Biomolecular Recognition Event Detection via Applied Electrical Engineering Theories

A Thesis
Presented to
The Academic Faculty

by

David Price Thompson

In Partial Fulfillment
Of the Requirements for the Degree
Bachelor of Science in the
School of Electrical and Computer Engineering

Georgia Institute of Technology
April 2020

Biomolecular Recognition Event Detection via Applied Electrical Engineering Theories

Approved by:

William D. Hunt, Ph.D.
School of Electrical and Computer Engineering
Georgia Institute of Technology

Ryan S. Westafer, Ph.D.
Georgia Tech Research Institute
Georgia Institute of Technology

Date Approved: [April 29, 2020]

Acknowledgements

I am thankful for everyone who has helped and aided me along my academic path. I would like to thank my advisor and research lead, Dr. William Hunt, for providing the opportunity to be a part of his research team as well as his continual teaching and mentoring. His foundational research lays an easy course to follow with endless development possibilities ahead. I thank Ph.D. candidate Daniel Canales and his unending patience as I work both for and alongside him on the team. Thanks to Elaine Rhoades, always keeping the team on track, along with the rest of the team for always being a team.

I offer thanks to Dr. Ryan Westafer for the duality of critiques and conversing as a reviewer and a scientist. I wish to thank Dr. Tobias Wilson-Bates, bringing excitement and joy into teaching otherwise dry research writing. Thanks to Dr. Thomas Gaylord and librarian Bette Finn with the hours they spend honing and refining research engineers. A thank you to Allyson Tant who would take back a 1st place award when appropriate. Special thanks to Director Rick Clark and the Georgia Tech Undergraduate Admissions department for believing one more chance was worth extending.

Last, I proffer never-ending thanks to my wife Frances without whose love, support, and care I would not be where I am today, and to our daughter Elizabeth whose empathy always shows no bounds.

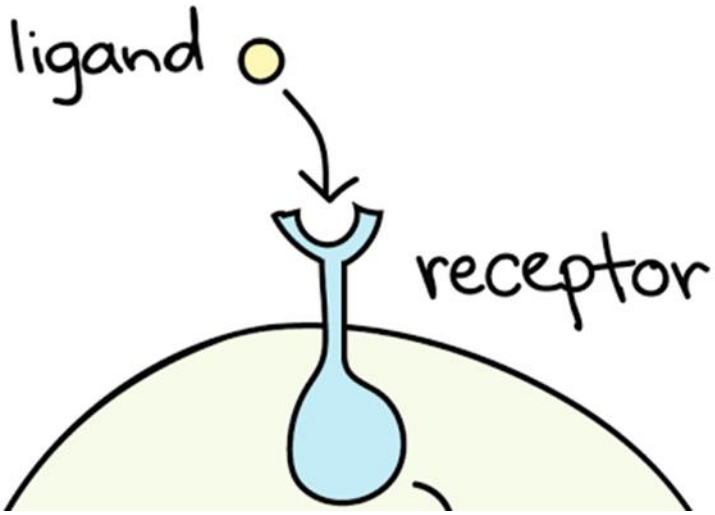
Table of Contents

Abstract

Multiple methods are used today to recognize biomolecular interactions, often utilizing thermodynamic and chemical procedures. These methods have been well established over decades of use. However, they tend to require numerous interactions and therefore may miss interactions such as those between a single ligand and receptor cell. Application of electrostatic and electrical engineering theories may provide a novel approach to identification of biomolecular recognition events. At a very basic level every molecule is comprised of individual atoms where each atom has a specific electrical charge. Observing these charges from points at short distances outside the molecule generates an electrostatic field pattern. This pattern may provide a unique potential signature to any given molecule and therefore biomolecular interactions as well. Electrical engineering, through applied radar signaling theory, provides methods to detect specific electromagnetic patterns against a background of noise. If we can identify and characterize a unique electrostatic potential pattern for a given interaction, then utilization of radar theory through matched filters may one day permit recognizing individual biomolecular events, eliminating the current requirement of numerous interactions for biomolecular recognition.

Introduction

Cellular assembly, gene expression, disease transmission, and drug interactions all occur through the non-covalent bonding between a ligand and a receptor. These bindings are known as biomolecular recognition events and constitute a key part of life as we understand it [1]. Learning more about biomolecular recognition events helps us to know how, when, and why these cellular processes occur. Identification and characterization of these events can



help tailor drugs to increase potency while reducing side effects, to better understand how cells carry out basic processes such as replication, or to understand why some are immune to a disease that others remain susceptible to. To aid these efforts, numerous methods for identifying biomolecular recognition events exist today, spanning the fields of biology, chemistry, and thermodynamics.

While numerous identification methods exist, most of them are limited through the requirements of either a significant quantity of the system under evaluation or modification of the system such as with added markers [2]. Recent efforts utilizing the realms of molecular dynamics, electrochemical, electrodynamic, and computer simulations have developed identification methods beyond the existing chemical, biological, and physical ones. Yet many of these newer methods incur their own limitations such as requiring thermal excitation or direct contact [2-4].

Electrical engineering may seem far removed from the fields of biology and chemistry. However, basic field theories regarding electrostatic potentials and applied techniques involving pattern matching might lead to biomolecular recognition event detection at a distance. Somewhat similar to how radar uses matched filters to detect an object at a significant distance, we may be able to detect molecules based on their constituent electric fields. A search of available research has uncovered scant investigation of electrostatic characteristics as a distinguishable trait for identifying biomolecular recognition events. Combined, applying matched filters to a molecule's electric charge field could theoretically provide a unique, identifiable pattern which might be observable at a distance.

Every molecule has a specific chemical structure which may be optimized to a ground energy state, utilizing the least amount of energy to maintain its composition. This ground state provides a stable set of atomic locations, bond lengths, and the angles between these bonds. The resultant geometric pattern of a molecule's individual atoms can be further associated with known atomic electric charge.

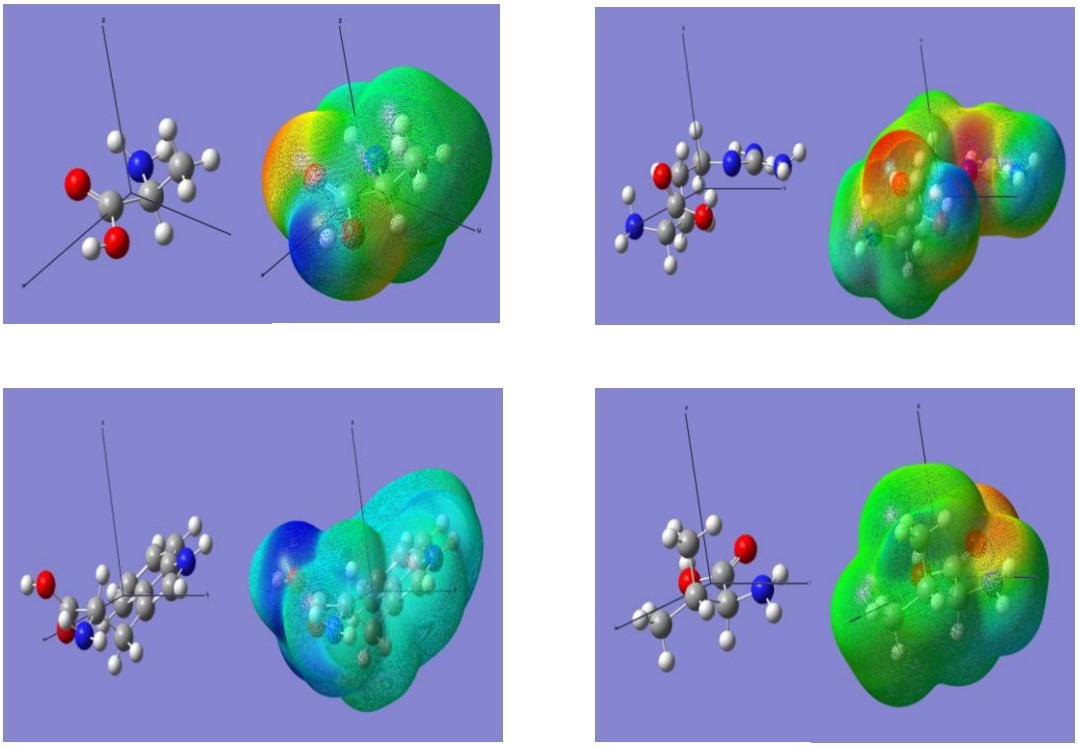


Figure 1: GaussView electrostatic potential plots of alanine, arginine, tryptophan, and valine.

To investigate whether such a concept is feasible requires an initial confirmation that an electrostatic pattern for a given molecule exists and is unique. For this confirmation, the proteinogenic amino acids provide an

extensive, well researched body of molecules with many known characteristics and traits. Proteinogenic amino acids are important building blocks in many chemical and biological processes within the body. As examples, the body uses the amino acid tryptophan to synthesize the neurotransmitter serotonin and glycene is known to bind with a number of ionotropic and metabotropic receptors. This subgroup of molecules can be used to derive electrostatic potential patterns and is considered of sufficient size and similarity to confirm this new identifier uniquely matches the specific amino acid under characterization.

This study will complete a geometric optimization to a ground energy state for each of the 21 proteinogenic amino acids including specification of a molecular center to orient around. This electrostatic geometric pattern will be used to construct a visual 'constellation map' which provides a unique representation that identifies the individual amino acid characterized. Future use of this constellation map should permit application of pattern matching from radar theory [3, 4] to identify biomolecular recognition events at a distance.

Identification of a unique electrostatic pattern observable at a distance, albeit likely a small distance measured in angstroms, should eliminate many of the limitations inherent in existing methods of biomolecular recognition event

detection. Detection at a distance would eliminate the need for direct contact, specifically, "binding." Use of electrostatic field theory removes the requirements for added markers or thermal excitation which each modify the system under observation. In theory, if a detection system can be built around these concepts, such would permit more accurate identification of biomolecular recognition events. This system under consideration would prove essential in investigations into relatively unknown biological processes, general understanding of disease transmission, and rapid evaluation of new drugs that both increase potency while reducing undesirable side effects.

Literature Review

A key aspect of biological life, biomolecular recognition occurs when a pair of cells complete non-covalent binding [1] - such as when a ligand and a receptor on two proteins match and bind together. Ligand and receptor bindings directly result in gene expression, cellular assembly, disease transmission, and drug interaction among other cellular processes and identification of these binding events permits further investigation into their resultant biologic processes. Given their importance, numerous fields across biology, chemistry, and physics have developed methods for identifying biomolecular recognition between cells

including use of thermodynamics, spectroscopy, nuclear magnetic resonance, and microarray techniques.

When observing a system, modification of that system is typically avoided due to potential effects resultant from such changes. Sadik et al. reports the majority of methods used in identifying biomolecular recognition events require adding a form of marker which can affect the interactions [2]. These markers include the use of radio-, enzymatic-, or fluorescent-labeling to detect biomolecular recognition events. The use of such labelling permits ultrasensitive detection while potentially changing the event itself in ways not immediately known or predictable. Labelling also requires additional steps thereby increasing the potential for error.

More recent efforts have since spanned fields including molecular dynamics, electrochemical, and computer simulations. However, while often leading to improvements in biomolecular recognition event detection capabilities, these methods still include undesirable effects upon the systems observed.

Considering specifically electrochemical methods we find efforts utilizing detection of potential (voltage), current, and resistance with all methods referenced requiring direct contact [2]. Hunt et al. report utilization of electrical

engineering techniques including digital radio derived signal pattern matching for biomolecular identification with experimental results generated from direct contact with chip-based sensors [3, 4]. This method eliminates concerns with marking however the direct contact requirements still limit detection capabilities as each is selective to the molecule(s) detected, the quantity required for detection, and biomaterial build-up decreasing sensor reliability. More recent experimentation utilizing nanomaterials has led to reduced concerns regarding direct contact through improved biocompatibility, greater detection surface areas, and sensor density [2, 5]. Where use of nanotechnology mitigates concerns, most limitations from direct-contact requirements remain.

Electrodynamics theory potentially provides identification of biomolecular recognition events at a distance through cellular resonance. A paper by Hejase de Trad, Fang, and Cosic reports discrete signal techniques based upon the resonant recognition model (RRM) [6]. The RRM underwent notable evaluation and research about two decades ago [6, 7] and a few years ago Preto et al. completed a mathematical exploration of whether electromagnetic resonances can extend to long-distances as a part of biomolecular interactions. The latter model utilized terahertz wavelengths permitting detection beyond short-distances. A notable limitation to this method is that the oscillation is shown not to occur at thermal

equilibrium and requires excitation of the biomolecules desired for identification [8]. Each of these methods require detection of discrete oscillations which may not be detectable at thermal equilibrium.

Electrostatic potential exists due to atomic charges within biomolecules without the need for specific frequency oscillations, eliminating the excitation required by electrodynamic methods. Perhaps for distance-based detection, electrostatic forces may perform differently from electrodynamic forces. If so, then application of modern electrical engineering theory to biochemical interaction could present a novel method for characterizing biomolecular recognition events.

In this study , we intend to extend the radio theory work of Hunt et al. [3, 4] in the detection of atomic scale electrostatic potentials. We will evaluate the electrostatic potential of proteinogenic amino-acids to determine whether unique identifiers exist for each molecule, suitable for pattern matching efforts. Confirmation of such identifying patterns could lead to distance-based identification of biomolecular recognition without significantly modifying the biosystem under observation.

Methods & Materials

To investigate whether distance-based identification of biomolecular recognition events is feasible, we first consider our overall goal. The general concept is to determine whether a molecule's electrostatic potential field provides a unique identifier for that molecule. Additionally, the intent is to use such a unique identifier later as a matched filter. We therefore determined to start out with characterizing a set of molecules through their electrostatic properties to build a dataset for a uniqueness comparison. As a set of molecules, we selected the 21 proteinogenic amino acids. We expect that from selecting to use these protein-creating amino acids that future derived research will benefit from the existent large base of information and data for verification of comparative simulation results and in setting up real-world experiments. Once characterized, we then apply properties of spherical harmonics to generate what we have termed a "constellation map" as a visual identifier of each molecule's electrostatic configuration.

Our first step is to obtain the molecular structure for each amino acid we wish to characterize. MolView [9], an open-source web-based data visualization platform, completes molecular data searches through multiple compound, protein, and spectral databases including The PubChem Project [10] and the RCSB

Protein Data Bank [11]. We searched MolView for a given amino acid, for example tryptophan, and selected the primary result. Via the online Tools menu we exported a .MOL file which includes 3D spatial coordinates for each atom within the molecule and the bonds between the atoms.

The resultant atomic geometric structure included in a compound's .MOL file is unlikely to be in an optimized state. The software tool Gaussian [12] can read in .MOL file data and compute an optimized molecular structure per parameters input into the tool. Using Gaussian 16, we opened the GaussView tool and opened a given .MOL file as its input. We modified the Gaussian Calculation Setup defaults as follows:

- Use the Optimization job type for the minimum
- Set density function theory (DFT) for the ground state
- Basis set 6-31G(d) with + diffuse function
- Merz-Kollman partial charge distributions
- Save ESP charges for later
- Calculate the electrostatic properties with additional keyword "prop"

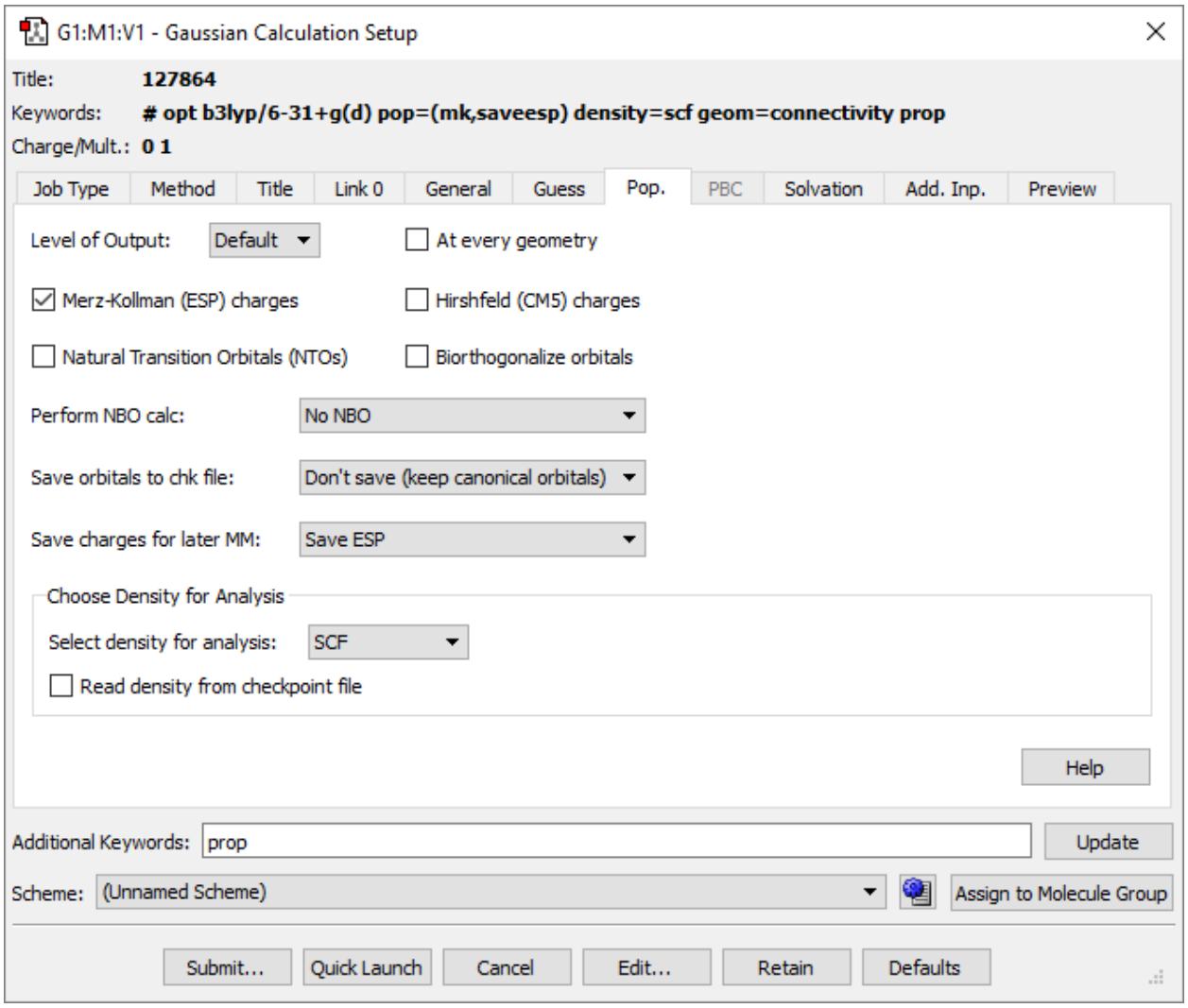


Figure 2: Example of Gaussian calculation setup screen.

Once run, Gaussian outputs a .LOG file with data representing an atomic geometry for the molecule in its computed minimum energy state.

The .LOG contains data including calculated atomic locations representative to the other atoms within the molecule. Additionally, the .LOG file provides the electrostatic charge found at each atomic location.

To utilize the .LOG file data, the team developed a series of custom MATLAB [13] scripts. The first of these scripts parses the atomic geometry and electrostatic charge for each atom. A second set of scripts algorithmically solves for the q_l^m values of a molecule using the charge distribution read from the Gaussian log-file and an input degree to calculate, *l*. The script sets the q_l^m order *m* as -*l* to *l*. With the q_l^m value calculations, a final set of scripts generates the visual "constellation map" of q_l^m vs *m* vs *l* and returns data files containing both the algorithmic calculated and Gaussian-reported dipole moments (x, y, z, normal) and a matrix of the data points used to construct the constellation diagram ($l,\ m, q_l^m,\ |q_l^m|,\ and\ phase(q_l^m)$).

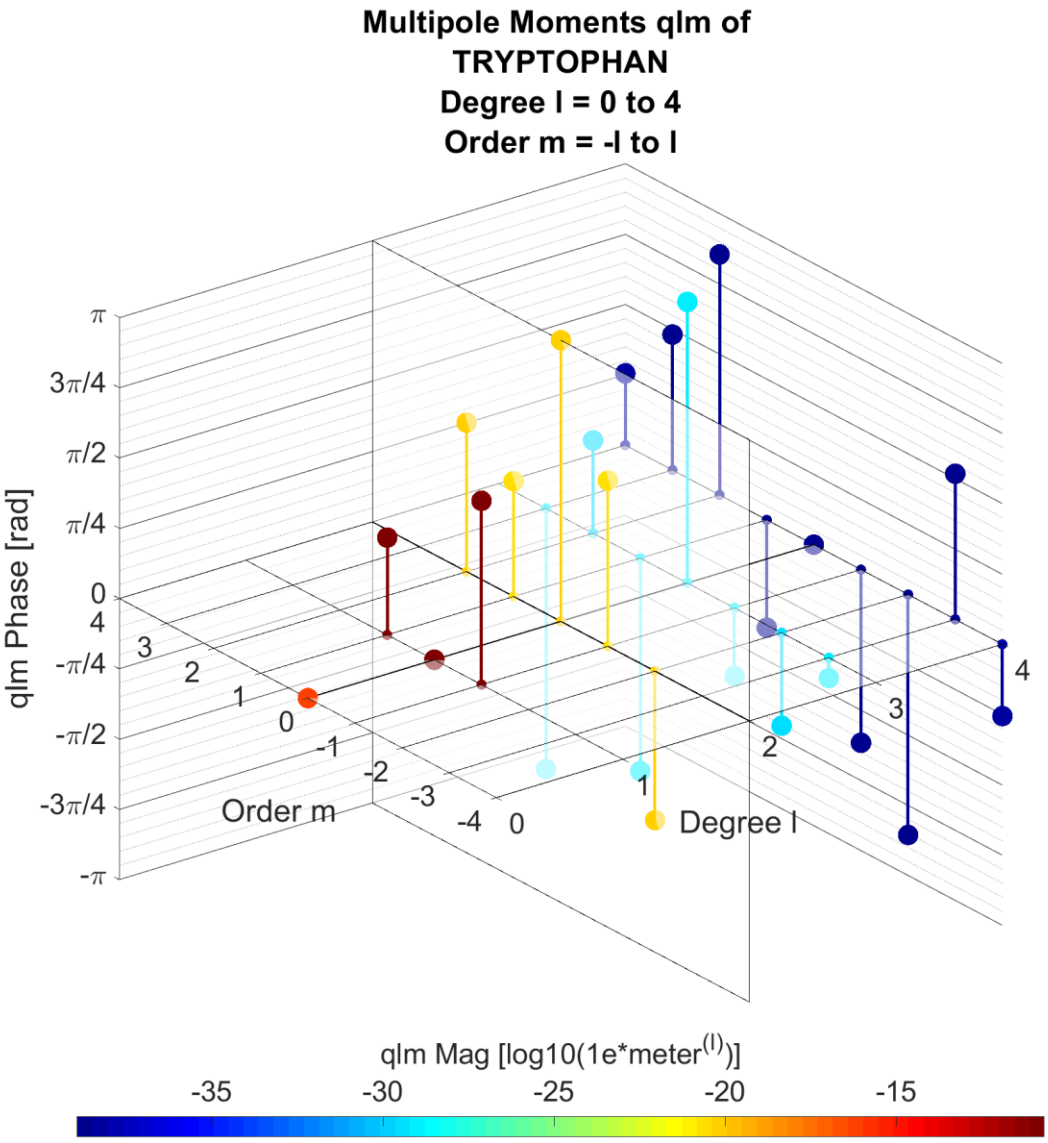


Figure 3: Constellation map for tryptophan amino acid.

Following this process, we then proceeded to generate constellation maps for each of the 21 proteinogenic amino acids as selected for direct comparison.

Findings

To characterize the amino acids, we generated constellation maps with the degree *l* set to 4 and the order *m* follows as -*l* to *l*. Through this method, every molecule was reduced to an individual characterization dataset containing 45 datapoints. Each of these datasets, however, contain 20 zero values leaving 25 non-zero values. This is readily visible in the constellation map as seen for tryptophan in Figure 3 and is an expected result of the algorithmic calculations. For comparison, a tryptophan molecule at the ground state contains 27 atoms each requiring three additional positional values (x/y/z) for a total of 108 non-zero values whereas the electrostatic potential map as displayed in Figure 1 includes thousands of datapoints. Characterizing these amino acids through q_l^m values alone appears to significantly reduce the dataset size necessary to identify a molecule.

The final step in the process in generating constellation maps provides graphical representations of the q_l^m datasets for the proteinogenic amino acids characterized in this study. Review of these maps confirmed through quick visual comparison that each molecule resolved to a unique set of q_l^m values. Each of the 21 maps are clearly distinguishable from the other 20. As desired, for this sample

set of 21 molecules, the constellation maps provide a unique characterization for each amino acid.

As an example set, the constellation maps for four molecules are included in Figure 4. Whereas visually alanine, tryptophan, and valine each appear to share the same result for degree *l*=0, arginine is already distinctly different. Valine then immediately diverges at degree *l*=1 in comparison to alanine and tryptophan. At degree *l*=2, these remaining two molecules diverge from one another. None of the entire set of 21 constellation maps appear visually matched for either *l* =3 or *l* =4. Numerical analysis of the q_l^m values confirms the visual inspection results across the entire set of proteinogenic amino acids.

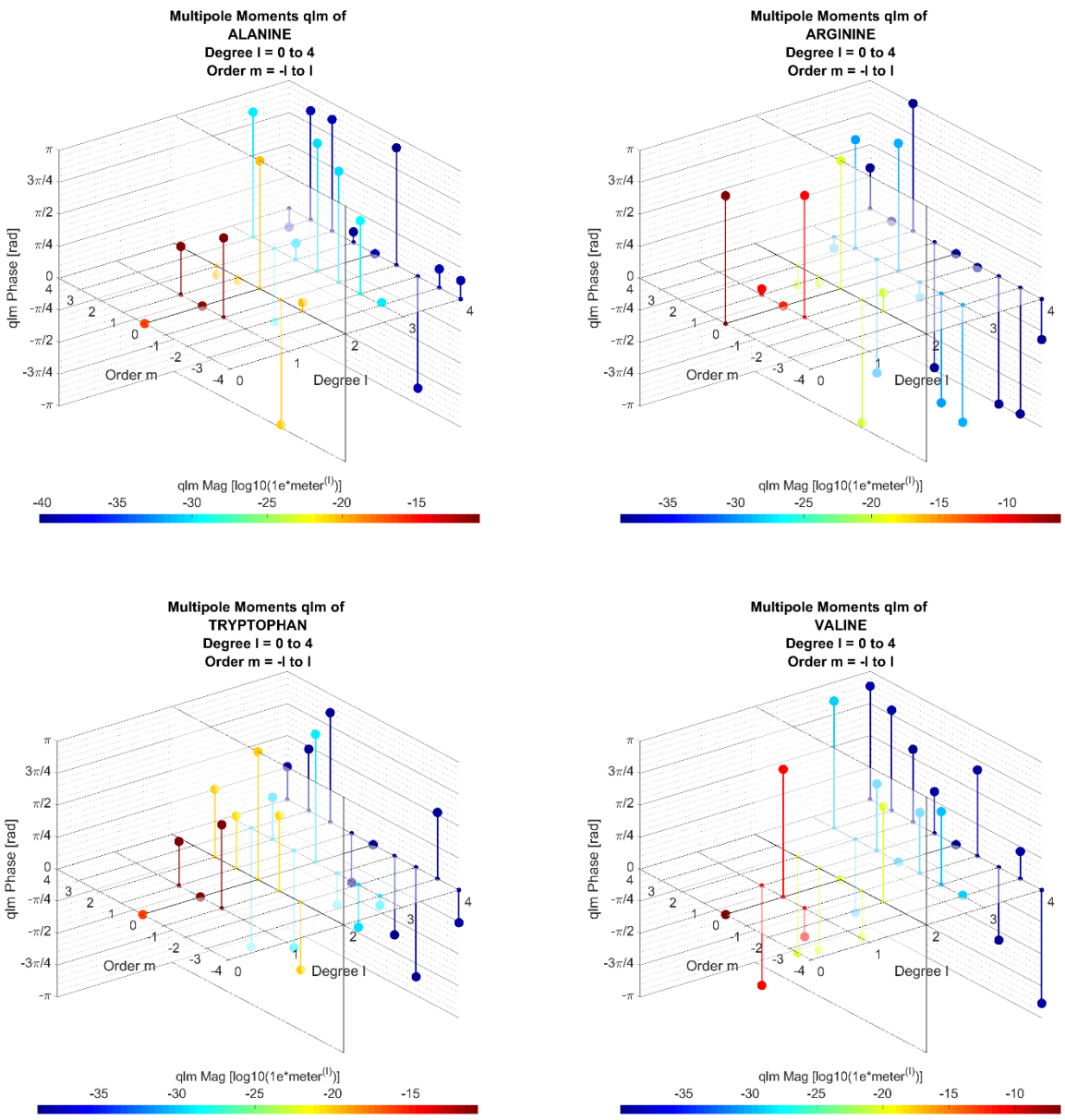


Figure 4 - Constellation plots of alanine, arginine, tryptophan, and valine.

As an initial confirmation using the proteinogenic amino acids as a subset of all molecules, this method generates a unique, condensed constellation map for each molecule.

Discussion

Applying theories derived within electrical engineering to the task of biomolecular recognition event detection is a multi-step process. At the very least, utilizing radar theory for recognition at a distance first requires a pattern to match. Further research may identify whether the constellation map generated with this method may provide a useful pattern. Ultimately, whether using electrodynamics to identify biomolecular recognition events proves of use remains a ways off. Today's calculations might result in physical observations in time. As a first step, the results of this study are promising towards confirming a unique, identifiable dataset generated based solely upon an atom's electric fields.

Continuing in this research path should continue to resolve whether the constellation map delivers unique identifications. Generating and reviewing maps for a greater variety of molecules would be one task, to extend the range beyond amino acids. Utilizing the amino acids as a subset provided several molecules, such as isoleucine and threonine, which exhibit similarities through matching chiral centers. However, reviewing true chiral pairs like limonene and carvone would help extend verification into the constellation map's capabilities for unique characterizations. Along similar lines, running this process for 'known' molecules

set into different ground states or viewed from differing orientations should help further knowledge regarding the method's use within different constraints.

In its current state, the constellation map does result in an apparently unique pattern for characterization of the proteinogenic amino acids. Much like how each amino acid's written name uniquely identifies the molecule, these compact q_l^m value representations also identify molecules. Yet the underpinning of the constellation map is the molecule's intrinsic electrostatic charge as opposed to letters of the English alphabet. Perhaps one day this physical underpinning, observing electrostatic charges at a distance, may prove the key to identifying molecules, and thereby, biomolecular recognition events, without unduly altering the systems of which they are a part.

References

[1] J. A. McCammon, "Theory of biomolecular recognition," *Curr. Opin. Struct. Biol.,* vol. 8, pp. 245-249, Apr. 1998.

[2] O. A. Sadik, A. O. Aluoch, and A. Zhou, "Status of biomolecular recognition using electrochemical techniques," *Biosens. Bioelectron.,* vol. 24, pp. 2749-2765, May 15, 2009.

[3] W. D. Hunt, P. J. Edmonson, D. D. Stubbs, and S.-H. Lee, "Molecular recognition in the digital radio domain," *Jpn. J. Appl. Phys.,* vol. 49, p. 07HA02 (9 pp.), Jul. 20, 2010.

[4] W. D. Hunt, S.-H. Lee, D. D. Stubbs, and P. J. Edmonson, "Clues from digital radio regarding biomolecular recognition," *IEEE Trans. Biomed. Circuits Syst.,* vol. 1, pp. 50-5, Mar. 2007.

[5] W.-Z. Shen, S. Cetinel, and C. Montemagno, "Application of biomolecular recognition via magnetic nanoparticle in nanobiotechnology," *J. Nanopart. Res.,* vol. 20, pp. 1-8, May 5, 2018.

[6] C. Hejase de Trad, Q. Fang, and I. Cosic, "The resonant recognition model (RRM) predicts amino acid residues in highly conserved regions of the hormone prolactin (PRL)," *Biophys. Chem.,* vol. 84, pp. 149-157, Apr. 14 2000.

[7] I. Cosic, *The resonant recognition model of macromolecular bioactivity : theory and applications*. Basel; Boston; Berlin: Birkhäuser Basel, 1997.

[8] J. Preto, M. Pettini, and J. A. Tuszynski, "Possible role of electrodynamic interactions in long-distance biomolecular recognition," *Phys. Rev. E,* vol. 91, p. 052710 (18 pp.), May 22, 2015.

[9] MolView, The MolView Open-Source Project, Netherlands, http://molview.org/.

[10] S. Kim, P. A. Thiessen, E. E. Bolton, J. Chen, G. Fu, A. Gindulyte, L. Y. Han, J. E. He, S. Q. He, B. A. Shoemaker, J. Y. Wang, B. Yu, J. Zhang, and S. H. Bryant, "PubChem substance and compound databases," *Nucleic Acids Res.,* vol. 44, pp. D1202-13, Jan. 4, 2016.

[11] H. M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov, and P. E. Bourne, "The Protein Data Bank," *Nucleic Acids Res.,* vol. 28, pp. 235-242, Jan. 1, 2000.

[12] Gaussian, Gaussian, Inc., Wallingford, CT, gaussian.com.

[13] MATLAB, Mathworks, Inc., Natick, MA, www.mathworks.com.