

MICE-PES: An Algorithm for Accurate Conformational Analysis and its Implementation to Natural Products

by
Muhammad Ali Hashmi



A thesis
submitted to Victoria University of Wellington
in fulfillment of the
requirements for the degree of
Doctor of Philosophy
in Chemistry.

Victoria University of Wellington
2018

Abstract

Secondary metabolites from natural sources have revolutionized the modern drug industry by acting as lead compounds. Many commercial drugs have evolved originally from natural molecules before being synthesized in the laboratory for commercialization. Because of the importance of natural molecules, it is crucial to determine their structural properties carefully as it is essential for their synthesis and studying their pharmacological behaviour. Many natural molecules have flexible structures and can adopt many different conformations in solution at room temperature. Hence, the determination of their relative configuration is a challenging task with the available experimental techniques. For structural analysis of natural molecules and to study their properties, all conformers which might be responsible for their chemical properties have to be considered.

Theoretical chemistry has been very helpful in absolute structure determination of complex and conformationally flexible natural molecules by calculating their theoretical nuclear magnetic resonance, ultraviolet, infra red, and circular dichroism spectra etc. There are a number of software tools that offer conformational analysis by utilizing different molecular mechanics approaches. They produce a large number of possible conformers and are not general purpose, thus compromising accuracy. Apart from that, different force fields available for conformational analysis and minimization have been designed for specific molecular classes and do not produce good results beyond their scope.

In the past, there have been reports about a “build-up procedure” for predicting the low energy conformations of peptides by optimising smaller fragments of the molecule under study and then joining them while minimizing their energies using force fields. Later on, this method was extended to predict the structure of DNA from sequences. This method used force field methods and did not gain much popularity due to its various limitations.

Here, *MICE-PES* (Method for the Incremental Construction and Exploration of the Potential Energy Surface) is presented, an algorithm which performs a conformational analysis using high level quantum chemical calculations by building the molecule incrementally from its smallest possible analogue whose conformational degrees of freedom are very well separated than the rest of the molecule. *MICE-PES* has been validated through studies on known biomolecule 3-*epi*-xestoaminol whose absolute configuration has been determined already by experimental and theoretical methods. *MICE-PES* has also been used to assign the relative configuration of a natural product (meroter-

phenol C) whose configuration could not be established experimentally. Overall, the development of *MICE-PES* will be very helpful in solving problems in the study of conformationally flexible systems, in all aspects of organic chemistry.

Dedicated

To My Mother

A strong and gentle soul who taught me to trust in Allah, believe in hard work, and that so much can be done with a little

To My Father

For earning an honest living for us. For always supporting and encouraging me to believe in myself

To My Wife and Kids

For sacrifice of spending these years without me and being supportive all the time

To My Siblings

For their prayers and their care and love for me and my family in my absence

Acknowledgements

All the virtues and praise to Almighty Allah, the most Compassionate, and Merciful. Blessings of Allah on our holy Prophet *Muhammad (peace be upon him)* whose teachings have served as guiding light for humanity in hours of despair and darkness.

It was around four years ago when I decided to do my PhD in one of the windiest and most beautiful cities in the world, Wellington. The journey of my PhD here at Victoria University of Wellington has been a great time of my life filled up with adventures. These three years would never have been so great without the help of some people who are mentioned below.

Undoubtedly, the first person I would like to express my deepest gratitude and sincere thanks to is my supervisor, *Dr. Matthias Lein* for his valuable advices, kindness, continued interest, guidance, and inspiration throughout the course of my work here. I have no hesitation to say that he is the best supervisor I have had until now. I got enormous fruitful help from him not only in completion of this thesis, but also constant support and mentoring whenever I needed. May he be blessed with progress and prosperity he deserves. I also pay my special thanks to my secondary supervisor *Dr. Robert A. Keyzers* who has been very kind and cooperative to me during the course of my PhD.

Thanks to *Sarah Andreassend* for helping me with optical rotation measurement. I also thank *Galen Eakins* and *Omar Alsager* for taking me to Ferrier Research Institute for measuring the ECD of echivulgarine. I am grateful to *Prof. Peter Schwerdtfeger* for useful tips for my research during his visit to my office.

I owe a debt of gratitude to my mother, father, wife, kids, and siblings for their encouragement and support, both spiritual and material, and for putting up so much time with patience throughout the successful completion of my studies. The love and care of my parents and siblings towards my wife and kids made me able to pursue my PhD so far from home. The prayers of my parents always bring light for me in the dark. And I don't have words to thank my wife for standing beside me in all thick and thin and taking care of our family in my absence. It was only possible with her cooperation that I completed my degree here.

Special thanks to the computer science students (*Dragos Bercea, Hamish Brown, Malachi McIntosh, David Phillips, Michael Sirvid, and Jian Zhu*) for their great contribution in the scripts to automate the addition of the functional groups using the xyz files. That

script saved me a lot of time. It would be injustice if I don't mention the name of *Kevin Buckley* who have been very helpful in using the *SciFac-HPC* and helping on all the issues regarding high performance computing. Thanks are extended to our faculty advisor *Patricia Stein* who welcomes everyone to her office with a smiling face and cooperated enormously for any issues. I am grateful to my thesis examiners *Dr. Ralf Toner*, *Dr. Joanne Harvey*, and *Dr. Jóhannes Reynisson* for their valuable comments to improve my thesis in its current form.

Heartiest thanks are extended to my good friends both in Pakistan and New Zealand, especially my research group mates for their nice company and support. Thank you *Julia* for teaching me photography and gnuplot because of which I was able to produce so good graphs for my research. Special thanks to *Kimberley* for helping me in Maths and Python. I would also like to thank *Wong*, *Dani*, *Lukas*, *Yasir*, *Ayesha*, *Abubakar*, *Nofal*, and *Dr. Alzeer* for their nice company and moral support for my research work. Special thanks for *Dr. Muhammad Hanif* who introduced this scholarship to me. I also want to acknowledge some useful discussions with *Dr. Khurshid Ayub* about my research work during my visit to Pakistan.

And at last but not the least, I greatly acknowledge the funding (Victoria Doctoral Scholarship) from Victoria University of Wellington. I also acknowledge the additional computer time provided by the high performance computing facilities *Heisenberg*, *SciFac-HPC*, and *NeSI* for my research.

Muhammad Ali Hashmi

Contents

Abstract	ii
Dedications	iv
Acknowledgements	v
Table of Contents	vii
List of Figures	x
List of Schemes	xiii
List of Tables	xv
Glossary	xviii
1 Introduction	1
1.1 Conformational Analysis	2
1.1.1 Conformational Sampling	4
1.1.1.1 Systematic Search	4
1.1.1.2 High Temperature Molecular Dynamics	6
1.1.1.3 Monte Carlo Conformational Search	6
1.1.1.4 Distance Geometry	7

1.1.1.5	Random Search	9
1.1.2	Conformational Build-Up Procedures	10
1.2	Force Fields	11
1.2.1	AMBER	14
1.2.2	CHARMM	15
1.2.3	GROMOS	16
1.2.4	OPLS	18
1.2.5	CFF	19
1.2.6	MM3	20
1.2.7	MMFF	21
1.2.8	UFF	22
1.2.9	Dreiding	23
1.2.10	Quantum Mechanically Derived Force Field	24
1.2.11	Wellington Fast Assessment of Reactions Force Field	26
1.2.12	Comparison of Functional Forms in Different Force Fields	26
1.3	Conformational Analysis for the Prediction of NMR Data	27
1.4	Conformational Analysis for Assigning Absolute Configurations	29
1.5	Different Software Tools Available for Conformational Analysis	30
2	Methods	32
2.1	Computational Methods	32
2.2	Experimental Details	34
3	Results and Discussion	35
3.1	Assignment of Stereochemical Configuration of Echivulgarine using Experimental and Computational Techniques	35
3.2	Need for a Conformational Analysis Algorithm	43

3.3	Method for Incremental Construction and Exploration of Potential Energy Surface (<i>MICE-PES</i>)	44
3.4	Accurate Prediction of Optical and Spectroscopic Properties of Sphingolipids using <i>MICE-PES</i>	45
3.4.1	<i>MICE-PES</i> Methodology	46
3.4.2	Validating the Algorithm	51
3.4.2.1	Analysis of the Percentile Ranks of Conformers	56
3.4.3	Validation of the Results with the Experiment	58
3.4.3.1	Comparison of Optical Rotation	58
3.4.3.2	Comparison of NMR Data	59
3.5	Conformational Analysis of Meroterphenol C using <i>MICE-PES</i>	61
3.5.1	Analysis of the ranks of Conformers	69
3.5.2	Validation of the Computed Results with Experimental Data	73
3.5.2.1	Comparison of Optical Rotation	73
3.5.2.2	Comparison of NMR Data	73
3.6	Conformational Analysis of Rimarikiamide A using <i>MICE-PES</i>	76
3.6.1	Validation of the Computed Data with Experiment	82
4	Conclusions and Future Prospects	83
	Appendices	86
A		87
A.1	Relaxed Scans of the Smallest Homologue (2) of 3- <i>epi</i> -xestoaminol C (12)	87
A.2	Relaxed Scans of the Smallest Homologue (13) of Meroterphenol C (26)	88
A.3	Relaxed Scans of the Smallest Homologue (27) of Rimarikiamide (43)	88
	Bibliography	90

List of Figures

1.1	A hypothetical arrangement of two molecules to show the terms in equation 1.1. The first molecule consists of atoms 1-4 while atom 5 belongs to a second molecule.	13
1.2	General workflow of WellFARe-FF.	25
3.1	Structure of <i>7R,8R,2'R,3'S</i> -echivulgarine isolated from <i>Echium vulgare</i>	36
3.2	Four possible side chain diastereomers of echivulgarine. (1a). <i>7R,8R,2'R,3'R</i> -echivulgarine, (1b). <i>7R,8R,2'R,3'S</i> -echivulgarine, (1c). <i>7R,8R,2'S,3'R</i> -echivulgarine, and (1d). <i>7R,8R,2'S,3'S</i> -echivulgarine with relative energies of 9.11, 0.00, 3.55, and 2.89 kcal/mol. The numbering shown in 1b is followed for all the NMR assignments and discussion throughout the chapter. The 3D structures are the lowest energy conformer of the respective diastereomer.	37
3.3	Comparison of the experimental (dichloromethane) and calculated ECD spectra of 1a (<i>2'R,3'R</i> -EV), 1b (<i>2'R,3'S</i> -EV), 1c (<i>2'S,3'R</i> -EV), and 1d (<i>2'S,3'S</i> -EV). EV is short for echivulgarine.	39
3.4	Comparison of the experimental (dichloromethane) and calculated ECD spectra of <i>7R,8R,2'R,3'S</i> -echivulgarine (1b).	39
3.5	Lowest-energy conformer predicted for <i>7R,8R,2'R,3'S</i> -echivulgarine, 1b showing key nOe correlations. The blue coloured numbers on arrows show the distance (Å) between the atoms for the nOe correlations.	40
3.6	Potential energy surfaces at the PBE0-D3BJ/def2-TZVP/SMD _{CH₂Cl₂} level of theory that identify all 23 conformers of 2 . Y-axis represents the relative energies of the conformers. X- and Z-axis show the rotational angles of the starting structure. The details of these angles and rotations can be found in the Appendix A. Optimisations were carried out for a total of 3750 individual structures. The numbers on the energy scale on the right are in kcal/mol. a–f are the relaxed scans on different conformations of structure 2 to ensure that the PES has been covered well.	47

3.7	The 23 lowest energy conformers of 2 optimized at the PBE0-D3BJ/def2-TZVP/SMD _{CH₂Cl₂} level of theory. Note that the first two conformers (2a and 2b) are stabilised by intramolecular hydrogen bonding.	48
3.8	Correlation analysis for all conformers of 4 whose antecedents do not rise in energy more than 10 percentage points. There is 95% correlation between the two. It is clear from the graph that low energy conformers of 4 descended from the low energy conformations of 3 most of the time. There are also many occurrences of low energy conformers of 3 becoming high in energy in 4	57
3.9	Naturally occurring meroterphenol C isolated from the algae <i>Sargassum yezoense</i>	61
3.10	Potential energy surfaces at the PBE0-D3BJ/def2-TZVP/SMD _{CH₃CN} level of theory that identify all 28 conformers of 13 . Y-axis represents the relative energies of the conformers. X- and Z-axis show the rotational angles of the starting structure. The details of these angles and rotations can be found in the Appendix A.2. Optimisations were carried out for a total of 5625 individual structures. The numbers on the energy scale on the right are in kcal/mol. a–i are the relaxed scans on different conformations of structure 13 to ensure that the PES has been covered well.	63
3.11	The 28 lowest energy conformers of 13 optimized at the PBE0-D3BJ/def2-TZVP/SMD _{CH₃CN} level of theory. Note that structures 13a , 13b , 13c , 13d , 13e , 13f , 13ab , and 13ac have intra-molecular hydrogen bonding.	64
3.12	Absolute configuration of naturally occurring meroterphenol C isolated from the algae <i>Sargassum yezoense</i>	73
3.13	Rimarikiamide A isolated from the sponge <i>Latrunculia</i> sp. The squiggly bonds in the figure represent the chiral centres whose configuration is as yet unknown.	76
3.14	Potential energy surfaces at the PBE0-D3BJ/def2-TZVP/SMD _{CHCl₃} level of theory that identify all 8 conformers of 27 . Y-axis represents the relative energies of the conformers. X- and Z-axis show the rotational angles of the starting structure. The details of these angles and rotations can be found in the Appendix A.3. Optimisations were carried out for a total of 2500 individual structures. The high energy maxima in c and d show the high energy structure of taurine having a base (NH ₂) and the neighbouring acid (SO ₃ H) while the energy minima in these two (c and d) are for the low energy structures where taurine is found as Zwitterionic structure (NH ₃ ⁺ and SO ₃ ⁻). The numbers on the energy scale on the right are in kcal/mol. a–d are the relaxed scans on different conformations of structure 27 to ensure that the PES has been covered well.	78

3.15	The 15 lowest energy conformers of 27 optimized at the PBE0-D3BJ/def2-TZVP/SMD _{CHCl₃} level of theory.	80
A.1	Relaxed scans on six manually generated conformers of 2 optimised at the PBE0-D3BJ/def2-TZVP/SMD _{CH₂Cl₂} level of theory. Relaxed scans were performed by rotating two sets of dihedral angles. First dihedral is between atoms 1-2-3-4 while the second one is denoted by atoms 1'-2'-3'-4'.	87
A.2	Relaxed scans on four different conformers of 27 optimised at the PBE0-D3BJ/def2-TZVP/SMD _{CHCl₃} level of theory. Relaxed scans were performed by rotating two sets of dihedral angles. In PES scans in Figure 3.10, plot a corresponds to the dihedral angles between atoms 1-2-3-6 and 5-3-6-7. Plot b , corresponds to the rotation of dihedrals 5-3-6-7 and 8-6-9-10, plot c originated by rotating the dihedrals 5-3-6-7 and 8-6-9-11, and plot d came from the rotation of dihedral angles 2-3-6-7 and 8-6-9-11.	88
A.3	Relaxed scans on four different conformers of 27 optimised at the PBE0-D3BJ/def2-TZVP/SMD _{CHCl₃} level of theory. Relaxed scans were performed by rotating two sets of dihedral angles. In PES scans in Figure 3.14, plot a corresponds to the dihedral angles between atoms 1-2-3-6 and 5-3-6-7. Plot b , corresponds to the rotation of dihedrals 5-3-6-7 and 8-6-9-10, plot c originated by rotating the dihedrals 5-3-6-7 and 8-6-9-11, and plot d came from the rotation of dihedral angles 2-3-6-7 and 8-6-9-11.	89

List of Schemes

- 3.1 The 11 compounds that were considered in this investigation. (2*S*,3*S*)-3-amino-2-butanol (**2**), (2*S*,3*S*)-2-amino-3-pentanol (**3**) etc. Note that **12** is the naturally occurring 3-*epi*-xestoaminol C (2*S*,3*S*-2-amino-tetradecan-3-ol). 45
- 3.2 Construction of conformers through step-wise lengthening of the alkyl chain. The lowest energy isomer of **12** is shown at the bottom. 50
- 3.3 An analysis of the 13th conformer of **2** and its descendants along with their propagation in subsequent steps. ΔE is the relative energy of the conformer with respect to the lowest energy conformer in the series. **BD** is the Boltzmann percentage distribution of the conformer in its total ensemble. Conformer number at the bottom of each structure shows its rank in its conformational ensemble while the number in parentheses shows the homologue whose conformer it is. 52
- 3.4 An analysis of the lowest energy conformer of **2** and its descendants along with their propagation in subsequent steps. ΔE is the relative energy of the conformer with respect to the lowest energy conformer in the series. **BD** is the Boltzmann percentage distribution of the conformer in its total ensemble. Conformer number at the bottom of each structure shows its rank in its conformational ensemble while the number in parentheses shows the homologue whose conformer it is. 54
- 3.5 Schematic representation of the conformational analysis of meroterphenol C (**26**) using *MICE-PES*. Structures **13-25** are the homologues of **26**. Note that **26** is one of the two possible enantiomers of the naturally occurring meroterphenol C. The bold groups in magenta colour are the ones which are being changed at that step. The lowest energy conformer of each homologue is shown on the right in 3D. 62
- 3.6 Construction of conformers through step-wise lengthening of the molecule by the addition of substituents using *MICE-PES*. The lowest energy isomer of 3'-*R*-meroterphenol C (**26**) is shown at the bottom. 66

3.7	Addition of double bond in conformers construction using <i>MICE-PES</i> . A methyl group is added on each of the three hydrogens of the previous methyl. In step-I, another methyl group is added on each of the three hydrogens of the existing methyl. In step-II the C–C bond length is shortened for a double bond, hydrogens are deleted on both carbons, and then hydrogens and other groups are added again to the terminal carbon.	67
3.8	An analysis of the 18 th conformer of 13 and its descendants along with their propagation in subsequent steps. ΔE is the relative energy of the conformer with respect to the lowest energy conformer in the series. BD is the Boltzmann percentage distribution of the conformer in its total ensemble. Conformer number at the bottom of each structure shows its rank in its conformational ensemble while the number in parentheses shows the homologue whose conformer it is.	69
3.9	An analysis of the 4 th lowest energy conformer of 13 and its descendants along with their propagation in subsequent steps. ΔE is the relative energy of the conformer with respect to the lowest energy conformer in the series. BD is the Boltzmann percentage distribution of the conformer in its total ensemble. Conformer number at the bottom of each structure shows its rank in its conformational ensemble while the number in parentheses shows the homologue whose conformer it is.	71
3.10	The 17 compounds that were considered in this investigation. 27 (2-aminoethane-1-sulfonic acid), 28 (2-acetamidoethane-1-sulfonic acid) etc. Note that 43 is one arbitrarily selected diastereomer of the naturally occurring rimarikiamide A. The atoms in blue colour represent the place of operation/addition of next atoms in the subsequent step. The lowest energy conformer of each homologue is shown in 3D.	77
3.11	Construction of conformers through step-wise lengthening of the alkyl chain using <i>MICE-PES</i> . The lowest energy isomer of 43 is shown at the bottom.	79

List of Tables

1.1	Comparison of the functional forms used in common force fields. The torsional energy E_{tors} in all cases is given as a Fourier series in the torsional angle. Notation: P_n : Polynomial of order n ; $P_n(\cos)$: polynomial of order n in cosine to the angle; $\cos(n\theta)$: Fourier term(s) in cosine to the angle; Exp^{-6} : <i>exponential</i> + R^{-6} ; $n - m$: $R^{-n} + R^{-m}$ Lennard-Jones type potential; <i>Imp.</i> : improper torsional angle; <i>ss</i> : stretch-stretch; <i>bb</i> : bend-bend; <i>sb</i> : stretch-bend; <i>st</i> : stretch-torsional; <i>bt</i> : bend-torsional; <i>btb</i> : bend-torsional-bend.	27
3.1	Boltzmann weights and relative energies (kcal/mol) of the selected lowest energy conformers of all four side chain diastereomers of echivulgarine (1a-1d)	38
3.2	Experimental (δ_{exp}) and calculated (δ_{calc}) ^{13}C -NMR data of echivulgarine diastereomers (1a-1d). MAE is the mean absolute error and RMSE is the root mean square error calculated for the chemical shift difference ($\Delta\delta$) from experimental values.	42
3.3	Experimental (δ_{exp}) and calculated (δ_{calc}) ^1H -NMR data of echivulgarine diastereomers (1a-1d). MAE is the mean absolute error and RMSE is the root mean square error calculated for the chemical shift difference ($\Delta\delta$) from experimental values.	43
3.4	The relative Gibbs free energies (ΔG in kcal/mol) and Boltzmann percentage distributions of the 23 minimum energy conformers of 2 at the PBE0-D3BJ/def2-TZVP/SMD $_{\text{CH}_2\text{Cl}_2}$ level of theory	49

- 3.5 \mathbf{n}^{conf} is the total number of conformers generated in a step. \mathbf{n}^{contr} is the total number of distinct conformers considered after duplicates removal (fully optimised at the PBE0-D3BJ/def2-TZVP/SMD_{DCM} level of theory). \mathbf{n}^{imp} is the number of important conformers (i.e. the number of conformers that have to be included to account for $\geq 90.0\%$ of the Boltzmann distribution at 298.15 K). \mathbf{B} is the Boltzmann percentage of all the \mathbf{n}^{imp} conformers. \mathbf{c}_0 is the percentage contribution of the lowest lying conformer at 298.15 K. \mathbf{P} is the total value of the partition function at 298.15 K. $\Delta\mathbf{G}$ is the difference in Gibbs free energy between the energetically lowest conformer and the energetically highest lying conformer (of all \mathbf{n}^{contr} conformers; in kcal/mol). $[\alpha]_D$ is the predicted value of the optical rotation (based on all \mathbf{n}^{contr} conformers) in DCM. . 50
- 3.6 A detailed analysis of all 23 conformers of **2** and their rank in subsequent descendants. First column shows conformers of **2** with their Boltzmann percentage distribution in parentheses. All the other columns show the ranks of the previous conformers' first descendant in current homologue. The numbers in parentheses show the Boltzmann percentage distribution of the respective conformer. A '-' indicates that the conformers has been eliminated from being propagated further due to being higher in energy than the selection cut-off in the previous step. Note that in each row the rank of the lowest energy descendant is shown. There are usually three descendants but the other two are not shown here. 55
- 3.7 First six lowest energy conformers of **12** with their relative energies and other details. *Conformer* is the name of the conformer from the last 339 conformers of **12**. $\Delta\mathbf{E}$ is the relative energy of each conformer with respect to the lowest energy conformer. $[\alpha]_D$ is the optical rotation of the conformer computed in CH₂Cl₂. \mathbf{B} is the Boltzmann percentage distribution of the conformer at 298.15 K. Weighted $[\alpha]_D$ is the Boltzmann weighted optical rotation of the conformer based on all the contributing conformers. 58
- 3.8 Comparison of Experimental and Theoretical NMR data for **12**. C. No. represents the carbon number while δ_{exp} and δ_{calc} represent the experimental and theoretical chemical shifts (based on Boltzmann average of all the conformers), respectively. $\Delta\delta$ is the difference between the experimental and calculated chemical shift at the PBE0-D3BJ/def2-TZVP/SMD_{CH₂Cl₂} level. MAE is the mean absolute error and RMSE is the root mean square error calculated for the chemical shift difference ($\Delta\delta$) from experimental values. 60
- 3.9 The relative Gibbs free energies ($\Delta\mathbf{G}$ in kcal/mol) and Boltzmann percentage distributions of the 28 minimum energy conformers of **13** at the PBE0-D3BJ/def2-TZVP/SMD_{CH₃CN} level of theory. Note that most of the conformers with identical energies are enantiomeric pairs. 65

- 3.10 n^{conf} is the total number of conformers generated in a step. n^{contr} is the total number of distinct conformers considered after duplicates removal (fully optimised at the PBE0-D3BJ/def2-TZVP/SMD_{CH₃CN} level of theory). n^{imp} is the number of important conformers (*i.e.* the number of conformers that have to be included to account for $\geq 90.0\%$ of the Boltzmann distribution at 298.15 K). B is the Boltzmann percentage of all the n^{imp} conformers. c_0 is the percentage contribution of the lowest lying conformer at 298.15 K. P is the total value of the partition function at 298.15 K. ΔG is the difference in Gibbs free energy between the energetically lowest conformer and the energetically highest lying conformer (of all n^{contr} conformers; in kcal/mol). $[\alpha]_D$ is the predicted value of the optical rotation (based on all n^{contr} conformers). 68
- 3.11 A detailed analysis of all 28 conformers of **13** and their rank in subsequent descendants. First column shows conformers of **13** with their Boltzmann percentage distribution in parentheses. All the other columns show the ranks of the previous conformers' first descendant in current homologue. The numbers in parentheses show the Boltzmann percentage distribution of the respective conformer. $R(n)$ means Rank in n . A '-' indicates that the conformers has been eliminated from being propagated further due to being higher in energy than the selection cut-off in the previous step. Note that in each row the rank of the lowest energy descendant is shown. There are usually three descendants but the other two are not shown here. 72
- 3.12 Comparison of Experimental and Theoretical NMR data for **26**. C. No. represent the carbon number, C. Type represents the type of carbon (*i.e.* quaternary (C), methine (CH), methylene (CH₂), and methyls (CH₃)), while δ_{exp} and δ_{calc} represent the experimental and theoretical chemical shifts (based on Boltzmann average of all the conformers), respectively. $\Delta\delta$ is the difference between the experimental and calculated chemical shift at the PBE0-D3BJ/def2-TZVP/SMD_{CH₃CN} level. MAE is the mean absolute error and RMSE is the root mean square error calculated for the chemical shift difference ($\Delta\delta$) from experimental values. 74
- 3.13 The relative Gibbs free energies (ΔG in kcal/mol) and Boltzmann percentage distributions of the 15 minimum energy conformers of **27** at the PBE0-D3BJ/def2-TZVP/SMD_{CH₃OH} level of theory. 79

- 3.14 \mathbf{n}^{conf} is the total number of conformers generated in a step. \mathbf{n}^{contr} is the total number of distinct conformers considered after duplicates removal (fully optimised at the PBE0-D3BJ/def2-TZVP/SMD_{CHCl₃} level of theory). \mathbf{n}^{imp} is the number of important conformers (*i.e.* the number of conformers that have to be included to account for $\geq 90.0\%$ of the Boltzmann distribution at 298.15 K). \mathbf{B} is the Boltzmann percentage of all the \mathbf{n}^{imp} conformers. \mathbf{c}_0 is the percentage contribution of the lowest lying conformer at 298.15 K. \mathbf{P} is the total value of the partition function at 298.15 K. $\Delta\mathbf{G}$ is the difference in Gibbs free energy between the energetically lowest conformer and the energetically highest lying conformer (of all \mathbf{n}^{contr} conformers; in kcal/mol). $[\alpha]_D$ is the predicted value of the optical rotation (based on all \mathbf{n}^{contr} conformers). 81

Glossary

AA	All-atom
AMBER	Assisted model building with energy refinement
CD	Circular dichroism
UFF	Consistent force field
CHARMM	Chemistry at Harvard molecular mechanics
COMPASS	Condensed-phase optimized molecular potentials for atomistic simulation studies
DFT	Density functional theory
ECD	Electronic circular dichroism
EV	Echivulgarine
FF	Force field
GROMOS	Groningen molecular simulation
IEFPCM	Integral equation formalism polarizable continuum model
IR	Infra red
MAE	Mean absolute error
MMFF	Merck molecular force field
MICE-PES	Method for the incremental construction and exploration of the potential energy surface
NMR	Nuclear magnetic resonance
nOe	Nuclear Overhauser effect
OPLS	Optimized potential for liquid simulations
ORD	Optical rotatory dispersion
PCM	Polarizable continuum model
PES	Potential energy surface
QM	Quantum mechanical
QMDF	Quantum mechanically derived force field
TDDFT	Time dependent density functional theory
UFF	Universal force field
UV	Ultra violet
VCD	Vibrational circular dichroism
VUW	Victoria university of Wellington
WellFARe-FF	Wellington fast assessment of reactions force field

Chapter 1

Introduction

Many natural products possess valuable therapeutic properties and their use as a source of drugs dates back to ancient human civilizations. The discovery of modern drugs is beholden to the crucial role played by the natural products in their development. Biomolecules derived from nature have been used to cure various ailments since the emergence of medicine.^{1,2} Secondary metabolites from natural sources have revolutionized drug development to fight diseases during the 20th century.³ Many natural products share common structural and pharmacological features with commercial drugs. “Discovery oriented synthesis” is a term used for synthesizing many drugs after inspiration by the biological significance of many natural molecules⁴⁻⁸ and has lead to the synthesis of natural molecular analogues and their approval as synthetic medicines.⁹⁻¹¹ It is a well-established fact that if one isomer of a molecule has excellent biological properties, the other may be inactive or toxic.^{12,13} Therefore, assignment of the correct structure and stereochemistry is an indispensable part of the research in this area and a complete knowledge of the physico-chemical and structural properties of naturally occurring molecules is essential for corresponding drug development process.¹⁴

Many naturally occurring molecules are conformationally flexible and can attain different conformations in solution at room temperature.¹⁵ These different conformations

interconvert by rotations around single bonds and lead to different minima or maxima on a potential energy surface (PES).¹⁶ Structure determination of these naturally occurring flexible molecules is a very challenging task with increasing number of flexible bonds and chiral centres. X-ray crystallography is a very useful tool in absolute structure determination but in the case of large and highly flexible molecules, most of them are not crystalline and prevent X-ray crystallography. Even for crystalline natural products, X-ray has a disadvantage in that it locks the molecule in a single conformation while the compound in solution usually exists as a mixture of different low energy conformations. Also, the locked conformation determined by X-ray might not be the lowest energy conformation in solution. Nowadays, advancements in theoretical chemistry and quantum chemistry has significantly aided the structure determination of many small sized complex and conformationally flexible natural molecules by calculating their theoretical nuclear magnetic resonance (NMR), infrared (IR), ultraviolet (UV), electronic circular dichroism (ECD), and vibrational circular dichroism (VCD) spectra.^{17–20} Quantum mechanical (QM) calculation of different properties of compounds under investigation such as UV, NMR parameters etc. are also dependent on all of the contributing conformational structures which those compounds may adopt.¹⁶

1.1 Conformational Analysis

The most important concern in biological, natural product, and medicinal chemistry and research is the determination of stereo structure. Only by knowing the correct structure can other physico-chemical properties and the mechanism of biological action be studied. The characteristics of a molecule are subject to the conformations that it can adopt. The study of the different forms (conformations) a molecule can adopt in solution, and their impact on its physical characteristics, is termed as conformational analysis. Barton is considered to be the inventor of modern conformational analysis who, for the first time, showed that the reactivity of substituted cyclohexanes changes by changing

the position of axial or equatorial functional group substituents.²¹ Conformations of a molecule may therefore be defined as different orientations of its atoms that may be interconverted only by rotation around single bonds and without breaking or making bonds.²²

The main purpose of conformational analysis is to study the physical properties of flexible molecules and the relationship between their flexibility and their behaviour or function. Conformational analysis finds applications in drug discovery by aiding design of targeted molecules for fitting into a potential binding site.²³ Molecular docking approaches are then used to dock these molecule into the binding sites to model their behaviour and interactions with proteins found inside the body. For rigid molecules, there is no issue with these docking studies but when flexible molecules have to be dealt with, it becomes much more complicated.²⁴ In such cases the role of conformational analysis becomes significant and it has to be performed before different conformations can be docked separately. Beyond these applications, conformational analysis may be used to study the optimization of future lead compounds in drug discovery.

In case of structural determination of highly flexible natural products, which can adopt different conformations in solution, computational chemistry can aid structural elucidation with spectroscopic techniques like NMR, ECD, and VCD spectroscopy. To deduce accurate structural information, a detailed conformational analysis is a vital part which must cover most of the minimum energy conformations on the PES to calculate the chemical properties with great confidence. If the number of energy minima is too large for it to be practical to identify all of them, a conformational search should list all thermally accessible minima. Then, the relative contribution or population of each conformer is calculated by the Boltzmann distribution. It is important to make sure that the Boltzmann weighted average involves the contributions from all the degrees of freedom including electronic energies, vibrations, rotations, as well as solvation effects.²⁵

The conformational analysis study mainly comprises three major steps which are: (1) Exploring the conformational space (conformational sampling), (2) optimization of

conformations, and (3) analysing the data or conformational analysis.

1.1.1 Conformational Sampling

Conformational sampling is a process used to explore the conformational space by generating different conformations; different approaches will be described in this section. Ideally, all stable local minima should be explored for a thorough conformational analysis. However, depending on the complexity of the molecule under study and its degrees of freedom, the number of conformations increases exponentially with increasing molecular size and covering the whole PES becomes impractical, therefore the needs of different conformational sampling procedures and techniques comes into practice. The majority conformational sampling methods operate in the following way. First of all, a crude starting structure is generated that is followed by energy minimization by molecular mechanics. The resultant energy minimized structure is then compared with already generated ones to avoid duplicates. If the structure is different from the previously discovered ones, it is added to the found list of conformers and the method is continued. The search is terminated according to the instruction sets of the different sampling procedures.²⁶ Some of the most commonly used conformational sampling procedures are as follows.

1.1.1.1 Systematic Search

Systematic conformational search is principally the most detailed search method to cover the whole conformational space. It is, therefore, predictable and works in a systematic manner, while covering the entire range of a molecule's torsional angles with fixed parameters. Bond angles and bond lengths are not distorted in a systematic search because they need more energy to be distorted as compared to the dihedral angles. This process is also known as the grid search method. In this search, first of all the dihedral angles in the molecule are identified. Then every single dihedral angle is rotated through

360° with an increment value specified by the user. The resulting trial conformations are then minimized for energy and this process stops when all possible combinations of torsional angles have been identified. If the increment value is small enough, the grid search would generate almost all possible conformers. There are several disadvantages which make systematic search less popular.²⁷ The main issue with this method is that the number of conformers increases exponentially with each additional dihedral angle, and even small molecules generate a lot of conformers and it is impractical to apply grid search to larger systems or biomolecules. For example, if a molecule with six dihedral angles is subjected to grid search with the increment value set to 30°, then the number of its conformers will be almost 3 million. Another drawback is that for larger molecules, a grid search would also produce physically impossible and high energy structures with different parts of the molecule crossing over each other. That is why this method is only applicable to small molecules (*e.g.* small peptides etc.²⁸).

Several approaches have been made to make the grid search better but many of them focus on filtering out certain criteria (*e.g.* sensible selection of initial dihedral angles, transannular distances, steric interactions, and rejection of degenerate structures etc.) and because of that, reasonable portions of the conformational space get eliminated.²⁷ One such approach was the use of artificial intelligence techniques and expert systems for systematic conformational search without being as rigid as the grid search and added flexibility of bond angles and bond lengths as well as dihedral angles*.²⁹ Another modification to the grid search was the “build-up procedure” for determining the low energy conformers of polypeptides. In this method, the molecule is split into smaller fragments and then a grid search is applied to each fragment. Afterwards, the lowest energy trial conformations of the fragments are joined together followed by energy minimization to build-up the molecule from the fragments.³⁰ This method was more useful for macromolecules than systematic search methods but it also had a disadvantage that its screening procedure may reject the optimum conformations of some fragments in the built-up molecule because of the density of low-energy states.

**e.g.* WIZARD software

1.1.1.2 High Temperature Molecular Dynamics

Molecular dynamics is a method where physical movements of atoms and molecules are studied using computer simulations. It is used to study the conformational rearrangements of molecules and their interactions with other molecular species in different environments. Usually in molecular dynamics, the model produces a trajectory of trial conformations that are confined to a specific energy or temperature. The problem with these simulations is that a finite-timed simulation cannot sample all of the conformational space as they cannot cross high energy barriers. To solve this problem, Bruccoleri and Karplus simply increased the modelled temperature of the molecular dynamic simulations.³¹ In this way it becomes possible for the simulation to cross high energy barriers and cover a broad range of conformational space. An advantage of this method is that regardless of the higher temperature, the bonds on the molecules are not broken and the integrity of the molecule remains intact. The purpose of using the higher temperature is to allow the molecule to cross high energy barriers and broaden the scope of conformational space.

There is no specific limitation or rule for using “high temperature” but normally temperatures up to 1000K are used for peptides and proteins.³¹ The advantage associated with this method of conformational sampling is that the information obtained from molecular forces is employed to search regions where energy minima are located. The main disadvantage of this method is that it not only covers the regions of interest but also ones that are not available at room temperature so the obtained trial conformations cannot be used directly prior to their energy minimization.

1.1.1.3 Monte Carlo Conformational Search

This method is basically a statistical technique based on probability and random numbers for conformational sampling. It was first used by Metropolis and Ulam who gave it this name.³² Monte Carlo simulations have been applied extensively in different fields

including physics and biochemistry.

In using the Monte Carlo technique for conformational sampling, the algorithm decides randomly at each step of the conformational search about which conformations to keep and which ones to discard. It generates a trial conformation randomly and then decides on the basis of energy by comparing it to the list of conformers if the new trial conformation should be added to the list of conformers or discarded.³³ The computer changes different torsional angles or Cartesian coordinates to generate new trial conformations based on random values. If the energy of the newly generated structure is lower than the current conformation, it is kept, but if the energy of the newly generated structure is higher than the current structure, then its acceptance or rejection is decided on the basis of the Boltzmann factor. If the Boltzmann factor is higher than a certain cut-off value, it is kept, otherwise it is discarded.³⁴ If a trial conformation is rejected, the procedure will repeat by randomly creating new trial structures from the current structure. If a low energy structure is found and accepted then it becomes the ‘current’ conformation and the searching continues from it.²³

In Monte Carlo conformational search, there is a possibility of finding conformers that lie in a completely different energy landscape on the PES. Temperature is an important factor here, which can be varied to allow crossing of higher energy barriers. These search methods are popular due to their ease of use and unbiased convergence properties. Its low acceptance ratios make it less popular for biomolecules which reduces the efficacy of the method. Some advancements have also been made to enhance the efficiency of Monte Carlo conformational search.^{35,36}

1.1.1.4 Distance Geometry

In this method the conformations of a molecule are not represented by Cartesian or internal coordinates but described as the distance between all pairs of atoms. There are a total of $N(N - 1)/2$ distances between atoms in a molecule (where N represents

the total number of atoms) which can be easily represented by an $N \times N$ symmetrical matrix.³⁷ In this method, the conformational space is explored by randomly generating many such distance matrices and then converting them to Cartesian coordinates. The constraints are defined by upper and lower bounds and can be qualitative or approximate. Distance geometry is particularly useful when experimental data for inter or intramolecular distances is available. The matrix can be filled by the distance information from experimental data, *e.g.* nOe's (nuclear Overhauser effects) or a hypothesis. The technique can be described in four steps which are as follows:

- (i) Calculation of a matrix of upper and lower distance bounds for interatomic distances. The bounds contain permitted maximum and minimum values for each interatomic distance in the molecule under study.
- (ii) Random assignment of values between the upper and lower bounds to each interatomic distance.
- (iii) Diagonalization of the matrix and its conversion to a trial set of Cartesian coordinates through a process called embedding.
- (iv) Refining the resulting coordinates so that they satisfy the initial distance bounds and storing the trial conformation in the list of generated conformers.

The process is repeated by generating new random distance matrices within the same bounds and new trial conformations are generated. The resolution of conformer sampling depends on the tightness of the bounds.²⁵ Often it is also desirable to add chiral constraints in the procedure while dealing with chiral stereoisomers to maintain the correct configuration.³⁸

Distance geometry proved to be very useful in determining conformations of small molecules³⁹ and where NMR data is available.^{40,41} It was also found to be helpful in modelling bioactive compounds from known biomolecules.⁴² It was also suited for

analysing conformations of macromolecules in solution.⁴³ A major success of this technique was the conformational analysis of the α -amylase inhibitor (Hoe-467A) in aqueous solution. In this study, the distance matrix was constructed from NMR data. The resultant four conformations were shown as the main contributing conformers.⁴⁴ An X-ray crystallographic study showed that the proposed backbone structure was consistent with the original structure.⁴⁵ Besides all these successes, this technique is not always appropriate for general molecules unless the required conformational space is limited by some constraints. It is limited to moderate sized molecules as matrix manipulation is a computationally expensive process.⁴⁶

1.1.1.5 Random Search

Random search, as the name implies, is based on an algorithm which generates conformers randomly. Only a systematic conformational search is guaranteed to locate the global minimum but in the case of the random search method, finding the global minimum is random as the name promises. It may or may not be found as progress is not systematic nor is there a defined end point. A random search can jump from one point on the PES to a completely different place in one step. Random search can operate in two different ways, *i.e.* by changing the Cartesian coordinates by a random amount or the dihedral angles as in systematic search. In the Cartesian method, a random increment is made to the x, y, and z coordinates of all the atoms in the molecule under study^{47,48} while in the dihedral method, the new trial conformations are generated by rotating the molecule through dihedral angles by a random amount, while keeping the bond lengths and bond angles static.^{49,50}

The application of this search method is simple. A user provided structure is taken as input and then either Cartesian coordinates or the dihedral angles are changed by an arbitrary amount. After that, a geometry minimization step is performed. Then the resulting structure is compared to already generated conformers and, if it is unique, it is stored and the search continues until a selected number of iterations have been

performed. In the Cartesian method, the energy minimization is a crucial step as the randomly generated initial structure might be distorted and higher in energy. The dihedral method is advantageous but extreme care is needed for the application of this method to molecules containing rings as they may be broken in the resulting structures. The structure selection for each next iteration can also be done in different ways. It can be either the structure obtained in the last step or it can be the lowest energy structure of all the generated conformers. In this search method, there is no specifically defined endpoint so a specific number of iterations can be given to the algorithm or the search can be continued until no new structure can be found. The second option can take longer than a systematic search.²⁵

1.1.2 Conformational Build-Up Procedures

As the name implies, build-up procedures build the conformational space of the molecule systematically and gradually from smaller fragments, thus reducing the total computational cost considerably. The basic idea was presented by Scheraga *et al.* for the first time when they calculated the conformational space of the membrane-bound portion of the peptide melittin from its building blocks, *i.e.* amino acids. They modeled the main non-polar residues from the low energy conformations of their component peptide building blocks. These few low energy segments were then joined for constructing the larger portions of the molecule.⁵¹ Later on, the same group extended this build-up procedure to peptides in aqueous solutions, making it a five-step procedure to build up the molecule. It starts with combining the low energy structures of the two fragments that need to be joined, followed by energy minimization. If there are overlapping atoms, the energy minimization is repeated with a pseudo energy function. The pseudo energy function retains the original potential for all the interatomic distances that exceed a suitably chosen value, but when a distance is smaller than that value, it substitutes a function for the corresponding potential that increases steadily with the decreasing interatomic distance and remains finite when the atoms coincide. Then only those

resulting conformations which differ significantly are retained and the ones with similar side chains are discarded. After that, a solvation term is incorporated and the energies are recomputed.⁵²

Later on Broyde *et al.* presented a build-up algorithm to predict the structure of the DNA from its sequences. This algorithm made energy minimization trials on smaller nucleosides and then joined them together followed by energy minimization.⁵³

1.2 Force Fields

According to the physicist Dirac, “the underlying physical laws of a large part of physics and the whole of chemistry are completely known, and the difficulty is only that the exact application of these laws leads to equations much too complicated to be soluble. It therefore becomes desirable that approximate practical methods of applying quantum mechanics should be developed, which can lead to an explanation of the main features of complex atomic systems without too much computation.”⁵⁴ To solve the time-independent Schrödinger wave equation for elements other than hydrogen, we need to make certain approximations and the most important of these is the Born-Oppenheimer approximation. According to this approximation, there is a strong separation of time scales between the motions of electrons and nuclei, since the nuclei are several thousand times heavier than the electrons. Due to this fact, we can separate the motions of nuclei and the electrons. Nuclei move much more slowly than the electrons due to being much heavier. This approximation is very important in quantum chemistry and allows computation of the wavefunction in two less complicated steps. While it is a very useful approximation but in the systems where nuclear motion is highly coupled to the electronic motion, it breaks down, for example in graphene,⁵⁵ certain diatomic molecules,^{56,57} and calculation of electric hyperpolarizabilities.⁵⁸

Force fields are useful to study large sized molecules at a reasonable level of

accuracy. A force field (FF) is a collection of energy functions and other sets of different parameters used to calculate the PES of a molecule. At the advent of computational chemistry, due to the large sizes of biomolecules, it was not possible to use quantum chemical methods to calculate their dynamic properties, therefore initially, empirical energy functions were used to calculate the dynamic properties of the biomolecules as these were computationally inexpensive due to using atoms as the smallest particles of the system instead of using the electrons and nuclei as in QM calculations. The application of these empirical energy functions to organic molecules was called molecular mechanics.⁵⁹ These empirical energy functions were converted to the potential energy functions which were used to compute the potential energy of a structure as a function of its three dimensional structure.^{60,61} The total energy of a system with a definite three dimensional (3D) structure, (E_{total}) may be represented by its bonding and non-bonding interactions as shown in equation 1.1 below:

$$E_{\text{total}} = E_{\text{str}} + E_{\text{bend}} + E_{\text{tors}} + E_{\text{el}} + E_{\text{vdW}} + E_{\text{cross}} \quad (1.1)$$

where E_{Str} represents the energy function due to stretching of a bond among two atoms, E_{bend} represents the energy needed to bend a bond angle, E_{tors} denotes the torsional potential for rotation around a single bond, E_{vdW} represents the van der Waals interactions, E_{el} is the energy due to electrostatic interactions (Figure 1.1), and E_{cross} is the relationship between the first three terms that describes the coupling of the internal variables of these terms, such as angles and bond lengths. This equation covers both bonding and nonbonding interactions. The minima on the PES represent the stable conformers which may be found by minimizing the E_{total} as a function of the nuclear coordinates.⁶² When this energy function is known, the geometries and relative energies can easily be calculated. For biological systems, the nonbonding energy terms in eq. 1.1 are considered to be very important. When the three dimensional structure along with all the information regarding connectivities are known, one can use eq. 1.1 to calculate the total potential energy of that system. Combinations of these potential energy functions in eq. 1.1 with addition of different other parameters gave rise to simple force fields

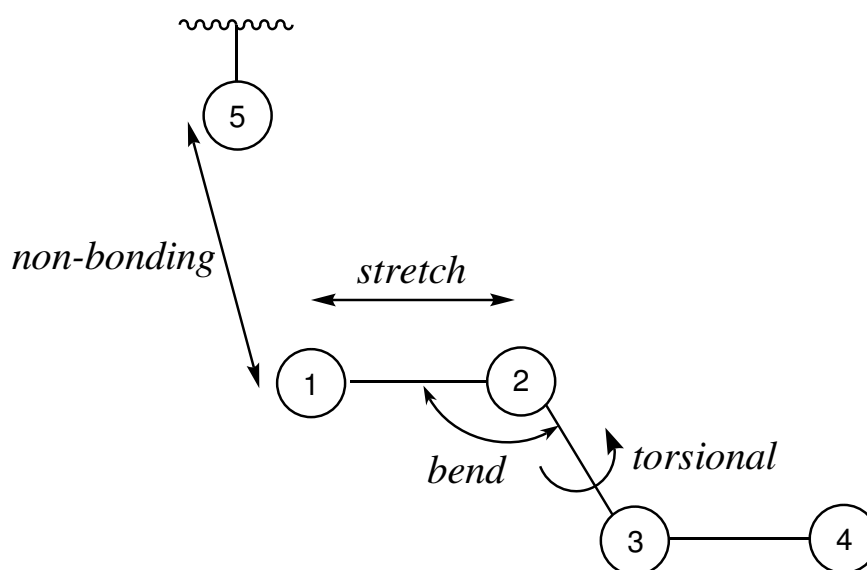


Figure 1.1: A hypothetical arrangement of two molecules to show the terms in equation 1.1. The first molecule consists of atoms 1-4 while atom 5 belongs to a second molecule.

like AMBER,⁶³ CHARMM,⁶⁴ and GROMOS.⁶⁵

Numerous force fields exist to study different types of molecules. Some of them have been designed specifically for biomolecular simulations while others have been designed for small organic molecules that can be used to study biomolecules as well. The majority of software tools use the AMBER, CHARMM, or GROMOS packages for studying nucleic acids and proteins.⁶⁶⁻⁶⁹ A variety of general purpose force fields can be used to study an extensive variety of molecules and these force fields have been optimized to work well with small sized gas phase molecules. These include the optimized potential for liquid simulations (OPLS) force field which was originally established for liquid simulations of organic molecules,^{70,71} MM3 which has also been used for organic molecules as well as proteins,^{72,73} consistent force field⁷⁴ for various small organic molecules, Merck Molecular force field (MMFF)⁷⁵ for a number of small molecules, Universal force field (UFF)⁷⁶, and Dreiding force field⁷⁷ which were developed for a large variety of compounds including inorganic compounds as well. A brief summary of these force fields is described here.

1.2.1 AMBER

The AMBER (“Assisted Model Building with Energy Refinement”) force field was developed at first by Kollman *et al.* for nucleic acids and proteins⁷⁸ which was recalibrated later using experimental parameters for these molecules.⁷⁹ It was extended to other organic molecules in its second generation.⁶⁹ AMBER is also a molecular dynamics program which implements the AMBER and other force fields.^{80,81} AMBER is no longer a single force field, but a group of different force fields. The general functional form is the same, and each member of the group has its own name and different values of the parameters of the general form.

Kollman *et al.* revised their force field again in 1996 after Beachy *et al.* performed some high level calculations on peptides⁸² using *ab initio* methods and showed different outcomes. This new model was named C96.⁸³ In this model, the derivation of atomic charges was done by fitting them to the electrostatic potentials from the QM calculations on appropriate compounds. The parameters for dispersion interactions were derived from the OPLS-FF.⁸⁴ This work was extended to include the restrained electrostatic potential (RESP) from QM calculations and using it for calculating the conformational energies of organic and biomolecules.⁸⁵ This extension of the AMBER force field was named Parm99. Another modification to the AMBER force field was made later on and it was described as “general Amber force field” (GAFF) that was also made compatible for arbitrary organic molecules made of C, H, O, N, S, P, and halogen atoms.⁸⁶ It included two models, *i.e.* empirical and heuristic models for the determination of force constants and proved to be better than the existing MMF94⁸⁷ and CHARMM⁶⁴ force fields. The Parm99 model for organic and biomolecules was modified later for nucleic acids for the accurate illustration of their α/γ concerted rotations and named as parmbsc0 force field.⁸⁸ For this force field, the fitting of parameters was done using the data from very high quality QM calculations and comparison to a lot of calculated as well as experimental data. Overall, the main focus of the AMBER force fields has been on proteins and nucleic acids and it produces mostly good results.

1.2.2 CHARMM

Martin Karplus and his team at Harvard developed a flexible computer program package named CHARMM (“Chemistry at HARvard Molecular Mechanics”) to model, perform molecular dynamics simulations, and energy minimize macromolecules in the gas phase.⁶⁴ This program was given a new force field concept for the empirical energy function. No directional properties or internal degrees of freedom were assigned to atoms. CHARMM includes the bonding potential, bonding angle potential, dihedral angle potential, improper torsions, dispersion interactions in the molecules, electrostatic potential, contribution of hydrogen bonding towards the total energy, atomic constraints, and dihedral constraints in the molecule as contributors towards the overall energy of the system.⁶⁴ After validation of extensive experimental data from different techniques like X-ray crystallography, IR and Raman spectroscopy, and *ab initio* calculations on test nucleic acid molecules, the same group at Harvard developed various parameters for the CHARMM program force field called as CHARMM22.⁶⁷ The optimized and revised parameters included bond lengths, bond angles, improper dihedral, and torsional angles terms. One parameter for proteins was also included in this revision to study the interactions of nucleic acids with proteins. Soon after this development, new parameters were made available in the all atom CHARMM force field for proteins.⁶⁸ This time, the parameters were optimized for bonding and non-bonding interactions and solvent-solute interactions were also introduced. The peptide backbone bonding parameters were also optimized according to the data obtained from experimentation.

Due to some limitations of CHARMM22 with the molecular dynamics of DNA double helical structures in water, it was completely re-optimized to produce the new CHARMM27 force field.^{89,90} In this new force field, dihedral and non-bonding parameters were especially improved on the basis of QM data and compatibility with water solvation was increased. In this successive series of development of the CHARMM force fields, it was optimized by Brooks *et al.* for different small proteins and

improvements were made for representing the β -sheet and helical structures of proteins and polarization effects were added.⁹¹ Afterwards, improvements for peptide backbones were introduced by the same group and parameters were improved using high level quantum chemical data.⁹² All previously developed versions of CHARMM had parameters for nucleic acids, proteins, carbohydrates, and lipids. Mackerell *et al.* produced an extension of this force field for drugs and organic molecules including heterocyclic compounds and termed it as CHARMM General Force Field (CGenFF).⁹³ This approach extended the scope of CHARMM to medicinal compounds and drugs. A more recent update of the CHARMM force field has been produced by Pastor and co-workers who made significant modifications for optimizing it for phospholipid bilayers and revised it to the all-atoms CHARMM lipid force field.⁹⁴ It has been anticipated to be of use for pure lipid systems as well as heterogeneous systems with membrane proteins embedded in them. Over the last three decades, the CHARMM molecular simulations software has been developed and updated for biological molecules including nucleic acids, proteins, carbohydrates, lipids, and small organic molecules. It contains both gaseous and solution environments for molecules and provides a number of different force fields and optimization procedures.⁹⁵

1.2.3 GROMOS

GRONingen MOlecular Simulation (GROMOS) is the name of a force field and the software suite which was developed nearly three decades ago by the Computer-Aided Chemistry Group at the University of Groningen and ETH Zurich.^{96,97} Kouwijzer *et al.* extended this force field for carbohydrates and simulated the crystal structures of monosaccharides with it correcting a previously determined structure.⁹⁸ The GROMOS96 software package was introduced afterwards for molecular modelling and molecular dynamics studies of biological molecules using the GROMOS force field.⁹⁹ In this program the molecular dynamics and energy minimization algorithms were completely rewritten and improved.⁶⁵ Ott and Meyer modified the GROMOS force field while

working on simulations of maltose and added parameters for exo-anomeric effects.¹⁰⁰ Further improvements in the force field were done for carbohydrates and the problem of energy difference for inversion of chair conformations in β -D-glucopyranoside was solved. The torsional barriers for the conformers having hydroxymethyl groups were also fixed.¹⁰¹

The journey of improvements in the GROMOS force field continued and it kept improving and changing. It was used for peptides and proteins until the year 2000. The developers of the force field then added parameters for its applications to lipids and validated it with known molecules. This new version was known as GROMOS96 parameters set 45A3.¹⁰² This allowed the use of the force field for lipid aggregates, aliphatic systems, and polymers. This parameter was later shown to have some issues while working with sugars and the double helical structure of DNA, so it was improved and the torsional angle parameters for the nucleotide backbone and the charge distribution in nucleoside bases were optimized based on quantum chemical data. This update was released as parameter set 45A4.^{103,104} In a continuation of the successive parametrizations of this force field, the free energies of hydration and non-polar solvation were provided for a variety of molecules. These new parameters, *i.e.* 53A5 and 53A6 were optimized for thermodynamic properties of liquids and fixation of charges for hydration free enthalpies in water, respectively.¹⁰⁵ The new parameter set 53A6 has also been validated by simulating different test sets of compounds and comparing the results to experimental data.¹⁰⁶

After all the above mentioned improvements and developments in the GROMOS force field and program package, the authors released a new version of the software suite in 2005 called as GROMOS05 which included all the improvements made until then.¹⁰⁷ More recently, new parameters sets were introduced for the GROMOS force field called 54A7 and 54B7.¹⁰⁸ These parameters included various modifications to the previous version including correction of torsional terms for helical systems, modifications in the amine and carbonyl repulsions, free hydration energies of Na^+ and Cl^- ions, and

calculations of free energy which involve a chirality change. The new parameter set 54A7 of the GROMOS force field has been validated by simulation studies of two folding β -peptides and comparison to NMR data.¹⁰⁹ Oostenbrink *et al.* improved the existing parameter set 57A7 for charged amino acid chains and re-calibrated the non-bonding interactions for the aforementioned molecules to produce the current parameter set 54A8.¹¹⁰ The latest version of the program is GROMOS11 that has been also functionalized to couple to and work with the quantum chemistry software package TURBOMOLE¹¹¹ and offers more flexibility in this working environment.^{112,113}

1.2.4 OPLS

The Optimized Potential for Liquid Simulations (OPLS) force field was originally developed to study liquid and hydration dynamics of proteins.^{84,114} It was later extended for nucleotide bases by fitting the parameters from *ab initio* calculations.¹¹⁵ The agreement of its results with the experimental data was good. To generalize it, this force field was extended to saturated hydrocarbons as well, but it was unable to find wide usage unlike CHARMM and AMBER.⁷⁰ Soon after that, an OPLS all-atom (AA) force field was developed for conformational analysis of organic liquids. For this, most of the angle bending and bond stretching parameters were derived from the AMBER force field.⁷¹ OPLS-AA-FF has since been extended to include carbohydrates. The torsional parameters for sugars were fitted by performing *ab initio* calculations on test molecules. The resulting force field was reported to perform very well for pyranoses.¹¹⁶ Soon after that, OPLS-AA was extended towards heterocycles and nitrogenous bases like pyridine, diazenes etc. Quite a number of parameters have also been implemented from AMBER, like bond stretching and angle bending parameters.¹¹⁷ Bruice and Kahn parametrized OPLS-AA to take into account some functional groups that are found in polyketides. They improved the existing parameters in the force field on the basis of high level QM calculations (at MP2/aug-cc-pVTZ level of theory) on model systems.¹¹⁴ Van Gunsteren and co-workers improved their previously produced OPLS-AA-FF, which was

for carbohydrates only, by including some extra scaling factors for the electrostatic interactions.¹¹⁸ Another improvement was also made to this force field by Acevedo and Sambasivarao who developed parameters for ionic liquids by performing high level *ab initio* calculations on 68 different ionic liquids.¹¹⁹ There is another modified version of this force field named as OPLS-2005 which is available in a closed source package hence the details of the implementation are not published[†].^{120,121}

1.2.5 CFF

The Consistent Force Field (CFF) is not a single force field but a series of force fields that were developed with the goal of treating a wider range of small organic compounds and also includes parameters for large molecules, *i.e.* peptides. In these force fields, most of the parameters like conformational energies, vibrations, and enthalpies were derived from optimisation of the internal terms by a least-squares algorithm to give the best agreement to a large amount of experimental data.¹²² CFF was applied for the first time to the family of alkanes, including cycloalkanes. Comparison between the experimental and calculated data was used as a basis for the selection of energy functions and successive determination of further parameters. Crystal data was also added, where available. CFF is able to calculate enthalpies of sublimation, molecular and lattice frequencies, and unit cell parameters.¹²³ In further studies, CFF has been applied to amides and lactams for calculating their conformations and vibrational frequencies.¹²⁴ The next class of compounds to be added and validated was non-conjugated alkenes. The CFF method was used to calculate the conformations, frequencies, and heats of hydrogenations of alkenes.¹²⁵ Successive studies on amides, proteins and peptides made CFF available to calculate the conformations and hydrogen bonding interactions for these compounds.¹²⁶ In continual expansion of CFF, hydrogen bound crystals of carboxylic acids, amides, and hydrogen bonded carbonyl compounds were optimized by least-square fitting of calculated to experimental data.¹²⁷ This data was then

[†]Schrödinger Macromodel

benchmarked against other available force fields and its utility was demonstrated.¹²⁸

An extension of CFF called COMPASS (Condensed-phase Optimized Molecular Potentials for Atomistic Simulation Studies) was designed for condensed phase applications after some modification to the non-bonding and valence parameters of CFF. COMPASS was produced after an extensive revision of the issues in CFF and made general for a wide range of compounds.^{129–132} It has since been optimised for nitrate esters, 100 inorganic compounds (He, Ne, Ar, Kr, Xe, H₂, O₂, N₂, CO, NO, CO₂, NO₂, SO₂, and CS₂) in liquid phases,¹³³ and compounds with aliphatic azide chains.¹³⁴

1.2.6 MM3

The molecular mechanics force field (MM3) was developed for aliphatic hydrocarbons in 1989. This force field enabled users to calculate the energies of molecules including enthalpy of formation, energies of conformers, and rotational barriers to interconvert between different conformers.⁷³ In the course of its development, the vibrational frequencies were also fitted with 213 experimental values. The calculation of torsional frequencies enabled the calculation of entropies for various alkanes and cycloalkanes.¹³⁵ The van der Waal's interactions between hydrogen and carbon in aliphatic and aromatic hydrocarbons were also improved from a previous version of this force field (MM2) and the new parameters were produced after fitting them with crystal parameters.¹³⁶ This was a popular force field for hydrocarbons as the other force fields available were not specifically made for hydrocarbons and were parametrized only for proteins, nucleic acids, carbohydrates etc. Soon after its development, it was also parametrized for alcohols and ethers,¹³⁷ alkenes,¹³⁸ amides, polypeptides, and proteins.⁷² The vibrational spectra of the representative compounds of these classes were also fitted to make calculations of electrostatic interactions more accurate and reliable. Bowen *et al.* extended the scope of the MM3 force field for sulphonamide and its alkyl derivatives. They calculated the geometries of these compounds at the *ab initio* MP2/6-31+G* level

and included the geometric and vibrational parameters from these quantum chemical calculations to make the simulations of this type of compounds accurate.¹³⁹ Subsequent parametrization of the MM3 force field produced the MM3(96) force field for group 1A and 2A complexes having ligands with conjugated ethers as donors.¹⁴⁰ The authors of MM3 calculated hydrogen bonding interactions for a variety of compounds containing C, N, O, and Cl on *ab initio* calculations and optimized the MM3(96) force field accordingly.¹⁴¹ Hay and co-workers extended the MM3(96) force field for dynamics of complexes of amides in which the amide oxygen is coordinated to a metal centre.¹⁴² Another improvement in this force field was the addition of parameters for Pd-olefin complexes but this was limited to just this class of complexes.¹⁴³ In the last decade, the parametrization of the MM3 force field was done by Schmid *et al.* for nanoporous metal-organic frameworks (MOFs). They used experimental data along with density functional theory calculations to parametrize the potentials of the force field.¹⁴⁴ This force field is available in a number of software packages[‡].

1.2.7 MMFF

Merck Molecular Force Field (MMFF) is a group of force fields that were developed and owned by Merck Research Laboratories. The backbone is the MM3 force field and it is not specified or optimized for a single class of compounds, rather it was designed to cover a wide range of organic compounds. The first published version of this force field was MMFF94, which was derived from the MM3 force field, but the main difference was that it was designed for a variety of organic compounds starting from small organic molecules to bigger protein structures.⁸⁷ It has also been optimized using quantum chemistry calculated data of more than 2800 structures of different types and all the parameters were obtained from the quantum chemical data to make it versatile and accurate. It has also been optimized for van der Waal's and electrostatic interactions including non-bonding interactions,¹⁴⁵ molecular geometries, vibrational

[‡]*e.g.* Schrödinger Software Suite, TINKER Molecular Modelling etc.

frequencies, and conformational geometries of organic compounds.¹⁴⁶ Later it was extended as MMFF94s for a broader range of systems and hydroxide ions, alkali metals, alkaline earth metals, halogens, other metal ions including Zn^{2+} , Ca^{2+} , Cu^{2+} , Fe^{2+} , and Fe^{3+} , oxyphosphorus compounds, and sulphonamides were also been added.¹⁴⁷ This force field has also been added to a number of molecular modelling software[§].

1.2.8 UFF

The Universal Force Field (UFF) is a full periodic table force field that has been built on the idea of universality and not for specific structural classes or elements. In UFF, the parameters were calculated using general rules for elements, their molecular hybridization and connectivities.⁷⁶ UFF has been tested for a wide variety of organic compounds and the results were compared to the results of MM2 and MM3 force fields. The results demonstrated that UFF predicted the structures of hydrocarbons, alkenes, saturated amines, silanes, aromatic systems, saturated ethers, phosphines, ketones, imines, and nitriles well. The maximum bond angles deviations were found to be 3° and the bond distance deviations were below 0.02\AA .¹⁴⁹ It also showed good results for conformational analysis of uncharged compounds. UFF has also been validated for main group compounds and in this case the bond length errors were larger than those for organic molecules. The calculated bond lengths were in good agreement to experimental data but for positively charged centres, where multiple electronegative substituents are present, showed deviations of up to 0.1\AA and the bond angles for hypervalent atoms had even larger errors.¹⁵⁰ While testing the UFF for metallic complexes, it was observed that M–C bond distances were calculated well and the errors in bond lengths for M–Y polar bonds were generally smaller than 0.05\AA , apart from a few exceptions.¹⁵¹ UFF has also been tested on a number of cobalamines and the results were compared to MM2 calculations. The results of UFF were comparable to those from the MM2 force field.¹⁵²

[§]Avogadro, Schrödinger Software Suite, RDKit¹⁴⁸, and TINKER Molecular Modelling etc.

UFF1 has been published as an extension of UFF and designed especially for dealing with drug discovery related problems. Dipole and quadrupole moments and van der Waal's interactions were parametrized on the basis of *ab initio* calculations on a large number of reference compounds.¹⁵³ Validation of UFF has been done through its comparison to density functional tight binding (DFTB) calculations of MOFs. These calculations showed that the quality of UFF's calculated bonding interactions is comparable to DFTB calculations at a much smaller computational cost.¹⁵⁴ UFF works well with general MOFs but for MOFs of the first transition series, it has certain limitations that have been eliminated in a recently published version of the universal force field known as UFF4MOF.¹⁵⁵ This force field has been reparametrized to include the transition elements that form MOFs and to calculate structures that were not possible with the standard UFF. This force field UFF4MOF has been made available in different software packages[¶].

1.2.9 Dreiding

The Dreiding force field was proposed in 1990 for a range of biological, organic, and main group compounds. The concept behind this force field was to use generalized parameters for all atoms based on hybridizations of atoms instead of other complex parameters for each type of atoms and their combinations.⁷⁷ Hence, there was only one force constant for each of the parameters of bonds, bond angles, and six different values for barriers to rotations around torsions. This force field has not found general applicability and extensive developmental study like other force fields discussed earlier. The Dreiding force field has been implemented in the Marvin Beans suite of programs.¹⁵⁹

[¶]*e.g.* deMon2k,¹⁵⁶ GULP,¹⁵⁷ and Amsterdam Density Functional (ADF).¹⁵⁸

1.2.10 Quantum Mechanically Derived Force Field

Recently, Grimme presented a new force field which is parametrized *in silico* specifically for a single conformer from quantum mechanically computed data. He named it Quantum Mechanically Derived Force Field (QMDFF) because it derives all potentials from QM data.¹⁶⁰ He gave a black-box type procedure that takes as input the QM calculation of any molecule of interest and uses the information, *i.e.* the equilibrium structure, atomic charges, and bond orders from this input and derives a force field that is specific for the molecule under study. The total energy resembles very closely to equilibrium geometry given by quantum chemical calculations and the force constants for stretching, bending, and torsions are fitted internally to the quantum chemical data. QMDFF can be applied to covalent or non-covalent molecules of arbitrary structures.

The general functional form of the total energy in QMDFF is given in equation 1.2:

$$E_{\text{total}} = E_{\text{e,QM}} + E_{\text{intra}} + E_{\text{NCI}} \quad (1.2)$$

Where E_{total} is the total energy, $E_{\text{e,QM}}$ is the electronic energy calculated by QM calculation of the input structure, E_{intra} is the energy due to bonding interactions, and E_{NCI} refers to non-covalent interactions. The details of all these terms can be found in Grimme’s work.¹⁶⁰ The basic restriction of this force field is that the input structure should already be a minimum on the potential energy surface (PES) and not a maximum because the force field relies on the supplied input from the QM calculation and tries to map the PES around that minimum energy structure. Various benchmarks for different types of molecules have been performed which show excellent results by QMDFF. These benchmarks included calculations of vibrational frequencies, conformational energies, atomization energies, gas phase structures of organic molecules, non-covalent interactions, and supramolecular interactions for test sets of different molecules including alkanes, peptides, sugars, amino acids, transition metal complexes, and supramolecular complexes from various databases and experimental datasets.

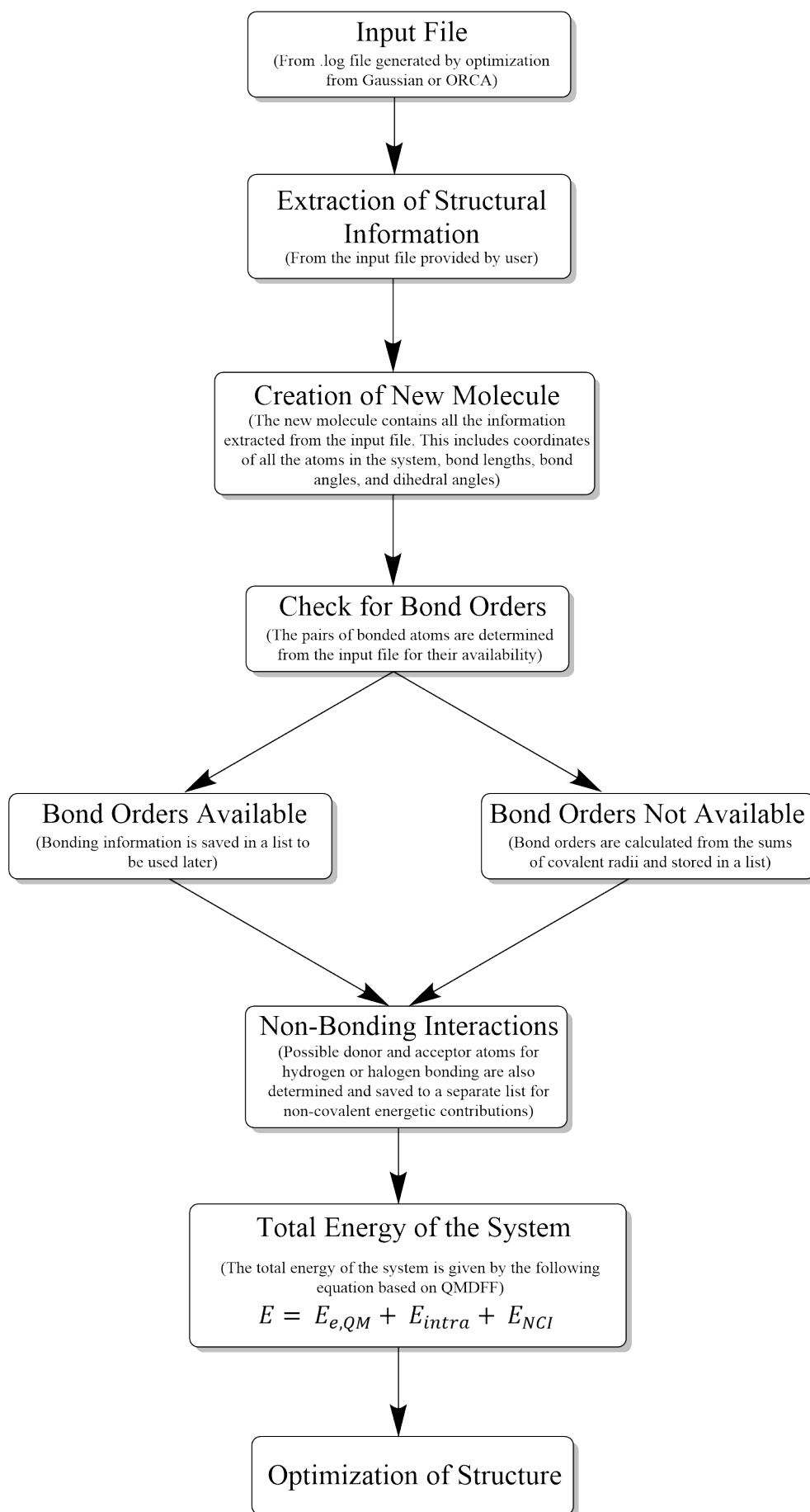


Figure 1.2: General workflow of WellFARe-FF.

1.2.11 Wellington Fast Assessment of Reactions Force Field

Recently the Lein group at VUW (the OnLein Lab) developed a new force field called Wellington Fast Assessment of Reactions Force Field (WellFARe-FF), using the basic idea of QMDFF but with greater flexibility and other options which are available for the user to choose between. This force field has been implemented to map the PES in order to find the transition states of reactions in a systematic way. WellFARe-FF allows the user to choose from many available energy terms to calculate and is freely available to distribute. The overall workflow of this force field is shown in Figure 1.2.

1.2.12 Comparison of Functional Forms in Different Force Fields

All of the above described force fields have different functional forms for the potentials described in eq. 1.1. These force fields differ from each other in functional forms of the energy terms used, number of cross terms included, and the types of information used for fitting the parameters. Generally, if the force field is designed for larger systems like proteins or DNA, the functional forms are kept very simple. In these force fields E_{str} and E_{bend} are only represented by harmonic functions or sometimes these terms may even be omitted thus forcing constant bond lengths and angles. E_{cross} is also not used in these cases. Such force fields are often called “Class I” force fields. For more accuracy and for small- to medium-sized molecules, a number of cross terms are also included along with cubic or quartic expansions of E_{str} and E_{bend} . These force fields are not only good at reproducing good geometries and energies but also at vibrational frequencies and are termed as “Class II” force fields. Further refinements by allowing different parameters to depend on neighbouring atom types, *e.g.* for modelling hyperconjugation etc. are termed as “Class III” force fields. Here, a comparison of the functional forms of all of them has been given in Table 1.1. From this table, we can see the similarities and differences between them and compare them effectively. This table has been adapted from Jensen’s Introduction to Computational Chemistry¹⁶¹ for the force fields described

above.

Table 1.1: Comparison of the functional forms used in common force fields. The torsional energy E_{tors} in all cases is given as a Fourier series in the torsional angle.

Notation: P_n : Polynomial of order n ; $P_n(\cos)$: polynomial of order n in cosine to the angle; $\cos(n\theta)$: Fourier term(s) in cosine to the angle; Exp^{-6} : *exponential* + R^{-6} ; $n - m$: $R^{-n} + R^{-m}$ Lennard-Jones type potential; *Imp.*: improper torsional angle; *ss*: stretch-stretch; *bb*: bend-bend; *sb*: stretch-bend; *st*: stretch-torsional; *bt*: bend-torsional; *btb*: bend-torsional-bend.

Force Field	E_{str}	E_{bend}	$E_{\text{oop}}/E_{\text{tors}}$	E_{vdW}	E_{el}	E_{cross}	Molecule Types
AMBER	P_2	P_2	<i>Imp</i>	12-6 12-10	Charge	None	Proteins, nucleic acids
CHARMM	P_2	P_2	<i>Imp</i>	12-6	Charge	None	Proteins
GROMOS	P_2	P_2	$P_2(\text{Imp})$	12-6	Charge	None	Proteins, nucleic acids
OPLS	P_2	P_2	<i>Imp</i>	12-6	Charge	None	Proteins, nucleic acids
CFF	P_4	P_4	P_2	9-6	Charge	<i>ss,bb,st, sb,bt,btb</i>	General
MM3	P_4	P_6	P_2	Exp^{-6}	Dipole/ Charge	<i>ss,bb,st</i>	General (all elements)
MMFF	P_4	P_3	P_2	14-7	Charge	<i>sb</i>	General
UFF	P_2 or Morse	$\cos(n\theta)$	<i>Imp</i>	12-6	Charge	None	All elements
Dreiding	P_2 or Morse	$P_2(\cos)$	$P_2(\cos)$	12-6 or Exp^{-6}	Charge	None	General
QMDF	Generalized Lennard Jones	P_2	$P_n(\cos)$	D3BJ	Charge	None	General

1.3 Conformational Analysis for the Prediction of NMR

Data

There is no doubt that NMR spectroscopy is an indispensable technique for the characterization of a very wide variety of molecules, both organic and inorganic. It gives a wealth of information about the structural properties of a compound, *i.e.* the direct connectivities of different atoms in a molecule, the neighbouring atoms of a particular centre in molecule, information about different functional groups, and the 3D spatial arrangement of atoms and their relationships.¹⁶² Another advantage of NMR is that one can study different types of molecules, *i.e.* impure samples to identify a specific

substance,^{163,164} reaction intermediates,^{165,166} and the progress of the reaction by analysing product formation directly inside an NMR tube^{167,168} without the need for purification or crystallization procedures. It is also remarkably quantitative in response factor. Flexible organic molecules exist in more than one interchangeable conformations in the solution state so their NMR spectrum is the averaged spectrum of all the low lying conformers at the temperature of the NMR experiment.

In the last decade, quantum chemistry has become greatly involved in solving structural problems as comparison with experimental spectroscopic techniques.^{169,170} Improvements in computational resources and the development of sophisticated methods has enabled experimental chemists to augment their confidence in assigning relative configurations to molecules, using quantum chemical calculations to verify experimental techniques.^{1,162,171–173} A number of methods have been established to distinguish between stereoisomers if only one set of experimental data is available.^{18,174,175} Calculation of NMR chemical shifts with smaller sized or rigid molecules is a fairly easy task but for complex molecules and/or conformationally flexible ones, the task becomes daunting. To predict the NMR chemical shifts of flexible molecules, it is necessary to identify their low energy conformers on the potential energy surface and calculate their NMR chemical shifts, followed by finding the Boltzmann averaged chemical shifts of all conformers according to their relative energies.^{16,176} There are a number of software tools available that offer conformational searching and finding the lower energy conformations of a molecule, but as the number of rotatable bonds increases in a molecule, the number of conformations also increases exponentially and finding all energy minima becomes impractical. There are a number of reports on calculation of NMR data of various organic and biomolecules as a useful tool in structural determination.^{162,177–182} QM calculations have also been used to calculate the ¹³C-NMR spectrum and its comparison with the experimental data to establish the relative configuration of conicasterol F, a sponge-derived sterol.¹⁸³ Correct assignments of systems containing cyclic rings in their structure have also been studied with DFT to compute their ¹³C-NMR chemical shifts and assign their configuration.^{184,185} Coupling constant calculations have also been reported

to assist determining the relative configuration of various natural molecules.^{186,187}

1.4 Conformational Analysis for Assigning Absolute Configurations

Assigning absolute configuration to flexible organic molecules through comparison of experimental spectral data with quantum chemically calculated data has also been becoming more and more practical.¹⁸⁸ The use of quantum chemical calculations for predicting NMR, electronic circular dichroism (ECD), vibrational circular dichroism (VCD) spectra, optical rotatory dispersion (ORD), and optical rotation and their comparison with the experimental data to assign absolute configuration has become a common tool for organic, especially natural product, chemists.^{189–197}

For complex molecules and when there are multiple chiral centres, ECD and VCD can be used to determine both the relative and absolute configuration with the help of theoretical calculations. These techniques are especially helpful when there is a pair of enantiomers that cannot be differentiated by NMR only so one can perform ECD or VCD calculations and compare only one experimental data set with all the possibilities using theoretical calculations. Similar to NMR, circular dichroism (CD) spectroscopy also requires accurate mapping of the PES and taking into account all the contributing conformers to get a Boltzmann averaged ECD or VCD spectrum that correlates strongly with the experimental one.

There are a large number of reports showing the use of time-dependent (TD)-DFT to calculate ECD and optical rotation data for assigning the absolute configuration of naturally occurring molecules.^{198–211,211–213} Sherer *et al.* developed a systematic approach for calculating VCD spectra of flexible compounds using three different Merck ‘in house’ software tools. They described that VCD spectroscopy can be effectively used with computational measurements in drugs and pharmaceutical molecules.²¹⁴ In a

comparative study of ECD, ORD, and VCD to determine the absolute configuration of a relatively flexible compound by differentiating between its different diastereomers, it was shown that VCD is more advantageous over the other two for the study of diastereomers.²¹⁵ The absolute configuration of a naturally occurring antiprotozoal lactone, klaivanolide, has been determined by comparison of its experimental VCD spectrum to the calculated one at the B3PW91/TZ2P level of theory.²¹⁶ Absolute configurations of the schizozygane alkaloids and some iridoids have also been assigned on the basis of density functional theory calculations of their VCD spectra with comparison to the experimental data.^{217,218}

Very recently, our research group also conducted a study to establish the absolute configuration of echivulgarine, isolated from the pollen of *Echium vulgare*. Theoretical calculations of ECD and ¹³C-NMR spectra were employed to predict the actual stereoisomer determined experimentally.²¹⁹

1.5 Different Software Tools Available for Conformational Analysis

There are a number of software tools available in the market which offer conformational analysis of flexible molecules using various classical force fields and different conformational sampling methods. Macromodel²²⁰ and Maestro²²¹ are popular software programs included in the Schrödinger software suite.²²² It provides various force fields like MM2, MM3, MMF94s, AMBER, and OPLS with systematic and Monte Carlo conformer search algorithms. Spartan is another famous quantum chemistry software which offers MMFF, MMFF94, MMFF94s, and SYBYL force fields using a systematic search.²²³ The Marvin Beans suite is available free of charge for academic use which offers the Dreiding force field for conformational search.²²⁴ Vega ZZ is a freely available tool which employs the SP4 force field with the AMMP molecular

mechanics search method.²²⁵ The AMBER software package uses the AMBER force field and offers different conformational sampling methods.⁸⁰

Currently it is a very challenging task to perform an accurate and efficient conformational analysis of molecules with a growing number of flexible bonds using the available tools and software. Quite a few software tools offer good conformational search but they are expensive, closed source, offer classical force fields, and only accept certain variables according to which the number of conformers may be chosen from the conformational search.²²² There are other free software tools available which also offer conformational search but they also have limitations, *e.g.*, they work only with smaller molecules, they do not have a tool to select reasonable conformers from a large number of generated conformers during conformational search, etc.²²⁴ The other limitation of all the available software is that they only offer a range of common classical force fields like AMBER,⁷⁸ CHARMM,⁶⁴ MM3,⁷³ UFF,⁷⁶ MMFF94⁸⁷ etc. which are specific for certain sets of molecules and cannot be widely applied to all unknown or new molecules. After generation of conformers from any of these force fields, the selection process of conformations from a conformational search is improvised and needs experienced users to select those conformers with proper Boltzmann weighting.²¹⁴

This research focused on the development of an algorithm to perform an accurate conformational search in a relatively short period of time by using density functional theory (DFT) at a higher level of accuracy. The other main objective was to automate the whole process using Python and bash scripts to save human effort and time. The developed algorithm is capable of mapping the PES and then removing duplicates and redundant structures, thus yielding a small number of contributing conformers with greater than 90% of the Boltzmann distribution. Prior to the development of the algorithm, a manual conformational analysis was performed using Vega ZZ²²⁵ to confirm the absolute configuration of echivulgarine (**12**), which has been published.²¹⁹ The difficulties and hard work behind that manual conformational analysis aided in the motivation to devise a smart algorithm to map the PES.

Chapter 2

Methods

The scripts for *MICE-PES* have been written using either version 3 of Python programming language or bash.

2.1 Computational Methods

All the calculations for the compounds under study were performed using the Gaussian09 suite of programs (Revision D.01)²²⁶ and the visualization of the results was done with GaussView5 or CYLView 1.0.²²⁷ The PES plots were obtained using GNU Plot.²²⁸ All calculations were carried out with density functional theory (DFT) using Adamo's hybrid²²⁹ version of the Perdew, Burke and Ernzerhof functional (PBE0)^{230,231} with Grimme's empirical D3 correction with Becke-Johnson damping (D3BJ).^{232–234} According to a study by Grimme, the hybrid functionals are unable to describe the energy between different conformers very well. The main reason for that is the bad description of the correlation energy by them, which interestingly, also describes the dispersion. Due to the use of D3BJ correction with a hybrid functional like PBE0 make it suffer much less and improve the energy.²³⁵ For compounds **2-43**, all calculations were carried out with Alrich's triple ζ basis set def2-TZVP²³⁶ while the calculations

on compound **1** were performed using the def2-SVP basis set in combination with the Polarizable Continuum Model (PCM) using the integral equation formalism variant (IEFPCM)²³⁷ with the SMD parameter set established by Cramer and Truhlar²³⁸ to model the effects of the solvent. The PCM solvation model performs well in terms of speed and accuracy in cases where there is no explicit interaction between the solvent molecules and the compound under study. However, if the solvent can interact with the molecule, which could lead to the disruption of intramolecular hydrogen bonds, the PCM model is not a very good choice. In these cases explicit solvent molecules have to be added to the quantum mechanical part of the system in addition to using the PCM model. Dichloromethane was used as a solvent in all calculations for compounds **1-12**, chloroform for compounds **27-43**, and acetonitrile was used for compounds **13-26**. The choice of the solvents for the calculations was based on the solvents used for studying the spectroscopic properties of the compounds experimentally.

All conformers were confirmed to be minima on the potential energy surface through the calculation of harmonic force constants at the same level of theory that was used for the optimisations. Usual statistical mechanics was used to calculate Gibb's free energies as implemented in Gaussian 09.²²⁶ NMR calculations were performed as implemented in Gaussian²³⁹⁻²⁴¹ using the same hybrid functional (PBE0) with Alrich's triple ζ basis set def2-TZVP for compounds **2-43**, and def2-SVP basis set for compound **1** while including D3BJ dispersion correction.. A multi-standard approach was used to reference the NMR chemical shift values as described by Sarotti and Pellegrinet¹ because of its better accuracy and precision than using TMS as a reference standard. Methanol was used as a reference standard for sp^3 carbon atoms while benzene was used for referencing sp and sp^2 carbons. Theoretical UV-Vis and ECD spectra were calculated using time-dependent density functional theory (TD-DFT) with the same density functional and basis set as the optimisations of the molecules (24 singlet excited states).²⁴² The conformational analysis for all diastereomers of **1** (**1a-1d**) was performed using the Boltzmann Jump method (SP4, 3000 steps, 300 K, 60° RMSD) as implemented in the AMMP package with the Vega ZZ graphical interface.²²⁵ The DP4 analysis¹⁸ was

employed for distinguishing between the different diastereomers of echivulgarine (**1**). The scripts used for automation of the process are available at github under a GNU general public license.²⁴³

2.2 Experimental Details

NMR spectra of 3-*epi*-xestoaminol C (**12**) were recorded in house by Sarah Andreassend using a Varian DirectDrive spectrometer equipped with a triple resonance HCN cryogenic probe operating at 600 MHz and 150 MHz for ^1H and ^{13}C nuclei, respectively. Residual solvent peaks were used as an internal reference for chemical shifts (CD_2Cl_2 : δ_{H} 5.32, δ_{C} 53.84).²⁴⁴ NMR spectra were obtained using 3 mm Norell® Select Series™ and 5 mm WILMAD® NMR tubes, respectively. Deuterated solvents were purchased from Cambridge Isotope Laboratories, Inc. Optima® grade dichloromethane for measuring the optical rotation was purchased from Fisher Scientific. For echivulgarine (**1**) CD data was obtained by Galen Eakins while the NMR data was obtained by Jonathan Singh.

Optical rotations of 3-*epi*-xestoaminol C were measured in house using a Rudolph Research Analytical Autopol II polarimeter at the sodium D-line (589 nm). The CD spectrum for **1** was recorded on a Chirascan CD spectrometer (Applied Photophysics, UK) using a 1cm quartz cell at Robinson Research Institute, Victoria University of Wellington.

Chapter 3

Results and Discussion

3.1 Assignment of Stereochemical Configuration of Echivulgarine using Experimental and Computational Techniques

Echivulgarine (**1**) is a pyrrolizidine alkaloid isolated previously from plants of the genus *Echium*. Unfortunately, its structure has not been confirmed spectroscopically in the past.²⁴⁵ Because only a small amount of **1** was obtained, it has not yet been tested for biological activity. However, while some pyrrolizidine alkaloids are toxic to human beings and livestock,^{246,247} they have shown good antimicrobial potential against various microorganisms.^{248,249} It could be expected that echivulgarine behaves similarly. During this study, echivulgarine (**1**, Figure 3.1) was re-isolated from the pollen of *Echium vulgare* and its structure was established on the basis of extensive NMR techniques, circular dichroism (CD) spectroscopy, and theoretical calculations of NMR and electronic circular dichroism (ECD) spectroscopy.²¹⁹

Computations of NMR and ECD spectra were performed by carrying out a manual conformational analysis of the molecule with the help of Vega ZZ to increase the confidence

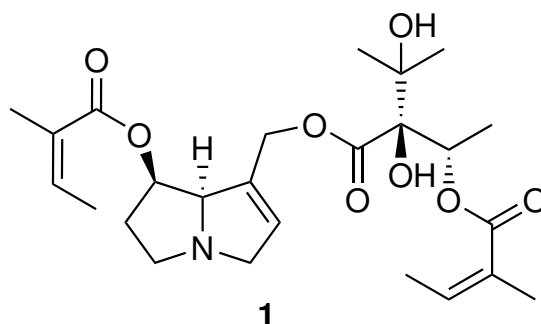


Figure 3.1: Structure of 7*R*,8*R*,2'*R*,3'*S*-echivulgarine isolated from *Echium vulgare*.

of assigning the absolute configuration.

The experimental NMR and CD spectral data were obtained by Dr. Jonathan Singh and Dr. Galen Eakins (VUW), respectively, and to compare that with the theoretical calculations, it was necessary to study all four possible side chain diastereomers of **1** to determine their stereochemistry by comparison to the experimental data. There were technically 2⁴ diastereomers of **1** but we could establish the relative configuration of the rings²¹⁹ and therefore reduce the number of diastereomers to those of the side chain. Once the relative configuration of the side chain in relation to the rings was established, then the absolute configuration could be assigned. Therefore, all the four diastereomers (**1a**, **1b**, **1c**, and **1d**) were modelled as shown in Figure 3.2. Because echivulgarine (**1**) is a flexible molecule, a conformational analysis was performed using the Boltzmann Jump search algorithm as implemented in VEGA ZZ.²²⁵

This conformational search produced 500 conformers for each diastereomer. After careful manual removal of duplicates, the energetically lowest lying 40, 49, 48, and 49 conformers (providing >99% Boltzmann distribution to the ensemble) of **1a**, **1b**, **1c**, and **1d**, respectively, were then optimized using DFT at the PBE0-D3BJ/def2-SVP/SMD_{CH₂Cl₂} level of theory. After the optimization, the SCF energies were used to determine the Boltzmann distribution and the relative contributions of all conformers for all diastereomers. This method showed the percentage contribution of different conformers in the solution form that were contributing to the real structure. Then, the highest contributing or lowest energy conformers (accounting for >99% of the ensemble) were selected to

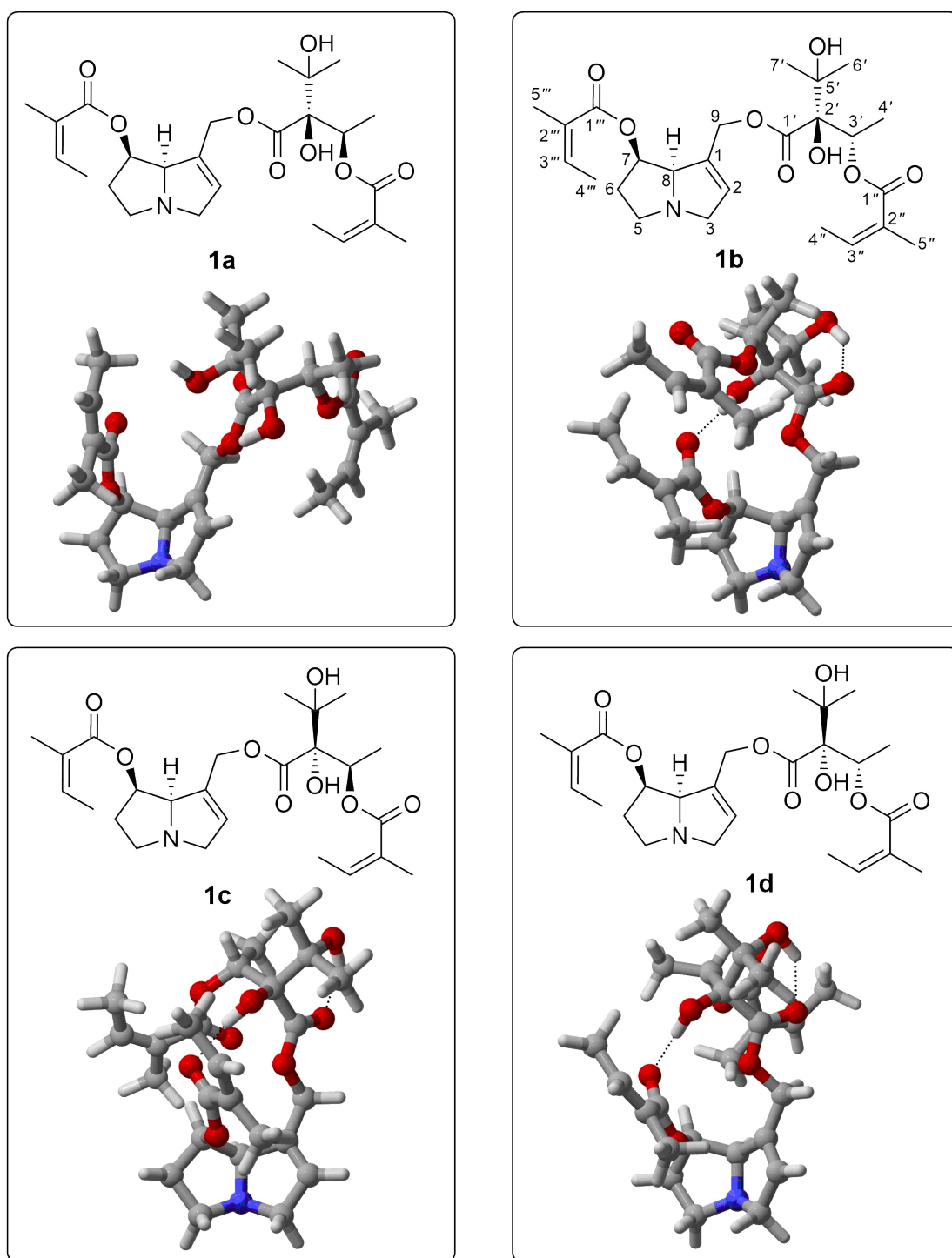


Figure 3.2: Four possible side chain diastereomers of echivulgarine. **(1a)**. *7R,8R,2'R,3'R*-echivulgarine, **(1b)**. *7R,8R,2'R,3'S*-echivulgarine, **(1c)**. *7R,8R,2'S,3'R*-echivulgarine, and **(1d)**. *7R,8R,2'S,3'S*-echivulgarine with relative energies of 9.11, 0.00, 3.55, and 2.89 kcal/mol. The numbering shown in **1b** is followed for all the NMR assignments and discussion throughout the chapter. The 3D structures are the lowest energy conformer of the respective diastereomer.

use for the NMR and ECD property calculations. These selected conformers along with their relative energies and Boltzmann weighted percentage are given in Table 3.1.

Table 3.1: Boltzmann weights and relative energies (kcal/mol) of the selected lowest energy conformers of all four side chain diastereomers of echivulgarine (**1a-1d**)

Conformer No.	Relative Energy	Boltzmann Weighted Percentage Contribution
<i>7R,8R,2'R,3'R</i> -echivulgarine (1a)		
1	0.00	36.45
2	0.54	14.58
3	0.64	12.36
4	0.66	11.9
5	0.92	7.73
6	1.07	6.03
7	1.27	4.27
8	1.53	2.76
9	1.7	2.08
10	1.98	1.29
<i>7R,8R,2'R,3'S</i> -echivulgarine (1b)		
1	0.00	92.74
2	1.52	7.08
3	4.49	0.05
4	4.54	0.04
5	4.73	0.03
6	5.46	0.01
7	5.57	0.01
<i>7R,8R,2'S,3'R</i> -echivulgarine (1c)		
1	0.00	45.75
2	0.34	25.93
3	0.55	18.19
4	1.57	3.21
5	1.68	2.69
6	1.85	2.01
7	2.31	0.93
8	2.72	0.46
<i>7R,8R,2'S,3'S</i> -echivulgarine (1d)		
1	0.00	73.24
2	1.06	12.33
3	1.36	7.40
4	2.02	2.43
5	2.14	1.98
6	2.28	1.55
7	2.98	0.47
8	3.11	0.38

The data from ECD calculations on all the selected conformers were submitted to the SpecDis²⁵⁰ software which summed them according to the their Boltzmann distribution and percentage contribution to the ensemble. A comparison of the ECD spectra of **1a**,

1b, **1c**, and **1d** with the experimental CD spectrum is presented in Figure 3.3.

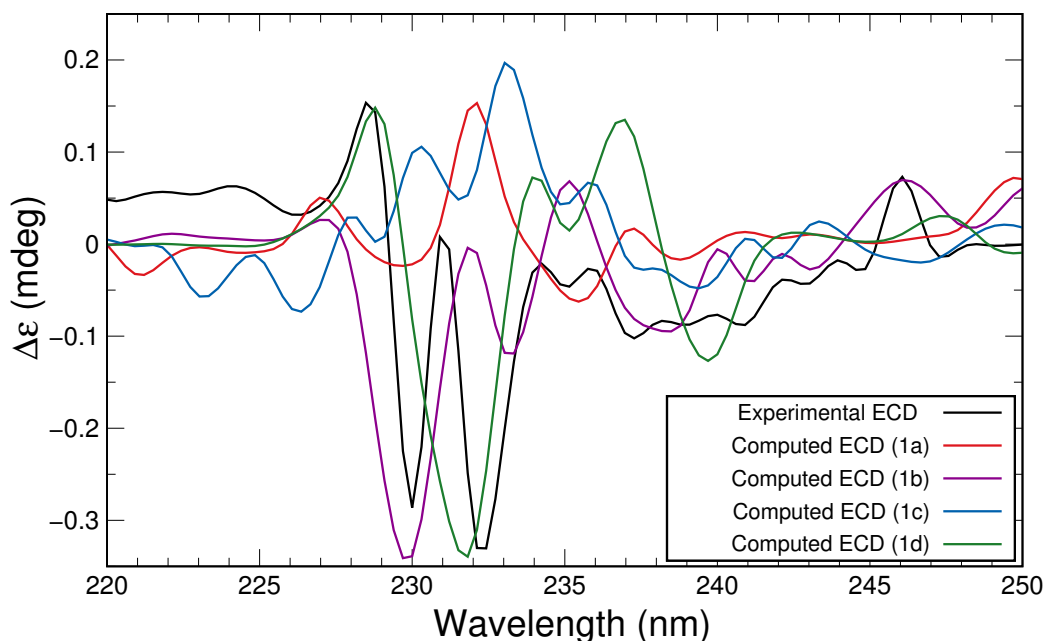


Figure 3.3: Comparison of the experimental (dichloromethane) and calculated ECD spectra of **1a** (*2'R,3'R*-EV), **1b** (*2'R,3'S*-EV), **1c** (*2'S,3'R*-EV), and **1d** (*2'S,3'S*-EV). EV is short for echivulgarine.

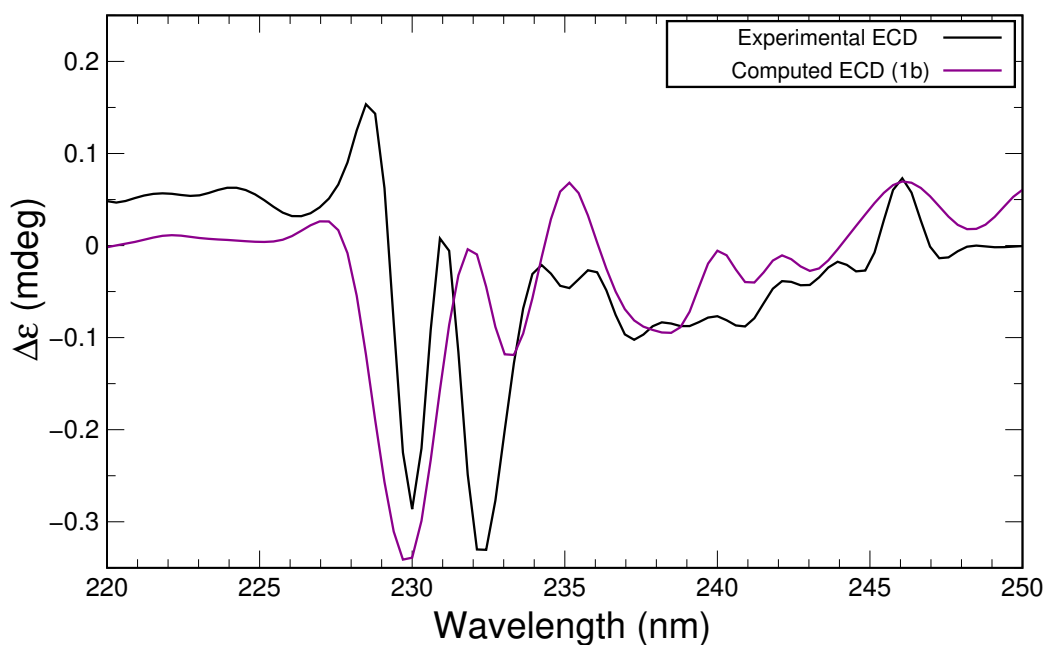


Figure 3.4: Comparison of the experimental (dichloromethane) and calculated ECD spectra of 7*R*,8*R*,2'*R*,3'*S*-echivulgarine (**1b**).

After comparison of the mean absolute deviation (MAD) of all the spectra (MADs:

1a = 0.006689, **1b** = 0.004306, **1c** = 0.008306, **1a** = 0.005104) from the experimental data and visual analysis of Figure 3.3, the best matching calculated ECD to the experimental CD spectrum was that of *7R,8R,2'R,3'S*-echivulgarine (**1b**) (Figure 3.4). Both of these spectra show negative Cotton effects at 230 nm. The lowest energy conformer of **1b** has 92.74% Boltzmann contribution to the ensemble and is shown in Figure 3.5.

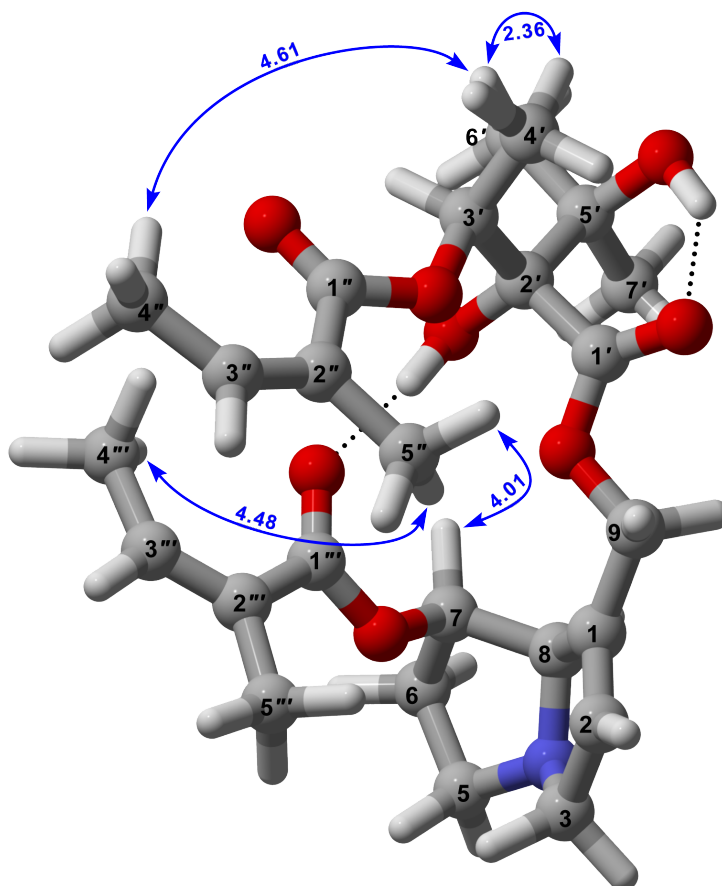


Figure 3.5: Lowest-energy conformer predicted for *7R,8R,2'R,3'S*-echivulgarine, **1b** showing key nOe correlations. The blue coloured numbers on arrows show the distance (Å) between the atoms for the nOe correlations.

This conformer has a twisted-geometry and is stabilised by intramolecular hydrogen bonds, the first one being in between hydrogen OH-2' and the oxygen at C-1'''. Because of this intramolecular hydrogen bond and twist in the geometry, the two angelyl fragments in the molecule are facing each other in a parallel fashion, thus orienting the isopropanol moiety attached to C-2' at almost 180° with respect to the hydrogen bonded –OH at C-2'. As a result of this, the –OH of the isopropyl moiety (attached at C-5') is stabilised through another intramolecular hydrogen bond with the oxygen

atom present at C-1'. These intramolecular hydrogen bonding interactions force the geometry of echivulgarine into the twisted form shown in Figure 3.5. The stabilising nature of these hydrogen bonding interactions is clearly evident from 92% Boltzmann percentage contribution of the lowest energy conformer (see Table 3.1). When all conformers of the four diastereomers are compared, **1b** is the only diastereomer that has the most favourable intramolecular stabilising interactions, thus its first five conformers have 99.95% Boltzmann percentage contribution to the ensemble.

After finding this geometry of the lowest energy conformer of **1b** I decided to look for the key nOe interactions (in the NMR data obtained by the experimental colleagues) which should be detected if it resembles the experimental structure because the nOe interactions recorded and discussed by the experimental colleagues did not include the ones shown in Figure 3.5. When the structure was initially reported, NMR assignments were based off a two-dimensional structure where many of the nOe correlations could not be rationalised due to uncertainty of the diastereomer present. With the new three-dimensional structure that was obtained from calculations and the confirmation of the diastereomer isolated, it was now possible to reassess the original NMR data. Careful comparison of the experimental and computational results supported the computationally identified structure. All nOe signals that would be expected from the theoretical prediction, were able to be identified in the experimental data and the determined geometry was fully supported by the nOe correlations between H-4'' and H-6', H-4''' and H-5'', H-5'' and H-7, and H-4' and H-6' (Figure 3.5). These nOe correlations supported the calculated structure of **1b** a great deal. On the basis of this evidence, the structure of echivulgarine (**1**) was confirmed as 7*R*,8*R*,2'*R*,3'*S*-echivulgarine (**1b**) as shown in Figure 3.1 with its lowest energy and mainly contributing conformer shown in Figure 3.5.

As an additional step of validation, a DP4 analysis²⁵¹ as reported by Goodman *et. al.* to differentiate between diastereomers when experimental NMR data for an individual isomer is available, was performed. For the NMR calculations the same conformers

selected for optical rotation calculations were subjected to calculations of NMR chemical shifts at the PBE0-D3BJ/def2-SVP/SMD_{CH₂Cl₂} level of theory.

Table 3.2: Experimental (δ_{exp}) and calculated (δ_{calc}) ¹³C-NMR data of echivulgarine diastereomers (**1a-1d**). MAE is the mean absolute error and RMSE is the root mean square error calculated for the chemical shift difference ($\Delta\delta$) from experimental values.

Position	δ_{exp} (ppm)	1a	1b	1c	1d
		δ_{calc} (ppm)			
1	132.7	137.6	136.1	136.0	135.7
2	126.7	127.9	130.4	134.1	138.1
3	61.5	62.3	62.8	62.0	61.7
5	53.9	52.5	51.3	52.2	51.8
6	34.3	34.2	36.1	34.9	35.2
7	72.7	71.5	74.3	71.3	72.3
8	76.3	77.3	76.8	78.0	77.5
9	61.7	62.3	62.7	62.4	62.6
1'	173.6	170.5	175.6	174.4	177.9
2'	83.5	83.7	81.2	80.3	84.0
3'	72.3	69.5	75.3	76.5	68.3
4'	15.4	15.5	17.8	16.1	17.7
5'	73.6	73.8	72.9	73.1	72.4
6'	26.7	27.6	28.1	28.1	27.2
7'	25.2	25.9	24.2	24.0	24.6
1''	166.2	163.3	163.1	164.1	163.0
2''	127.5	127.3	128.3	129.8	129.3
3''	139.0	147.7	149.8	148.7	146.4
4''	15.9	17.8	16.9	17.6	17.1
5''	20.6	23.5	23.5	23.1	22.7
1'''	166.5	163.1	164.8	164.8	166.7
2'''	126.5	129.1	128.4	127.8	128.7
3'''	141.0	148.2	149.7	151.4	151.6
4'''	16.0	18.4	17.6	17.9	18.0
5'''	20.7	22.4	21.5	22.3	22.0
MAE		0.90	1.56	1.63	1.70
RMSE		2.98	3.40	3.67	3.92

The comparison of experimental and theoretically calculated Boltzmann averaged ¹³C-NMR data of all the four diastereomers is presented in Table 3.2. . The comparison of experimental and theoretically calculated Boltzmann averaged ¹H-NMR data of all the four diastereomers is presented in Table 3.3.

After adding all the ¹³C- and ¹H-NMR data to the DP4 applet, it showed that the most probable candidate of all four diastereomers was **1b** (100% confidence level).

After combining all results, the final configuration of echivulgarine was established as 7*R*,8*R*,2'*R*,3'*S*-echivulgarine (**1b**) and this data has been published.²¹⁹

Table 3.3: Experimental (δ_{exp}) and calculated (δ_{calc}) ^1H -NMR data of echivulgarine diastereomers (**1a-1d**). MAE is the mean absolute error and RMSE is the root mean square error calculated for the chemical shift difference ($\Delta\delta$) from experimental values.

Position	δ_{exp} (ppm)	1a	1b	1c	1d
		δ_{calc} (ppm)			
2	5.82	5.95	5.86	6.02	6.09
3a	3.53	3.75	3.74	3.75	3.77
3b	4.29	3.21	3.31	3.07	3.31
5a	2.87	2.92	2.86	2.9	2.91
5b	3.78	2.41	2.42	2.4	2.44
6a	2.24	1.86	1.81	1.85	1.88
6b	2.30	1.72	1.60	1.64	1.69
7	5.57	5.01	5.34	5.03	5.50
8	4.82	3.97	3.99	4.03	3.98
9a	4.62	4.64	4.83	4.84	4.50
9b	4.65	3.89	4.41	4.31	4.20
3'	5.49	5.21	5.21	4.79	5.92
4'	1.38	1.15	1.28	1.44	1.30
6'	1.38	1.11	1.04	0.97	0.86
7'	1.24	1.28	1.23	1.13	1.05
3''	6.04	5.79	6.18	6.51	6.27
4''	1.93	2.24	2.15	2.13	2.01
5''	1.77	1.69	1.36	1.69	1.81
3'''	6.14	6.36	6.27	6.40	2.94
4'''	1.96	1.83	1.81	1.99	2.16
5'''	1.78	1.65	1.65	3.03	1.74
MAE		0.29	0.25	0.18	0.35
RMSE		0.52	0.48	0.61	0.85

3.2 Need for a Conformational Analysis Algorithm

After this study for finding the correct structure of echivulgarine (**1**), it was clear that there is a need for a robust and automatic procedure for sampling the conformational space of molecules so that the properties of flexible molecules can be studied computationally with greater confidence and ease.

Therefore it was decided to develop an algorithm for this purpose, keeping in mind the

idea of build-up procedures discussed previously in section 1.1.2.

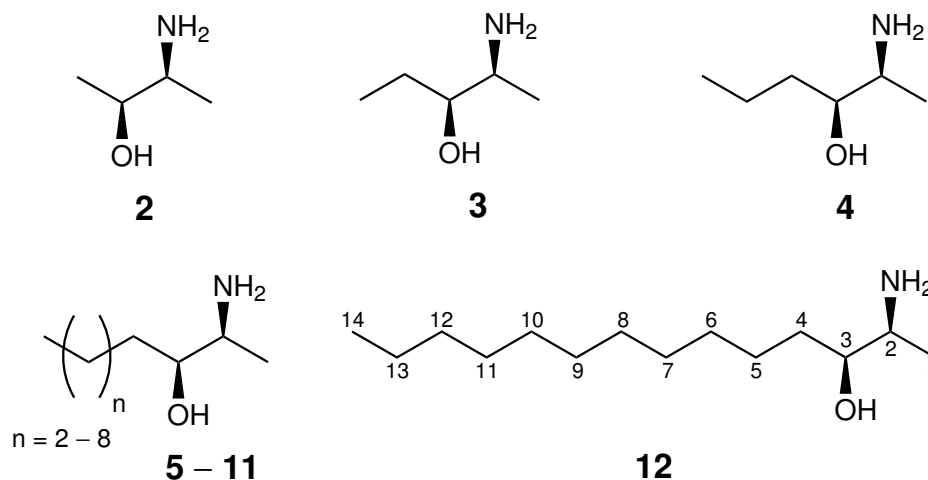
3.3 Method for Incremental Construction and Exploration of Potential Energy Surface (*MICE-PES*)

Performing a complete systematic conformational analysis of large and flexible molecules is almost impossible, particularly above the force fields level. To overcome the issues of accurate conformational sampling for larger and flexible molecules, the idea of constructing the conformational space successively from the smallest possible homologues of the molecule under study was developed. This idea has some similarity to the literature reports of the conformational build-up procedures. In all the reports of build-up procedures described, attempts were made to break the molecules into pieces, find energy minima and combine the individual fragments to construct the complete molecule which was inefficient because of losing many conformational possibilities while doing so. Therefore, it was planned to modify the traditional build-up procedure approaches while taking the smallest homologue of the molecule with important functionality and extensively map its potential energy surface using high level quantum chemical calculations and then build-up the molecule step by step adding one group at each step. This can be done very efficiently because the conformational degrees of freedom of the smallest homologue are very well separated when compared to those of the bigger molecule. The process therefore starts with an exhaustive mapping of the PES of the smallest homologue of the original molecule and then the lowest energy conformers are selected to proceed further for the construction of the full molecule. Then the next carbon atom is added at each of the three terminal hydrogens of the methyl group, or the two hydrogens in the case of a methylene. This process is repeated until the full length of the molecule is reached by removing duplicates at each step and passing the conformers with >90% Boltzmann contribution to the ensemble to the next step. One important thing to note here is that enantiomers are not discarded, because they have a

chance to become different diastereomers in following steps. This procedure is called *MICE-PES* (Method for the Incremental Construction and Exploration of the Potential Energy Surface). A detailed description of *MICE-PES* is given in section 3.4.1.

MICE-PES was validated on three different sized natural products, *i.e.* 3-*epi*-xestoaminol C (**12**), meroterphenol C (**26**), and rimarikiamide (**43**). The first case is an example of a simple compound with a long aliphatic chain with high flexibility while the other two represent more complex cases where intramolecular hydrogen bonding locks the PES into a rather narrow 3D space. As a note of caution, the systems where the Born-Oppenheimer approximation fails cannot be treated with *MICE-PES*.

3.4 Accurate Prediction of Optical and Spectroscopic Properties of Sphingolipids using *MICE-PES*



Scheme 3.1: The 11 compounds that were considered in this investigation. (2*S*,3*S*)-3-amino-2-butanol (**2**), (2*S*,3*S*)-2-amino-3-pentanol (**3**) etc. Note that **12** is the naturally occurring 3-*epi*-xestoaminol C (2*S*,3*S*-2-amino-tetradecan-3-ol).

Recently, 3-*epi*-xestoaminol C (**12**, Scheme 3.1), belonging to the class of 1-deoxysphingoids, was isolated and characterized by other members of the research group.²⁵² It was evaluated biologically and showed great antimicrobial, anti-tuberculosis and anticancer potential with specificity against mammalian HL-60 cancer cell lines.

The relative configuration was assigned with the help of NMR spectroscopy and derivatisation, while the absolute configuration was determined using Mosher's method.²⁵³ This compound was therefore a good test case for the present research work because it possesses high structural flexibility due to the long hydrocarbon chain and can adopt many different conformations in solution²⁵⁴ yet its complete structure is known and confirmed by synthesis.²⁵⁵ If a systematic conformational search is applied, the number of resulting conformers could reach over 170,000, which is almost impossible to utilise even at the semi-empirical level, therefore **12** was a very good test case for the validation of *MICE-PES*.

3.4.1 *MICE-PES* Methodology

MICE-PES is a new semi-automatic algorithm developed as the major component of this project and used to identify the most important conformers of a molecule in solution in a relatively short time.²⁵⁶ It combines the accuracy of QM calculations with an efficient conformer search method and can be used for conformationally flexible compounds. Herein, the use of *MICE-PES* on compound **12** is described in detail.

The PES mapping of **12** was done with *MICE-PES* using high level QM calculations by successively constructing the molecule from its smallest possible homologue (Structure **2**, Scheme 3.1). The first step to choose the smallest homologue of the molecule under study having important functional groups in the molecule is rather intuitive more than having a strict criterion. The reason for beginning with **2** is that the conformational degrees of freedom that stem from the first four carbon atoms (including NH₂ and OH functional groups) are very well separated compared from the ones due to the linear alkyl chain that comprises the rest of the molecule.

The examination of the PES of **2** started by performing a sequence of relaxed potential energy scans that mapped the PES by rotating all dihedral angles of **2** in steps of 15° (Figure 3.6). After careful examination of the resulting energy landscapes, a total of

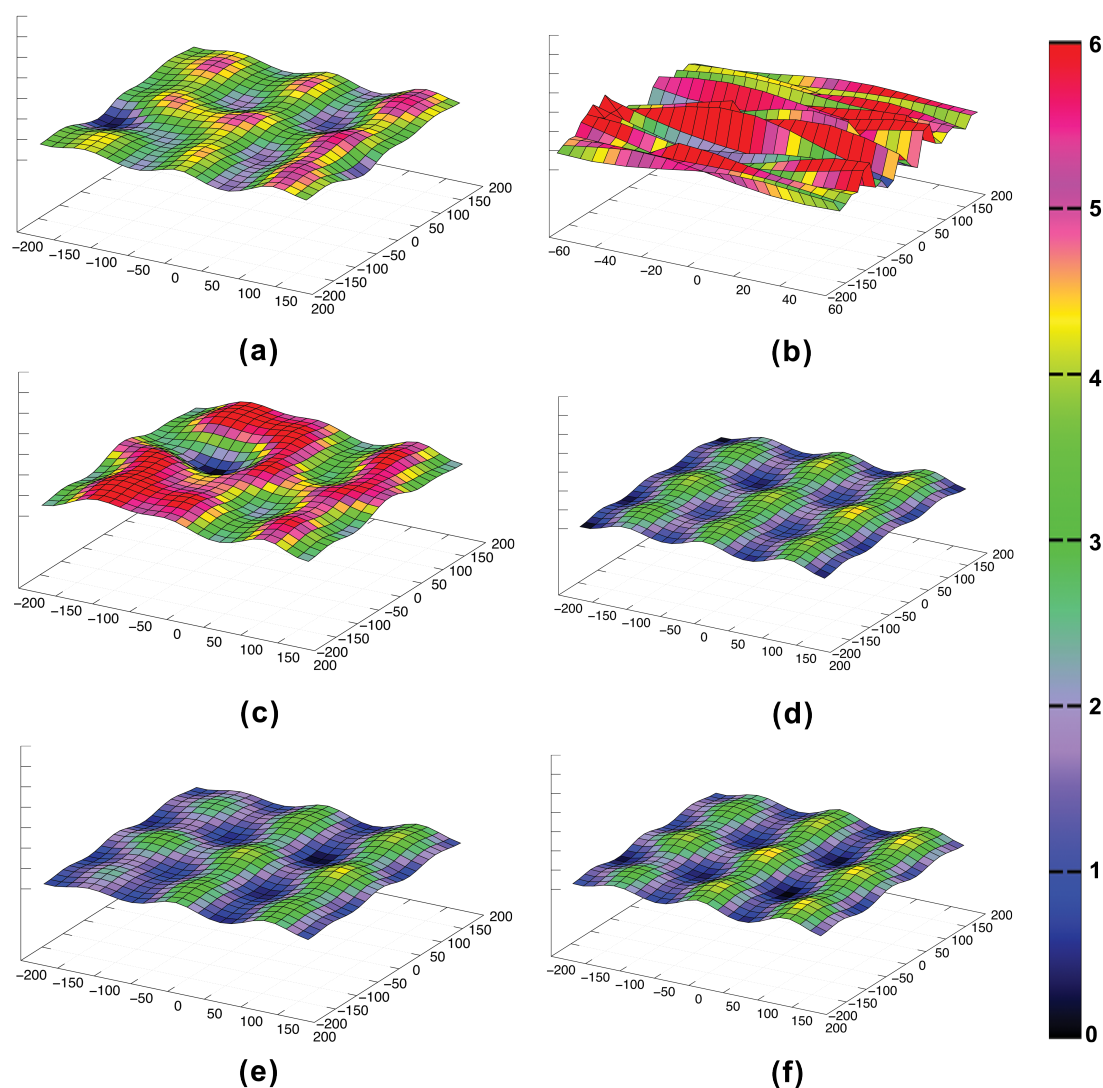


Figure 3.6: Potential energy surfaces at the PBE0-D3BJ/def2-TZVP/SMD_{CH₂Cl₂} level of theory that identify all 23 conformers of **2**. Y-axis represents the relative energies of the conformers. X- and Z-axis show the rotational angles of the starting structure. The details of these angles and rotations can be found in the Appendix A. Optimisations were carried out for a total of 3750 individual structures. The numbers on the energy scale on the right are in kcal/mol. **a–f** are the relaxed scans on different conformations of structure **2** to ensure that the PES has been covered well.

23 distinct conformers (Figure 3.7) were identified and confirmed as energy minima by unconstrained relaxation. Their relative Gibbs free energies are given in Table 3.4. The search was restricted to the most important contributions only based on >85-90% Boltzmann distribution, thus discarding high energy conformers at each stage of the molecular construction.

The difference in relative ΔG values between the global minimum and the structure

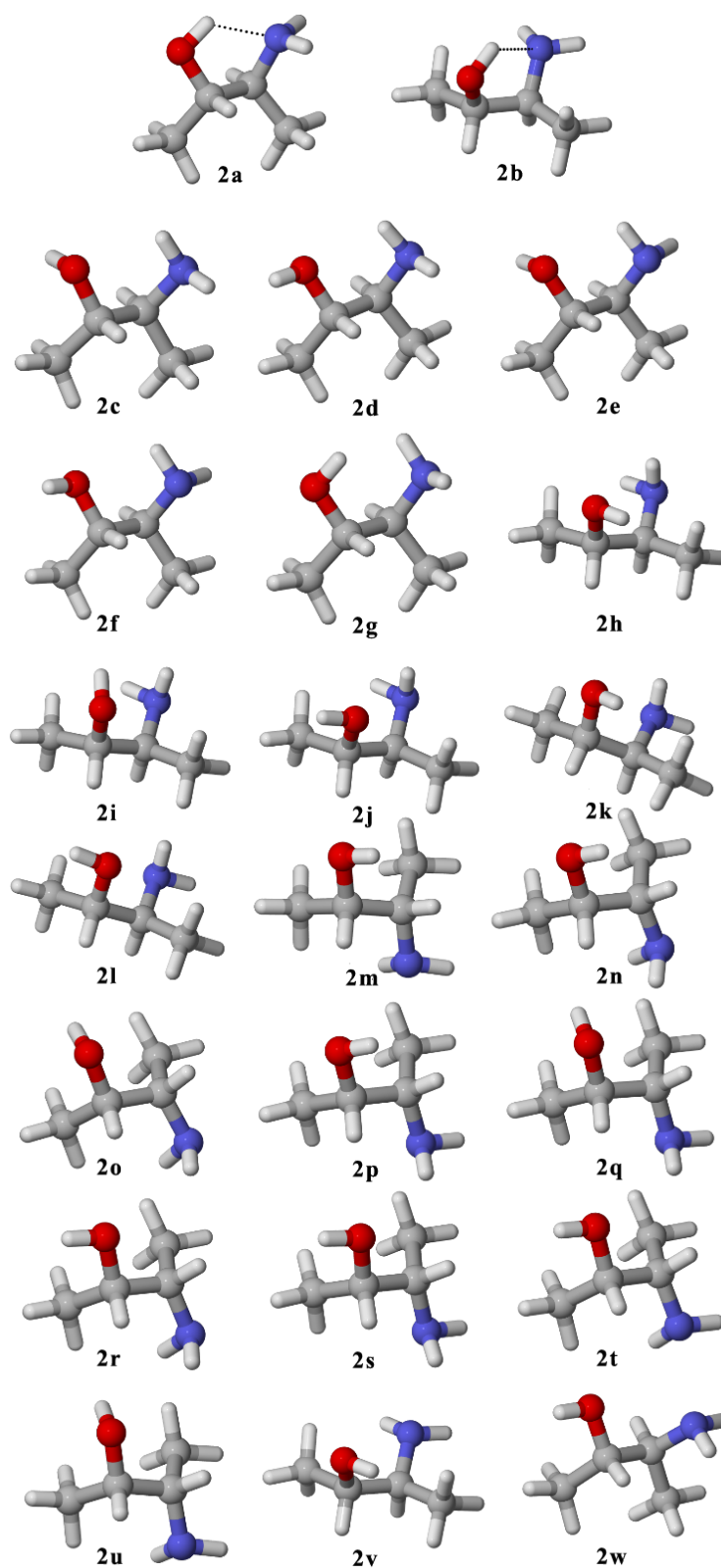


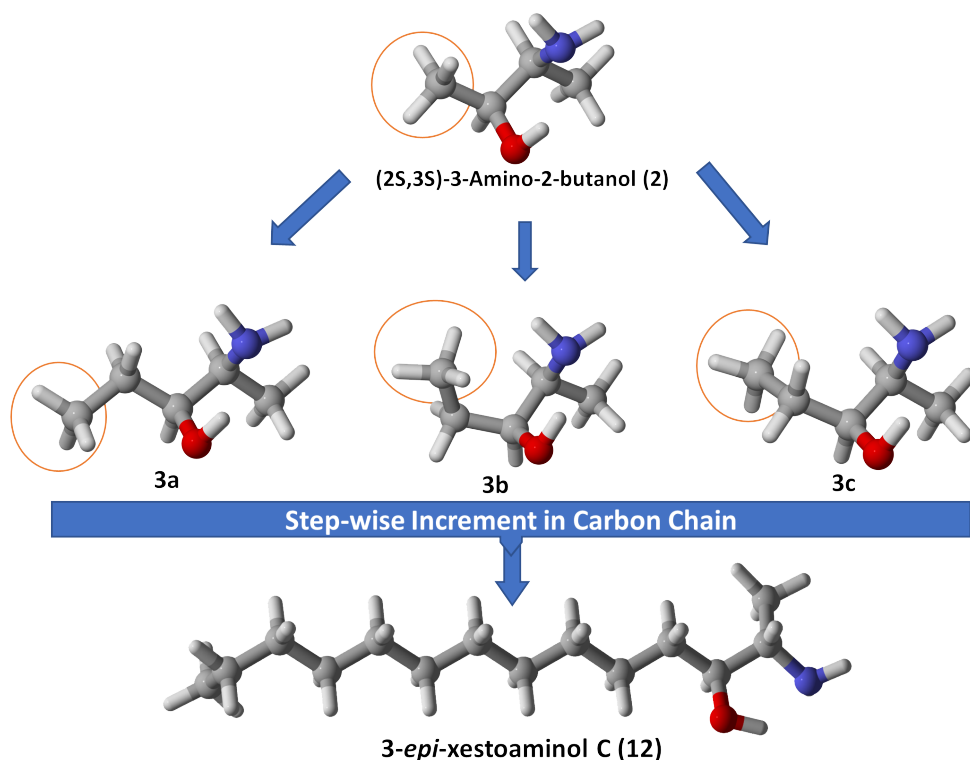
Figure 3.7: The 23 lowest energy conformers of **2** optimized at the PBE0-D3BJ/def2-TZVP/SMD_{CH₂Cl₂} level of theory. Note that the first two conformers (**2a** and **2b**) are stabilised by intramolecular hydrogen bonding.

with highest energy was only 4.7 kcal/mol (Table 3.4). An important thing to note here is that conformers **2a** and **2b** are stabilised by intramolecular hydrogen bonds. After the identification of these 23 conformers, the next step was to increase the carbon chain length by one carbon stepwise using all of these 23 conformers, therefore conformers of (2*S*,3*S*)-2-amino-3-pentanol (**3**) were constructed from the 23 identified lowest energy conformers of **2**, by substituting each of the three hydrogens on C4 with a methyl group one by one, on the terminal methyl group of all these conformers (Scheme 3.2). Hence, 69 conformers for structure **3** were obtained, which after optimisation at the same QM level described previously and removal of duplicates, yielded 67 conformers for **3**.

Table 3.4: The relative Gibbs free energies (ΔG in kcal/mol) and Boltzmann percentage distributions of the 23 minimum energy conformers of **2** at the PBE0-D3BJ/def2-TZVP/SMD_{CH₂Cl₂} level of theory

Conformer	Relative ΔG	Boltzmann Distribution (%)	Conformer	Relative ΔG	Boltzmann Distribution (%)
2a	0.00	66.96	2m	2.91	0.49
2b	1.02	11.97	2n	2.92	0.49
2c	1.68	3.92	2o	3.06	0.38
2d	1.89	2.76	2p	3.08	0.37
2e	2.01	2.27	2q	3.09	0.37
2f	2.11	1.89	2r	3.13	0.34
2g	2.12	1.87	2s	3.28	0.26
2h	2.23	1.55	2t	3.32	0.25
2i	2.39	1.19	2u	3.42	0.21
2j	2.50	0.98	2v	3.75	0.12
2k	2.63	0.79	2w	4.67	0.03
2l	2.83	0.57			

The same methodology was used to construct the conformers of the other homologues **4-12**. All conformers were considered to propagate into next step until compound **3** after the removal of duplicates. From **4** onwards, only those conformers which contributed to $\geq 85.0\%$ of the Boltzmann distribution were selected for propagation into the next step due to the increasing size of the molecule and increasing number of conformers possible. From **4-8** the Boltzmann percentage of the conformers propagated into next step was $>90\%$ but from **9-11** where the number of contributing conformers increased above 300, the propagating conformers' Boltzmann percentage distribution



Scheme 3.2: Construction of conformers through step-wise lengthening of the alkyl chain. The lowest energy isomer of **12** is shown at the bottom.

Table 3.5: \mathbf{n}^{conf} is the total number of conformers generated in a step. \mathbf{n}^{contr} is the total number of distinct conformers considered after duplicates removal (fully optimised at the PBE0-D3BJ/def2-TZVP/SMD_{DCM} level of theory). \mathbf{n}^{imp} is the number of important conformers (i.e. the number of conformers that have to be included to account for $\geq 90.0\%$ of the Boltzmann distribution at 298.15 K). **B** is the Boltzmann percentage of all the \mathbf{n}^{imp} conformers. \mathbf{c}_0 is the percentage contribution of the lowest lying conformer at 298.15 K. **P** is the total value of the partition function at 298.15 K. $\Delta\mathbf{G}$ is the difference in Gibbs free energy between the energetically lowest conformer and the energetically highest lying conformer (of all \mathbf{n}^{contr} conformers; in kcal/mol). $[\alpha]_D$ is the predicted value of the optical rotation (based on all \mathbf{n}^{contr} conformers) in DCM.

Molecule	\mathbf{n}^{conf}	\mathbf{n}^{contr}	\mathbf{n}^{imp}	B	\mathbf{c}_0	P	$\Delta\mathbf{G}$	$[\alpha]_D$
2	23	23	23	100.0	67.0	1.49	4.7	+19.4
3	69	67	67	100.0	38.1	2.62	6.8	−7.00
4	210	188	65	98.65	27.1	3.69	9.4	−20.5
5	195	190	80	94.98	14.6	6.84	6.2	−21.3
6	240	240	80	90.17	11.3	8.83	6.0	−22.4
7	240	238	84	90.17	7.2	13.82	5.5	−19.4
8	252	251	100	90.16	13.2	7.5	6.0	−10.6
9	300	299	105	85.15	9.3	10.7	5.9	−29.3
10	315	314	111	85.09	6.9	14.5	5.9	−16.2
11	333	332	113	85.10	4.9	20.5	6.1	−19.1
12	339	339	255	99.14	21.4	4.7	6.7	−20.0

was restricted to $\geq 85.0\%$ of the total ensemble.

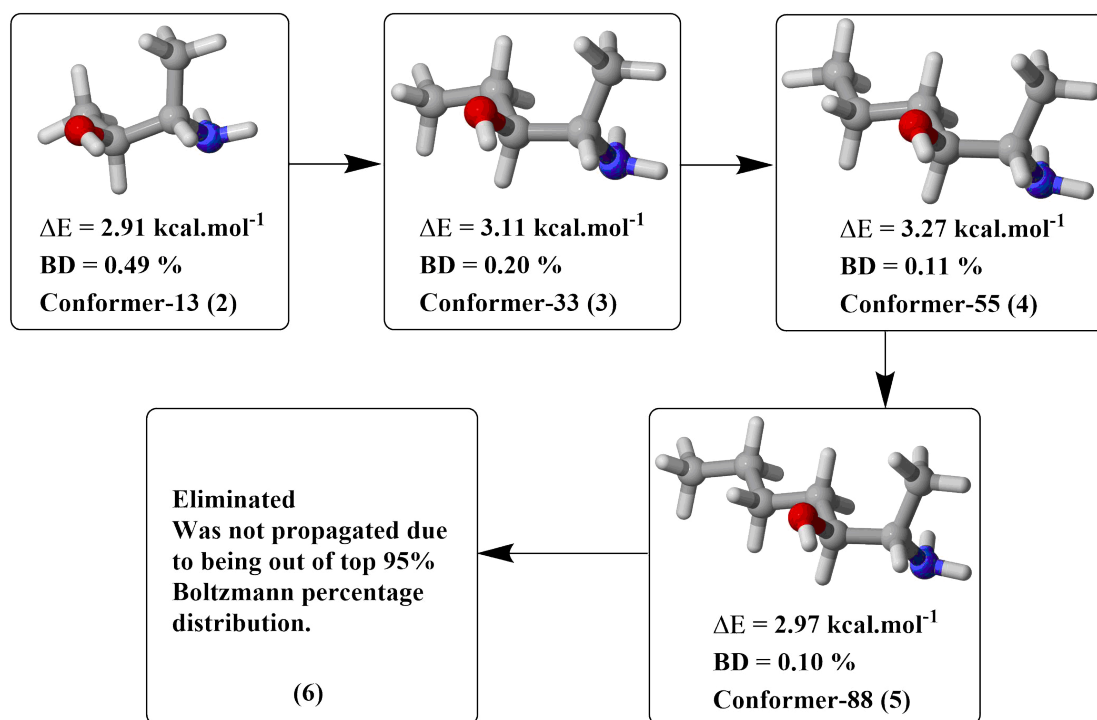
For 3-*epi*-xestoaminol C (**12**), a total of 339 conformers were obtained and optimised out of which only the lowest energy 127 conformers contribute to $>90\%$ of the total partition function (Table 3.5). For compound **12** (Table 3.5) 127 conformers were enough to represent the compound and its properties but due to all the calculations available, the Boltzmann percentage distribution considered in this case was $>99\%$.

3.4.2 Validating the Algorithm

MICE-PES proved to be convenient for accurately mapping the conformational space of the compound under study (**12**) and is predicted to work well for similar compounds, but there might be a question about the conformers that were excluded systematically at each step as they might propagate as lower energy conformers in the next step, if included. To justify this, the relative energies of all the conformers of compounds **2**, **3**, and **4** were examined without eliminating any of the conformers. Fortunately, it was found that the high energy conformers in one step kept increasing in energy in the subsequent steps and did not change to lower energy conformers with increasing chain length. This is because of the energetics of the conformational space becoming dominated by steric repulsion and hence the alkyl chain does not find any opportunity to interact with other parts of the molecule in a stabilizing fashion. The initial proposition at this stage was that this method (*MICE-PES*) is limited to straight chain compounds of similar nature that do not have intra-molecular hydrogen bonding or other stabilising interactions in their conformations (except for at one end of molecule without the presence of any other group which might interact with that end) that might render the molecule to bend, thus changing its geometry. Subsequent studies on other compounds showed that the method does actually cover hydrogen bonding and intra-molecular interactions and preserves them (see section 3.5).

In order to justify the systematic exclusion of the high energy conformers from

propagation into the next step, the relative energies of all the conformers of **2**, **3**, and **4** were examined where all the conformers were propagated into next step without elimination. It was essential to verify that the high energy conformations in one compound do not become low energy in subsequent steps, thus contributing much to the ensemble. In this study of the fate of the descendants of the high energy conformers, it was found that mostly there are many descendants of the low-energy conformers of structure **2** that are relatively high in energy in the conformers of **3** but there are very few descending conformations of high-energy conformers of **2** which become low energy in **3**. For example, upon randomly dividing the conformers of **2** into eleven “high-energy” (with 3.30% Boltzmann distribution) and twelve “low-energy” (with 96.70% Boltzmann distribution) conformers (Table 3.4), it was evident that the first conformer (**2m**) of the 11 “high-energy” conformers (Figure 3.7) became the 33rd lowest energy conformer in **3** and contributes only 0.20% to the total ensemble. In the next iteration, this effect was



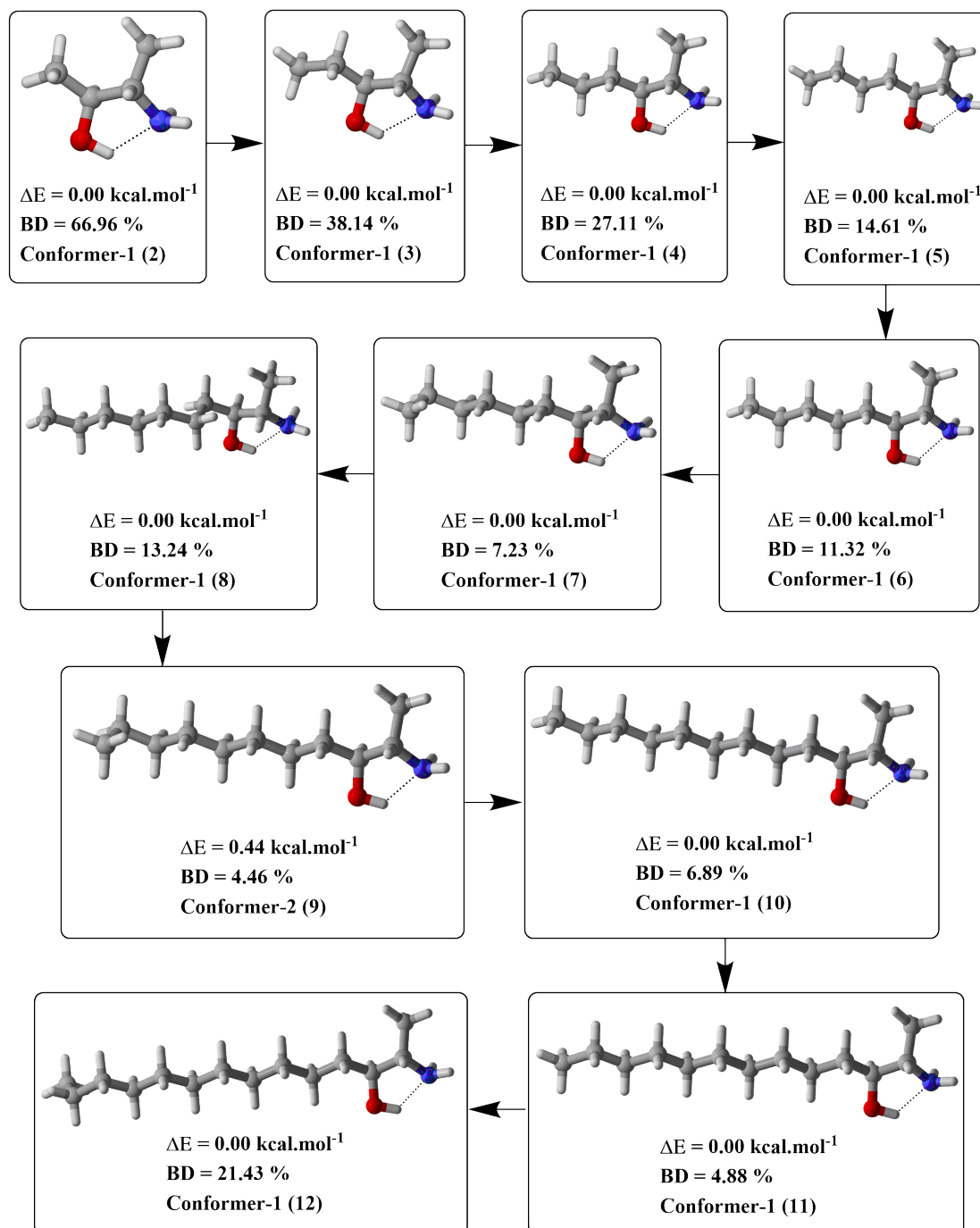
Scheme 3.3: An analysis of the 13th conformer of **2** and its descendants along with their propagation in subsequent steps. ΔE is the relative energy of the conformer with respect to the lowest energy conformer in the series. **BD** is the Boltzmann percentage distribution of the conformer in its total ensemble. Conformer number at the bottom of each structure shows its rank in its conformational ensemble while the number in parentheses shows the homologue whose conformer it is.

intensified while analysing the descendants of the conformers of **3** in **4**. Hence, the first descendant of those 11 “high-energy” conformers of **2** becomes only the 55th lowest energy conformer of **4** (Scheme 3.3). Further going down, it becomes 88th lowest energy conformer in **5** and hence does not propagate further due to being too higher in energy to be propagated to **6**.

By looking at the first lowest energy conformer (**2a**, 66.96% Boltzmann percentage distribution) of twelve “low energy” conformers, and its propagation into subsequent steps, it is verified that most low energy conformers of one homologue are also low in energy in subsequent steps unlike the “higher energy” conformers which become higher and higher in energy and eventually get excluded sooner or later during the incremental molecular construction. This is clarified in Scheme 3.4 where the lowest energy descendants of the lowest energy conformer (**2a**) are also the lowest energy conformers overall in all steps. The exception to this is step nine, where the descendant of **2a** is the second lowest in energy.

A further detailed analysis of the conformer ranks during propagation into their subsequent homologues was performed for all 23 conformers of **2** and is presented in Table 3.6. This analysis provided detailed insights and interesting facts about conformational analysis through *MICE-PES*. This analysis proved that the rank of a conformer with relatively high energy and less percentage contribution to the total Boltzmann ensemble becomes higher and higher while propagating in subsequent steps/homologues which leads to the systematic exclusion of that conformer at a later stage during the analysis. There are, of course, some exceptions to this trend at some individual steps (*e.g.* rank of **2b** (11.97% Boltzmann percentage distribution) in its homologues **8**, **9**, **10**, and onwards) but still those exceptions did not change the overall trend and follow it at later stages. The other interesting fact unveiled through this analysis was that all the 21 conformers of **2** except the first two (**2a** and **2b**) were excluded before or at homologue **10** and did not contribute to the conformational space of 3-*epi*-xestoaminol C (**12**). Only the first two conformers of **2** (**2a** and **2b**) and their descendants dominated

the conformational space of **12**. This might be attributed to the stabilising intramolecular hydrogen bonding in those conformers (**2a** and **2b**) that was preserved, particularly in a solvent like CH₂Cl₂ which does not interfere with these interactions.



Scheme 3.4: An analysis of the lowest energy conformer of **2** and its descendants along with their propagation in subsequent steps. ΔE is the relative energy of the conformer with respect to the lowest energy conformer in the series. **BD** is the Boltzmann percentage distribution of the conformer in its total ensemble. Conformer number at the bottom of each structure shows its rank in its conformational ensemble while the number in parentheses shows the homologue whose conformer it is.

Table 3.6: A detailed analysis of all 23 conformers of **2** and their rank in subsequent descendants. First column shows conformers of **2** with their Boltzmann percentage distribution in parentheses. All the other columns show the ranks of the previous conformers' first descendant in current homologue. The numbers in parentheses show the Boltzmann percentage distribution of the respective conformer. A '-' indicates that the conformers has been eliminated from being propagated further due to being higher in energy than the selection cut-off in the previous step. Note that in each row the rank of the lowest energy descendant is shown. There are usually three descendants but the other two are not shown here.

Conf. of 2	Rank in 3	Rank in 4	Rank in 5	Rank in 6	Rank in 7	Rank in 8	Rank in 9	Rank in 10	Rank in 11	Rank in 12
2a (66.96)	1 (38.14)	1 (27.11)	1 (14.61)	1 (11.32)	1 (7.23)	1 (13.23)	2 (4.46)	1 (6.89)	1 (4.88)	1 (21.43)
2b (11.97)	2 (17.44)	3 (10.34)	3 (10.86)	3 (5.21)	3 (6.34)	5 (3.25)	1 (9.31)	7 (2.22)	8 (2.47)	8 (1.85)
2c (3.92)	6 (1.65)	10 (1.23)	27 (0.61)	35 (0.56)	51 (0.40)	73 (0.27)	96 (0.23)	315 (0.00)	-	-
2d (2.76)	10 (1.38)	13 (1.00)	20 (0.92)	52 (0.32)	77 (0.22)	138 (0.10)	-	-	-	-
2e (2.27)	8 (1.61)	20 (0.63)	33 (0.50)	34 (0.59)	74 (0.24)	85 (0.23)	137 (0.16)	-	-	-
2f (1.89)	21 (0.43)	16 (0.88)	21 (0.89)	60 (0.26)	81 (0.19)	144 (0.09)	-	-	-	-
2g (1.87)	13 (1.10)	14 (0.98)	28 (0.59)	45 (0.42)	57 (0.35)	102 (0.18)	-	-	-	-
2h (1.55)	9 (1.56)	8 (1.54)	16 (1.39)	6 (3.08)	29 (0.78)	50 (0.42)	115 (0.18)	-	-	-
2i (1.19)	5 (2.06)	11 (1.11)	11 (2.26)	39 (0.49)	12 (1.91)	84 (0.23)	19 (1.01)	108 (0.20)	-	-
2j (0.98)	11 (1.27)	22 (0.59)	22 (0.79)	57 (0.30)	42 (0.51)	111 (0.17)	-	-	-	-
2k (0.79)	16 (0.76)	15 (0.90)	23 (0.69)	21 (1.11)	63 (0.31)	89 (0.22)	204 (0.06)	-	-	-
2l (0.57)	15 (0.84)	25 (0.50)	32 (0.53)	85 (0.15)	-	-	-	-	-	-
2m (0.49)	33 (0.20)	50 (0.11)	88 (0.10)	-	-	-	-	-	-	-
2n (0.49)	25 (0.31)	34 (0.35)	41 (0.29)	102 (0.13)	-	-	-	-	-	-
2o (0.38)	32 (0.22)	42 (0.25)	79 (0.12)	202 (0.02)	-	-	-	-	-	-
2p (0.37)	27 (0.28)	40 (0.27)	77 (0.12)	112 (0.11)	-	-	-	-	-	-
2q (0.37)	24 (0.35)	38 (0.29)	76 (0.12)	179 (0.04)	-	-	-	-	-	-
2r (0.34)	23 (0.37)	39 (0.29)	69 (0.14)	44 (0.42)	85 (0.18)	-	-	-	-	-
2s (0.26)	30 (0.24)	44 (0.22)	83 (0.11)	-	-	-	-	-	-	-
2t (0.25)	36 (0.11)	49 (0.17)	81 (0.11)	-	-	-	-	-	-	-
2u (0.21)	35 (0.16)	54 (0.12)	98 (0.09)	-	-	-	-	-	-	-
2v (0.12)	34 (0.16)	59 (0.10)	97 (0.09)	-	-	-	-	-	-	-
2w (0.03)	44 (0.02)	106 (0.01)	-	-	-	-	-	-	-	-

3.4.2.1 Analysis of the Percentile Ranks of Conformers

For further clarification of this scenario, an analysis of the percentile ranks (in terms of the relative energy) of the conformers of **3** was performed to see their reflection in the percentile ranks of their descendants in **4**. It was found that there is an 83% value for the correlation between the ranking of a precursor structure and the ranking of its descendants. More importantly, once all conformers whose percentile rank in **4** has increased by at least 10 percentage points with respect to their antecedent's percentile rank (i.e. those whose relative energy is markedly higher in **4** and are hence of lower importance for the mixture of conformers) have been removed, the correlation improves to 95% (Figure 3.8). This demonstrates that the descendants of high-energy conformers become increasingly higher and higher in energy themselves and are thus of less and less importance to the observed mixture of conformations. It is important, however, to note that while the correlations found are high, there are nevertheless several exceptions to this trend that can affect the reliability of the data obtained from the truncated set of conformers. This means there is no guarantee that no important conformers are discarded, but merely that the likelihood of the discarded conformers giving rise to important descendants is relatively low. This is demonstrated by the detailed analysis given below which documents the fate of the descendants of the low and high energy conformers. It has been shown that the discarded conformers had very few descendants of low energy conformers and their contribution was never more than 0.4% to the ensemble. It is of course possible to increase the threshold that is applied to the truncation if a higher accuracy (at the expense of a larger computational cost) is desired.

After that, the fate of the top ten "highest-energy" conformers of **4-11** was analysed in their subsequent step, i.e. in their propagation to **5-12**. It was found that most of the high-energy conformers when propagated to the next step, resulted in correspondingly high-energy conformers. Actually, out of the 30 conformers of **5** which descended from the selected "10 highest-energy conformers" of **4**, there were only two conformers which had low enough energy that they could be propagated in the next step but their

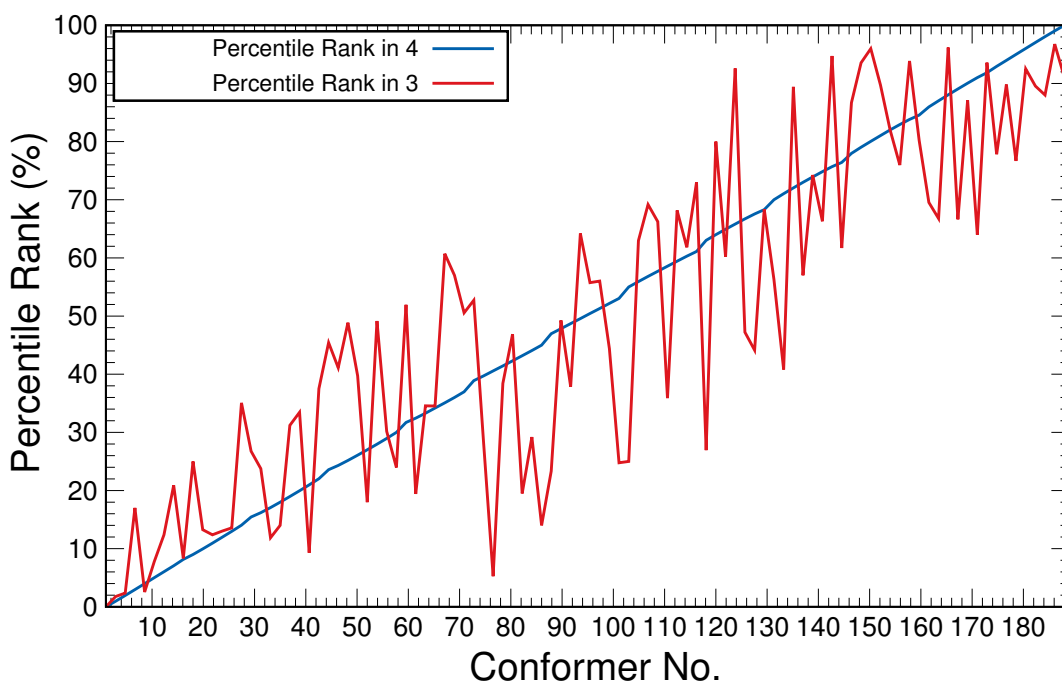


Figure 3.8: Correlation analysis for all conformers of **4** whose antecedents do not rise in energy more than 10 percentage points. There is 95% correlation between the two. It is clear from the graph that low energy conformers of **4** descended from the low energy conformations of **3** most of the time. There are also many occurrences of low energy conformers of **3** becoming high in energy in **4**.

contributions were only 0.11% and 0.12% to the ensemble, respectively. To summarize, at every step of the propagation, we never observed more than four out of the thirty descending conformers of the selected “10 highest-energy conformers” that were low enough in energy to be propagated to the next step and none had a contribution to the overall ensemble of more than 0.4%. The important thing to note here was that all the conformers of structures **10**, **11**, and **12** were descendants of the conformers **2a** and **2b** (Figure 3.7). This implies that the descending structures of these conformers (**2a** and **2b**) without the stabilizing intramolecular hydrogen bond were very high in energy. In summary, the sheer number of relatively low-energy descendants of **2a** and **2b** eliminated the other conformers and their descendants after the 8th stage of the algorithm developed.

By comparison, a full systematic conformational search of **12** by scanning the angles in steps of 15 degrees, would have generated 177,147 conformers but with *MICE-PES*,

only 339 optimised conformers were reached. Using *MICE-PES* around 2500 total optimisations were performed from the first step until compound **12** was reached and most of them were optimisations on very small molecules (Table 3.5). In contrast for a full systematic search on the same level of theory, 177,147 optimisations on conformers of **12** would have taken about 626 years using the same resources (12 cores with 24 GB memory for an optimisation of 3-*epi*-xestoaminol C (**12**)). To put it into context, *MICE-PES* took around four months using the VUW high performance computing (HPC) facility.

3.4.3 Validation of the Results with the Experiment

In order to test the validity of the conformational space covered, the optical rotation and NMR data of **12** were compared to the theoretically calculated data.

3.4.3.1 Comparison of Optical Rotation

The theoretically calculated Boltzmann averaged value of the individual DFT calculations of optical rotation of 339 conformers of **12** was $-20.0 \text{ deg dm}^{-1}\text{cm}^3\text{g}^{-1}$ which

Table 3.7: First six lowest energy conformers of **12** with their relative energies and other details. *Conformer* is the name of the conformer from the last 339 conformers of **12**. ΔE is the relative energy of each conformer with respect to the lowest energy conformer. $[\alpha]_D$ is the optical rotation of the conformer computed in CH_2Cl_2 . **B** is the Boltzmann percentage distribution of the conformer at 298.15 K. Weighted $[\alpha]_D$ is the Boltzmann weighted optical rotation of the conformer based on all the contributing conformers.

<i>Conformer</i>	ΔE	$[\alpha]_D$	B	Weighted $[\alpha]_D$
<i>xa-conf-01</i>	0.00	-51.5	21.43	-11.1
<i>xa-conf-02</i>	0.81	+35.8	5.42	+1.9
<i>xa-conf-03</i>	0.89	-15.9	4.78	-0.8
<i>xa-conf-04</i>	0.92	+43.8	4.55	+2.0
<i>xa-conf-05</i>	1.09	+37.6	3.38	+1.3
<i>xa-conf-06</i>	1.19	-55.2	2.86	-1.6
Total	-	-5.4	42.42	-5.1

was in very good agreement with the experimentally determined value of $[\alpha]_D = -25.3$ deg dm⁻¹cm³g⁻¹ (c 0.76 mg/mL, in CH₂Cl₂ at 26.2°C, variance: 5.54). It shows that the important conformational space was covered by the calculations. Table 3.7 shows the first six lowest energy conformers contributing 42.42% Boltzmann percentage distribution to the total ensemble of 339 conformers for **12**, with some of their properties.

The overall optical rotation of **12** in CH₂Cl₂ was calculated as an additional piece of evidence for the validity of the approach. The experimental report of the optical rotation value (CHCl₃) of **12** was -6.19 deg dm⁻¹cm³g⁻¹ which is still in very close agreement to the theoretically calculated value of -13.89 deg dm⁻¹cm³g⁻¹ on the same set of 339 conformers obtained previously for **12**.²⁵⁶ This is apart from the fact that the calculation of the optical rotation in CHCl₃ was performed on the ensemble of conformers originally optimized in CH₂Cl₂. It means that it could yield even better agreement once all the conformers were reoptimized with CHCl₃ as a solvent.

3.4.3.2 Comparison of NMR Data

The Boltzmann averaged ¹H- and ¹³C-NMR data for the 339 conformers of **12** were calculated and compared to those observed experimentally as shown in Table 3.8. It is obvious that the agreement between the experimental and the theoretical values is very good. Upon closer inspection, mean absolute errors (MAE) of only 0.24 ppm for ¹H-NMR and 1.00 ppm for ¹³C-NMR chemical shifts were obtained. This also supports the validity of *MICE-PES* to map the potential energy surface and the important conformational space.

MICE-PES worked nicely for 3-*epi*-xestoaminol C and gave good comparison between experimental and calculated data but it was a fairly simple molecule as a test. To really test the algorithm a more complicated example involving other factors influencing the geometry like remote intramolecular hydrogen bonding was needed. Next chapter presents meroterphenol C (**26**) as a complex molecule that has different functional

groups and the remote intramolecular interactions which lock the conformation to a rather narrower space, thus twisting the geometry.

Table 3.8: Comparison of Experimental and Theoretical NMR data for **12**. C. No. represents the carbon number while δ_{exp} and δ_{calc} represent the experimental and theoretical chemical shifts (based on Boltzmann average of all the conformers), respectively. $\Delta\delta$ is the difference between the experimental and calculated chemical shift at the PBE0-D3BJ/def2-TZVP/SMD_{CH₂Cl₂} level. MAE is the mean absolute error and RMSE is the root mean square error calculated for the chemical shift difference ($\Delta\delta$) from experimental values.

C. No.	¹³ C-NMR			¹ H-NMR		
	δ_{exp}	δ_{calc}	$\Delta\delta$	δ_{exp}	δ_{calc}	$\Delta\delta$
1	16.1	18.9	2.8	1.27	0.90	−0.37
2	53.5	53.6	0.1	3.09	2.21	−0.88
3	73.2	75.3	2.1	3.44	2.80	−0.64
4	34.7	34.9	0.2	1.40	1.47	−0.07
				1.56	0.90	−0.66
5	26.4	27.9	1.5	1.39	1.57	+0.18
				1.54	1.01	−0.53
6	30.8	31.4	0.6	1.30	1.18	−0.12
				1.30	1.07	−0.23
7	30.7	31.7	1.0	1.30	1.19	−0.11
				1.30	1.15	−0.15
8	30.6	31.5	0.9	1.30	1.14	−0.16
				1.30	1.17	−0.13
9	30.5	31.5	1.0	1.30	1.15	−0.15
				1.30	1.14	−0.16
10	30.5	31.6	1.1	1.30	1.15	−0.15
				1.30	1.11	−0.19
11	30.5	30.4	−0.1	1.30	1.24	−0.06
				1.30	1.04	−0.26
12	33.1	32.7	−0.4	1.29	1.18	−0.11
				1.29	1.11	−0.18
13	23.8	23.3	−0.4	1.31	1.19	−0.12
				1.31	1.15	−0.16
14	14.5	11.8	−2.7	0.91	0.77	−0.14
-	MAE		1.00	MAE		0.25
-	RMSE		1.34	RMSE		0.32

3.5 Conformational Analysis of Meroterphenol C using *MICE-PES*

Meroterphenol C (**26**, Figure 3.9) is a meroditerpenoid with phenolic substituent. It belongs to the plastoquinone class of compounds and acts as a peroxisome proliferator-activated receptor γ -agonist. It was isolated by Kim and co-workers from the algae *Sargassum yezoense* in Korea.²⁵⁷ It has shown very good type-2 antidiabetic potential *in vitro*. Due to its important biological activity, this compound was an interesting example to study its conformational analysis.

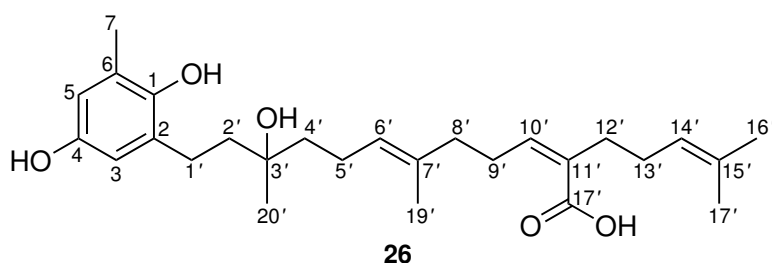
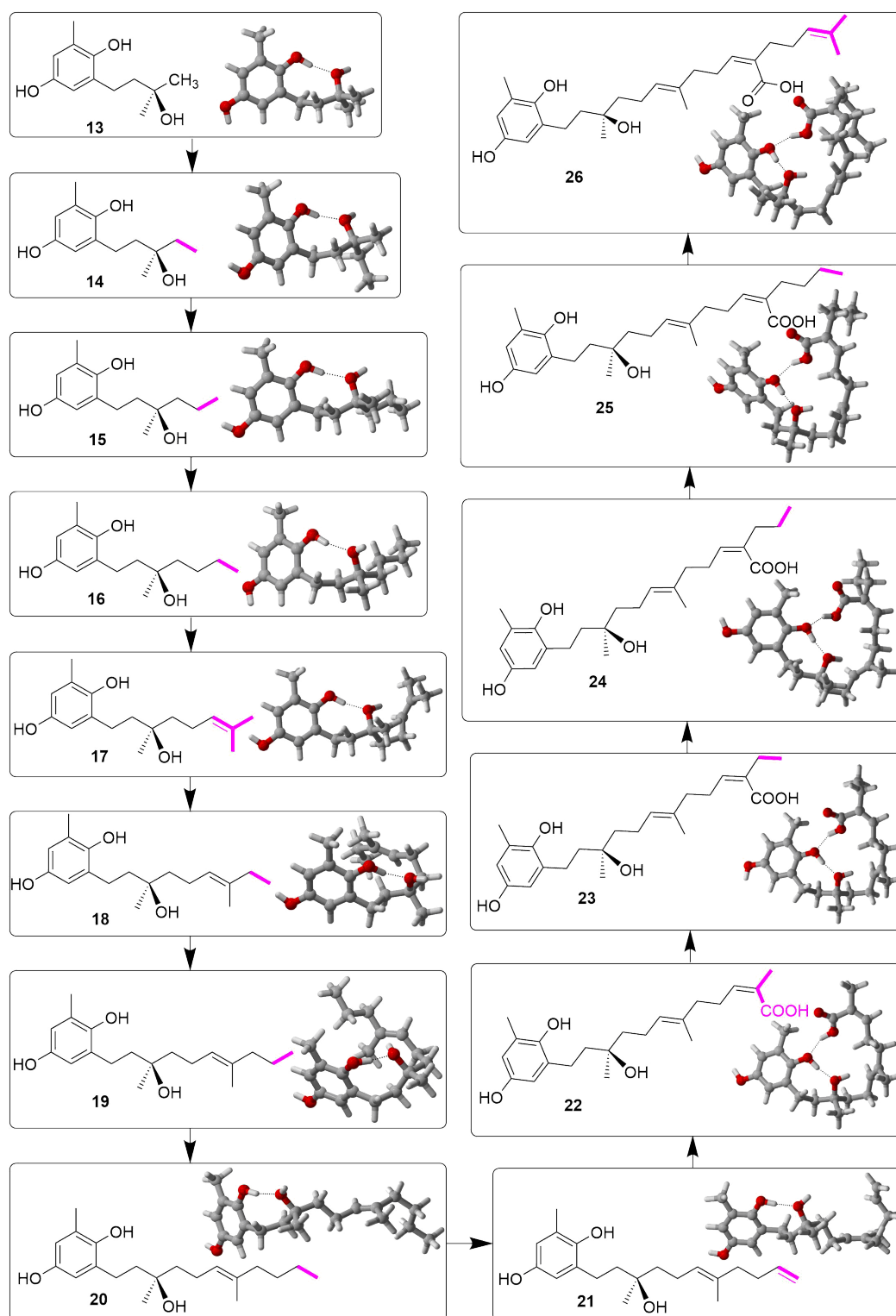


Figure 3.9: Naturally occurring meroterphenol C isolated from the algae *Sargassum yezoense*.

It's structure was determined using mass spectrometry, ^1H -NMR, ^{13}C -NMR, and 2D-NMR spectroscopic techniques although it's absolute configuration has not been established so it is one of two possible enantiomers, because of its single chiral centre. **26** is an interesting case to study with *MICE-PES* because it would not only validate *MICE-PES* but also confirm the absolute configuration of the chiral centre in the molecule. Therefore this molecule was selected from the literature to study with *MICE-PES* because its experimental data was available and there is also the possibility of intra-molecular hydrogen bonding. This is because for long chain systems like **12**, the energetics of the conformational space are dominated by the steric repulsion due to which the alkyl chain cannot interact with other parts of the molecule in a stabilising fashion. Meroterphenol (**26**) is different and has a long chain with other functional groups like phenolic, hydroxy, and carboxylate groups which can interact with the other distant parts of the molecule in a stabilising manner, thus giving rise to different

geometries of the compound (structures **22**, **23**, **24** etc. in Scheme 3.5).



Scheme 3.5: Schematic representation of the conformational analysis of meroterphenol C (**26**) using *MICE-PES*. Structures **13-25** are the homologues of **26**. Note that **26** is one of the two possible enantiomers of the naturally occurring meroterphenol C. The bold groups in magenta colour are the ones which are being changed at that step. The lowest energy conformer of each homologue is shown on the right in 3D.

Beginning with the basic methodology of *MICE-PES*, the smallest possible homologue of **26** with the important phenolic chromophore (**13**) was selected. The PES of **13** was mapped by performing a sequence of relaxed potential energy scans on its nine different conformers by rotating all dihedral angles of **13** in steps of 15° each (Figure 3.10). The details of the method were described previously in Section 3.4.1.

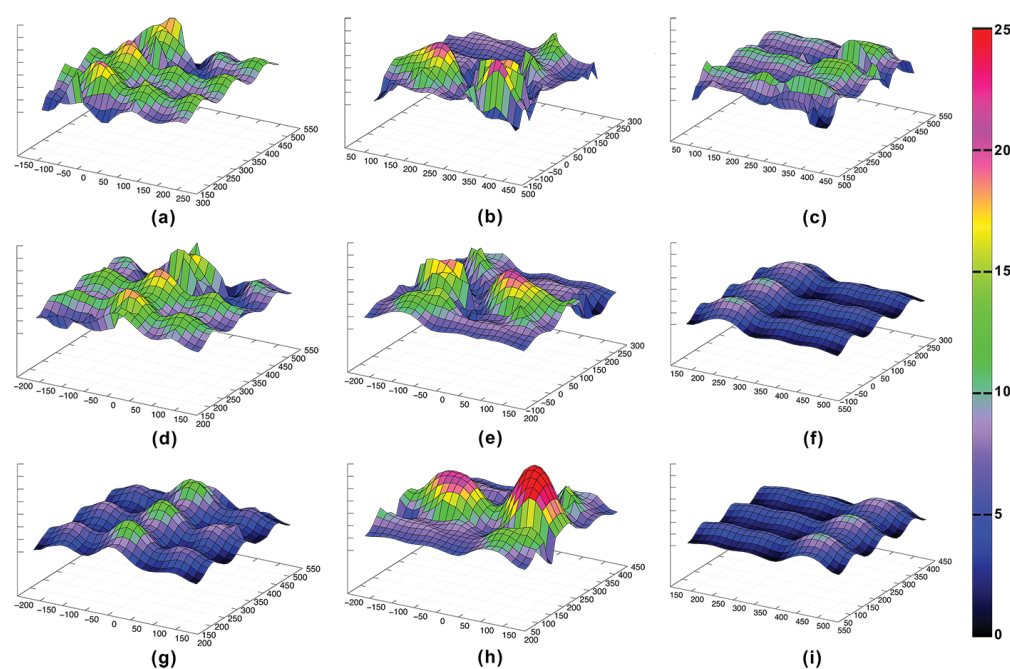


Figure 3.10: Potential energy surfaces at the PBE0-D3BJ/def2-TZVP/SMD_{CH₃CN} level of theory that identify all 28 conformers of **13**. Y-axis represents the relative energies of the conformers. X- and Z-axis show the rotational angles of the starting structure. The details of these angles and rotations can be found in the Appendix A.2. Optimisations were carried out for a total of 5625 individual structures. The numbers on the energy scale on the right are in kcal/mol. **a–i** are the relaxed scans on different conformations of structure **13** to ensure that the PES has been covered well.

These resulting energy landscapes were then examined very carefully to select 28 distinct lowest energy conformers (Figure 3.11) which were then confirmed as minima on the PES by unconstrained relaxation. The relative Gibbs free energies of these 28 minimum energy conformers are given in Table 3.9.

The difference in relative ΔG values between the global minimum and the structure with the highest energy was only 3.99 kcal/mol. The interesting thing to note here is that there are several structures with stabilising intra-molecular hydrogen bonding interactions among the 28 lowest energy conformers that were propagated to the next

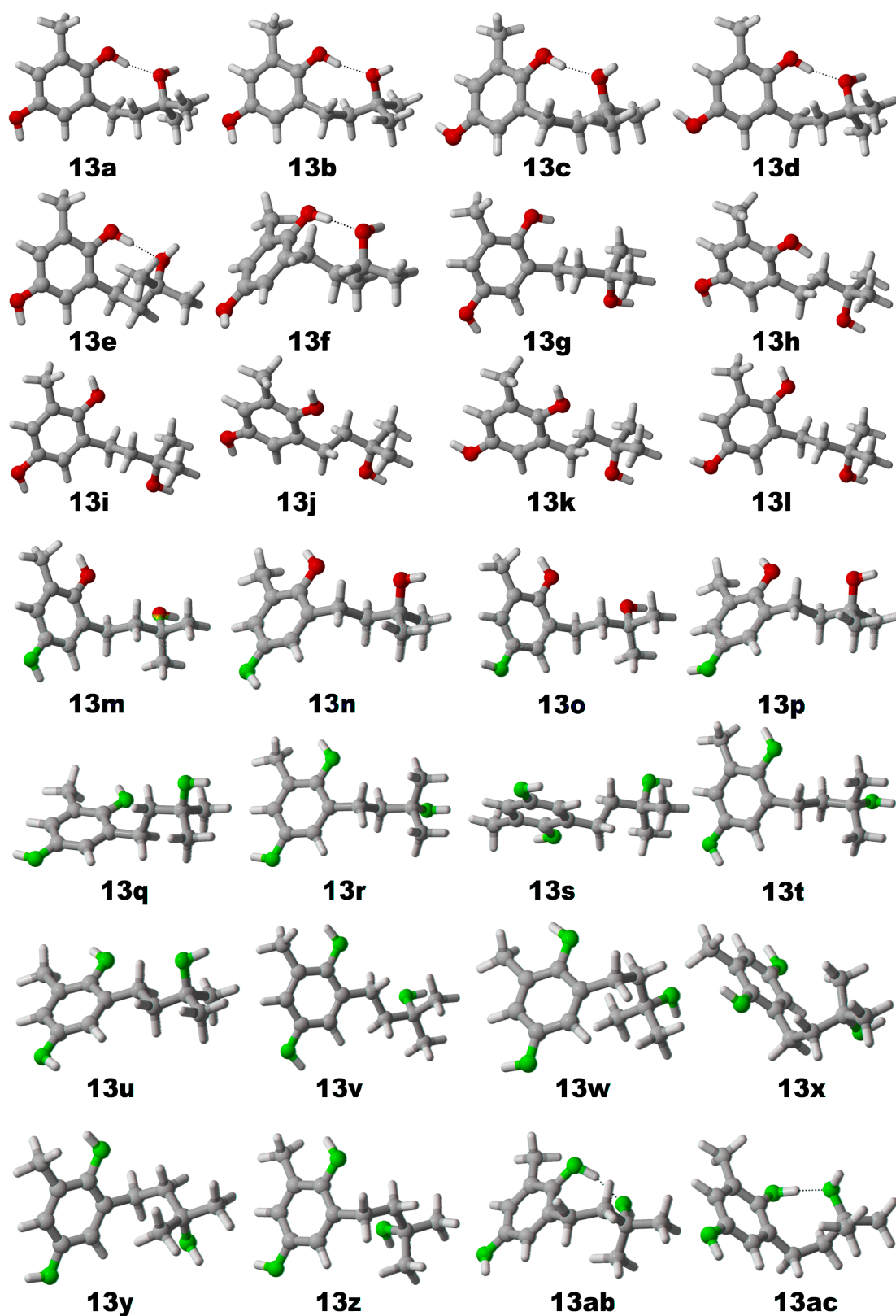


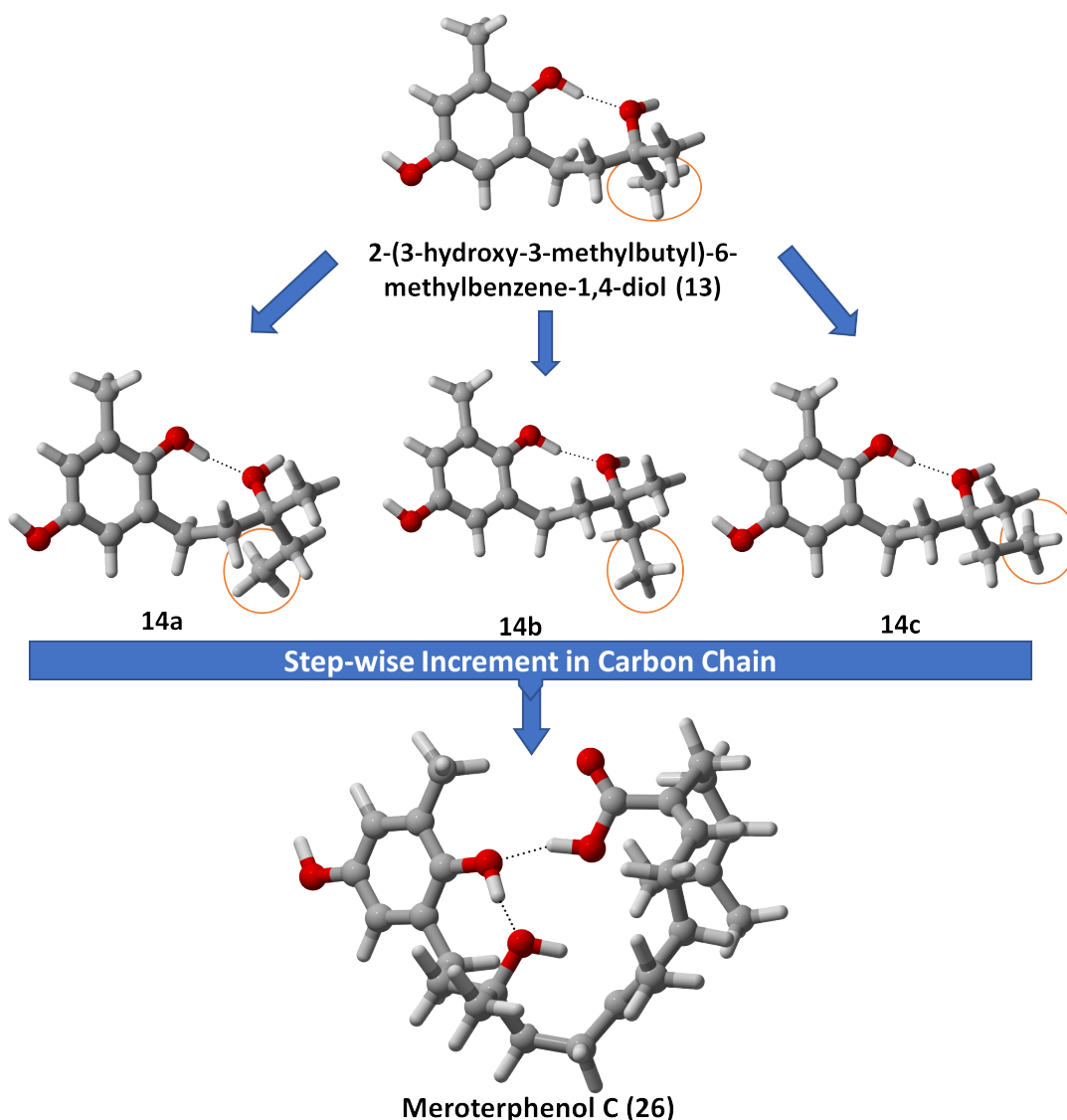
Figure 3.11: The 28 lowest energy conformers of **13** optimized at the PBE0-D3BJ/def2-TZVP/SMD_{CH₃CN} level of theory. Note that structures **13a**, **13b**, **13c**, **13d**, **13e**, **13f**, **13ab**, and **13ac** have intra-molecular hydrogen bonding.

Table 3.9: The relative Gibbs free energies (ΔG in kcal/mol) and Boltzmann percentage distributions of the 28 minimum energy conformers of **13** at the PBE0-D3BJ/def2-TZVP/SMD_{CH₃CN} level of theory. Note that most of the conformers with identical energies are enantiomeric pairs.

Conformer	Relative ΔG	Boltzmann Distribution (%)	Conformer	Relative ΔG	Boltzmann Distribution (%)
13a	0.00	17.47	13o	1.78	0.87
13b	0.00	17.47	13p	1.78	0.87
13c	0.06	15.83	13q	1.84	0.78
13d	0.06	15.83	13r	1.84	0.78
13e	0.71	5.27	13s	1.88	0.73
13f	0.71	5.27	13t	1.88	0.73
13g	1.37	1.74	13u	2.26	0.67
13h	1.37	1.74	13v	2.26	0.67
13i	1.47	1.45	13w	3.35	0.38
13j	1.47	1.45	13x	3.35	0.38
13k	1.52	1.34	13y	3.63	0.06
13l	1.52	1.34	13z	3.63	0.06
13m	1.74	0.93	13ab	3.99	0.04
13n	1.74	0.93	13ac	3.99	0.04

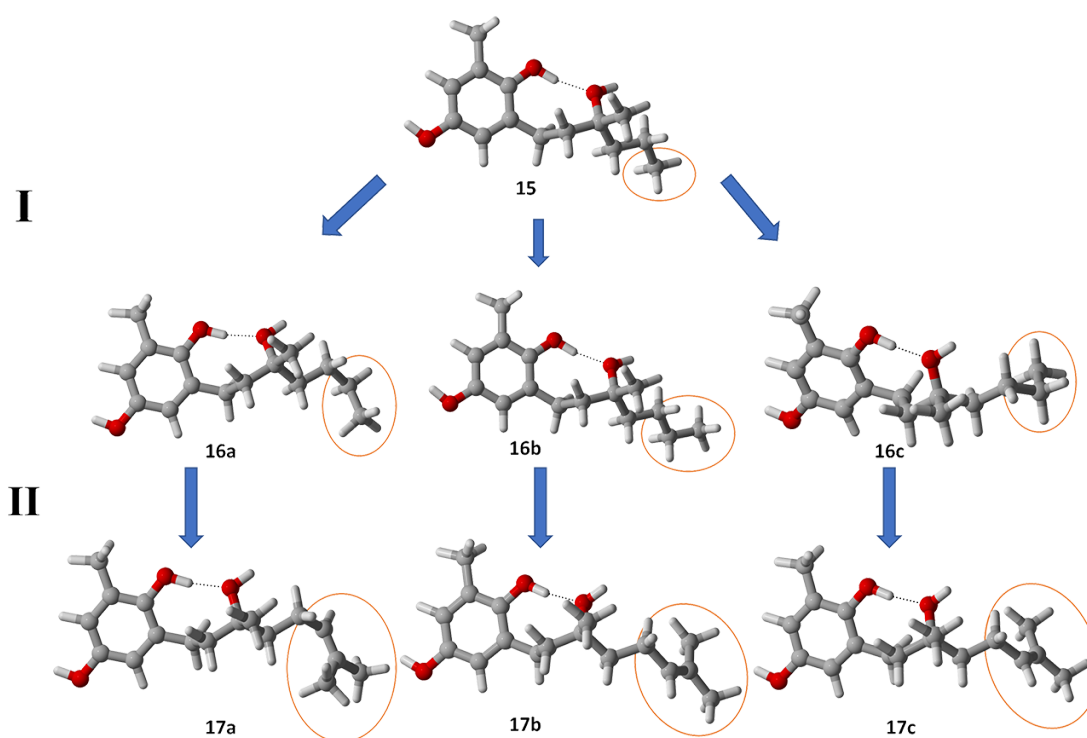
incremental molecular build-up steps. This was a good starting point to observe if these conformers remained low energy ones in the next steps and if the intra-molecular interactions would persist throughout the conformational search or not. After the identification of these 28 conformers, the incremental molecular build up was started. All these conformers of **13** were propagated into the next step (structure **14**) by the addition of a methyl group on the terminal hydrogens one by one (Scheme 3.6). This yielded 84 conformers for structure **14** which were optimized at the PBE0-D3BJ/def2-TZVP/SMD_{CH₃CN} level of theory and yielded total 83 conformers after removal of duplicates.

The same *MICE-PES* methodology as applied to 3-*epi*-xestoaminol C was used to construct the conformers of all the other homologues **15-26**. The addition of methyl groups at each stage was straightforward but for double bonds it is not enough to add a double bond directly on a group because it restricts the conformational preference to a narrow space. To overcome this, the following procedure was developed: first a methyl group was added (which becomes C1 of the double bond later on) on each of the three



Scheme 3.6: Construction of conformers through step-wise lengthening of the molecule by the addition of substituents using *MICE-PES*. The lowest energy isomer of 3'*R*-meroterphenol C (**26**) is shown at the bottom.

hydrogens of the previous methyl group followed by the optimisation of resultant structures. In the next step, another methyl group is subsequently added (which becomes C2 of the double bond later on) to each of the previously added methyl group hydrogens, one by one, following an optimisation. This ensures the complete exploration of the conformational space in the region of the double bond. After this, the hydrogens on the last two added methyls are deleted followed by decreasing the C–C distance. Then, the hydrogens are added to both of these carbons around the double bond followed by the optimization (Scheme 3.7). To reduce the number of generated structures at each



Scheme 3.7: Addition of double bond in conformers construction using *MICE-PES*. A methyl group is added on each of the three hydrogens of the previous methyl. In step-I, another methyl group is added on each of the three hydrogens of the existing methyl. In step-II the C–C bond length is shortened for a double bond, hydrogens are deleted on both carbons, and then hydrogens and other groups are added again to the terminal carbon.

step and the related computational cost and human effort, all conformers were propagated (after duplicate removal) for **13** but after that only these conformers contributing $\geq 90.0\%$ of the Boltzmann distribution to the ensemble were propagated into subsequent steps. One important thing to note here is that all the stereoisomers were considered including the enantiomeric pairs without being discarded because all the stereoisomers are important to consider for chemical properties and with molecules having more than one chiral centre, enantiomeric pairs can lead to different diastereomers at later stages after the addition of other atoms/substituents (not in the case of **26** because it has only one chiral centre). Hence for meroterphenol C (**26**), a total of 77 conformers were obtained and optimised out of which only the lowest energy 39 conformers contribute to $>95\%$ of the total partition function (Table 3.10).

If a full systematic conformational search of **26** would have been performed by scanning

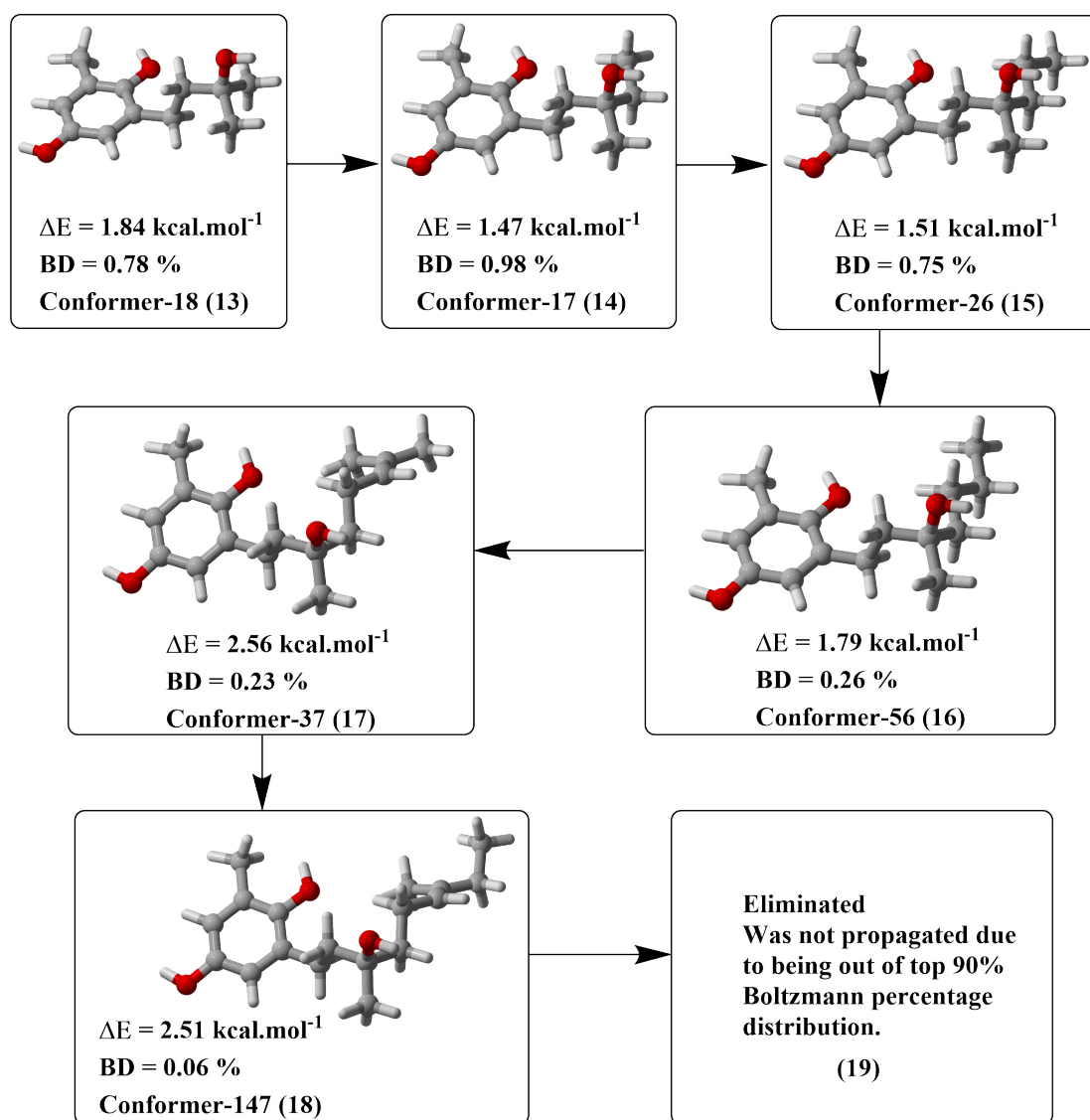
Table 3.10: \mathbf{n}^{conf} is the total number of conformers generated in a step. \mathbf{n}^{contr} is the total number of distinct conformers considered after duplicates removal (fully optimised at the PBE0-D3BJ/def2-TZVP/SMD_{CH₃CN} level of theory). \mathbf{n}^{imp} is the number of important conformers (*i.e.* the number of conformers that have to be included to account for $\geq 90.0\%$ of the Boltzmann distribution at 298.15 K). \mathbf{B} is the Boltzmann percentage of all the \mathbf{n}^{imp} conformers. \mathbf{c}_0 is the percentage contribution of the lowest lying conformer at 298.15 K. \mathbf{P} is the total value of the partition function at 298.15 K. $\Delta\mathbf{G}$ is the difference in Gibbs free energy between the energetically lowest conformer and the energetically highest lying conformer (of all \mathbf{n}^{contr} conformers; in kcal/mol). $[\alpha]_D$ is the predicted value of the optical rotation (based on all \mathbf{n}^{contr} conformers).

Molecule	\mathbf{n}^{conf}	\mathbf{n}^{contr}	\mathbf{n}^{imp}	\mathbf{B}	\mathbf{c}_0	\mathbf{P}	$\Delta\mathbf{G}$	$[\alpha]_D$
13	28	28	28	100.0	17.46	5.73	3.99	+1.4
14	84	83	55	99.22	11.70	8.50	7.15	+3.9
15	165	162	76	98.79	9.54	10.48	5.38	-0.1
16	228	217	131	98.01	5.35	18.69	7.15	-0.3
17	131	108	70	98.09	17.69	5.65	5.48	+5.2
18	210	199	100	92.56	3.91	25.60	4.63	+5.3
19	300	294	130	92.31	7.39	13.53	6.35	+19.4
20	390	382	196	90.50	3.92	25.53	5.68	+28.3
21	196	159	110	95.40	14.57	6.87	4.01	+90.2
22	110	104	85	99.13	17.15	5.83	4.77	+80.7
23	255	246	80	95.70	25.87	3.87	7.28	+77.9
24	238	236	71	95.39	10.64	9.40	6.49	+37.1
25	213	197	77	95.12	18.92	5.29	9.32	-9.4
26	77	71	54	99.04	24.63	4.06	5.05	-22.4

the angles in steps of 15 degrees, it would have generated 3.65×10^{16} conformers but with *MICE-PES*, only 71 optimised conformers were obtained. Using *MICE-PES* 2625 total optimisations from the first step were performed till reaching compound **26** and most of them were optimisations on very small molecules (Table 3.10) and it took around seven months using the VUW HPC facility and performing all the calculations on a triple ζ quality basis set. In contrast for a full systematic search on the same level of theory, 3.65×10^{16} optimisations on **26** would have taken about 1.99×10^{15} years using the same resources (12 cores with 24 GB memory for an optimisation of meroterphenol C (**26**)).

3.5.1 Analysis of the ranks of Conformers

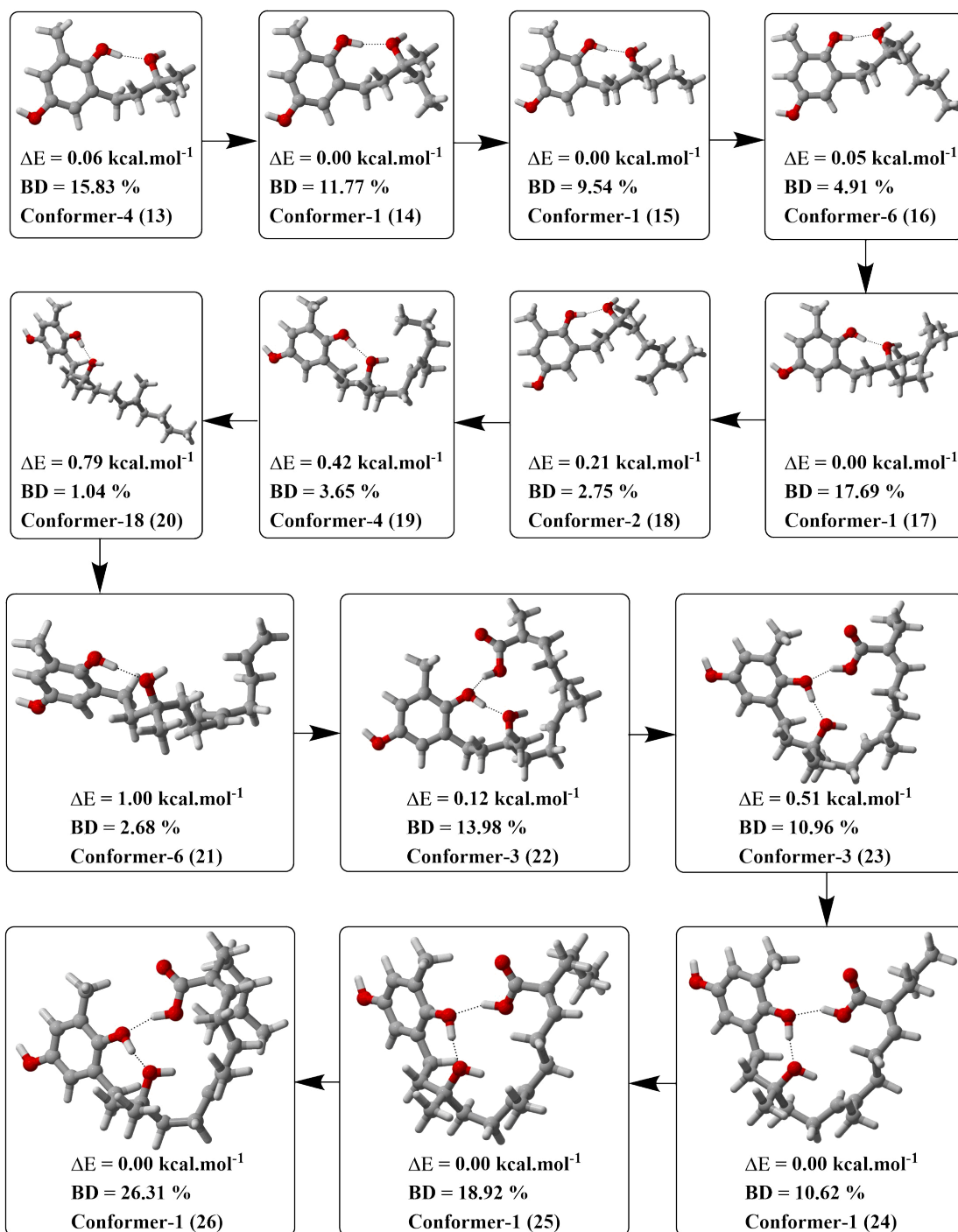
An analysis of the ranks of all the conformers of **13** and their descendants was performed and the relative energies and Boltzmann percentage distribution of the first descendant in each case were noted in Table 3.11. It was a continuation of the analysis performed previously for 3-*epi*-xestoaminol C (**12**) which was a fairly simple compound. The analysis for **13** showed similar trends that the high energy conformations in one



Scheme 3.8: An analysis of the 18th conformer of **13** and its descendants along with their propagation in subsequent steps. ΔE is the relative energy of the conformer with respect to the lowest energy conformer in the series. **BD** is the Boltzmann percentage distribution of the conformer in its total ensemble. Conformer number at the bottom of each structure shows its rank in its conformational ensemble while the number in parentheses shows the homologue whose conformer it is.

compound do not become low energy in subsequent steps with very few exceptions that still do not change the overall trend. Interestingly, only first four conformers (**13a-13d**) of **13** could make it to the last homologue (**25**) of meroterphenol C and all the others (**13e-13ac**) were eliminated during the incremental construction of the conformational space of **26**. If the conformers of **13** are divided into eleven “high-energy” (making up 4.69% of the Boltzmann distribution) and seventeen “low-energy” (making up 95.31% of the Boltzmann distribution) conformers, it is clearly evident that the first conformer (**13r**) of the eleven “high-energy” conformers (Figure 3.11) became the 17th lowest energy conformer in **14** and contributed only 0.98% to the total ensemble. This effect was intensified in the next iteration going from **14** to **15**. Hence, the first descendant of those eleven “high-energy” conformers of **13** became only the 26th lowest energy conformer of **15** with 0.75% Boltzmann percentage contribution to the total ensemble (Scheme 3.8). Further going down, it became 56th lowest energy conformer of **16**, then 37th lowest energy conformer of **17**, and finally the 147th lowest energy conformer of **18** where it did not end up in top 90% contributing conformers and was eliminated from the ensemble.

By looking at the first four lowest energy conformers (**13a-13d**) of seventeen “low-energy” conformers, and their propagation into subsequent steps, it is again verified that mostly the lowest energy conformers of one homologue become lower in energy in subsequent steps. As an interesting fact, the 4th lowest energy conformer of **13** became the lowest energy conformer of meroterphenol C (**26**) instead of the 1st one. This can be explained as the first four lowest energy conformers of **13** have almost identical relative energies and Boltzmann percentage contributions to the total ensemble and it is not surprising that 4th ranked conformer became the lowest energy while reaching the last incremental step. This 4th lowest energy conformer of **13** and its propagation into subsequent steps has been shown in Scheme 3.9.



Scheme 3.9: An analysis of the 4th lowest energy conformer of **13** and its descendants along with their propagation in subsequent steps. ΔE is the relative energy of the conformer with respect to the lowest energy conformer in the series. **BD** is the Boltzmann percentage distribution of the conformer in its total ensemble. Conformer number at the bottom of each structure shows its rank in its conformational ensemble while the number in parentheses shows the homologue whose conformer it is.

Table 3.11: A detailed analysis of all 28 conformers of **13** and their rank in subsequent descendants. First column shows conformers of **13** with their Boltzmann percentage distribution in parentheses. All the other columns show the ranks of the previous conformers' first descendant in current homologue. The numbers in parentheses show the Boltzmann percentage distribution of the respective conformer. **R(n)** means Rank in **n**. A '-' indicates that the conformers has been eliminated from being propagated further due to being higher in energy than the selection cut-off in the previous step. Note that in each row the rank of the lowest energy descendant is shown. There are usually three descendants but the other two are not shown here.

Conf. of 13	R (14)	R (15)	R (16)	R (17)	R (18)	R (19)	R (20)	R (21)	R (22)	R (23)	R (24)	R (25)	R (26)
13a (17.47)	5 (6.98)	5 (6.36)	2 (5.35)	7 (4.69)	5 (2.26)	3 (4.30)	5 (2.01)	5 (2.76)	7 (2.18)	19 (0.51)	12 (2.78)	23 (0.88)	45 (0.17)
13b (17.47)	3 (7.10)	3 (7.35)	1 (5.35)	6 (4.69)	4 (2.26)	2 (4.30)	2 (2.04)	2 (14.14)	2 (16.67)	1 (25.87)	3 (6.55)	4 (3.95)	3 (6.38)
13c (15.83)	4 (7.03)	6 (5.72)	3 (5.27)	5 (5.06)	1 (3.91)	1 (7.39)	1 (3.92)	10 (1.84)	1 (17.15)	2 (20.34)	4 (6.37)	3 (5.11)	8 (2.85)
13d (15.83)	1 (11.77)	1 (9.54)	6 (4.91)	1 (17.69)	2 (2.75)	4 (3.65)	18 (1.04)	6 (2.68)	3 (13.98)	3 (10.96)	1 (10.62)	1 (18.92)	1 (26.31)
13e (5.27)	12 (1.78)	16 (1.10)	35 (0.72)	11 (3.10)	3 (2.40)	11 (1.47)	9 (1.33)	1 (14.57)	30 (0.48)	38 (0.18)	76 (0.12)	-	-
13f (5.27)	13 (1.57)	14 (1.69)	37 (0.67)	28 (0.63)	77 (0.40)	61 (0.49)	103 (0.28)	149 (0.05)	-	-	-	-	-
13g (1.74)	11 (2.32)	9 (4.40)	4 (5.17)	12 (2.75)	57 (0.60)	19 (0.98)	24 (0.97)	11 (1.83)	67 (0.14)	64 (0.10)	91 (0.06)	-	-
13h (1.74)	54 (0.16)	15 (1.17)	18 (1.44)	64 (0.10)	182 (0.02)	-	-	-	-	-	-	-	-
13i (1.45)	32 (0.42)	28 (0.67)	42 (0.51)	33 (0.33)	106 (0.19)	-	-	-	-	-	-	-	-
13j (1.45)	31 (0.43)	36 (0.47)	46 (0.39)	35 (0.24)	73 (0.43)	95 (0.30)	267 (0.06)	-	-	-	-	-	-
13k (1.34)	28 (0.44)	24 (0.80)	13 (2.11)	44 (0.18)	86 (0.31)	247 (0.02)	-	-	-	-	-	-	-
13l (1.34)	25 (0.47)	21 (0.91)	43 (0.49)	26 (0.79)	88 (0.30)	201 (0.04)	-	-	-	-	-	-	-
13m (0.93)	37 (0.39)	29 (0.66)	44 (0.43)	38 (0.23)	90 (0.27)	119 (0.20)	152 (0.18)	146 (0.06)	-	-	-	-	-
13n (0.93)	27 (0.47)	44 (0.34)	54 (0.27)	48 (0.15)	100 (0.20)	250 (0.01)	-	-	-	-	-	-	-
13o (0.87)	26 (0.47)	43 (0.34)	53 (0.27)	43 (0.19)	101 (0.20)	-	-	-	-	-	-	-	-
13p (0.87)	23 (0.48)	47 (0.31)	67 (0.22)	57 (0.12)	150 (0.05)	-	-	-	-	-	-	-	-
13q (0.78)	24 (0.48)	46 (0.31)	66 (0.22)	56 (0.12)	153 (0.05)	-	-	-	-	-	-	-	-
13r (0.78)	17 (0.98)	26 (0.75)	56 (0.26)	37 (0.23)	147 (0.06)	-	-	-	-	-	-	-	-
13s (0.73)	16 (0.98)	25 (0.75)	55 (0.26)	36 (0.23)	146 (0.06)	-	-	-	-	-	-	-	-
13t (0.73)	19 (0.98)	13 (2.93)	50 (0.33)	39 (0.23)	145 (0.06)	-	-	-	-	-	-	-	-
13u (0.67)	18 (0.98)	12 (2.93)	49 (0.33)	40 (0.23)	144 (0.06)	-	-	-	-	-	-	-	-
13v (0.67)	48 (0.23)	65 (0.16)	27 (0.93)	34 (0.31)	115 (0.15)	-	-	-	-	-	-	-	-
13w (0.38)	58 (0.11)	-	-	-	-	-	-	-	-	-	-	-	-
13x (0.38)	57 (0.11)	-	-	-	-	-	-	-	-	-	-	-	-
13y (0.06)	63 (0.03)	-	-	-	-	-	-	-	-	-	-	-	-
13z (0.06)	67 (0.01)	-	-	-	-	-	-	-	-	-	-	-	-
13ab (0.04)	65 (0.01)	-	-	-	-	-	-	-	-	-	-	-	-
13ac (0.04)	68 (0.01)	-	-	-	-	-	-	-	-	-	-	-	-

3.5.2 Validation of the Computed Results with Experimental Data

In order to test the validity of the conformational space covered, the optical rotation and NMR data from the literature was compared to the theoretically calculated data.

3.5.2.1 Comparison of Optical Rotation

Optical rotation is a very sensitive property to compute. Very minute changes in geometry change its value. While comparing the computed optical rotation to the experimentally determined one, the most important observation is the same sign of rotation. The magnitude of the value usually appears in a fair agreement which is considered a good agreement. There are examples in literature showing the comparisons as good agreements where the magnitude of the value differs in order of tens of magnitude.^{258–261} The theoretically calculated Boltzmann averaged value of the optical rotation of 71 conformers of **26** was $-22.4 \text{ deg dm}^{-1}\text{cm}^3\text{g}^{-1}$ which was in good agreement with the experimentally found value of $[\alpha]_D = +6.0 \text{ deg dm}^{-1}\text{cm}^3\text{g}^{-1}$ (c 0.035 mg/0.1 mL, in CH_3CN at 29.0°C). The opposite sign shows that the isolated compound is the other enantiomer than the computed one. It confirms that the structure of meroterphenol C determined by Kim and co-workers is 3'S-meroterphenol C as shown in Figure 3.12.

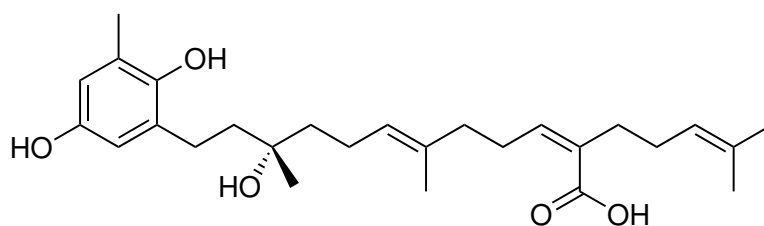


Figure 3.12: Absolute configuration of naturally occurring meroterphenol C isolated from the algae *Sargassum yezoense*.

3.5.2.2 Comparison of NMR Data

The Boltzmann averaged ^1H - and ^{13}C -NMR data for the 71 conformers of **26** was calculated and compared to the experimental NMR data as shown in Table 3.12.

Table 3.12: Comparison of Experimental and Theoretical NMR data for **26**. C. No. represent the carbon number, C. Type represents the type of carbon (*i.e.* quaternary (C), methine (CH), methylene (CH₂), and methyls (CH₃)), while δ_{exp} and δ_{calc} represent the experimental and theoretical chemical shifts (based on Boltzmann average of all the conformers), respectively. $\Delta\delta$ is the difference between the experimental and calculated chemical shift at the PBE0-D3BJ/def2-TZVP/SMD_{CH₃CN} level. MAE is the mean absolute error and RMSE is the root mean square error calculated for the chemical shift difference ($\Delta\delta$) from experimental values.

C. No.	C. Type	¹³ C-NMR			¹ H-NMR		
		δ_{exp}	δ_{calc}	$\Delta\delta$	δ_{exp}	δ_{calc}	$\Delta\delta$
1	C	145.8	146.0	−0.2			
2	C	130.6	132.5	−1.8			
3	CH	113.8	111.3	+2.5	6.42	5.57	+0.85
4	C	149.8	150.2	−0.4			
5	CH	114.8	112.5	+2.3	6.43	5.57	+0.86
6	C	125.8	128.5	−2.7			
7	CH ₃	16.0	15.5	+0.5	2.15	2.06	+0.09
1'	CH ₂	24.7	23.4	+1.3	2.58	2.34	+0.24
					2.58	1.90	+0.68
2'	CH ₂	41.7	42.3	−0.6	1.67	1.63	+0.04
					1.67	1.67	0.00
3'	C	72.3	75.0	−2.7			
4'	CH ₂	41.8	36.0	+5.8	1.49	1.16	+0.33
					1.49	1.68	−0.19
5'	CH ₂	22.6	23.3	−0.7	2.06	1.82	+0.24
					2.06	2.71	−0.65
6'	CH	125.5	129.6	−4.1	5.19	5.67	−0.48
7'	C	134.3	145.2	−10.9			
8'	CH ₂	38.9	35.9	+3.0	2.09	1.98	+0.11
					2.09	1.74	+0.35
9'	CH ₂	27.9	30.6	−2.7	2.53	1.64	+0.89
					2.53	2.15	+0.38
10'	CH	142.1	144.8	−2.7	5.91	6.07	−0.16
11'	C	131.5	136.1	−4.6			
12'	CH ₂	34.7	36.3	−1.6	2.24	1.75	+0.49
					2.24	1.93	+0.31
13'	CH ₂	27.7	26.8	+0.9	2.13	1.94	+0.19
					2.13	1.98	+0.15
14'	CH	123.6	125.8	−2.2	5.12	5.28	−0.16
15'	C	132.1	139.9	−7.8			
16'	CH ₃	17.0	16.3	+0.7	1.60	1.41	+0.19
17'	CH ₃	25.0	21.6	+3.4	1.69	1.47	+0.22
18'	C=O	168.6	172.6	−4.0			
19'	CH ₃	15.1	15.2	−0.1	1.64	1.67	−0.03
20'	CH ₃	26.2	22.3	+3.9	1.18	1.28	−0.10
-	-	MAE		1.83	MAE		0.26
-	-	RMSE		3.64	RMSE		0.42

A multi-standard approach was used to reference the NMR chemical shift values as described by Sarotti and Pellegrinet¹ because of its better accuracy and precision than using TMS as a reference standard. Methanol was used as a reference standard for sp^3 carbon atoms while benzene was used for referencing sp and sp^2 carbons. The agreement between the experimental and theoretical values was very good with MAEs of only 0.25 ppm for ^1H -NMR and 1.83 ppm for ^{13}C -NMR chemical shifts. Thus, the NMR data does not contradict the configurational data because enantiomers have identical NMR chemical shifts. This supports the validity of *MICE-PES* to map the potential energy surface and the important conformational space of more functionally complicated molecules than just simple sphingoid lipids like 3-*epi*-xestoaminol C (**12**, Scheme 3.1).

On the basis of these findings, the absolute configuration of the molecule is confirmed as shown in Figure 3.12.

3.6 Conformational Analysis of Rimarikiamide A using *MICE-PES*

Rimarikiamide A (**43A**, Figure 3.13) is a taurinated diterpenoid derivative that possesses antimicrobial potential. It was isolated by Dr Nathaniel Dysan of the VUW marine natural products research group.²⁶² Its structure was established using mass spectrometry and NMR spectroscopic techniques. Unfortunately, neither the relative nor the absolute configuration could be established due to the lack of availability of a CD spectrophotometer. Due to its small amounts, the authors only tested it for mammalian cell line assays and found interesting results against HL-60 cell lines even at lower concentration. Given the severe lack of available material, coupled with the flexibility of the chain, identifying the relative configuration between the methyl and hydroxyl is impossible without synthesis or possibly computational methods. Rimarikiamide A (**43A**) is another very good example to study using *MICE-PES* for intra-molecular interactions because it is polyfunctional and the longer chain makes it suitable for potential stabilising intra-molecular interactions. Determination of the NMR properties of two possible enantiomeric pairs would allow the relative configuration to be determined and comparison of optical rotation or CD data would determine the absolute configuration to be established.

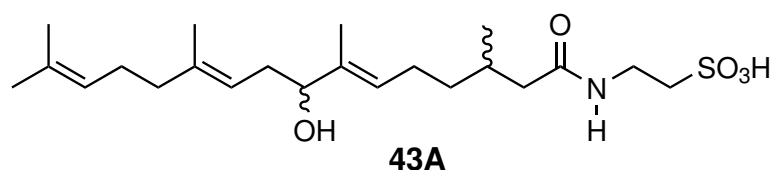
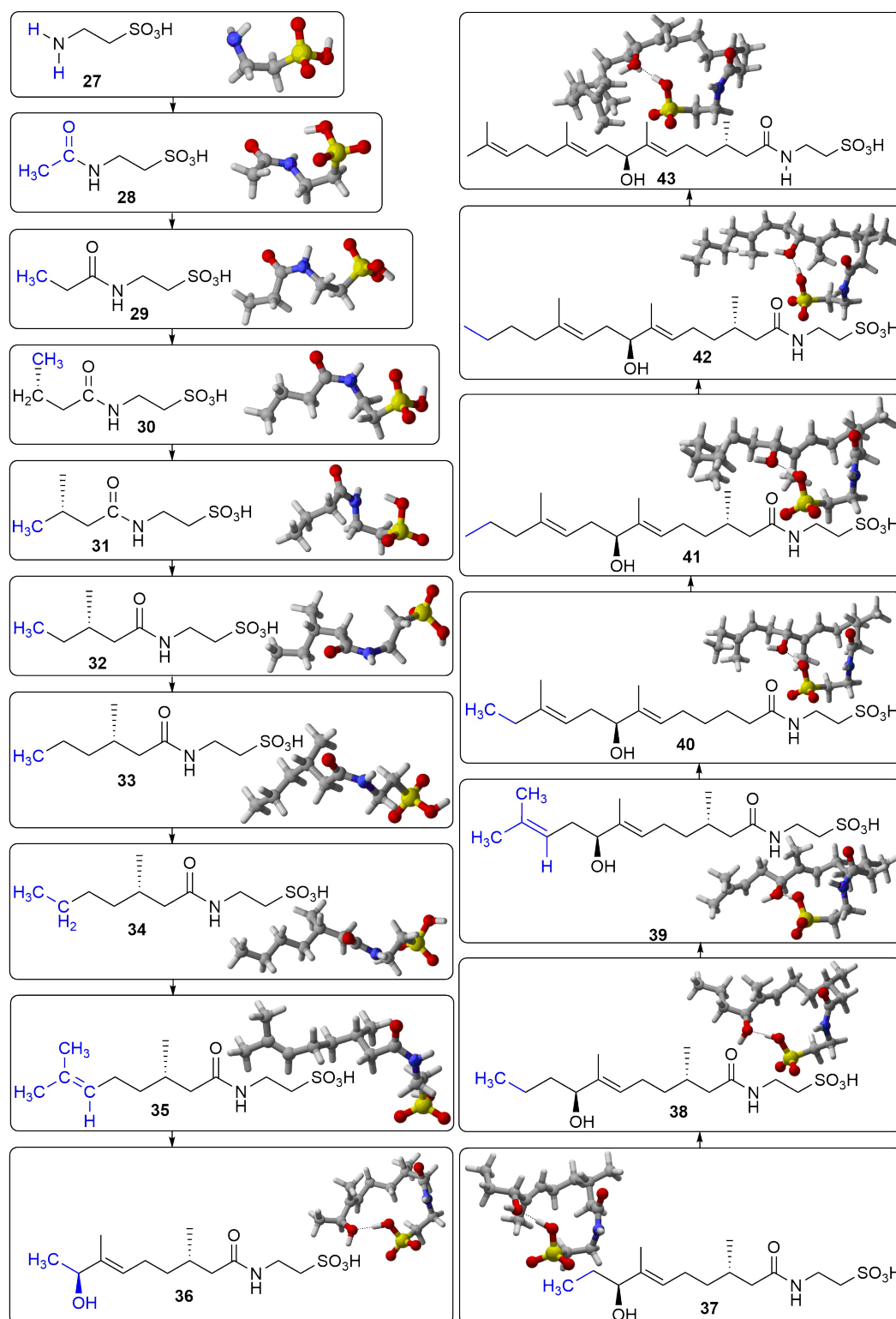


Figure 3.13: Rimarikiamide A isolated from the sponge *Latrunculia* sp. The squiggly bonds in the figure represent the chiral centres whose configuration is as yet unknown.

The PES of **43** was mapped with *MICE-PES* using high level QM calculations by successively constructing the molecule from its smallest possible homologue (taurine **27**, Scheme 3.10) as previously described in Section 3.4.1. Again, the selection was kept restricted to the important contributions only (>90% Boltzmann distribution), thus

discarding high energy conformers at each stage of the molecular construction.



Scheme 3.10: The 17 compounds that were considered in this investigation. **27** (2-aminoethane-1-sulfonic acid), **28** (2-acetamidoethane-1-sulfonic acid) etc. Note that **43** is one arbitrarily selected diastereomer of the naturally occurring rimarikiamide A. The atoms in blue colour represent the place of operation/addition of next atoms in the subsequent step. The lowest energy conformer of each homologue is shown in 3D.

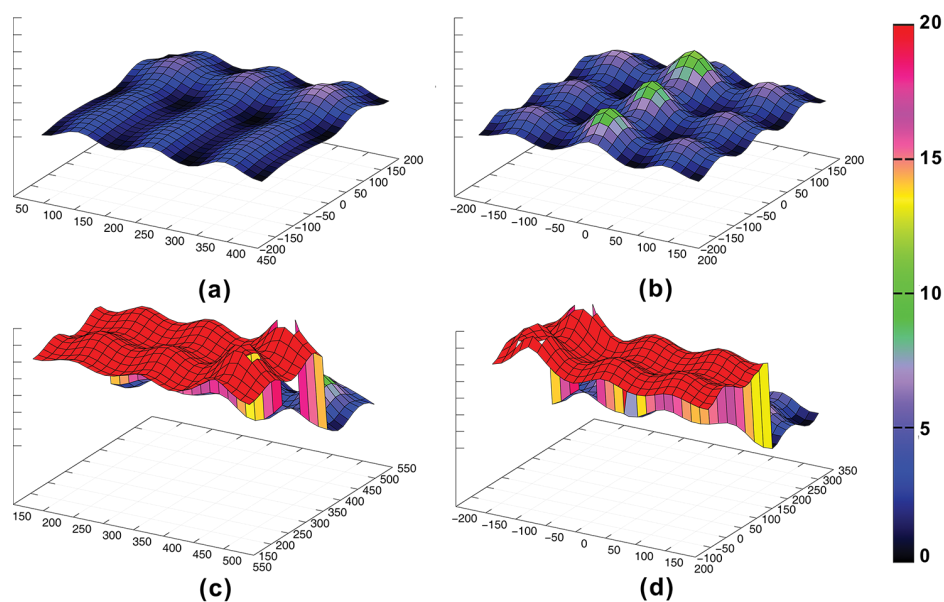


Figure 3.14: Potential energy surfaces at the PBE0-D3BJ/def2-TZVP/SMD_{CHCl₃} level of theory that identify all 8 conformers of **27**. Y-axis represents the relative energies of the conformers. X- and Z-axis show the rotational angles of the starting structure. The details of these angles and rotations can be found in the Appendix A.3. Optimisations were carried out for a total of 2500 individual structures. The high energy maxima in **c** and **d** show the high energy structure of taurine having a base (NH₂) and the neighbouring acid (SO₃H) while the energy minima in these two (**c** and **d**) are for the low energy structures where taurine is found as Zwitterionic structure (NH₃⁺ and SO₃⁻). The numbers on the energy scale on the right are in kcal/mol. **a–d** are the relaxed scans on different conformations of structure **27** to ensure that the PES has been covered well.

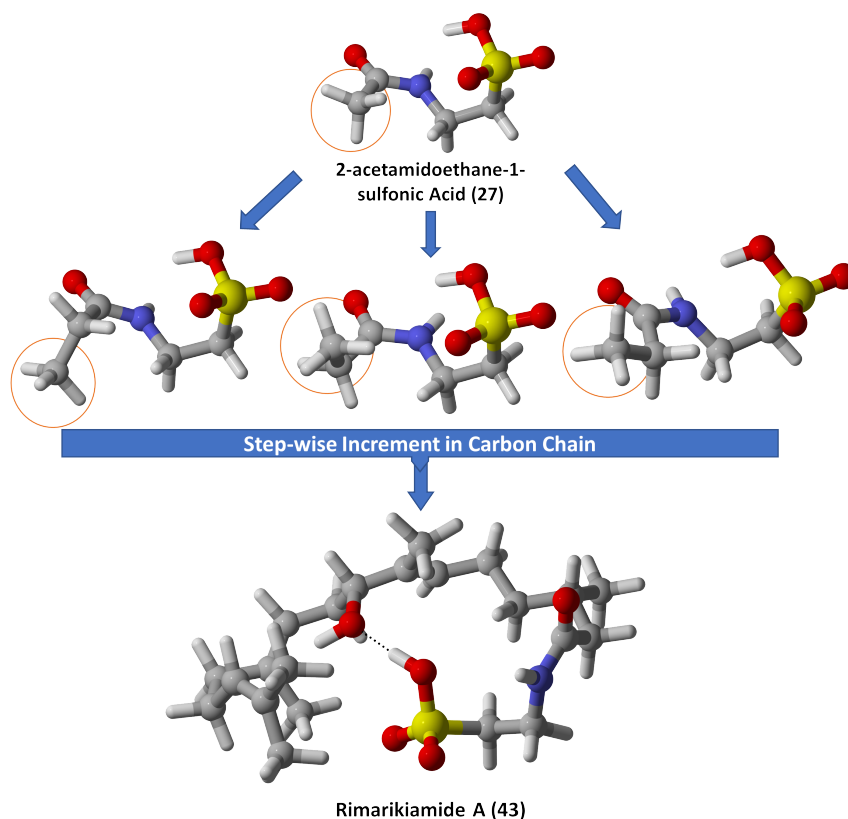
The examination of the PES of **27** was started by performing a sequence of relaxed potential energy scans on its four different conformers which mapped the PES by rotating all the dihedral angles of **27** in steps of 15° each (Figure 3.14). The resulting energy landscapes were then carefully examined to select 15 distinct conformers (Figure 3.15) which were then confirmed as minima by unconstrained relaxation. Their relative Gibb's free energies are given in Table 3.13.

In the conformational space obtained from these scans, all 15 minima contributed significantly because the difference between the highest and the lowest energy conformer was only 1.49 kcal/mole. According to the methodology of *MICE-PES*, sequential increments to the carbon chain were made. These 15 conformers of **27** were propagated into the next step to yield 30 conformers of **28** by the addition of an acetyl group at each of the amine hydrogens one by one (Scheme 3.11). In this molecule the double bonds

Table 3.13: The relative Gibbs free energies (ΔG in kcal/mol) and Boltzmann percentage distributions of the 15 minimum energy conformers of **27** at the PBE0-D3BJ/def2-TZVP/SMD_{CH₃OH} level of theory.

Conformer	Relative ΔG	Boltzmann Distribution (%)	Conformer	Relative ΔG	Boltzmann Distribution (%)
27a	0.00	12.15	27i	0.40	6.18
27b	0.18	8.90	27j	0.42	5.95
27c	0.22	8.38	27k	0.43	5.95
27d	0.23	8.37	27l	0.44	5.95
27e	0.25	8.37	27m	0.46	5.62
27f	0.37	6.46	27n	0.68	3.86
27g	0.38	6.45	27o	1.49	0.97
27h	0.39	6.45			

were not added in one step, rather a methyl group was added to each of the hydrogens in the first step. This procedure has been described in detail previously in section 3.5 with a graphical representation in Scheme 3.7. Note that carbonyl functional groups were added in a similar fashion to alkenes as described in section 3.5.



Scheme 3.11: Construction of conformers through step-wise lengthening of the alkyl chain using *MICE-PES*. The lowest energy isomer of **43** is shown at the bottom.

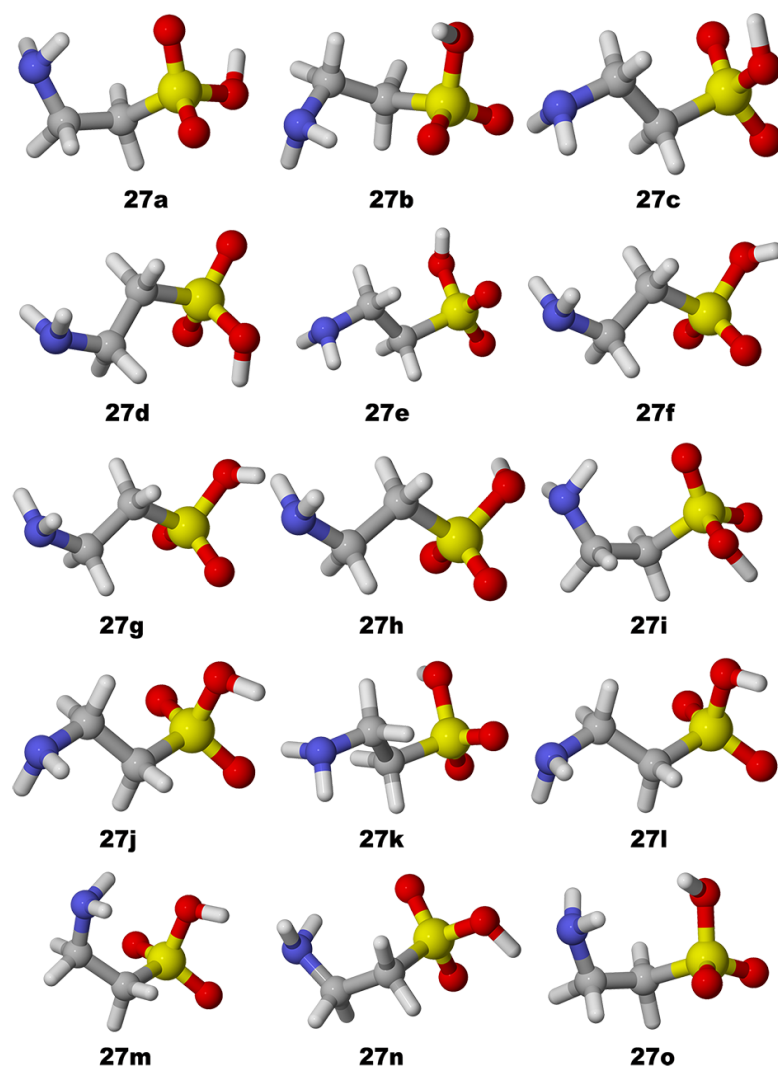


Figure 3.15: The 15 lowest energy conformers of **27** optimized at the PBE0-D3BJ/def2-TZVP/SMD_{CHCl₃} level of theory.

In the first two steps all the optimised conformers were propagated into the next step but after the second step, due to the increasing molecular size and increasing number of high energy conformers, only conformers with $\geq 90\%$ Boltzmann distribution proceeded to subsequent steps (Table 3.14). For rimarikiamide A (**43**), a total of 70 conformers were obtained out of which 67 conformers made 99.11% Boltzmann contribution to the ensemble. An important point to note is the case where there is intramolecular hydrogen bonding, that is when the --OH group was added to the conformers of homologue **35**, the number of important conformers suddenly reduced as the conformers with stabilising intramolecular hydrogen bonding dominate the entire conformational space. In this case out of 125 contributing conformers only 4 conformers contributed 99.94%

Boltzmann percentage contribution to the total ensemble. In this scenario the number of conformers propagated to the next step were 20, having 99.97% Boltzmann percentage contribution to the total ensemble, to test if the high energy conformers get better position in the ensemble in subsequent steps or not. In general, it can be stated that the descendants of conformers that have stabilising intramolecular interactions crowd outweigh the other structures.

MICE-PES proved itself efficient at capturing the intra-molecular hydrogen bonding interactions and it was quite surprising to see the stability of the molecules with favourable intra-molecular interactions and their increased Boltzmann percentage contribution to the ensemble.

Table 3.14: \mathbf{n}^{conf} is the total number of conformers generated in a step. \mathbf{n}^{contr} is the total number of distinct conformers considered after duplicates removal (fully optimised at the PBE0-D3BJ/def2-TZVP/SMD_{CHCl₃} level of theory). \mathbf{n}^{imp} is the number of important conformers (*i.e.* the number of conformers that have to be included to account for $\geq 90.0\%$ of the Boltzmann distribution at 298.15 K). \mathbf{B} is the Boltzmann percentage of all the \mathbf{n}^{imp} conformers. \mathbf{c}_0 is the percentage contribution of the lowest lying conformer at 298.15 K. \mathbf{P} is the total value of the partition function at 298.15 K. $\Delta\mathbf{G}$ is the difference in Gibbs free energy between the energetically lowest conformer and the energetically highest lying conformer (of all \mathbf{n}^{contr} conformers; in kcal/mol). $[\alpha]_D$ is the predicted value of the optical rotation (based on all \mathbf{n}^{contr} conformers).

Molecule	\mathbf{n}^{conf}	\mathbf{n}^{contr}	\mathbf{n}^{imp}	\mathbf{B}	\mathbf{c}_0	\mathbf{P}	$\Delta\mathbf{G}$	$[\alpha]_D$
27	15	15	15	100.0	12.15	8.23	1.49	+8.6
28	30	30	30	100.0	5.60	17.85	2.17	+11.1
29	90	90	60	98.83	12.16	8.22	3.60	+12.9
30	180	140	103	95.05	4.92	20.33	2.85	−0.4
31	103	99	60	95.33	8.69	11.51	4.48	+5.6
32	180	173	76	90.63	19.47	5.14	6.61	+0.4
33	228	225	81	90.52	17.15	5.83	6.49	+16.2
34	243	241	120	90.55	5.39	18.57	5.40	+20.7
35	120	109	67	90.31	11.21	8.92	3.35	+28.1
36	134	125	20	99.97	45.72	2.19	8.90	−36.0
37	60	60	15	99.99	32.47	3.08	9.62	−36.4
38	45	43	20	99.85	42.54	2.35	9.14	−19.2
39	20	17	15	99.96	79.28	1.26	5.75	−12.0
40	45	45	30	99.35	22.80	4.39	5.23	+2.5
41	90	90	55	99.08	33.05	3.03	6.90	+3.2
42	165	163	70	95.34	16.15	6.19	7.58	+9.7
43	70	67	54	99.11	20.44	4.89	5.32	−23.2

An important question to discuss here is the possibility of a different hydrogen bond linkage in some structures where the carbon chain is large enough to twist the molecule and interact to the same end of the molecule from behind. Looking at the lowest energy conformers of structures **36** and **37** (Scheme 3.10) makes it clear. Here it was tried to twist the carbon chain in different directions to see if the the —OH group could make a hydrogen bond to the —NH or the oxygen atoms of the —SO₃H moiety in a stabilising fashion. These attempts did not succeed as the structures were too strained and high in energy to be stable. This validated the reliability of *MICE-PES* in that it does not miss important conformations.

3.6.1 Validation of the Computed Data with Experiment

Because **43A** contains two chiral centres with an as yet indeterminate relative configuration, there are four possible diastereomers. Unfortunately, due to the small amount of the isolated compound an optical rotation measurement was not possible. The conformational analysis of one of its diastereomers (**43**) will provide a dataset of predicted NMR and ECD spectra. Experimental verification might become possible once rimarikiamide A has been successfully synthesised. All calculations were performed in CHCl₃ as a solvent due to the availability of spectroscopic data for synthetic analogues of **43**. However, the final total synthesis of rimarikiamide A has not yet been achieved and therefore final comparison of computational and experimental spectroscopic data is pending. The NMR data (DMSO) of isolated **43** cannot therefore be directly compared to the calculated data for one of the two possible diastereomeric pairs.^{262,263}

Chapter 4

Conclusions and Future Prospects

Natural products possess immense therapeutic potential and have been used since time immemorial as drugs. Secondary metabolites from different natural sources have played a vital role in modern drug development. Correct structural assignment is a vital part in natural products structure determination because a therapeutic natural product may also have a toxic stereoisomer. Many naturally occurring molecules are conformationally flexible and free rotation around single bonds leads to a variety of minima and maxima on the potential energy surface (PES). Absolute structure determination relies on the identification of all the lowest energy conformations present in solution. Quantum chemical calculations of physical properties like NMR, IR, UV, ECD, and VCD spectra are also dependent on the identification of all the minimum energy conformations in solution.

Herein, the development of a tool named *MICE-PES* (Method for Incremental Construction and Exploration of Potential Energy Surface) is described which maps the PES of a flexible compound efficiently with high level quantum chemical calculations in a relatively short amount of time. *MICE-PES* works stepwise in a systematic way as described below:

- (i) The smallest possible analogue of the compound under study is identified which

contains the most important chromophore or the portion having important functionality the molecule.

- (ii) An exhaustive PES scan is performed on different conformations of the smallest analogue identified in the previous step by rotating all of its dihedral angles in steps of 15° (adjustable as per desired accuracy).
- (iii) Lowest energy structures are selected from the PES scans followed by unconstrained optimization at a high level of theory.
- (iv) All low energy structures are then subjected to incremental increases in the chain length by the replacement of three terminal hydrogens (in case of a methyl) with a carbon group, or the two hydrogens in case of a methylene or carbonyl. This process is repeated until the full length of the molecule is reached by removing duplicates but keeping the stereoisomers at each step and proceeding with those conformers with >90% Boltzmann contribution to the ensemble to the next step.

MICE-PES was validated through studies on biomolecules including 3-*epi*-xestoaminol C and meroterphenol C, demonstrating very good comparison of the calculated properties to those observed experimentally. The scripts used for automation of the process are available at GitHub under a GNU general public license.²⁴³ *MICE-PES* is a smart algorithm which snakes its way through the important areas of the PES and finds the most important lowest energy conformers relatively quickly, while using state of the art DFT calculations which ensure accuracy of the method compared to classical force fields to do the conformational analysis. The available tools on the market (described previously in section 1.5) provide a conformational analysis using a range of available force fields which are typically less accurate. This is because force fields are parametrized for a specific class of molecules and perform well for only those compounds. The advantage of *MICE-PES* is that it is not specific to a class of compounds because it starts with a quantum chemical calculation of the starting compound which ensures the accuracy of molecular geometry and other chemical data. A limitation of *MICE-PES* is that it hasn't

been tested with cyclic systems but that function will be added in the near future.

The major future objective is expanding the studies to different classes of natural products and catalyst design. The following are the main objectives for future work:

- (i) Study the use of semi empirical calculations at intermediate incremental steps and only study the first PES mapping of the smallest analogue and the full molecule at a higher level of theory. This is to ensure the minimum usage of resources and time while keeping the accuracy at a reasonably high level.
- (ii) Semi empirical methods with dispersion correction will be used for the above mentioned studies.
- (iii) The development of a completely automatic procedure is also on the list of future prospects so that the program takes the first structure and does the rest automatically. This work is already in progress for a structure comparison algorithm to identify the unique structures out of a set of conformers obtained.
- (iv) Complex flexible biomolecules may also be an extension of the project in future.
- (v) Conformationally flexible catalysts design is also an important future goal of this project where *MICE-PES* will find its application.

Appendices

Appendix A

A.1 Relaxed Scans of the Smallest Homologue (2) of 3-*epi*-xestoaminol C (12)

Relaxed scans of selected the dihedral angles were performed for different manually generated conformers of **2** whose corresponding plots can be found in Figure 3.6 as **a-f**. The details of these dihedral angles have been provided in Figure A.1.

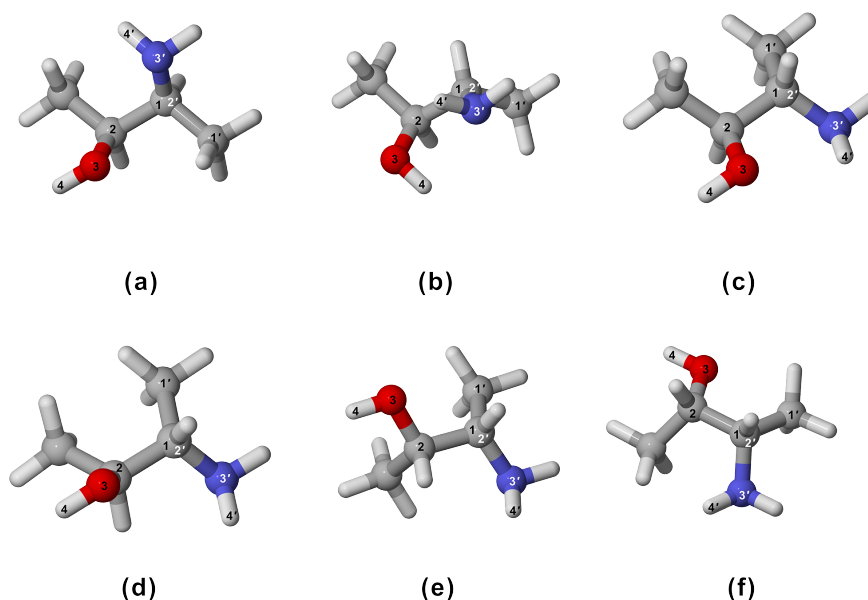


Figure A.1: Relaxed scans on six manually generated conformers of **2** optimised at the PBE0-D3BJ/def2-TZVP/SMD_{CH₂Cl₂} level of theory. Relaxed scans were performed by rotating two sets of dihedral angles. First dihedral is between atoms 1-2-3-4 while the second one is denoted by atoms 1'-2'-3'-4'.

A.2 Relaxed Scans of the Smallest Homologue (13) of Meroterphenol C (26)

Relaxed scans of the selected dihedral angles were performed for three different manually generated optimised conformers of **13** whose corresponding plots can be found in Figure 3.10 as **a-i**. The details of these dihedral angles have been provided in Figure A.2.

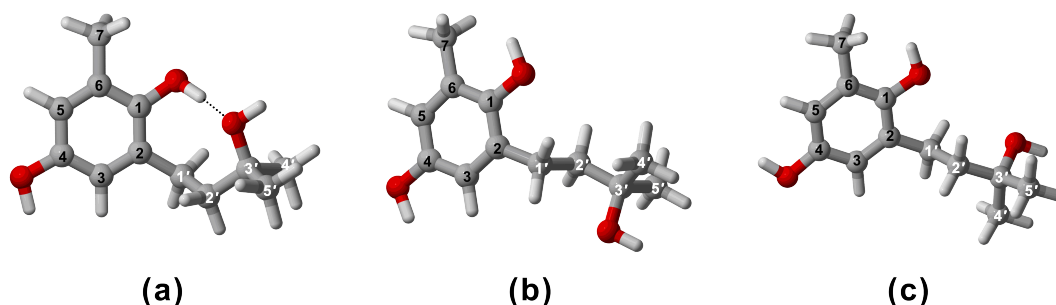


Figure A.2: Relaxed scans on four different conformers of **27** optimised at the PBE0-D3BJ/def2-TZVP/SMD_{CHCl₃} level of theory. Relaxed scans were performed by rotating two sets of dihedral angles. In PES scans in Figure 3.10, plot **a** corresponds to the dihedral angles between atoms 1-2-3-6 and 5-3-6-7. Plot **b**, corresponds to the rotation of dihedrals 5-3-6-7 and 8-6-9-10, plot **c** originated by rotating the dihedrals 5-3-6-7 and 8-6-9-11, and plot **d** came from the rotation of dihedral angles 2-3-6-7 and 8-6-9-11.

A.3 Relaxed Scans of the Smallest Homologue (27) of Rimarikiamide (43)

Relaxed scans of the selected dihedral angles were performed for four different manually generated optimised conformers of **13** whose corresponding plots can be found in Figure 3.14 as **a-d**. The details of these dihedral angles have been provided in Figure A.3.

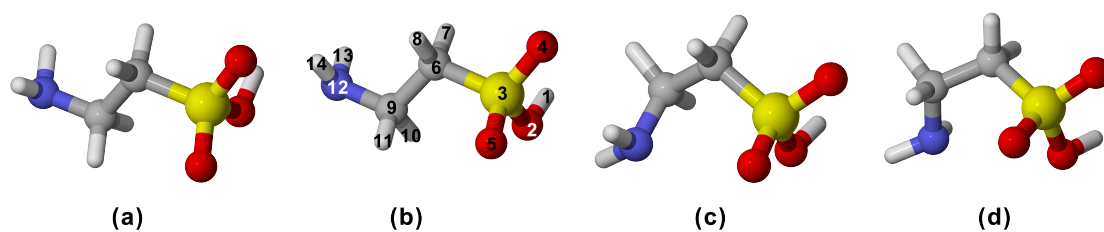


Figure A.3: Relaxed scans on four different conformers of **27** optimised at the PBE0-D3BJ/def2-TZVP/SMD_{CHCl₃} level of theory. Relaxed scans were performed by rotating two sets of dihedral angles. In PES scans in Figure 3.14, plot **a** corresponds to the dihedral angles between atoms 1-2-3-6 and 5-3-6-7. Plot **b**, corresponds to the rotation of dihedrals 5-3-6-7 and 8-6-9-10, plot **c** originated by rotating the dihedrals 5-3-6-7 and 8-6-9-11, and plot **d** came from the rotation of dihedral angles 2-3-6-7 and 8-6-9-11.

Bibliography

1. Sarotti, A. M.; Pellegrinet, S. C. *J. Org. Chem.* **2009**, *74*, 7254–7260.
2. Koehn, F. E.; Carter, G. T. *Nat. Rev. Drug Discov.* **2005**, *4*, 206–220.
3. Demain, A. J. *Ind. Microbiol. Biotechnol.* **2014**, *41*, 185–201.
4. Newman, D. J.; Cragg, G. M. *J. Nat. Prod.* **2007**, *70*, 461–477.
5. Spring, D. R. *Org. Biomol. Chem.* **2003**, *1*, 3867–3870.
6. Pelish, H. E.; Westwood, N. J.; Feng, Y.; Kirchhausen, T.; Shair, M. D. *J. Am. Chem. Soc.* **2001**, *123*, 6740–6741.
7. Burke, M. D.; Schreiber, S. L. *Angew. Chem. Int. Ed.* **2004**, *43*, 46–58.
8. Gan, Z.; Reddy, P. T.; Quevillon, S.; Couve-Bonnaire, S.; Arya, P. *Angew. Chem. Int. Ed.* **2005**, *44*, 1366–1368.
9. Cragg, G. M.; Newman, D. J.; Snader, K. M. *J. Nat. Prod.* **1997**, *60*, 52–60.
10. Newman, D. J.; Cragg, G. M.; Snader, K. M. *J. Nat. Prod.* **2003**, *66*, 1022–1037.
11. Newman, D. J.; Cragg, G. M. *J. Nat. Prod.* **2012**, *75*, 311–335.
12. Trébucq, A. *Int. J. Tuberc. Lung Dis.* **1997**, *1*, 12–15.
13. Kern, J. R. *J. Chromatogr. A* **1991**, *543*, 355–366.
14. Banerjee, P.; Erehman, J.; Gohlke, B.-O.; Wilhelm, T.; Preissner, R.; Dunkel, M. *Nucleic Acids Res.* **2015**, *43*, D935–D939.
15. Hoffmann, R. W. *Angew. Chem. Int. Ed.* **1992**, *31*, 1124–1134.
16. Micco, S.; Chini, M.; Riccio, R.; Bifulco, G. *Handbook of Marine Natural Products*; Springer Netherlands, 2012; Chapter 10, pp 571–599.
17. Stephens, P. J.; Pan, J. J.; Devlin, F. J.; Cheeseman, J. R. *J. Nat. Prod.* **2008**, *71*, 285–288.
18. Smith, S. G.; Goodman, J. M. *J. Am. Chem. Soc.* **2010**, *132*, 12946–12959.
19. Hashmi, M. A.; Khan, A.; Ayub, K.; Farooq, U. *Spectrochim. Acta A* **2014**, *128*, 225–230.

20. Nugroho, A.; Morita, H. *J. Nat. Med.* **2014**, *68*, 1–10.
21. Barton, D. H. R. *Experientia* **1950**, *6*, 316–320.
22. Moss, G. P. Basic terminology of stereochemistry (IUPAC Recommendations 1996). 1996; [//www.degruyter.com/view/j/pac.1996.68.issue-12/pac199668122193/pac199668122193.xml](http://www.degruyter.com/view/j/pac.1996.68.issue-12/pac199668122193/pac199668122193.xml).
23. Becker, O.; MacKerell, A. D.; Roux, B.; Watanabe, M. *Comput. Biochem. Biophys.*; Taylor Francis, 2001; pp 491–500.
24. Gschwend, D. A.; Good, A. C.; Kuntz, I. D. *J. Mol. Recognit.* **1996**, *9*, 175–186.
25. Leach, A. *Molecular Modelling: Principles and Applications*; Prentice Hall, 2001.
26. Saunders, M.; Houk, K. N.; Wu, Y. D.; Still, W. C.; Lipton, M.; Chang, G.; Guida, W. C. *J. Am. Chem. Soc.* **1990**, *112*, 1419–1427.
27. Howard, A. E.; Kollman, P. A. *J. Med. Chem.* **1988**, *31*, 1669–1675.
28. Bruccoleri, R. E.; Karplus, M. *Biopolymers* **1987**, *26*, 137–168.
29. Dolata, D. P.; Leach, A. R.; Prout, K. J. *Comput. Aided Mol. Des.* **1987**, *1*, 73–85.
30. Gibson, K. D.; Scheraga, H. A. *J. Comput. Chem.* **1987**, *8*, 826–834.
31. Bruccoleri, R. E.; Karplus, M. *Biopolymers* **1990**, *29*, 1847–1862.
32. Metropolis, N.; Ulam, S. *J. Am. Stat. Assoc.* **1949**, *44*, 335–341.
33. Frenkel, D.; Smit, B. *Understanding Molecular Simulation: From Algorithms to Applications*; Elsevier Science, 2001.
34. Tuckerman, M. *Statistical Mechanics: Theory and Molecular Simulation*; Oxford graduate texts; Oxford University Press, 2010.
35. Cao, J.; Berne, B. J. *J. Chem. Phys.* **1990**, *92*, 1980–1985.
36. Andricioaei, I.; Straub, J. E. *J. Chem. Phys.* **1997**, *107*, 9117–9124.
37. Crippen, G. M.; Havel, T. F. *Acta Crystallogr. A* **1978**, *34*, 282–284.
38. Crippen, G. M.; Havel, T. F. *Distance geometry and molecular conformation*; John Wiley and Sons, New York: Research Studies Press, Taunton, England, 1988.
39. Weiner, P. K.; Profeta Jr, S.; Wipff, G.; Havel, T.; Kuntz, I. D.; Langridge, R.; Kollman, P. A. *Tetrahedron* **1983**, *39*, 1113–1121.
40. Havel, T. F. *Prog. Biophys. Mol. Biol.* **1991**, *56*, 43–78.
41. Fesik, S. W.; Bolis, G.; Sham, H. L.; Olejniczak, E. T. *Biochemistry* **1987**, *26*, 1851–1859.
42. Sheridan, R. P.; Nilakantan, R.; Dixon, J. S.; Venkataraghavan, R. *J. Med. Chem.* **1986**, *29*, 899–906.

43. Kuntz, I. *Protein Eng.* **1987**, *1*, 147–148.
44. Kline, A. D.; Braun, W.; Wüthrich, K. *J. Mol. Biol.* **1986**, *189*, 377–382.
45. Pflugrath, J. W.; Wiegand, G.; Huber, R.; Vértessy, L. *J. Mol. Biol.* **1986**, *189*, 383–386.
46. Blaney, J. M.; Dixon, J. S. *Reviews in Computational Chemistry*; John Wiley and Sons Inc., 2007; pp 299–335.
47. Ferguson, D. M.; Raber, D. J. *J. Am. Chem. Soc.* **1989**, *111*, 4371–4378.
48. Saunders, M. *J. Am. Chem. Soc.* **1987**, *109*, 3150–3152.
49. Li, Z.; Scheraga, H. A. *Proc. Natl. Acad. Sci. U.S.A.* **1987**, *84*, 6611–6615.
50. Chang, G.; Guida, W. C.; Still, W. C. *J. Am. Chem. Soc.* **1989**, *111*, 4379–4386.
51. Pincus, M. R.; Klausner, R. D.; Scheraga, H. A. *Proc. Natl. Acad. Sci. U.S.A.* **1982**, *79*, 5107–5110.
52. Gibson, K. D.; Scheraga, H. A. *J. Comput. Chem.* **1987**, *8*, 826–834.
53. Hingerty, B. E.; Figueroa, S.; Hayden, T. L.; Broyde, S. *Biopolymers* **1989**, *28*, 1195–1222.
54. Dirac, P. A. M. *Proc. Royal Soc. Lond. A* **1929**, *123*, 714–733.
55. Pisana, S.; Lazzeri, M.; Casiraghi, C.; Novoselov, K. S.; Geim, A. K.; Ferrari, A. C.; Mauri, F. *Nat. Mater.* **2007**, *6*, 198.
56. Bunker, P. *J. Mol. Spectrosc.* **1973**, *45*, 151–158.
57. Bunker, P.; Moss, R. *Mol. Phys.* **1977**, *33*, 417–424.
58. Bishop, D. M.; Solunac, S. A. *Phys. Rev. Lett.* **1985**, *55*, 1986–1988.
59. Brooks, C.; Karplus, M.; Pettitt, B. *Advances in Chemical Physics, Proteins: A Theoretical Perspective of Dynamics, Structure, and Thermodynamics*; Wiley, 2009.
60. Hulburt, H. M.; Hirschfelder, J. O. *J. Chem. Phys.* **1941**, *9*, 61–69.
61. Halgren, T. A. *Curr. Opin. Struct. Biol.* **1995**, *5*, 205–210.
62. Blondel, A.; Karplus, M. *J. Comput. Chem.* **1996**, *17*, 1132–1141.
63. Pearlman, D. A.; Case, D. A.; Caldwell, J. W.; Ross, W. S.; Cheatham, T. E.; DeBolt, S.; Ferguson, D.; Seibel, G.; Kollman, P. *Comput. Phys. Commun.* **1995**, *91*, 1–41.
64. Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. *J. Comput. Chem.* **1983**, *4*, 187–217.

65. Scott, W. R. P.; Hünenberger, P. H.; Tironi, I. G.; Mark, A. E.; Billeter, S. R.; Fennen, J.; Torda, A. E.; Huber, T.; Krüger, P.; van Gunsteren, W. F. *J. Phys. Chem. A* **1999**, *103*, 3596–3607.
66. Schlenkrich, M.; Brickmann, J.; MacKerell, A. D.; Karplus, M. *Biological Membranes*; Birkhäuser Boston, 1996; Chapter 2, pp 31–81.
67. MacKerell, A. D.; Wiorkiewicz-Kuczera, J.; Karplus, M. *J. Am. Chem. Soc.* **1995**, *117*, 11946–11975.
68. MacKerell, A. D. et al. *J. Phys. Chem. B* **1998**, *102*, 3586–3616.
69. Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. *J. Am. Chem. Soc.* **1995**, *117*, 5179–5197.
70. Kaminski, G.; Duffy, E. M.; Matsui, T.; Jorgensen, W. L. *J. Phys. Chem.* **1994**, *98*, 13077–13082.
71. Jorgensen, W. L.; Maxwell, D. S.; Tirado-Rives, J. *J. Am. Chem. Soc.* **1996**, *118*, 11225–11236.
72. Lii, J.-H.; Allinger, N. L. *J. Comput. Chem.* **1991**, *12*, 186–199.
73. Allinger, N. L.; Yuh, Y. H.; Lii, J. H. *J. Am. Chem. Soc.* **1989**, *111*, 8551–8566.
74. Maple, J. R.; Hwang, M. J.; Jalkanen, K. J.; Stockfisch, T. P.; Hagler, A. T. *J. Comput. Chem.* **1998**, *19*, 430–458.
75. Schaefer, M.; Karplus, M. *J. Phys. Chem.* **1996**, *100*, 1578–1599.
76. Rappe, A. K.; Casewit, C. J.; Colwell, K. S.; Goddard, W. A.; Skiff, W. M. *J. Am. Chem. Soc.* **1992**, *114*, 10024–10035.
77. Mayo, S. L.; Olafson, B. D.; Goddard, W. A. *J. Phys. Chem.* **1990**, *94*, 8897–8909.
78. Weiner, S. J.; Kollman, P. A.; Case, D. A.; Singh, U. C.; Ghio, C.; Alagona, G.; Profeta, S.; Weiner, P. *J. Am. Chem. Soc.* **1984**, *106*, 765–784.
79. Weiner, S. J.; Kollman, P. A.; Nguyen, D. T.; Case, D. A. *J. Comput. Chem.* **1986**, *7*, 230–252.
80. Salomon-Ferrer, R.; Case, D. A.; Walker, R. C. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2013**, *3*, 198–210.
81. Case, D. A.; Cheatham, T. E.; Darden, T.; Gohlke, H.; Luo, R.; Merz, K. M.; Onufriev, A.; Simmerling, C.; Wang, B.; Woods, R. J. *J. Comput. Chem.* **2005**, *26*, 1668–1688.
82. Beachy, M. D.; Chasman, D.; Murphy, R. B.; Halgren, T. A.; Friesner, R. A. *J. Am. Chem. Soc.* **1997**, *119*, 5908–5920.
83. Kollman, P.; Dixon, R.; Cornell, W.; Fox, T.; Chipot, C.; Pohorille, A. *Computer Simulation of Biomolecular Systems*; Computer Simulations of Biomolecular Systems; Springer Netherlands, 1997; Vol. 3; Chapter 2, pp 83–96.

84. Jorgensen, W. L.; Tirado-Rives, J. *J. Am. Chem. Soc.* **1988**, *110*, 1657–1666.
85. Wang, J.; Cieplak, P.; Kollman, P. A. *J. Comput. Chem.* **2000**, *21*, 1049–1074.
86. Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. *J. Comput. Chem.* **2004**, *25*, 1157–1174.
87. Halgren, T. A. *J. Comput. Chem.* **1996**, *17*, 490–519.
88. Pérez, A.; Marchán, I.; Svozil, D.; Sponer, J.; Cheatham, T. E.; Laughton, C. A.; Orozco, M. *Biophys. J.* **2007**, *92*, 3817–3829.
89. Foloppe, N.; MacKerell, J. A. D. *J. Comput. Chem.* **2000**, *21*, 86–104.
90. MacKerell, A. D.; Banavali, N. K. *J. Comput. Chem.* **2000**, *21*, 105–120.
91. Patel, S.; Mackerell, A. D.; Brooks, C. L. *J. Comput. Chem.* **2004**, *25*, 1504–1514.
92. Mackerell, A. D.; Feig, M.; Brooks, C. L. *J. Comput. Chem.* **2004**, *25*, 1400–1415.
93. Vanommeslaeghe, K.; Hatcher, E.; Acharya, C.; Kundu, S.; Zhong, S.; Shim, J.; Darian, E.; Guvench, O.; Lopes, P.; Vorobyov, I.; Mackerell, A. D. *J. Comput. Chem.* **2010**, *31*, 671–690.
94. Klauda, J. B.; Venable, R. M.; Freites, J. A.; O'Connor, J. W.; Tobias, D. J.; Mondragon-Ramirez, C.; Vorobyov, I.; MacKerell, A. D.; Pastor, R. W. *J. Phys. Chem. B* **2010**, *114*, 7830–7843.
95. Brooks, B. R. et al. *J. Comput. Chem.* **2009**, *30*, 1545–1614.
96. Van Gunsteren, W.; Berendsen, H. *Library manual, Biomos, Groningen, The Netherlands* **1987**, 1–221.
97. van Gunsteren, W. F.; Daura, X.; Mark, A. E. *Encyclopedia of Computational Chemistry*; John Wiley and Sons, Ltd, 2002.
98. Kouwijzer, M. L. C. E.; van Eijck, B. P.; Kooijman, H.; Kroon, J. *Acta Crystallogr. B* **1995**, *51*, 209–220.
99. van Gunsteren, W.; Billeter, S. R.; Eising, A. A.; Hünenberger, P.; Krüger, P.; Mark, A.; Scott, W. R. P.; Tironi, I. *Biomolecular Simulation: The GROMOS96 Manual and User Guide*, Vdf Hochschulverlag AG an der ETH Zürich, Zürich, Switzerland; Hochschulverlag AG an der ETH Zürich, 1996.
100. Ott, K.-H.; Meyer, B. *J. Comput. Chem.* **1996**, *17*, 1068–1084.
101. Spieser, S. A. H.; Albert van Kuik, J.; Kroon-Batenburg, L. M. J.; Kroon, J. *Carbohydr. Res.* **1999**, *322*, 264–273.
102. Schuler, L. D.; Daura, X.; van Gunsteren, W. F. *J. Comput. Chem.* **2001**, *22*, 1205–1218.
103. Soares, T. A.; Hünenberger, P. H.; Kastenholz, M. A.; Kräutler, V.; Lenz, T.; Lins, R. D.; Oostenbrink, C.; van Gunsteren, W. F. *J. Comput. Chem.* **2005**, *26*, 725–737.

104. Lins, R. D.; Hünenberger, P. H. *J. Comput. Chem.* **2005**, *26*, 1400–1412.
105. Oostenbrink, C.; Villa, A.; Mark, A. E.; Van Gunsteren, W. F. *J. Comput. Chem.* **2004**, *25*, 1656–1676.
106. Oostenbrink, C.; Soares, T. A.; van der Vegt, N. F. A.; Van Gunsteren, W. F. *Eur. Biophys. J.* **2005**, *34*, 273–284.
107. Christen, M.; Hünenberger, P. H.; Bakowies, D.; Baron, R.; Bürgi, R.; Geerke, D. P.; Heinz, T. N.; Kastenholz, M. A.; Kräutler, V.; Oostenbrink, C.; Peter, C.; Trzesniak, D.; van Gunsteren, W. F. *J. Comput. Chem.* **2005**, *26*, 1719–1751.
108. Schmid, N.; Eichenberger, A. P.; Choutko, A.; Riniker, S.; Winger, M.; Mark, A. E.; van Gunsteren, W. F. *Eur. Biophys. J.* **2011**, *40*, 843–856.
109. Huang, W.; Lin, Z.; van Gunsteren, W. F. *J. Chem. Theory Comput.* **2011**, *7*, 1237–1243.
110. Reif, M. M.; Hünenberger, P. H.; Oostenbrink, C. *J. Chem. Theory Comput.* **2012**, *8*, 3705–3723.
111. Furche, F.; Ahlrichs, R.; Hättig, C.; Klopper, W.; Sierka, M.; Weigend, F. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2014**, *4*, 91–100.
112. Meier, K.; Schmid, N.; van Gunsteren, W. F. *J. Comput. Chem.* **2012**, *33*, 2108–2117.
113. Schmid, N.; Christ, C. D.; Christen, M.; Eichenberger, A. P.; van Gunsteren, W. F. *Comput. Phys. Commun.* **2012**, *183*, 890–903.
114. Kahn, K.; Bruice, T. C. *J. Comput. Chem.* **2002**, *23*, 977–996.
115. Pranata, J.; Wierschke, S. G.; Jorgensen, W. L. *J. Am. Chem. Soc.* **1991**, *113*, 2810–2819.
116. Damm, W.; Frontera, A.; Tirado-Rives, J.; Jorgensen, W. L. *J. Comput. Chem.* **1997**, *18*, 1955–1970.
117. Jorgensen, W. L.; McDonald, N. A. *J. Mol. Struct. THEOCHEM* **1998**, *424*, 145–155.
118. Kony, D.; Damm, W.; Stoll, S.; Van Gunsteren, W. F. *J. Comput. Chem.* **2002**, *23*, 1416–1429.
119. Sambasivarao, S. V.; Acevedo, O. *J. Chem. Theory Comput.* **2009**, *5*, 1038–1050.
120. Shivakumar, D.; Williams, J.; Wu, Y.; Damm, W.; Shelley, J.; Sherman, W. *J. Chem. Theory Comput.* **2010**, *6*, 1509–1519.
121. Shivakumar, D.; Harder, E.; Damm, W.; Friesner, R. A.; Sherman, W. *J. Chem. Theory Comput.* **2012**, *8*, 2553–2558.
122. Lifson, S.; Warshel, A. *J. Chem. Phys.* **1968**, *49*, 5116–5129.

123. Warshel, A.; Lifson, S. *J. Chem. Phys.* **1970**, *53*, 582–594.
124. Warshel, A.; Levitt, M.; Lifson, S. *J. Mol. Spectrosc.* **1970**, *33*, 84–99.
125. Ermer, O.; Lifson, S. *J. Am. Chem. Soc.* **1973**, *95*, 4121–4132.
126. Hagler, A. T.; Huler, E.; Lifson, S. *J. Am. Chem. Soc.* **1974**, *96*, 5319–5327.
127. Lifson, S.; Hagler, A. T.; Dauber, P. *J. Am. Chem. Soc.* **1979**, *101*, 5111–5121.
128. Hagler, A. T.; Lifson, S.; Dauber, P. *J. Am. Chem. Soc.* **1979**, *101*, 5122–5130.
129. Sun, H. *J. Phys. Chem. B* **1998**, *102*, 7338–7364.
130. Sun, H.; Ren, P.; Fried, J. R. *Comput. Theor. Polym. Sci.* **1998**, *8*, 229–246.
131. Sun, H.; Rigby, D. *Spectrochim. Acta A* **1997**, *53*, 1301–1323.
132. Rigby, D.; Sun, H.; Eichinger, B. E. *Polym. Int.* **1997**, *44*, 311–330.
133. Yang, J.; Ren, Y.; Tian, A.-m.; Sun, H. *J. Phys. Chem. B* **2000**, *104*, 4951–4957.
134. McQuaid, M. J.; Sun, H.; Rigby, D. *J. Comput. Chem.* **2004**, *25*, 61–71.
135. Lii, J. H.; Allinger, N. L. *J. Am. Chem. Soc.* **1989**, *111*, 8566–8575.
136. Lii, J. H.; Allinger, N. L. *J. Am. Chem. Soc.* **1989**, *111*, 8576–8582.
137. Allinger, N. L.; Rahman, M.; Lii, J. H. *J. Am. Chem. Soc.* **1990**, *112*, 8293–8307.
138. Allinger, N. L.; Li, F.; Yan, L. *J. Comput. Chem.* **1990**, *11*, 848–867.
139. Liang, G.; Bays, J. P.; Bowen, J. P. *J. Mol. Struct. THEOCHEM* **1997**, *401*, 165–179.
140. Hay, B. P.; Yang, L.; Lii, J.-H.; Allinger, N. L. *J. Mol. Struct. THEOCHEM* **1998**, *428*, 203–219.
141. Lii, J.-H.; Allinger, N. L. *J. Comput. Chem.* **1998**, *19*, 1001–1016.
142. Hay, B. P.; Clement, O.; Sandrone, G.; Dixon, D. A. *Inorg. Chem.* **1998**, *37*, 5887–5894.
143. Hagelin, H.; Svensson, M.; Åkermark, B.; Norrby, P.-O. *Organometallics* **1999**, *18*, 4574–4583.
144. Tafipolsky, M.; Amirjalayer, S.; Schmid, R. *J. Comput. Chem.* **2007**, *28*, 1169–1176.
145. Halgren, T. A. *J. Comput. Chem.* **1996**, *17*, 520–552.
146. Halgren, T. A. *J. Comput. Chem.* **1996**, *17*, 553–586.
147. Halgren, T. A. *J. Comput. Chem.* **1996**, *17*, 616–641.
148. Tosco, P.; Stiefl, N.; Landrum, G. *J. Cheminform.* **2014**, *6*, 37.

149. Casewit, C. J.; Colwell, K. S.; Rappe, A. K. *J. Am. Chem. Soc.* **1992**, *114*, 10035–10046.
150. Casewit, C. J.; Colwell, K. S.; Rappe, A. K. *J. Am. Chem. Soc.* **1992**, *114*, 10046–10053.
151. Rappe, A. K.; Colwell, K. S.; Casewit, C. J. *Inorg. Chem.* **1993**, *32*, 3438–3450.
152. Sirovatka, J. M.; Rappé, A. K.; Finke, R. G. *Inorganica Chim. Acta* **2000**, *300*–*302*, 545–555.
153. Yakovenko, O. Y.; Li, Y. Y.; Oliferenko, A. A.; Vashchenko, G. M.; Bdzhola, V. G.; Jones, S. J. M. *J. Mol. Model.* **2012**, *18*, 663–673.
154. Garberoglio, G.; Taioli, S. *Microporous Mesoporous Mater.* **2012**, *163*, 215–220.
155. Addicoat, M. A.; Vankova, N.; Akter, I. F.; Heine, T. *J. Chem. Theory Comput.* **2014**, *10*, 880–891.
156. Geudtner, G.; Calaminici, P.; Carmona-Espíndola, J.; del Campo, J. M.; Domínguez-Soria, V. D.; Moreno, R. F.; Gamboa, G. U.; Goursot, A.; Köster, A. M.; Reveles, J. U.; Mineva, T.; Vásquez-Pérez, J. M.; Vela, A.; Zúñiga-Gutierrez, B.; Salahub, D. R. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2012**, *2*, 548–555.
157. Gale, J. D. *J. Chem. Soc. Faraday Trans.* **1997**, *93*, 629–637.
158. te Velde, G.; Bickelhaupt, F. M.; Baerends, E. J.; Fonseca Guerra, C.; van Gisbergen, S. J. A.; Snijders, J. G.; Ziegler, T. *J. Comput. Chem.* **2001**, *22*, 931–967.
159. Marvin v.15.8.24, ChemAxon (<http://www.chemaxon.com>) **2015**,
160. Grimme, S. *J. Chem. Theory Comput.* **2014**, *10*, 4497–4514.
161. Jensen, F. *Introduction to Computational Chemistry*; John Wiley and Sons, 2007.
162. Bifulco, G.; Dambruoso, P.; Gomez-Paloma, L.; Riccio, R. *Chem. Rev.* **2007**, *107*, 3744–3779.
163. Kim, H. K.; Choi, Y. H.; Verpoorte, R. *Nat. Protoc.* **2010**, *5*, 536–549.
164. Krishnan, P.; Kruger, N. J.; Ratcliffe, R. G. *J. Exp. Bot.* **2005**, *56*, 255–265.
165. Haw, J. F.; Richardson, B. R.; Oshiro, I. S.; Lazo, N. D.; Speed, J. A. *J. Am. Chem. Soc.* **1989**, *111*, 2052–2058.
166. Li, D.; Keresztes, I.; Hopson, R.; Williard, P. G. *Acc. Chem. Res.* **2009**, *42*, 270–280.
167. Kühne, R. O.; Schaffhauser, T.; Wokaun, A.; Ernst, R. R. *J. Magn. Reson.* **1979**, *35*, 39–67.
168. Assink, R. A.; Kay, B. D. *Annu. Rev. Mater. Sci.* **1991**, *21*, 491–513.
169. Willoughby, P. H.; Jansma, M. J.; Hoye, T. R. *Nat. Protoc.* **2014**, *9*, 643–660.

170. Chini, M. G.; Riccio, R.; Bifulco, G. *Eur. J. Org. Chem.* **2015**, 2015, 1320–1324.
171. Sarotti, A. M.; Pellegrinet, S. C. *J. Org. Chem.* **2012**, 77, 6059–6065.
172. Andrews, K. G.; Spivey, A. C. *J. Org. Chem.* **2013**, 78, 11302–11317.
173. Lodewyk, M. W.; Siebert, M. R.; Tantillo, D. J. *Chem. Rev.* **2012**, 112, 1839–1862.
174. Wiitala, K. W.; Al-Rashid, Z. F.; Dvornikovs, V.; Hoye, T. R.; Cramer, C. J. *J. Phys. Org. Chem.* **2007**, 20, 345–354.
175. Smith, S. G.; Goodman, J. M. *J. Org. Chem.* **2009**, 74, 4597–4607.
176. Mendoza-Espinoza, J. A.; López-Vallejo, F.; Fragoso-Serrano, M.; Pereda-Miranda, R.; Cerda-García-Rojas, C. M. *J. Nat. Prod.* **2009**, 72, 700–708.
177. Di Micco, S.; Chini, M. G.; Riccio, R.; Bifulco, G. *Eur. J. Org. Chem.* **2010**, 2010, 1411–1434.
178. Barone, G.; Gomez-Paloma, L.; Duca, D.; Silvestri, A.; Riccio, R.; Bifulco, G. *Chem. Eur. J.* **2002**, 8, 3233–3239.
179. Barone, G.; Duca, D.; Silvestri, A.; Gomez-Paloma, L.; Riccio, R.; Bifulco, G. *Chem. Eur. J.* **2002**, 8, 3240–3245.
180. Rosselli, S.; Bruno, M.; Maggio, A.; Bellone, G.; Formisano, C.; Mattia, C. A.; Di Micco, S.; Bifulco, G. *Eur. J. Org. Chem.* **2007**, 2007, 2504–2510.
181. Dyson, B. S.; Burton, J. W.; Sohn, T.-i.; Kim, B.; Bae, H.; Kim, D. *J. Am. Chem. Soc.* **2012**, 134, 11781–11790.
182. Smith, S. G.; Paton, R. S.; Burton, J. W.; Goodman, J. M. *J. Org. Chem.* **2008**, 73, 4053–4062.
183. Chini, M. G.; Jones, C. R.; Zampella, A.; D'Auria, M. V.; Renga, B.; Fiorucci, S.; Butts, C. P.; Bifulco, G. *J. Org. Chem.* **2012**, 77, 1489–1496.
184. Tai, C.-K.; Yeh, P.-L.; Wu, Y.-S.; Shih, T.-L.; Wang, B.-C. *J. Mol. Struct.* **2014**, 1068, 84–93.
185. Yeh, P.-L.; Tai, C.-K.; Shih, T.-L.; Hsiao, H.-L.; Wang, B.-C. *J. Mol. Struct.* **2012**, 1018, 64–71.
186. Subramaniam, G.; Patel, V.; Karimi, S. *Spectrosc. Lett.* **2008**, 41, 349–353.
187. Steinmetz, W. E.; Robustelli, P.; Edens, E.; Heineman, D. *J. Nat. Prod.* **2008**, 71, 589–594.
188. Allenmark, S.; Gawronski, J. *Chirality* **2008**, 20, 606–608.
189. Stephens, P. J.; Devlin, F. J.; Cheeseman, J. R.; Frisch, M. J.; Rosini, C. *Org. Lett.* **2002**, 4, 4595–4598.
190. Polavarapu, P. L. *Chirality* **2002**, 14, 768–781.

191. Freedman, T. B.; Cao, X.; Oliveira, R. V.; Cass, Q. B.; Nafie, L. A. *Chirality* **2003**, *15*, 196–200.
192. Stephens, P. J.; Devlin, F. J.; Cheeseman, J. R.; Frisch, M. J.; Bortolini, O.; Besse, P. *Chirality* **2003**, *15*, S57–S64.
193. Zuber, G.; Goldsmith, M.-R.; Beratan, D. N.; Wipf, P. *Chirality* **2005**, *17*, 507–510.
194. Batista Jr, J. M.; Batista, A. N. L.; Rinaldo, D.; Vilegas, W.; Cass, Q. B.; Bolzani, V. S.; Kato, M. J.; López, S. N.; Furlan, M.; Nafie, L. A. *Tetrahedron: Asymmetry* **2010**, *21*, 2402–2407.
195. Polavarapu, P. L. *Chem. Rec.* **2007**, *7*, 125–136.
196. Autschbach, J. *Chirality* **2009**, *21*, E116–E152.
197. Zhao, S.-D.; Shen, L.; Luo, D.-Q.; Zhu, H.-J. *Curr. Org. Chem.* **2011**, *15*, 1843–1862.
198. He, F.; Nugroho, A. E.; Wong, C. P.; Hirasawa, Y.; Shiota, O.; Morita, H.; Aisa, H. A. *Chem. Pharm. Bull.* **2012**, *60*, 213–218.
199. Almeida, C.; Hemberger, Y.; Schmitt, S. M.; Bouhired, S.; Natesan, L.; Kehraus, S.; Dimas, K.; Gütschow, M.; Bringmann, G.; König, G. M. *Chem. Eur. J.* **2012**, *18*, 8827–8834.
200. Xue, M.; Zhang, Q.; Gao, J.-M.; Li, H.; Tian, J.-M.; Pescitelli, G. *Chirality* **2012**, *24*, 668–674.
201. Debbab, A.; Aly, A. H.; Edrada-Ebel, R.; Wray, V.; Pretsch, A.; Pescitelli, G.; Kurtan, T.; Proksch, P. *Eur. J. Org. Chem.* **2012**, *2012*, 1351–1359.
202. Sun, P.; Huo, J.; Kurtán, T.; Mándi, A.; Antus, S.; Tang, H.; Draeger, S.; Schulz, B.; Hussain, H.; Krohn, K.; Pan, W.; Yi, Y.; Zhang, W. *Chirality* **2013**, *25*, 141–148.
203. Hirasawa, Y.; Hara, M.; Nugroho, A. E.; Sugai, M.; Zaima, K.; Kawahara, N.; Goda, Y.; Awang, K.; Hadi, A. H. A.; Litaudon, M.; Morita, H. *J. Org. Chem.* **2010**, *75*, 4218–4223.
204. Kwit, M.; Rozwadowska, M. D.; Gawronski, J.; Grajewska, A. *J. Org. Chem.* **2009**, *74*, 8051–8063.
205. Taniguchi, T.; Martin, C. L.; Monde, K.; Nakanishi, K.; Berova, N.; Overman, L. E. *J. Nat. Prod.* **2009**, *72*, 430–432.
206. Bringmann, G.; Bruhn, T.; Maksimenka, K.; Hemberger, Y. *Eur. J. Org. Chem.* **2009**, *2009*, 2717–2727.
207. Grkovic, T.; Ding, Y.; Li, X.-C.; Webb, V. L.; Ferreira, D.; Copp, B. R. *J. Org. Chem.* **2008**, *73*, 9133–9136.

208. Tartaglia, S.; Padula, D.; Scafato, P.; Chiummiento, L.; Rosini, C. *J. Org. Chem.* **2008**, *73*, 4865–4873.
209. Kwit, M.; Sharma, N. D.; Boyd, D. R.; Gawronski, J. *Chirality* **2008**, *20*, 609–620.
210. Crawford, T. D.; Tam, M. C.; Abrams, M. L. *J. Phys. Chem. A* **2007**, *111*, 12057–12068.
211. Nagakura, Y.; Nugroho, A. E.; Hirasawa, Y.; Hosoya, T.; Rahman, A.; Kusumawati, I.; Zaini, N.; Morita, H. *J. Nat. Med.* **2013**, *67*, 381–385.
212. Tedesco, D.; Zanasi, R.; Wainer, I. W.; Bertucci, C. *J. Pharm. Biomed. Anal.* **2014**, *91*, 92–96.
213. Wang, Z.; Zhao, L.; Chen, Y.; Xu, W.; Sun, T. *Eur. J. Org. Chem.* **2014**, *2014*, 3814–3821.
214. Sherer, E. C.; Lee, C. H.; Shpungin, J.; Cuff, J. F.; Da, C.; Ball, R.; Bach, R.; Crespo, A.; Gong, X.; Welch, C. J. *J. Med. Chem.* **2014**, *57*, 477–494.
215. Gussem, E. D.; Bultinck, P.; Feledziak, M.; Marchand-Brynaert, J.; Stevens, C. V.; Herrebout, W. *Phys. Chem. Chem. Phys.* **2012**, *14*, 8562–8571.
216. Devlin, F. J.; Stephens, P. J.; Figadère, B. *Chirality* **2009**, *21*, E48–E53.
217. Stephens, P. J.; Pan, J.-J.; Devlin, F. J.; Urbanová, M.; Hájíček, J. *J. Org. Chem.* **2007**, *72*, 2508–2524.
218. Stephens, P. J.; Pan, J. J.; Devlin, F. J.; Krohn, K.; Kurtán, T. *J. Org. Chem.* **2007**, *72*, 3521–3536.
219. Cairns, E.; Hashmi, M. A.; Singh, A. J.; Eakins, G.; Lein, M.; Keyzers, R. *J. Agric. Food Chem.* **2015**, *63*, 7421–7427.
220. Schrödinger, *Schrödinger LLC New York, NY* **2014**,
221. Schrödinger, *Schrödinger, LLC, New York, NY* **2013**,
222. MacroModel version 10.9, *Schrödinger LLC New York, NY* **2015**,
223. Spartan 10, *Wavefunction Inc., Irvine. CA* **2015**,
224. Hanwell, M. D.; Curtis, D. E.; Lonie, D. C.; Vandermeersch, T.; Zurek, E.; Hutchison, G. R. *J. Cheminform.* **2012**, *4*, 17–17.
225. Pedretti, A.; Villa, L.; Vistoli, G. *J. Mol. Graph. Model.* **2002**, *21*, 47–49.
226. Frisch, M. J. et al. *Gaussian Inc.; Wallingford CT* **2010**,
227. CYLview, 1.0b,; Legault, C. Y. *Université de Sherbrooke* **2009**, (<http://www.cylview.org>).
228. Racine, J. *J. Appl. Econom.* **2006**, *21*, 133–141.
229. Adamo, C.; Barone, V. *J. Chem. Phys.* **1999**, *110*, 6158–6170.

230. Perdew, J. P.; Burke, K.; Ernzerhof, M. *Phys. Rev. Lett.* **1996**, 77, 3865–3868.
231. Perdew, J. P.; Burke, K.; Ernzerhof, M. *Phys. Rev. Lett.* **1997**, 78, 1396–1396.
232. Grimme, S. *J. Comput. Chem.* **2006**, 27, 1787–1799.
233. Grimme, S.; Antony, J.; Ehrlich, S.; Krieg, H. *J. Chem. Phys.* **2010**, 132, 154104.
234. Grimme, S.; Ehrlich, S.; Goerigk, L. *J. Comput. Chem.* **2011**, 32, 1456–1465.
235. Grimme, S. *Ange. Chem. Int. Ed.* **2006**, 45, 4460–4464.
236. Weigend, F.; Ahlrichs, R. *Phys. Chem. Chem. Phys.* **2005**, 7, 3297–3305.
237. Tomasi, J.; Mennucci, B.; Cammi, R. *Chem. Rev.* **2005**, 105, 2999–3094.
238. Marenich, A. V.; Cramer, C. J.; Truhlar, D. G. *J. Phys. Chem. B* **2009**, 113, 6378–6396.
239. Cheeseman, J. R.; Trucks, G. W.; Keith, T. A.; Frisch, M. J. *J. Chem. Phys.* **1996**, 104, 5497–5509.
240. Gauss, J. *Berichte der Bunsengesellschaft für physikalische Chemie* **1995**, 99, 1001–1008.
241. Gauss, J. *J. Chem. Phys.* **1993**, 99, 3629–3643.
242. Stratmann, R. E.; Scuseria, G. E.; Frisch, M. J. *J. Chem. Phys.* **1998**, 109, 8218–8224.
243. Hashmi, M. A.; Lein, M. *MICE-PES, Scripts and User Guide* **2018**, <https://github.com/matthiaslein/WellFARe>.
244. Fulmer, G. R.; Miller, A. J. M.; Sherden, N. H.; Gottlieb, H. E.; Nudelman, A.; Stoltz, B. M.; Bercaw, J. E.; Goldberg, K. I. *Organometallics* **2010**, 29, 2176–2179.
245. Boppré, M.; Colegate, S. M.; Edgar, J. A. *J. Agric. Food Chem.* **2005**, 53, 594–600.
246. Wiedenfeld, H.; Edgar, J. *Phytochem. Rev.* **2011**, 10, 137–151.
247. Mattocks, A. R. *Nature* **1968**, 217, 723–728.
248. Singh, B.; Sahu, P.; Singh, S. *Fitoterapia* **2002**, 73, 153–155.
249. da Silva Negreiros Neto, T.; Gardner, D.; Hallwass, F.; Leite, A. J. M.; de Almeida, C. G.; Silva, L. N.; de Araújo Roque, A.; de Bitencourt, F. G.; Barbosa, E. G.; Tasca, T.; Macedo, A. J.; de Almeida, M. V.; Giordani, R. B. *Biomed. Pharmacother.* **2016**, 83, 323–329.
250. Bruhn, T.; Schaumlöffel, A.; Hemberger, Y.; Bringmann, G. *Chirality* **2013**, 25, 243–249.
251. Goodman, J. M.; Still, W. C. *J. Comput. Chem.* **1991**, 12, 1110–1117.

252. Dasyam, N.; Munkacsi, A. B.; Fadzilah, N. H.; Senanayake, D. S.; O'Toole, R. F.; Keyzers, R. A. *J. Nat. Prod.* **2014**, 77, 1519–1523.
253. Sullivan, G. R.; Dale, J. A.; Mosher, H. S. *J. Org. Chem.* **1973**, 38, 2143–2147.
254. Bagno, A.; Saielli, G. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2015**, 5, 228–240.
255. Choi, J.; Ha, H.-J. *J. Korean Chem. Soc.* **2015**, 59, 203–204.
256. Hashmi, M. A.; Andreassend, S. K.; Keyzers, R. A.; Lein, M. *Phys. Chem. Chem. Phys.* **2016**, 18, 24506–24510.
257. Kim, M. C.; Kwon, H. C.; Kim, S. N.; Kim, H. S.; Um, B. H. *Chem. Pharm. Bull.* **2011**, 59, 834–838.
258. Yu, J.; Cao, Y.; Song, H.; Wang, X.; Yao, S. *J. Serb. Chem. Soc.* **2012**, 77, 887–898.
259. Simpson, S.; Autschbach, J.; Zurek, E. *J. Chem. Edu.* **2013**, 90, 656–660.
260. Baranowska-Łączkowska, A.; Łączkowski, K. Z. *J. Comput. Chem.* **2013**, 34, 2006–2013.
261. Autschbach, J. *ChemPhysChem* **2011**, 12, 3224–3235.
262. Dasyam, N. Identification of Anti-tubercular Compounds in Marine Organisms from Aotearoa. Ph.D. thesis, School of Chemical and Physical Sciences, Victoria University of Wellington, 2014; <http://researcharchive.vuw.ac.nz/handle/10063/3629>.
263. Asbury, A. N. Toward the Synthesis of Rimarikiamide A. M.Sc. thesis, School of Chemical and Physical Sciences, Victoria University of Wellington, 2017; <http://researcharchive.vuw.ac.nz/handle/10063/6443>.