quality.

V. CONCLUSION

Electronic structure calculations are usually performed to determine the relative stability of the different peptide conformations. However, the time complexity of this calculations to locate the minima structures depend on the level of theory needed and accuracy through spacing of the grid points. In every single point optimization, the best conformation in a particular neighborhood of initial constraints, it takes minutes to hours to converge to this minimum.

To reduce this time complexity, several studies looked onto representing the conformational potential energy surface, which contains several internal bond rotations, through multidimensional mathematical functions. However, the choice and complexity of this functions still affect both the accuracy and time needed to yield acceptable results, as more complex functions require more time to determine the optimal parameters that can describe the PES. In fact, previous studies, such as [13], used a significant number of parameters, which resulted to a large amount of time doing the mathematical fitting.

This paper provides a relatively simple mathematical model, which consists of a linear combination of a Fourier series and a mixture of Gaussian functions, that can provide location of two backbone angles, ϕ and ψ, with reasonable accuracy and much less computational time needed. Results yielded an average RMSE of 2.6343 kJ·mol^{-1} for the three amino acid diamides considered, which is within the acceptable error threshold of 3.4 kJ·mol^{-1} used in chemistry.

Furthermore, albeit the general periodic behavior of the conformational PES, this study also emphasized the importance of using mixture of Gaussian functions in improving the accuracy of the fit.

For future work, it is interesting to test the fitting procedure with a grid computed with a higher level of theory and basis set. A similar model can also be constructed for higher dimensional PES, which can describe the polypeptide relative conformational stabilities.

ACKNOWLEDGMENTS

John Justine S. Villar is supported by grants from Philippine Council for Industry, Energy and Emerging Technologies Research and Development through Engineering Research and Development for Technology HRDP and "Bridging the Human Resource Competency Gaps in Support of the National R&D Agenda" Program.

Part of this research was carried out using the HPC facilities under the CoARE Project, DOST Advanced Science and Technology Institute, Diliman, Quezon City, Philippines, and of the Department of Chemical Informatics, University of Szeged, Hungary.

REFERENCES

[1] L. Radom, W. J. Hehre, and J. A. Pople, "Molecular orbital theory of the electronic structure of organic compounds. XIII. fourier component analysis of internal rotation potential functions in saturated molecules," *J Am Chem Soc*, vol. 94, pp. 2371-2381, 1972.

[2] L. Radom, W. Lathan, W. Hehre, and J. A. Pople, "Molecular orbital theory of the electronic structure of organic compounds. XVII. internal rotation in 1,2-disubstituted ethanes," *J Am Chem Soc*, vol. 95, pp. 693-698, 1973.

[3] G. R. Demare, M. R. Peterson, I. G. Csizmadia, and O. P. Strausz, "Conformational energy surfaces of triplet-state isomeric methyloxiranes," *J Comp Chem*, vol. 1, pp. 141-148, 1980.

[4] R. Peterson, R. D. E. Mare, and A. F. D. Roosevelt, "Conformations formaldehyde, acetone of triplet carbonyl compounds: And the

reaction of 0 (3P) atoms with olefins proceeds via a C % Ci, triplet biradical [l], which can undergo intersystem crossing to the so state and be stabilized as an epoxide," *J Mol Struct (THEOCHEM)*, vol. 86, pp. 131-147, 1981.
[5] T. A. K. Kehoe, M. R. Peterson, G. A. Chass, B. Viskolcz, L. Stachod, and I. G. Csizmadia, "The fitting and functional analysis of a double rotor potential energy surface for the R and S enantiomers of 1-chloro-3-fluoro-isobutane," *J. Mol. Struct. (THEOCHEM)*, vol. 666-667, pp. 79-87, 2003.
[6] A. Ragyanszki, A. Suranyi, I. G. Csizmadia, A. Kelemen, S. J. K. Jensen, S. Y. Uysal, and B. Viskolcz, "Fourier type potential energy function for conformational change of selected organic functional groups," *Chem Phys Lett*, vol. 599, 169-174, 2014.
[7] M. A. McAllister, A. Perczel, P. Csaszar, W. Viviani, J. L. Rivail, and I. G. Csizmadia, "Peptide models IV: Topological features of molecular mechanics and ab-initio 2D-ramachandran maps. A conforma tional data base for For-L-Ala-NH2 and Ac-L-Ala-NH2Me," *J Mol Struct*, vol. 288, pp. 161-180, 1993.
[8] G. Endredi, C.-M. Liegner, M. McAllister, A. Perczel, J. Ladik, and I. Csizmadia, "Peptide models VIII. folding energetics of the "extended-like" or (bl)n or (c5)n conformation of alanine oligopeptides," *J Mol Struct*, vol. 306, pp. 1-7, 1994.
[9] A. Rodriguez, H. A. Baldoni, F. Suvire, R. Nieto-Vasquez, G. Zamarbide, R. D. Enriz, O. Farkas, A. Perczel, and I. Csizmadia, "Characteristics of ramachandran maps of alanine diamides as computed by various molecular mechanics, semiemprical and ab initio MO methods," *J Mol Struct (THEOCHEM)*, vol. 455, pp. 275-302, 1998.
[10] S. G. Stepanian, I. D. Reva, E. D. Radchenko, and L. Adamowicz, "Conformational behavior of α-alanine. Matrix-isolation infrared and theoretical DFT and ab initio study," *J Phys Chem A*, vol. 102, 1998.
[11] W. Viviani, J.-L. Rivail, A. Perczel, and I. G. Csizmadia, "Peptide models III. conformational potential energy hypersurface of formyl-L-valinamide," *J Am Chem Soc*, vol. 115, 1993.
[12] H. A. Baldoni, G. N. Zamarbide, R. D. Enriz, E. A. Jauregui, O. Farkas, A. Perczel, S. J. Salpietro, and I. G. Csizmadia, "Peptide models XXIX. cis-trans isomerism of peptide bonds: ab initio study on small peptide model compound; the 3D-Ramachandran map of formylglycinamide," *J Mol Struct (THEOCHEM)*, vol. 500, 2000.
[13] A. Ragyanszki, K. Z. Gerlei, A. Suranyi, A. Kelemen, S. J. K. Jensen, I. G. Csizmadia, and B. Viskolcz, "Big data reduction by _tting mathematical functions: A search for appropriate functions to fit Ramachandran surfaces," *Chem Phys Lett*, vol. 625, pp. 91-97, 2015.
[14] M. R. Peterson and I. G. Csizmadia, "Analysis of the topological features of the conformational hypersurface of n-butane," *J Am Chem Soc*, vol. 100, pp. 6911-6916, 1978.
[15] D. Autrey, N. Meinander, and J. Laane, "A two-dimensional potential energy surface and associated quantum states for the ring-puckering vibrations of two equivalent rings: a study of bicyclo[3.3.0]oct-1,5-ene," *J Phys Chem A*, vol. 108, pp. 409-416, 2004.
[16] M. J. Frisch *et al.*, *Gaussian09 Revision A.02*, 2009.
[17] G. A. Chass, M. A. Sahai, J. M. S. Law, S. Lovas, O. Farkas, A. Perczel, J.-L. Rivail, and I. G. Csizmadia, "Toward a computed peptide structure database: The role of a universal atomic numbering system of amino acids in peptides and internal hierarchy of database," *Int J Quant Chem*, vol. 90, pp. 933-968, 2002.
[18] K. Levenberg, "A method for the solution of certain problems in least squares," *Quart Appl Math*, vol. 2, pp. 164-168, 1944.
[19] D. Marquardt, "An algorithm for least-squares estimation of nonlinear parameters," *SIAM J Appl Math*, vol. 11, pp. 431-441, 1963.

John Justine S. Villar is a PhD student at the Department of Computer Science, University of the Philippines, Diliman, Quezon City Philippines, where he is an active member of the Scientific Computing Laboratory. He is currently a visiting research student at the Drug Discovery Research Center and Institute of Chemistry, Faculty of Material Sciences and Engineering, University of Miskolc, Hungary.

Anita Rágyanszki received her Ph.D. in theoretical chemistry in 2015 from University of Szeged, Hungary. Currently she is a Postdoctoral Researcher at the University of Miskolc, Hungary. She is a member of the Hungarian Drug Discovery Research Center. Her research interest includes peptide chemistry, chemical evolution, space chemistry, potential energy surfaces, mathematical modelling and theoretical chemistry.

David H. Setiadi completed his degree in engineering physics from McMaster University, Canada and his Ph. D in Medicinal Chemistry in from the University of Szeged, Hungary. He currently is the Chief Technology Officer at @Point of Care and Chief Information Officer at Projects in Knowledge. He is a founder of dgNetrix Communications. His research interests includes biological signaling, protein chemistry, cognitive sciences, machine learning and predictive algorithms.

Béla Viskolcz received his Ph.D. in chemistry in 1996 from University of Halle-Wittenberg, Germany and his Cs.C. in Chemistry in 1998 from Hungarian Academy of Sciences, Budapest, Hungary. Currently he is a head of the Department at the University of Miskolc, Hungary. He is a member of the Hungarian Drug Discovery Research Center. He has several publications in the fields of peptide chemistry, chemical evolution, molecular ageing, conformational analysis, computational chemistry and industrial chemistry.

Imre G. Csizmadia received his Ph.D. in chemistry in 1962 from University of British Columbia, Canada. Currently he is an emeritus professor at the University of Miskolc, Hungary and at the University of Toronto, Canada. He is a member of the Hungarian Drug Discovery Research Center and the Scientific Academy. His research interest includes peptide chemistry, chemical evolution, organic chemistry, conformational analysis, theoretical chemistry and mathematical modelling.

Adrian Roy L. Valdez is an assistant professor at the Department of Computer Science, College of Engineering, University of the Philippines Diliman, where he also heads the Scientific Computing Laboratory under the same department. His research interests include partial differential equations, mathematical modelling, bioinformatics, stochastic analysis and mathematical finance.