Dimension Reduction in Conformational Analysis: A Two-Rotor Mathematical Model of Amino Acid Diamide Conformational Potential Energy Surface

Journal:	<i>Canadian Journal of Chemistry</i>
Manuscript ID	cjc-2017-0124.R1
Manuscript Type:	Article
Date Submitted by the Author:	22-May-2017
Complete List of Authors:	Villar, John Justine; University of the Philippines Diliman, Department of Computer Science; Miskolci Egyetem, Institute of Chemistry Valdez, Adrian Roy; University of the Philippines Diliman, Department of Computer Science Setiadi, David H.; University of Toronto, Department of Chemistry Fiser, Bela; Miskolci Egyetem, Institute of Chemistry Viskolcz, Béla; Miskolci Egyetem, Institute of Chemistry Csizmadia, G. Imre; University of Toronto, Department of Chemistry; Miskolci Egyetem, Institute of Chemistry Rágyanszki, Anita; University of Toronto, Department of Chemistry; Miskolci Egyetem, Institute of Chemistry
Is the invited manuscript for consideration in a Special Issue?:	N/A
Keyword:	potential energy surfaces, analytical fitting, parameter search, conformational analysis

SCHOLARONE™
Manuscripts

Dimension Reduction in Conformational Analysis: A Two-Rotor Mathematical Model of Amino Acid Diamide Conformational Potential Energy Surface

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

Abstract: The conformational potential energy surface (PES) of a molecule provides insights into the relative stability of the possible foldamers. However, the time and space complexity of electronic structure calculations, commonly used to generate PES, increases exponentially with an increasing number of atoms.

The use of mathematical functions to model the topology of conformational PES is an alternative to more computer-intensive quantum chemical calculations, but the choice and complexity of functions used are crucial in achieving more accurate results.

This paper presents a method to illustrate the topology of 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 $4.9946 \text{ kJ}\cdot\text{mol}^{-1}$ for all fits, which suggest that these functions may be used to represent the topology of the PESs, with around twofold order of magnitude decrease in computational time, with respect to DFT electronic structure calculations. This study ultimately aims to provide a foundation for a framework on building polypeptide PES from individual amino acid PESs.

Key words: potential energy surfaces, analytical fitting, parameter search, conformational analysis.

John Justine S. Villar. Scientific Computing Laboratory, Department of Computer Science, University of the Philippines, Diliman, 1101 Quezon City, Philippines, and Institute of Chemistry, University of Miskolc, 3515 Miskolc-Egyetemváros, Hungary

Adrian Roy L. Valdez. Scientific Computing Laboratory, Department of Computer Science, University of the Philippines, Diliman, 1101 Quezon City, Philippines

David H. Setiadi. Department of Chemistry, University of Toronto, M5S 3H6 Toronto, Ontario, Canada

Béla Fiser and Béla Viskolcz. Institute of Chemistry, Faculty of Materials Science and Engineering, University of Miskolc, 3515 Miskolc-Egyetemváros, Hungary

Imre G. Csizmadia and Anita Rágyanszki.¹ Institute of Chemistry, University of Miskolc, 3515 Miskolc-Egyetemváros, Hungary, and Department of Chemistry, University of Toronto, M5S 3H6 Toronto, Ontario, Canada

¹Corresponding author (e-mail: a.ragyanszki@utoronto.ca).

1. Introduction

The protein folding problem, which is the process of how a linear sequence of amino acids acquires its biologically active three-dimensional shape, is still considered as one of the most difficult problems in biology, and it is yet not fully understood despite advancements in currently available experimental and computational methods. Attacking this problem computationally also entails a large amount of data which is deemed too complex, despite the advent of advanced computers, while empirical methods demand stringent experimental protocols which are usually expensive and tedious to perform.

One aspect of dissecting this problem is to observe 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. In principle, finding the stable foldamers of a protein requires an efficient sampling of the entire conformational space of the protein, to which there is an associated *potential energy surface* (PES). The global minimum of the surface may correspond to the energy of native fold.

Potential energy surfaces are important in the analysis of conformations because these are usually used to visualize the relationship between the energy profile and geometry of the molecules of interest, as well as to understand how structure prediction methods locate and characterize their preferred conformations. However, the time and space complexity of empirically produced PESs (through electronic structure calculations) increases exponentially with the number of atoms in the system.

An accurate characterization of all physical forces operating on the system between different atoms is only obtainable by applying quantum mechanical (QM) calculations. This is due to the fact that the forces resulting from the spatial distribution of electrons around atoms, which require QM calculations. However, this procedure is computationally expensive, and researchers resort to approximation of the real physical forces in order to simplify calculations. This refers to *molecular mechanics*, since it approximates the protein system using Newton's classical mechanics, i.e., all atoms of the systems are considered and the energy content is described using a set of mathematical equations called the *force field*, which treats atoms and covalent bonds as balls and springs, respectively.¹² These force fields have the form

$$V(x) = V_{length} + V_{angle} + V_{dihedrals} + V_{electrostatic} + V_{weak} \quad (1.1)$$

where x is the vector representing the conformation of the molecule. Each expression in the force field describes the *potential energy* resulting from a different interaction between any two atoms in the system.

Assuming that the force field is accurate, a calculation of the potential energy of the system allows us to determine the stability of one conformation. In principle, finding the native fold of the protein requires sampling the entire conformational PES efficiently.

Conformational analysis of organic molecules was initiated in the late 1970s, where simple hydrocarbons like ethane ($\text{H}_3\text{C}-\text{CH}_3$), propane ($\text{CH}_3-\text{CH}_2-\text{CH}_3$) and n-butane ($\text{CH}_3-\text{CH}_2-\text{CH}_2-\text{CH}_3$) exhibited such structural characteristics that provided the basis of conformational analysis.¹⁴ Methyl rotations ($-\text{C}-\text{CH}_3$) and ethyl rotations ($-\text{CH}_2-\text{CH}_3$) were very similar in a variety of compounds. In 1963, Ramachandran attempted to perform conformational analysis on simple peptides,²⁰ and the PES topology for the double rotor $\text{CONH}-\text{CH}_2-\text{CONH}$ turned out to be quite different from the propane surface. More recently, it turned out that even a compound as simple as n-pentane behaved in a non-ideal fashion.²³ Thus, it is reasonable to assume that surface topology is a function of two dihedral angles, ϕ and ψ , associated with the rotation about two bonds connected to the alpha-carbon of the amino acid.

One dimensional trigonometric fit to simple internal rotation energies was proposed by Pople et al in the early 1970s.^{16,17} For potential energy surfaces (3D) and hypersurfaces (n -dimensions, $n > 3$), functions of at least two (and at most n) variables are necessary, and these were achieved for

Villar et al.

relatively simple surfaces.^{7,11,15} More recently, the problem has been reinvestigated to determine how the mathematical complexity of Fourier series relate to the complexity of topology of the PEC or PES.¹⁸

This paper presents an improvement of a fitting procedure presented in a previous study,¹⁹ 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. This implies a further dimension reduction and less computational time. Results also show that this process significantly improves the goodness-of-fit, with respect to R^2 and RMSE values, as compared to fitting Fourier series to peptide conformational PES.

2. Preliminaries

Potential energy surfaces¹² are important because they aid us in visualizing and understanding the relationship between potential energy and molecular geometry, and in understanding how prediction methods locate and characterize structures of interest. However, the time needed to construct potential energy surfaces increases exponentially with the number of atoms in the system and the level of theory used.

A simpler mathematical model of the conformational potential energy surface may be used in order to provide a bottom-up framework to protein structure prediction. Consequently, fitting mathematical functions for computed grid points will lead to an approximate mathematical representation of a conformational PES. An example of a conformational PES is shown in Fig. 1.

Fig. 1. Ramachandran potential energy surface of N-Ac-Glycine-N-Methylamide ($\text{HCONH-CH}_2\text{-CONH}_2$)

Studies have used different set of functions, such as trigonometric,¹⁴ power series¹ or Gaussian functions,¹⁴ to describe a potential energy curve or surface. Power series can be used successfully to fit PES in the reaction subspace¹ while in the conformational subspace, trigonometric functions are preferred.¹⁴

Note that a conformational PES shows periodicity in general, but most of the amino acids are

asymmetrical (glycine being the exception). This yields dissymmetries in the corresponding PES. Considering this constraint, the fitting procedure was developed on the following two premises:¹⁹

- using global functions that cover a large region with “fairly good” accuracy, and
- using functions that perform well in a local region.

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

Subsequently, a linear combination of two functions, later referred to as $\Delta E_{f+g}(\phi, \psi)$, was used to fit the amino acid diamide potential energy surfaces, where $(\phi, \psi) \in [-\pi, \pi]^2$:

- a Fourier series of the form

$$\Delta E_f(\phi, \psi) = k + \sum_{i=1}^{N_f} (w_i \cos i\omega\phi + x_i \cos i\omega\psi + y_i \sin i\omega\phi + z_i \sin i\omega\psi), \quad (2.1)$$

where 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., $\omega = \frac{\pi}{180}$, and $\{w_i\}, \{x_i\}, \{y_i\}, \{z_i\} \in \mathbb{R}$, with $i = 1, \dots, N_f$, and

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

$$\Delta E_g(\phi, \psi) = \sum_{j=1}^{N_g} A_j \exp \left[-\frac{c_{\phi_j}(\phi - \bar{\phi}_j)^2}{2\sigma_{\phi_j}^2} - \frac{c_{\psi_j}(\psi - \bar{\psi}_j)^2}{2\sigma_{\psi_j}^2} \right], \quad (2.2)$$

where N_g be the multiplicity, $\{A_j\}$ be the amplitude, $\{(\bar{\phi}_j, \bar{\psi}_j)\}$ and $\{\sigma_{\phi_j}^2, \sigma_{\psi_j}^2\}$ be the center and axes of the ellipsoids, respectively, all with $j = 1, \dots, N_g$. Also, $\exp(\cdot)$ is the exponential function.

Thus,

$$\Delta E_{f+g}(\phi, \psi) = \Delta E_f(\phi, \psi) + \Delta E_g(\phi, \psi). \quad (2.3)$$

3. Methods

3.1. Benchmark Data

This paper considers three simple amino acid diamides, which include glycine, alanine and valine. The structures are illustrated in Fig. 2.

For each of these molecules, the energy $E(\phi, \psi)$ corresponding to the rotations of two peptide backbone dihedral angles, ϕ and ψ , were computed over the interval $[-180^\circ, 180^\circ]^2$, with a step size of 15° . These values were generated from DFT electronic structure calculations under the B3LYP^{2,22} hybrid functional and 6-31G(d) basis set⁹ implementation on GAUSSIAN09⁸ software package, and the input Z-matrices were constructed following.⁵ For every amino acid considered, the geometrical

Villar et al.

Fig. 2. Chemical Structures of 3 Amino Acid Diamides, and the Dihedral Angle (ϕ , ψ) Locations: (a) N-Acetyl-Glycine-N-methylamide, (b) N-Acetyl-Alanine-N-methylamide, (c) N-Acetyl-Valine-N-methylamide

parameters were fully relaxed except for the constrained torsion angles ϕ and ψ . The relative energy, given by the formula

$$\Delta E(\phi_i, \psi_i) = E(\phi_i, \psi_i) - \min_{(\phi, \psi) \in [-\pi, \pi]^2} E(\phi, \psi), \quad (3.1)$$

were also subsequently calculated, generating a 625-point grid surface, which was used as an input for the fitting procedure discussed in Section 3.2.

The goodness-of-fit was determined using the root mean square error (RMSE) and R^2 between the resulting fit and relaxed scan relative energy values.

For the comparative analysis, the molecular mechanics minimizations were performed with the sander module of the AMBER 12 software suite.^{4,21} Constraints were used to keep the dihedral angles of the peptides at the desired values $(\phi, \psi) \in [-\pi, \pi]^2$, while structures were defined using the ff99SB force field parameter set.¹⁰

3.2. Procedure for Analytical Fitting

Given a grid $G \subseteq [-\pi, \pi]^2$, and the corresponding relative energy ΔE values over G computed from DFT calculations, denoted by $\Delta E^*(G)$, the fitting procedure consists of three phases:

1. Enumerating the grid minima of $\Delta E^*(G)$
2. Nonlinear least squares (NLS) fit on $\Delta E_g(\phi, \psi)$
3. Nonlinear least squares fit on $\Delta E_{f+g}(\phi, \psi)$

The trust region reflective algorithm^{3,6} was used to obtain the optimal parameter values of ΔE_g and ΔE_{f+g} for the NLS fit procedures in the second and third phases. All calculations performed were implemented in the Python programming language.

Published by NRC Research Press

3.2.1. Enumerating the minima of $\Delta E(G)$

The first phase lists the minima of G with ΔE values of at most $\tau_g \max[\Delta E^*(G)]$, which are needed for the next phase, where τ_g is the threshold factor, also supplied by the user. The (ϕ, ψ) and the corresponding ΔE values in the above list will be used as in constructing the initial parameter vector for the second phase.

3.2.2. Nonlinear Least Squares Fit on $\Delta E_g(\phi, \psi)$

In this phase, a nonlinear least squares fit on $\Delta E_{f+g}(\phi, \psi)$ given $\Delta E^*(G)$ and an initial parameter vector g . The value for g is determined by setting the amplitude and ellipsoid centers to be the ΔE and (ϕ, ψ) values, respectively, which are obtained in the previous phase. The other parameters are set arbitrarily.

3.2.3. Nonlinear Least Squares Fit on $\Delta E_{f+g}(\phi, \psi)$

Starting with $N_f = 1$, an initial parameter f_{N_f} corresponding to the Fourier series component is generated, and an NLS fit is performed on $\Delta E_f(\phi, \psi)$, given $\Delta E^*(G)$ and a parameter vector f_{N_f} . The resulting (optimized) parameter vector f^* will be used in the following step.

The obtained parameter vectors $f_{N_f}^*$ and g^* is used to perform an NLS fit on ΔE_{f+g} , in which an initial parameter h_{N_f} is computed as a vector concatenation of $f_{N_f}^*$ and g^* . The fit will yield an optimal parameter vector $h_{N_f}^*$. The root mean square error is then computed between $\Delta E(G)$ and $\Delta E_{f+g}(\phi, \psi; h_{N_f}^*)$.

This process is repeated by increasing N_f until the relative RMSE is within ϵ_f , i.e.,

$$\left| RMSE \left[\Delta E(G), \Delta E_{f+g}(\phi, \psi; h_{N_f}^*) \right] - RMSE \left[\Delta E(G), \Delta E_{f+g}(\phi, \psi; h_{N_f-1}^*) \right] \right| \leq \epsilon_f, \quad (3.2)$$

where ϵ_f is the tolerance error supplied by the user. If the inequality above is satisfied, the optimal parameter vector $h_{N_f}^*$ will be considered as the parameter vector for the function $E_{f+g}(\phi, \psi)$ which will represent the PES.

4. Results and Discussion

4.1. Ramachandran-type Potential Energy Surfaces

Table 1 shows the number of parameters and the used to fit the coarse-grained Ramachandran PES obtained from DFT calculations. The conformation regions²⁰ provided in the table are also illustrated through Figure 3.

The associated contour maps in Figure 4 show that it had precisely captured all the prominent minima of the PES, with its (ϕ, ψ) positions comparable with the result in,¹⁹ 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 102 parameters only.

The overall fit procedure yields an average RMSE of 4.9946 kJ·mol⁻¹ with the glycine PES yielded the lowest RMSE of 4.8363 kJ·mol⁻¹. Note that the results obtained in this study is comparable to results in,¹⁹ which yielded an average RMSE of 5.733 kJ·mol⁻¹. Note that¹⁹ used a third function, i.e. $E(\phi, \psi) = E_f(\phi, \psi) + E_g(\phi, \psi) + E_x(\phi, \psi)$, where

$$\Delta E_x(\phi, \psi) = k + \sum_{i=1}^{N_x} (a_i \cos i\omega\phi \cos i\omega\psi + b_i \cos i\omega\phi \sin i\omega\psi + c_i \sin i\omega\phi \cos i\omega\psi + d_i \sin i\omega\phi \sin i\omega\psi). \quad (4.1)$$

Villar et al.

Table 1. Summary of Fit PES Results

Amino Acid	Parameters			Local Minima				RMSE (kJ·mol ⁻¹)	R ²
	N_f	N_g	Total	Region	ϕ	ψ	ΔE		
Gly	3	9	76	γ_L, γ_D	76.89	-61.36	2.1860	4.8363	0.9190
				β_L	174.58	174.58	2.2417		
				δ_L, δ_D	-118.58	22.70	13.5978		
Ala	4	10	87	γ_L	-82.90	79.29	3.2005	4.9544	0.9186
				β_L	-163.58	161.37	3.0173		
				γ_D	66.45	-61.47	10.4051		
				δ_L	-122.11	22.64	10.7066		
				α_D	62.65	22.38	24.9102		
				δ_D	-158.75	-27.02	26.3552		
Val	6	11	102	γ_L	-77.85	83.35	3.0478	5.1930	0.9280
				β_L	-137.31	156.51	5.2447		
				γ_D	63.41	-51.88	10.0331		
				δ_L	-119.70	22.06	10.6554		
				ϵ_L	-119.67	22.05	10.6538		
				α_L	-85.80	-21.02	17.1093		
				α_D	54.93	33.76	30.7688		
				δ_D	-158.02	-32.63	26.2996		
				ϵ_D	82.20	159.50	36.5746		

Fig. 3. Conformational Regions in the Amino Acid PES

The large error obtained in¹⁹ may be due to Equation 4.1 that may have contributed to the overfitting on $\Delta E^*(G)$. It is observed that Equation 4.1 is relatively computationally expensive to compute, which also contributed to more than twofold increase in computational time reported in.¹⁹ It is important to note that both papers used the same $\Delta E^*(G)$ points for analytical fitting.

Aiming for further dimension reduction in mathematical fitting, this study used only one form of Fourier series, instead of two, and that the (mathematical) interactions between the two axes can be captured by the mixture of Gaussian functions (Equation 2.2), which belongs to the family of exponential functions.

Furthermore, checking the grid maxima positions in the PES for the Gaussian mixture helped in

Fig. 4. Contour maps of PES obtained from the proposed method, with labeled DFT-optimized minima, from left to right: Glycine, Alanine, Valine

increasing the accuracy of the fit, as only prominent peaks are only used in the fitting procedure.

It is also worth noting that the fitting process is performed for less than an hour on a laptop computer, while DFT electronic structure calculations described in Section 3.1 took several days on the same machine. Thus, this mathematical model may aid in locating the approximate minima of the PES, and doing a further electronic structure calculation from the structure of the local minimum will give a shorter time, than creating a decent DFT-generated PES.

4.2. Location of Minima Points: Error Analysis with ff99sb Results

To check the deviations between the relative energies obtained from the analytical fitting and ff99sb force field calculations, Table 2 provides the values for the aforementioned methods, evaluated with respect to the dihedral angle values of optimized structures obtained from DFT calculations. The relative energy values from the analytical fit is calculated from the lowest energy value from $\Delta E^*(G)$.

From the table above, the analytical fit produced smaller differences with respect to the DFT calculations, as compared to ff99sb force field results. The analytical fitting results yielded an average deviation of 2.3172 kJ·mol⁻¹ for the given minima structures above, while the force field calculation returned 7.5594 kJ·mol⁻¹. It is also interesting to point out the relative energies of valine structures in α_L , α_D and ϵ_D conformation from the ff99sb results are too high that the structures are less probable to occur, as 40 kJ·mol⁻¹ is equivalent to twice the strength of a strong hydrogen bond.¹³ Note that the relative energy of valine in the ϵ_D conformation from the fitted result is also greater than 40 kJ·mol⁻¹ but the corresponding DFT relative energy value is around 40.48 kJ·mol⁻¹. This gives a small error difference, i.e., $\Delta\Delta E_{fit-dft}$, of 3.43 kJ·mol⁻¹.

Furthermore, the positions of the DFT-obtained minima structures are plotted into different PES generated through DFT calculations, analytical fit results, and force field calculations, and is shown in Figure 5.

Furthermore, to check the correlation of the values obtained from analytical fit and force field methods, with respect to the DFT calculations, the plots in Figure 6 is shown below.

It is shown that the fit function consistently provided higher R^2 values, as compared to the results from ff99sb force field calculations, where they provided an average R^2 of 0.9219 and 0.8414, respectively. This may imply that the analytical fit function ΔE_{f+g} decently captured the topology of the PES, even with just considering only the backbone torsional angles of the amino acid diamides considered in the study.

Villar et al.

Table 2. Comparison of potential energy values from analytical fitting results against DFT [B3LYP/6-31G(d)] optimized structures and ff99sb force field calculations. Relative energy ΔE values are in $\text{kJ}\cdot\text{mol}^{-1}$.

Amino Acid	Unique Minima	DFT Results			Fit Results		ff99sb Results	
		ϕ	ψ	ΔE_{dft}	ΔE_{fit}	$\Delta\Delta E_{fit-dft}$	ΔE_{ff}	$\Delta\Delta E_{ff-dft}$
Gly	γ_L, γ_D	82.15	-68.54	0.00	0.00	0.00	1.88	1.88
	β_L	179.99	180.00	2.58	2.01	-0.57	4.94	2.36
	δ_L, δ_D	-122.78	21.93	10.24	14.36	4.12	19.80	9.56
Ala	γ_L	-82.87	72.60	0.00	3.37	3.37	2.57	2.57
	β_L	-158.33	163.41	5.97	3.30	-2.67	4.24	-1.73
	γ_D	73.52	-57.39	10.88	11.06	0.18	11.74	0.86
	δ_L	-126.55	21.42	13.04	11.35	-1.69	16.60	3.56
	α_D	67.92	26.68	24.21	27.63	3.42	38.08	13.87
	δ_D	-165.06	-30.94	29.28	27.25	-2.03	30.95	1.67
Val	γ_L	-83.49	78.96	0.00	3.33	3.33	8.67	8.67
	β_L	-132.08	159.92	5.11	8.49	3.38	3.50	-1.61
	γ_D	73.51	-60.81	9.10	12.37	3.27	20.49	11.39
	δ_L	-122.94	18.74	10.99	13.87	2.88	17.91	6.92
	ϵ_L	-122.84	18.70	10.99	13.85	2.86	17.91	6.92
	α_L	-86.81	-22.80	22.46	20.05	-2.41	44.46	22.00
	α_D	62.93	39.08	29.08	30.28	1.20	51.98	22.90
	δ_D	-160.47	-36.22	29.76	27.09	-2.67	24.38	-5.38
	ϵ_D	77.03	142.69	40.48	43.91	3.43	54.20	13.72

4.3. Importance of the Use of Gaussian Functions in Improving Accuracy

Fourier series have been used in various force field functional forms to represent the torsional rotation of four consecutive atoms about a central bond. However, to account for different bonded and non-bonded interactions, it is important to find a suitable function that may account for dissymmetries in the PES. It has been shown from the previous sections that ΔE_{f+g} may be used to represent simple amino acid diamide conformational PESs, with respect to the resulting RMSE and R^2 values. It is interesting to compare the relative performance of the resulting analytical fit function, with the simple trigonometric fit function ΔE_f (eqn. 2.1). Table 3 summarizes the number of parameters, and corresponding RMSE values used in the different fit functions.

Table 3. RMSE Values of Two Fit Functions

Amino Acid	First Fit [$\Delta E_f(\phi, \psi)$]			Second Fit [$\Delta E_{f+g}(\phi, \psi)$]		
	No. Parameters	RMSE	R^2	No. Parameters	RMSE	R^2
Glycine	13	12.0071	0.4984	76	4.8363	0.9190
Alanine	17	12.5412	0.4876	87	4.9544	0.9186
Valine	24	13.0951	0.5498	102	5.1930	0.9280

From the preliminary analyses, it has been found out that $N_f = \{4, 5, 6\}$ gives a good fit, with the relative change in the RMSE on the three amino acid diamide PESs change only by at most 0.001 even with increased. Also, requiring more number of terms for the Fourier series yields a more ragged surface, which in turn gives worse RMSE.

The fitting using only the Fourier series ΔE_f yields an average RMSE of $12.5478 \text{ kJ}\cdot\text{mol}^{-1}$, while using ΔE_{f+g} returned an average RMSE of $4.9946 \text{ kJ}\cdot\text{mol}^{-1}$. Although ΔE_{f+g} uses around 2.5 times

Fig. 5. Contour maps of PES obtained from different methods (DFT, analytical fit and force field methods): (a) Glycine, (b) Alanine, (c) Valine. Labeled points correspond to the Horizontal and vertical axes pertain to dihedral angles ϕ and ψ values.

the number of parameters as ΔE_f , this has lessened the RMSE by around $7.5532 \text{ kJ}\cdot\text{mol}^{-1}$.

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.

5. 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

Villar et al.

Fig. 6. Correlation plots of the analytical fit function and ff99sb force field results against the DFT relaxed scan relative energy values: (a) Glycine, (b) Alanine, (c) Valine. R^2 values are written inside corresponding scatter plots.

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 studies¹⁹ used a significant number of parameters, which resulted to a large amount of time doing the mathematical fitting.

This paper provides a relatively simple two-rotor mathematical model to represent the topology of the conformational potential energy surface of selected amino acid diamides. This function is a linear combination of a Fourier series and a mixture of Gaussian functions, that can provide location of two backbone angles with reasonable accuracy and around twofold less computational time needed compared to DFT calculations. Results yielded an average RMSE of 4.9946 kJ·mol⁻¹ and R^2 of 0.9219 for the three amino acid diamides considered. It has also yielded lower deviation on the relative energy values with respect to DFT-optimized structures, as compared to corresponding ff99sb force field calculations. 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, where the use of Gaussian mixtures yielded a reduction of error of up to 2.5 times as compared to only using a Fourier series.

One of the ultimate goals of the study is to extend the results in,¹⁸ where they proposed a mathematical model with respect to a single dihedral rotation. Note that most force field models from classical molecular mechanics are defined with respect to only one dihedral rotation. This paper shows that a

Published by NRC Research Press

two-rotor mathematical model, with respect to the two backbone dihedral angles ϕ and ψ , is feasible and can also be used as a foundation to illustrate the PES of flexible molecules, such as peptides, with several internal bond rotations. Furthermore, this model may be used as well for doing optimizations near the approximate minima, which are obtained from the fitted results. As the proposed model returns a dihedral angle pair corresponding to the energy minima, one can compute for the more precise conformation by doing further optimization, but with a shorter amount of time as compared to an initial conformation estimate.

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

6. Acknowledgements

JJV is supported by Philippine Council for Industry, Energy and Emerging Technologies Research and Development through the 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 also 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.

This research is also supported by the European Union and the Hungarian State, co-financed by the European Regional Development Fund in the framework of the GINOP-2.3.4-15-201600004 project, aimed to promote the cooperation between the higher education and the industry.

References

- (1) Autrey, D.; Meinander, N.; Laane, J. *J. Phys. Chem. A* **2004**, *108*, 409–416.
- (2) Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 5648–52.
- (3) Branch, M. A.; Coleman, T. F.; Li, Y. *SIAM Journal on Scientific Computing* **1999**, *21*, 1–23.
- (4) Case, D. A. et al. AMBER 12., 2012.
- (5) Chass, G. A.; Sahai, M. A.; Law, J. M. S.; Lovas, S.; Farkas, O.; Perczel, A.; Rivail, J.-L.; Csizmadia, I. G. *Int. J. Quant. Chem.* **2002**, *90*, 933–968.
- (6) Coleman, T. F.; Li, Y. *SIAM Journal on Optimization* **1996**, *6*, 418–445.
- (7) Demare, G. R.; Peterson, M. R.; Csizmadia, I. G.; Strausz, O. P. *J Comp Chem* **1980**, *1*, 141–148.
- (8) Frisch, M. J. et al. Gaussian 09, Revision A.02., 2009.
- (9) Hehre, W. J.; Ditchfield, R.; Pople, J. A. *J. Chem. Phys.* **1972**, *56*, 2257–2261.
- (10) Hornak, V.; Abel, R.; Okur, A.; Strockbine, B.; Roitberg, A.; Simmerling, C. *Proteins* **2006**, *65*, 712–725.
- (11) Kehoe, T. A. K.; Peterson, M. R.; Chass, G. A.; Viskolcz, B.; Stachod, L.; Csizmadia, I. G. *J Mol Struct (THEOCHEM)* **2003**, *666-667*, 79–87.
- (12) Kessel, A.; Ben-Tal, N., *Introduction to Proteins: Structure, Function, and Motion*; CRC Press: 2010.
- (13) Markovitch, O.; Agmon, N. *The Journal of Physical Chemistry A* **2007**, *111*, 2253–2256.
- (14) Peterson, M. R.; Csizmadia, I. G. *J Am Chem Soc* **1978**, *100*, 6911–6916.
- (15) Peterson, R.; Mare, R. D. E.; Roosevelt, A. F. D. *J Mol Struct (THEOCHEM)* **1981**, *86*, 131–147.
- (16) Radom, L.; Hehre, W. J.; Pople, J. A. *J Am Chem Soc* **1972**, *94*, 2371–2381.

Villar et al.

- (17) Radom, L.; Lathan, W.; Hehre, W.; Pople, J. A. *J Am Chem Soc* **1973**, 95, 693–698.
- (18) Ragyanszki, A.; Suranyi, A.; Csizmadia, I. G.; Kelemen, A.; Jensen, S. J. K.; Uysal, S. Y.; Viskolcz, B. *Chem Phys Lett* **2014**, 599, 169–174.
- (19) Ragyanszki, A.; Gerlei, K. Z.; Suranyi, A.; Kelemen, A.; Jensen, S. J. K.; Csizmadia, I. G.; Viskolcz, B. *Chem Phys Lett* **2015**, 625, 91–97.
- (20) Ramachandran, G. N.; Ramakrishnan, C.; Sasisekharan, V. *J Mol Biol* **1963**, 7, 95–99.
- (21) Salomon-Ferrer, R.; Case, D. A.; Walker, R. C. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2013**, 3, 198–210.
- (22) Stephens, P. J.; Devlin, F. J.; Chabalowski, C. F.; Frisch, M. J. *J. Phys. Chem.* **1994**, 98, 11623–11627.
- (23) Tasi, G.; Nagy, B.; Matisz, G.; Tasi, T. S. *Comp. Theor. Chem.* **2011**, 963, 378–383.

Draft