

Conformational Cartography:

Investigating how QM methods capture pyranose ring conformational landscapes

June 19, 2017

1 Introduction

1.1 Significance

Carbohydrates are a ubiquitous biological molecule found in nature that are essential for biological functions involved with human health[1] and as an attractive feedstock for a new generation of biofuels.[2, 3] The most abundant carbohydrate in the world, cellulose, is comprised of repeating $\beta(1\rightarrow4)$ linked D-glucose monomeric subunits.[1] During biofuel production of cellulosic biomass, conversion of cellulose into β -D-glucose occurs by cleaving the highly stable, cellulose glycosidic bonds. These glycosidic bonds are very stable in water (half-life of millions of years), thus require elevated temperatures (fast pyrolysis), highly acidic conditions, or enzymes to be efficiently break.[3, 4] For improved biofuel production, extensive research is being used to understand the breakdown of cellulose using glycosidic hydrolases (GHs).[2, 5, 6]

Advanced computational simulations are now providing researchers an improved, fundamental understanding of the inner workings of GH enzymes. Quantum mechanics and molecular mechanics (QM/MM) simulations have revealed important molecular mechanisms associated with the depolymerization of carbohydrates using enzymes.[7] (CHECK Kurihara 2009 5–8) In the study of family 6 Glycoside hydrolase (GH6) enzymes, Mayes et. al demonstrated that a network of hydrogen bonding aligns two active site water molecules to transfer the proton to the D175 residual (catalytic base) and the conformational catalytic itinerary of the β -glucose, pryanose ring at the -1 position.[7] The benefit of QM/MM simulations is their ability to simulate a variety of time- and length-scales to offer increased understanding of molecular level phenomena of carbohydrates in the presence of enzymes. The QM/MM system is divided into two: the active site region (the cellulose substrate and important residues), and the spectator region (the surrounding protein). The spectator region, is modeled with less-costly MM methods while the active site is modeled with more-accurate QM methods (hence the term QM/MM). MM methods are based solely on Newton's Laws of Motion, making them computationally inexpensive compared to the QM methods, and are necessary to study molecular motion of the surrounding protein. However, MM methods lack the ability to treat fundamental processes such as bond breakage and formation, and charge fluctuations due to geometry.[8] QM methods capture the electronic structure of the cellulose substrate to provide computational modeling of bond formation and breakage, essential in studying carbohydrate-enzyme interactions on a molecular level.

Selecting an appropriate QM method is a non-trivial task; a plethora of QM methods with varying accuracy and computational cost exist. A discussion on the differences between the different methods...

I. *ab initio*

II. Density Functional Theory (DFT)

A. DFT serves as the main work-horse of modern quantum-chemical modeling of materials and molecules.[9]
The DFT methods were selected in our study were based on work from Medvedev et al. and INSERT THE COUNTER ARTICLE DISCUSSED BY MEDVEDEV.

III. Semi-Empirical (SE) Methods

A. Approximations reduce the computational cost...making them more feasible for QM methods. Fit using exper-

imental data, which allows for reduced computational cost.

B. SE methods are derived from Hartree-Fock theory

IV. Density Functional Tight Binding (DFTB) Methods

- A. Unlike SE methods which are fit to experimental data...DFTB methods are fit to DFT methods making that they don't suffer from some of the same pitfalls that are associated with the SE methods.

The importance of selecting an adequate method when modeling carbohydrates has been studied by a variety of research groups. Work by Marianski et al. compares the electronic energies of SE methods, DFT, and *ab initio* methods. After developing their own benchmark data set for β -glucose, α -glucose, and α -maltose, single point energy calculations were performed on each local minima structure using a variety of QM methods to provide a direct comparison in electronic energy.[10] Parody plots between SE and *ab initio* methods show no correlation between the methods, demonstrating how SE methods, on average, fail to reproduce accurately the reference energy of ‘gold standard’ *ab initio* methods.[10] Biarnés et al. used *ab initio* pseudopotentials (DFT) metadynamic simulations to study the conformational free energy landscape of β -D-glucose as a function of the Cremer-Pople parameters.[11] Metadynamics allows for enhanced sampling of the conformational free energy landscape by forcing the system to leave regions of space that have already been explored.[12, 13] The eight local minima conformations found on the potential energy surface show favorable features for efficient catalysis, suggesting that the factors governing the conformational landscape are intrinsic properties of a single β -D-glucose monomeric unit.[11] Using a model system of the active site region, Pereira et al. studied the accuracy of a plethora of QM methods (DFT, SCC-DFTB, and SE) for the retaining glycosidases mechanism.[14] Pereira et al. showed that SCC-DFTB provides inaccurate energy barriers, while SE methods provide inaccurate energy barriers and geometry optimizations.[14] Mayes et al. looked at the conformational landscape of five biologically relevant carbohydrates, performing high level of theory geometry optimizations to both local minima and transition state structures and *ab initio* energy calculations.[15] To date, the work of Mayes et al. serves as the most extensive search of the conformational landscape.

1.2 Research Questions

There currently exists a gap in our knowledge about the transition states conformations on the free energy landscape; previous research only focuses on the local minima conformations. Here, we present a comprehensive study comparing the conformational landscapes generated by different aforementioned methods. We first determined whether optimizations with *ab initio*, DFT, and semi-empirical methods produce the same stable ring geometries for both local minima and transition states for three pyranose rings. Then, the relative energies of the local minima and transition state structures were compared to ‘gold standard’ *ab initio* methods. Last, we discuss whether or not these trends are true across all of the studied pyranose rings. Our work provides a systematic approach to quantify how the potential energy landscape differs depending on the method selected.

1.3 Pyranose Conformational Landscapes

1.3.1 General

Cyclic sugar molecules containing one oxygen atom and five carbon atoms are classified as containing a pyranose ring, and adopt a variety of conformations (also called puckering). The nomenclature used to classify these puckered ring conformations was developed by Schwartz [REFERENCE](#) and adopted by the International Union of Pure and Applied Chemistry (IUPAC).[REFERENCE](#) A total of 38 canonical conformations represent the unique ways the ring can be in a chair (C), envelope (E), half-chair (H), skew (S), and boat (B), with the numbers associated with each letter representing certain atoms in a particular position (up or down). Puckered conformations, however, do not always correspond directly to the canonical conformations. The puckers are often found as intermediates between two of the canonical designations.[11] There exists a variety of transforms to reduce the 18 cartesian coordinates associated with the ring atoms into comparable parameters.[16, 17, 18, 19] The most popular transformation, developed by Cremer and Pople, decomposes the ring atom Cartesian coordinates into spherical coordinates (ϕ , θ , Q) called the Cremer-Pople (CP) parameters. Only the ϕ and θ CP parameters are necessary to describe the ring puckering conformational landscape.[16] Utilizing the Cremer-Pople (CP) parameters to classify ring puckers provides a quantitative measurement to puckered conformations.

The exo-cyclic groups attached to the ring carbon atoms dictates the potential puckered conformations the pyranose adopts. As exo-cyclic group complexity increases, the puckered conformational landscape increases in diversity. The puckered conformations cyclic sugar adopt is dependent on the exo-cyclic groups attached to the ring carbon atoms.

2 Computational Methods

2.1 Generating Reference Landscapes

Previous work by Mayes et al. performed an extensive search of the conformational landscape of similar monosaccharides obtaining geometries vibrational frequencies at the B3LYP/6–311+G(2df,p) level of theory with an ultrafine integration grid and tight convergence, and the electronic energies at the CCSD–(T)/6–311+G(d,p) level of theory.[15] The extensive search identified unique local minima and transition state structures for each monosaccharide within our study, representing a highly diverse benchmark data set.

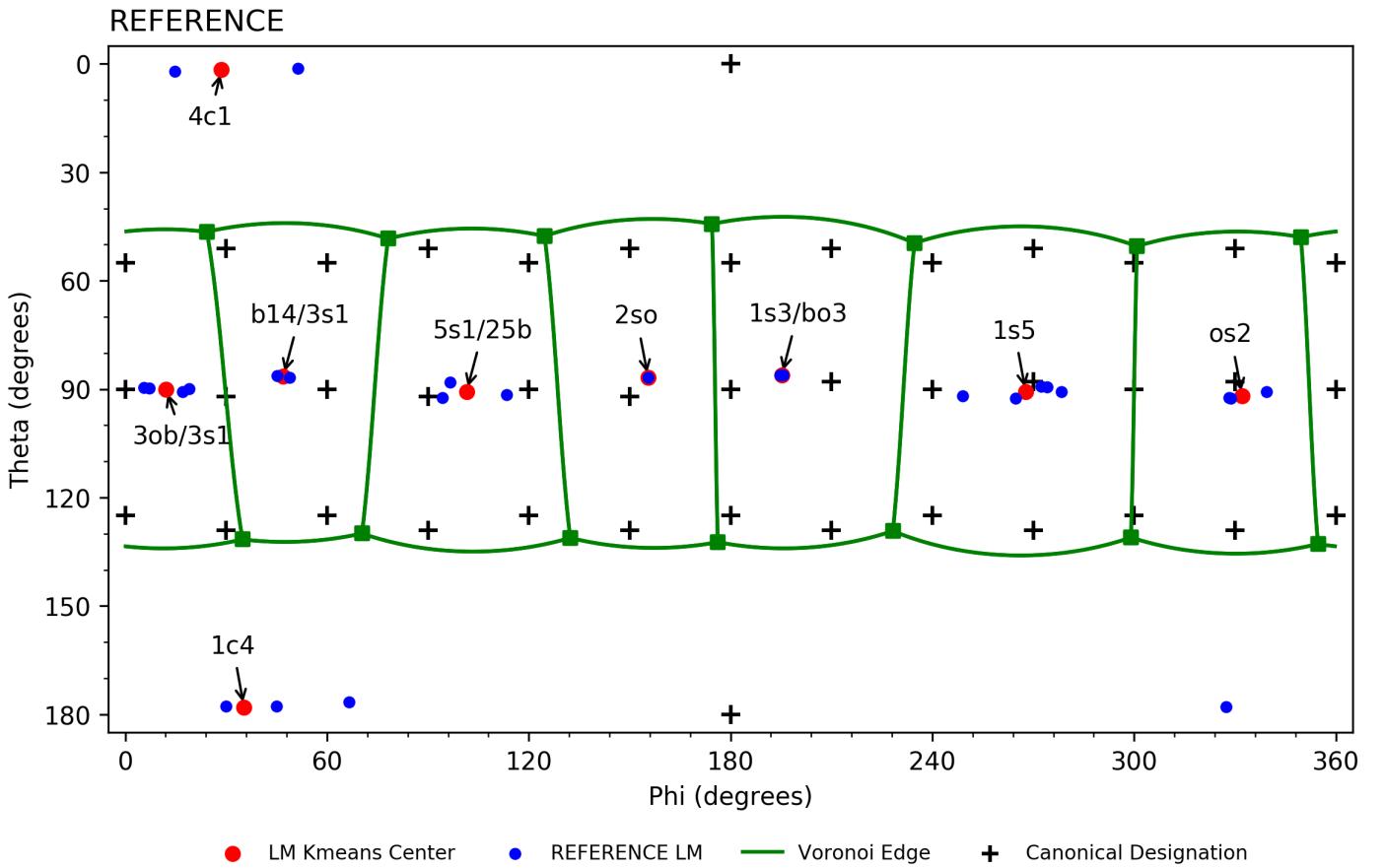


Figure 1: The LM reference landscape for β -xylose. The reference structures (blue) are shown with the resulting k-means clustering centers (red).

To compare the conformational landscapes generated by different methods, we utilized the conformational landscape generated by Mayes et al. [15] to assess the accuracy in producing stable local minima and transition state ring geometries structures and pathways by a particular QM method. Rather than making comparisons based on the canonical designations, clustering algorithms were imposed on the CP parameters of the stable ring geometries of Mayes et al. to generate reference clusters. The reference clusters served as the foundation for making comparisons between methods. For each molecule, a spherical k-means clustering algorithm[20] was first applied to the Mayes et al. local minima, resulting

in the local minima clusters for each molecule. Each spherical k-means center was assigned a canonical designation to identify that particular group. Figure 1 demonstrates the final LM k-means cluster groups for β -xylose. The voronoi edges correspond to points that are equidistant from two LM k-means centers allowing for the tessellation to be visualized.

Once the LM k-means centers were generated, the next phase involved clustering the TS on their associated LM. For a particular TS structure, the two LM were assigned to one of the LM k-means centers. The same LM k-means cluster algorithm[20] was then applied on each TS structure that connected two of the LM k-means centers (as shown in Figure 2). Figure 2 demonstrates that three unique pathways are available to convert between the 4c1 and 2so LM cluster groups. The same procedure was repeated for all possible combinations of LM groups to determine if there exists a pathway (or several) connecting the two. For a particular molecule, each of the LM k-mean centers and the TS k-mean centers can be found in the Supporting Information.

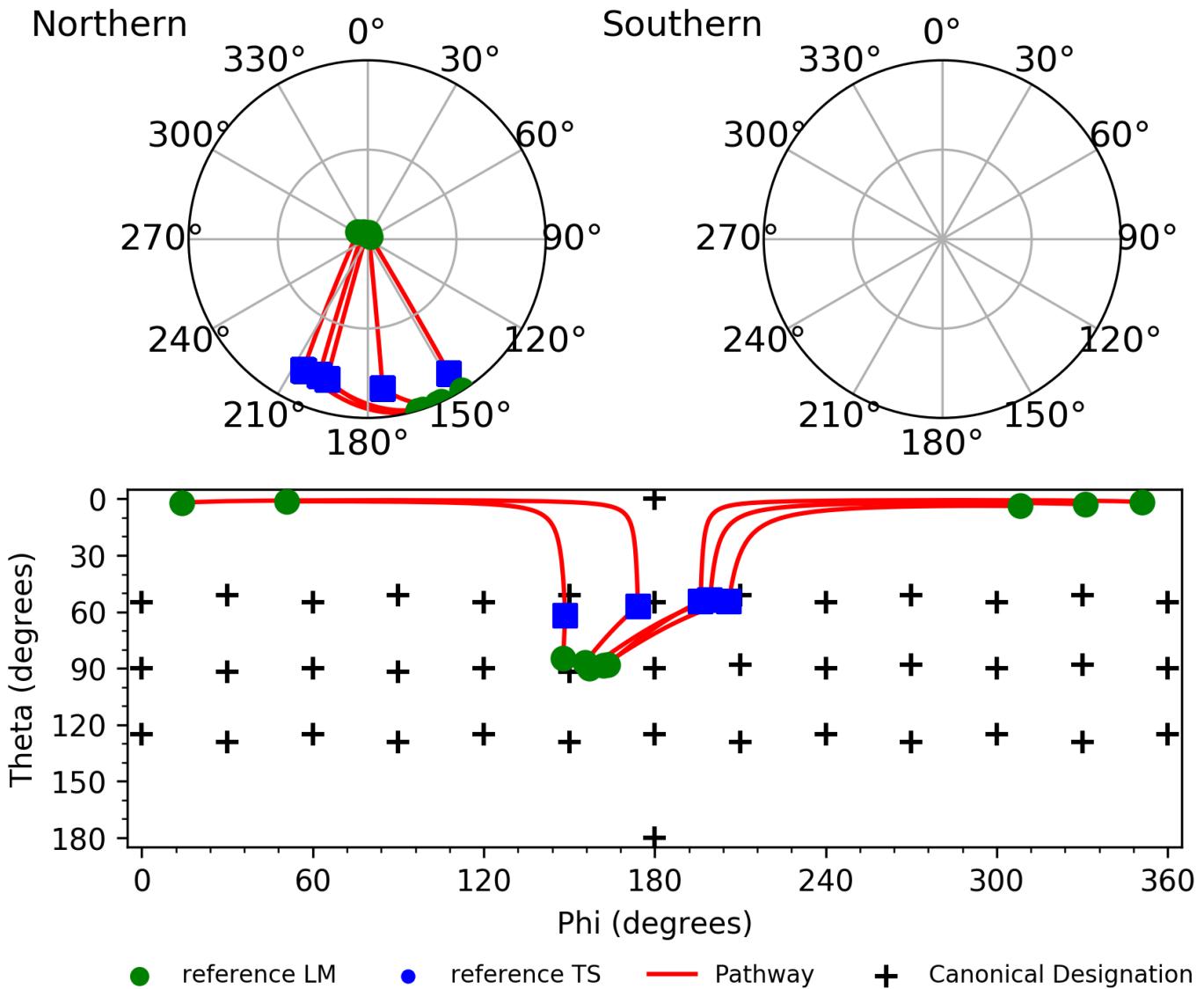


Figure 2: The TS reference landscape for β -xylose that connects the LM cluster groups 4c1 and 2so.

2.2 QM Calculations in Gaussian

The conformational landscapes of oxane, β -xylose, and β -glucose were determined using Gaussian 09 rev D.01.[21] To compare the landscapes across different methods, a consistent screening procedure was developed and maintained for all calculations. Rather than performing a similar exhaustive search for all molecules using each proposed QM method, the final geometries of the Mayes et al. served as our initial starting geometries for our work on oxane, β -xylose, and β -glucose.[15] The initial procedure involved optimizing the benchmark structures to local minima and transition state at each of the proposed QM methods. Verification of the local minima and transition state structures was performed by viewing the imaginary frequencies: local minima have no imaginary frequencies while transition state (first order saddle point) structures have exactly one imaginary frequency. For consistency, an ultrafine integration grid and tight convergence criteria was set for each method. For all DFT methods, a basis set of 6–31+G(d,p) was specified. The 6–31+G(d,p) basis set was selected for its polarization functions on all atoms and diffuse function on hydrogen atoms. Subsequent frequency calculations were performed to determine the thermodynamic quantities associated with each local minima and transition state structure. Python ([REFERENCE](#)) scripts were created to identify unique local minima and transition structures based on root mean square difference (RMSD) and dipole differences between structures. Duplicate local minima and transition state structures were removed to focus on unique structures and reduce computational cost.

The next phase of the screening process involved determining whether or not the transition state was a ring-puckering or exo-cyclic transition state using normal mode analysis in Gaussian.[21] Ring-puckering transition states were determined if the ring dihedral contribution to the imaginary normal mode exceeded a specific tolerance value. A definitive distinction between ring-puckering or exo-cyclic transition states was visible upon analysis of the contributions. Additionally, structures were checked using ChemCraft to visually verify the ring-puckering designation.[22] The exo-cyclic transition states were removed from further analysis to focus on the interconversion pathways between pucksers. With the ring-puckering transition states identified, intrinsic reaction coordinate (IRC) calculations were performed to find the pathways interconnecting local minima through a particular transition state.

2.3 Comparison Metrics

The conformational landscape generated by each QM method was compared to the extensive, benchmark conformational landscape by Mayes et al. The following metrics were used to quantify difference in conformation (structure) and energy for the various landscapes.

I. Root mean square difference (RMSD) within a cluster group.

A. Quantifies the variation in the structures generated by a particular method to the benchmark structures based solely on CP parameters. This method is less than ideal since the energy of the structures are not incorporated into the calculation.

B. Equation:

$$RMSD = \sqrt{\frac{1}{N} \sum_{i=1}^N (d_i)^2}$$

where i is one of N structures located within a particular clustering group and d is the arc length between point i and the k-means cluster center.

II. Weighted root mean square difference (WRMSD) with a cluster group.

- A. Quantifies the variation in the structures generated by a particular method to the benchmark structures based on the CP parameters and the energy of the structures. Low energy structures that are closer to the benchmark k-mean center are rewarded, whereas low energy structures that are further away from the benchmark k-mean center are penalized.

- B. Equation:

$$WRMSD = \sum_{i=1}^N \frac{W_i(d_i)^2}{N}$$

where i is one of N structures located within a particular clustering group, d is the arc length for structure i and the k-means cluster center, and W_i is the Boltzmann weighting factor for structure i .

III. Boltzmann Weighted Relative Gibbs Free Energies

- A. Calculation for the Boltzmann Weighted Relative Gibbs Free energies within each cluster group.

- B. Equation:

$$\Delta G_W = \sum_{i=1}^N \Delta G_i W_i = \sum_{i=1}^N \Delta G_i \left(\frac{e^{-\Delta G_i / (k_B T)}}{\sum_{j=1}^N e^{-\Delta G_j / (k_B T)}} \right)$$

where ΔG_i relative Gibbs Free energy for structure i , k_B is Boltzmann's constant, T is the temperature (in Kelvin), ΔG_j is the relative Gibbs Free energy for structure j of N structures within the cluster group.

IV. Root mean square difference within a cluster group (based on the Weighted Boltzmann Relative Gibbs Free energies)

- A. Quantifies the variation in the predicted Gibbs Free Energies to the reference Gibbs Free energies (within each cluster group).

- B. Equation:

$$GRMSD = \sum_{i=1}^N \frac{(\Delta G_{W_i} - \Delta G_{REF_i})^2}{N}$$

where ΔG_{W_i} is the weighted Gibbs Free Energy for the i cluster group, ΔG_{REF_i} is the reference Gibbs Free Energy for the i cluster group, and N represents the total number of clusters. Note that the RMSD calculations only incorporate matches; all false positives (the QM method finds a structure that the reference method missed) and false negatives (the QM method missed a structure that the reference method contains) are removed from the calculation.

3 Preliminary Results

3.1 Oxane

Oxane, the simplest pyranose ring, served as the foundation for our conformational investigation since it contains no exo-cyclic groups. All calculations were initially performed on oxane, and then expanded to the other molecules in our investigation.

3.1.1 Local Minima Reference Landscape

To compare the conformational landscapes across different methods, we initially generated a set of reference landscapes based on the work of Mayes et al. (as discussed in the Computational Methods: Generating Reference Landscapes section). Figure 3 shows the LM k-mean centers (red) for the reference structures as well as the voronoi edges (green) and the raw data from Mayes et al. (blue).

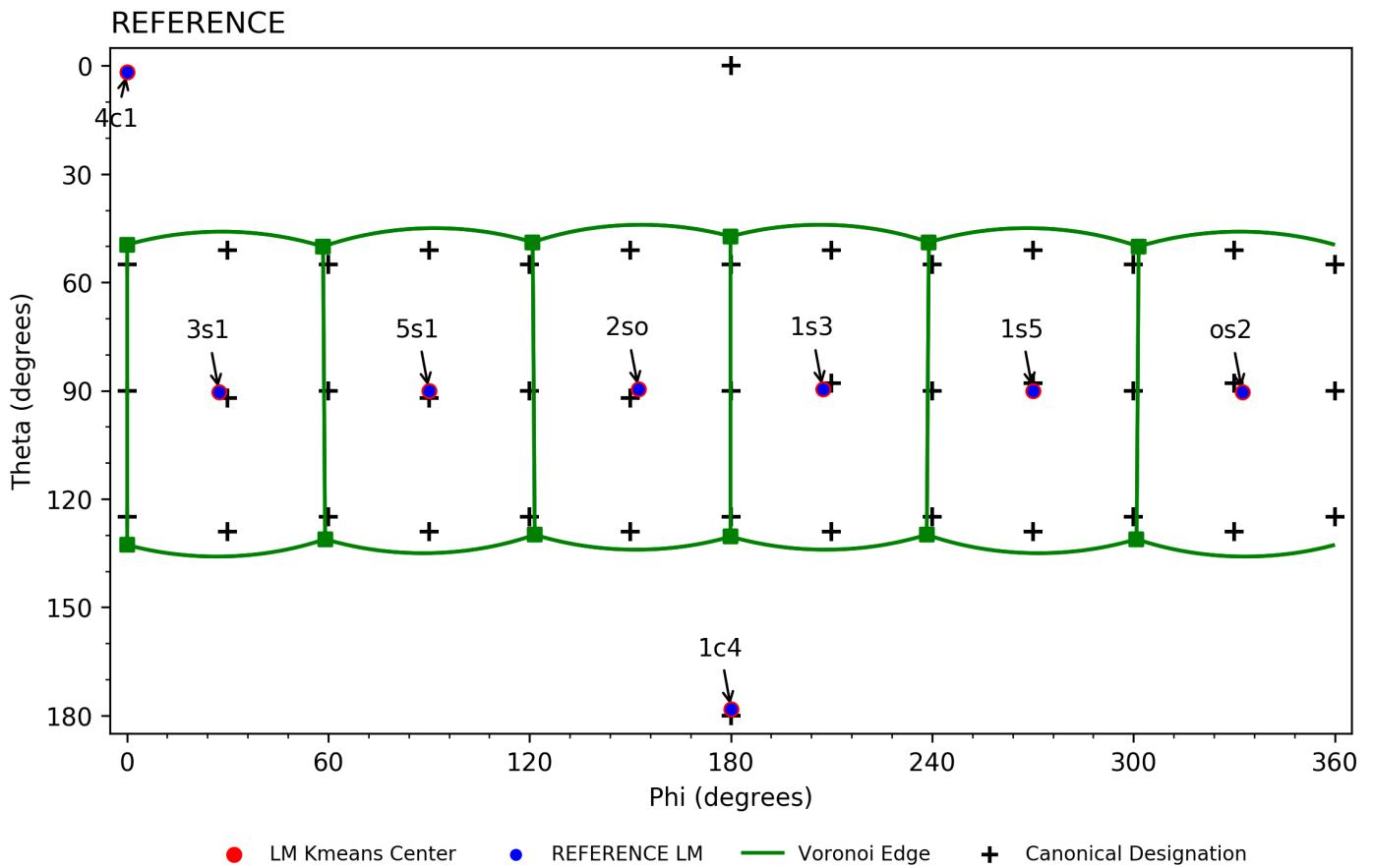


Figure 3: The LM reference landscape for oxane.

3.1.2 Transition State Reference Landscape

With the LM conformational landscape for oxane generated, the TS connecting each LM were then determined. Each TS found in Figure 4 represents a unique structure that connects two LM. To aid in visualization, Northern and Southern hemisphere plots were created to highlight key pathways from 4c1 and 1c4 conformations.

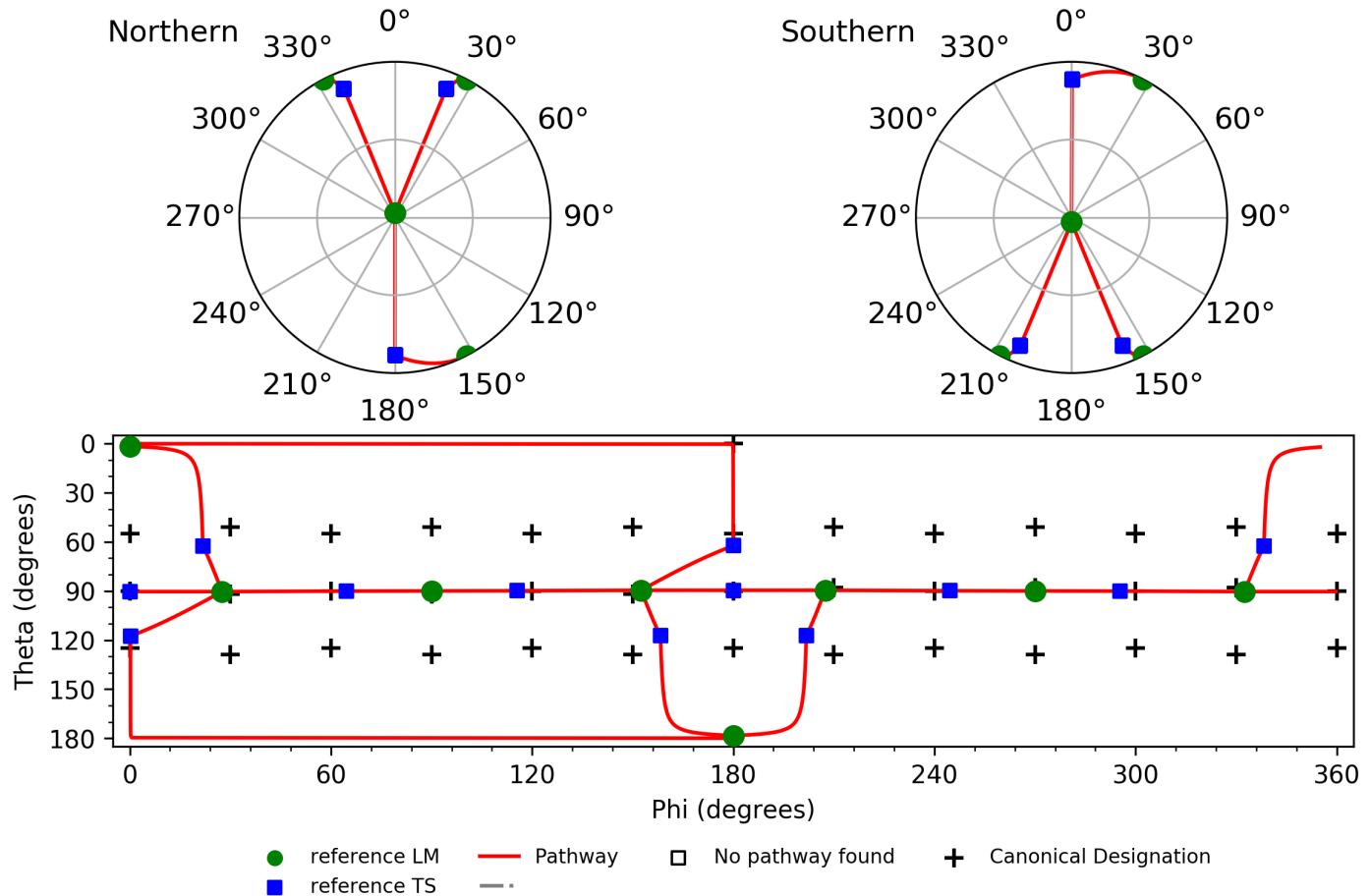


Figure 4: The TS reference landscape for oxane.

3.1.3 Local Minima Landscapes

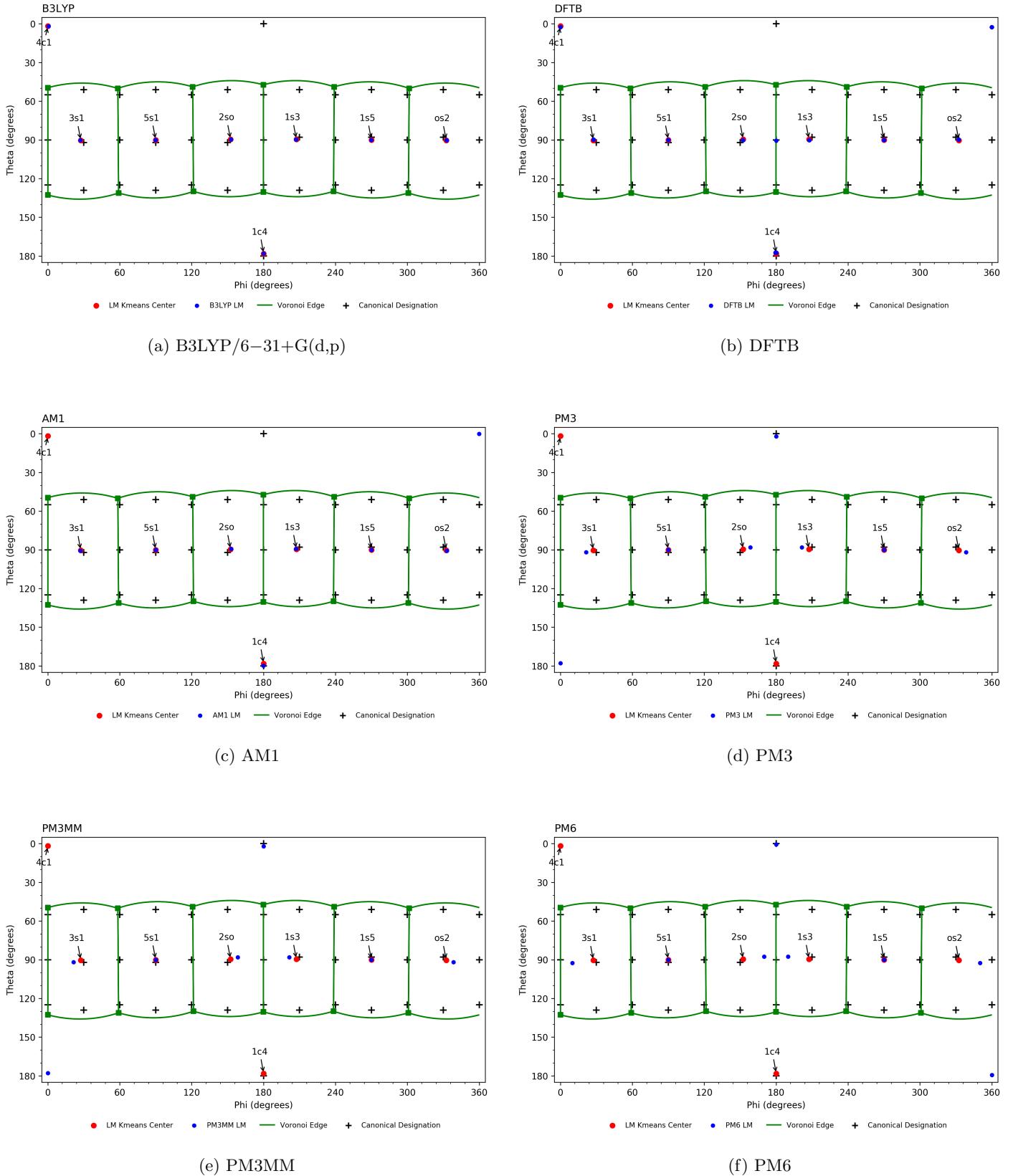


Figure 5: Insert Caption

3.1.4 Transition State Landscapes

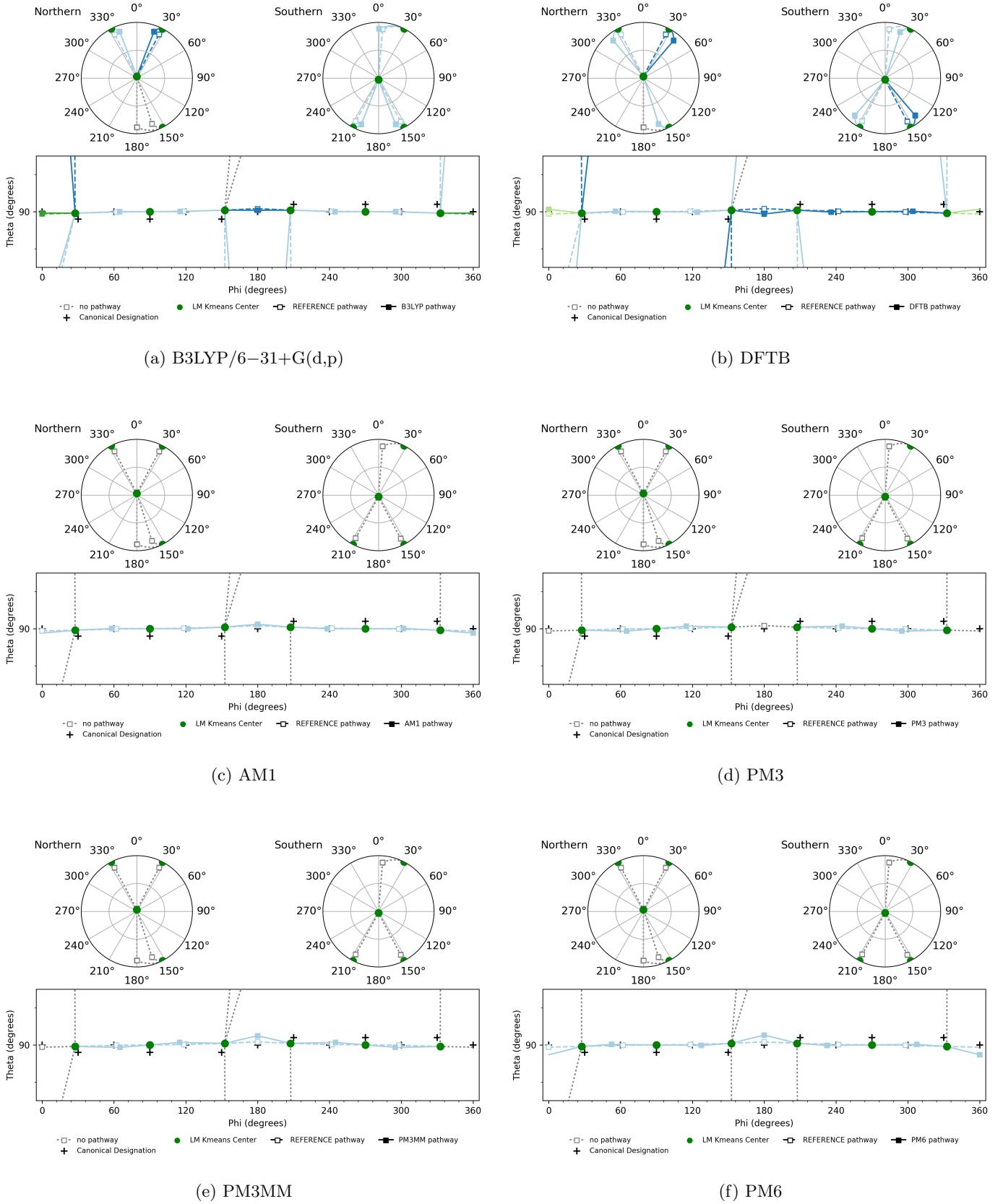


Figure 6: Insert Caption

3.1.5 WRMSD Table

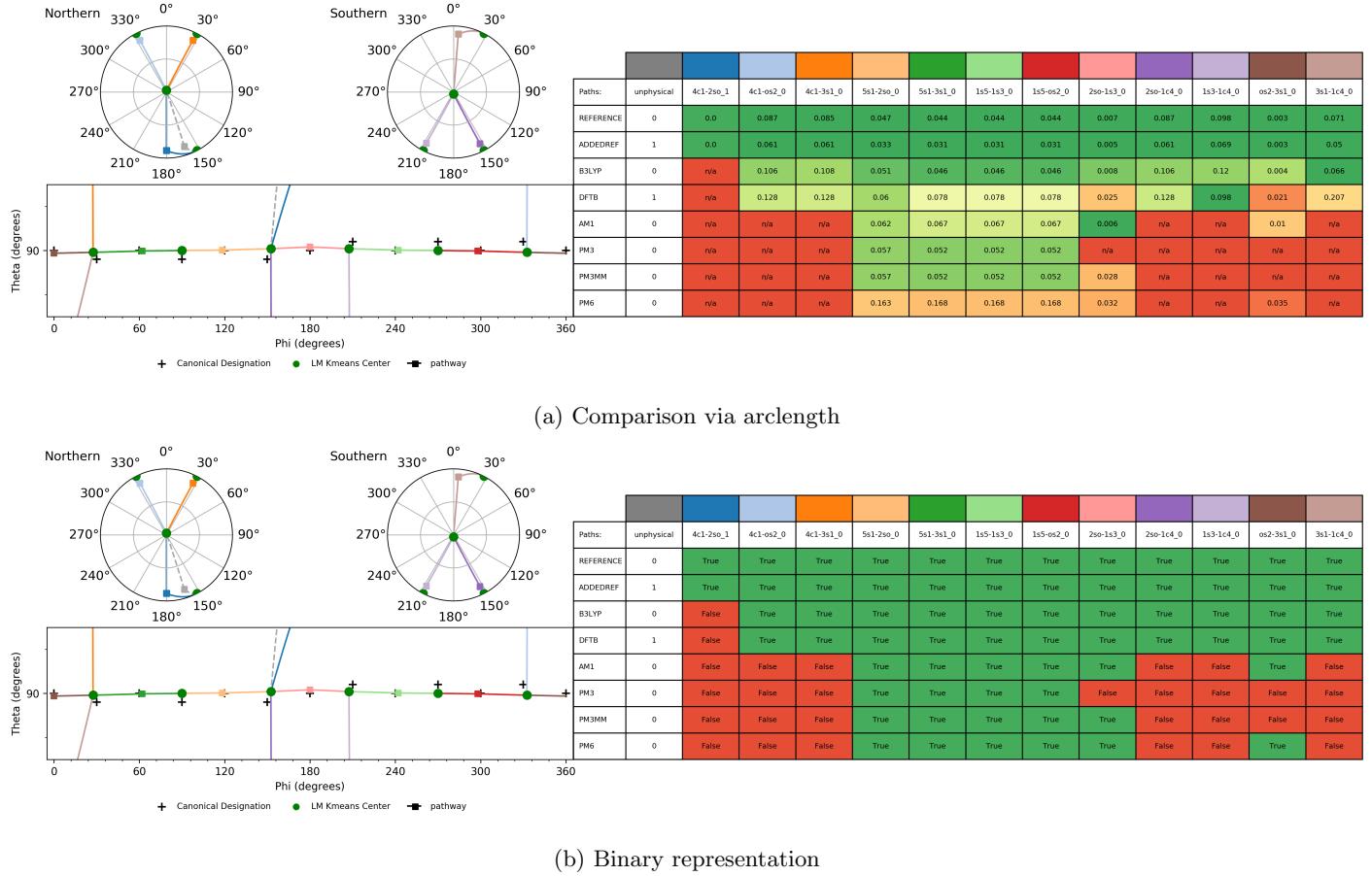


Figure 7: Insert Caption

3.1.6 Relative Gibbs Free Energy Comparison Table

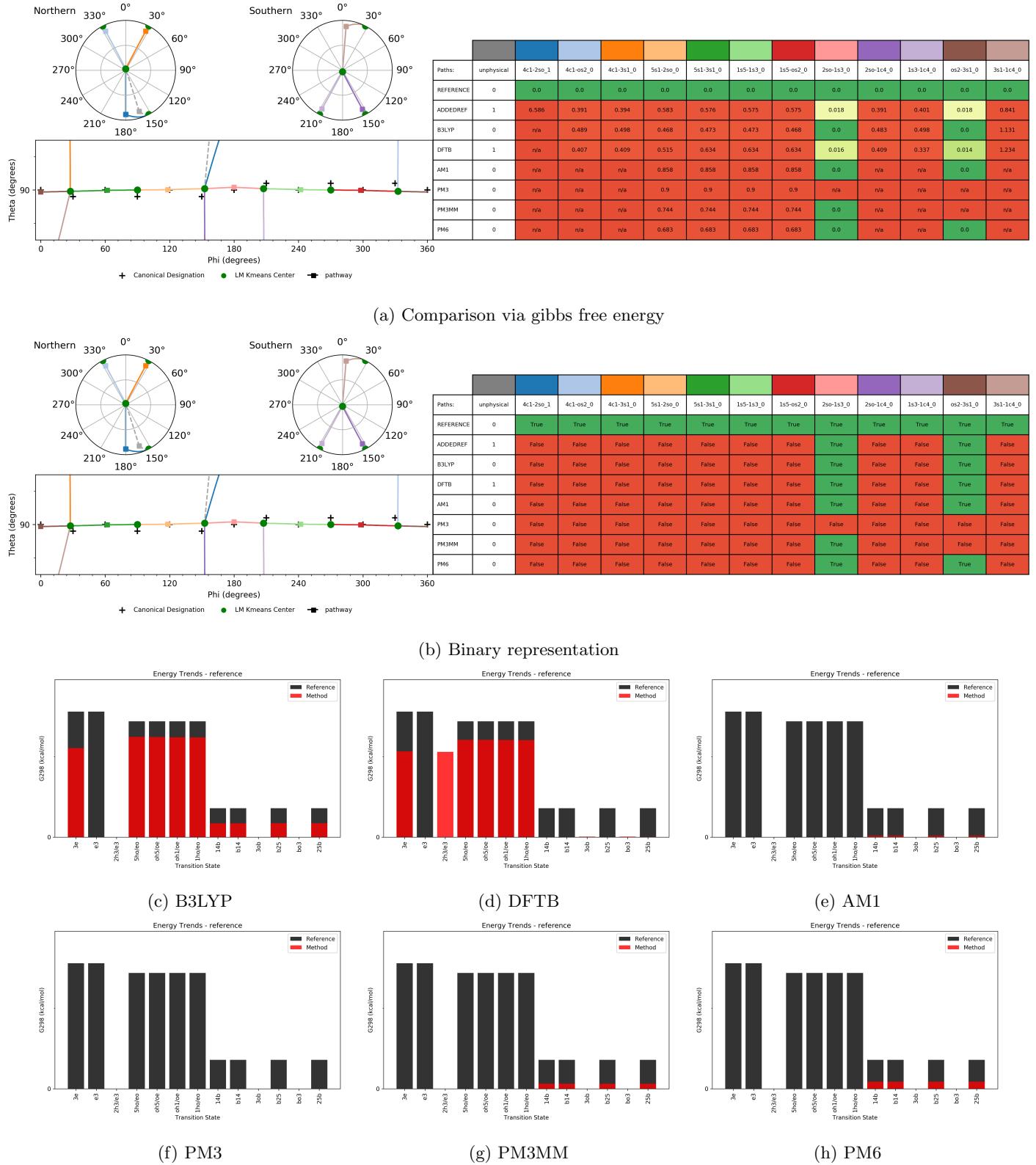


Figure 8: Insert Caption

3.2 β -Xylose

The next conformational landscape explored was β -xylose, which contains four hydroxyl groups. The addition of these hydroxyl groups allows for the ring to adopt a variety of conformations.

3.2.1 Local Minima Reference Landscape

To compare the conformational landscapes across different methods for β -xylose, we initially generated a set of reference landscapes based on the work of Mayes et al. (as discussed in the Computational Methods: Generating Reference Landscapes section). Figure 9 shows the LM k-mean centers (red) for the reference structures as well as the voronoi edges (green) and the raw data from Mayes et al. (blue). For β -xylose, nine LM k-means centers were used to capture the conformational landscape of Mayes et al. The max WRMSD for the reference landscape (5s1/25b) has a value of 0.092.

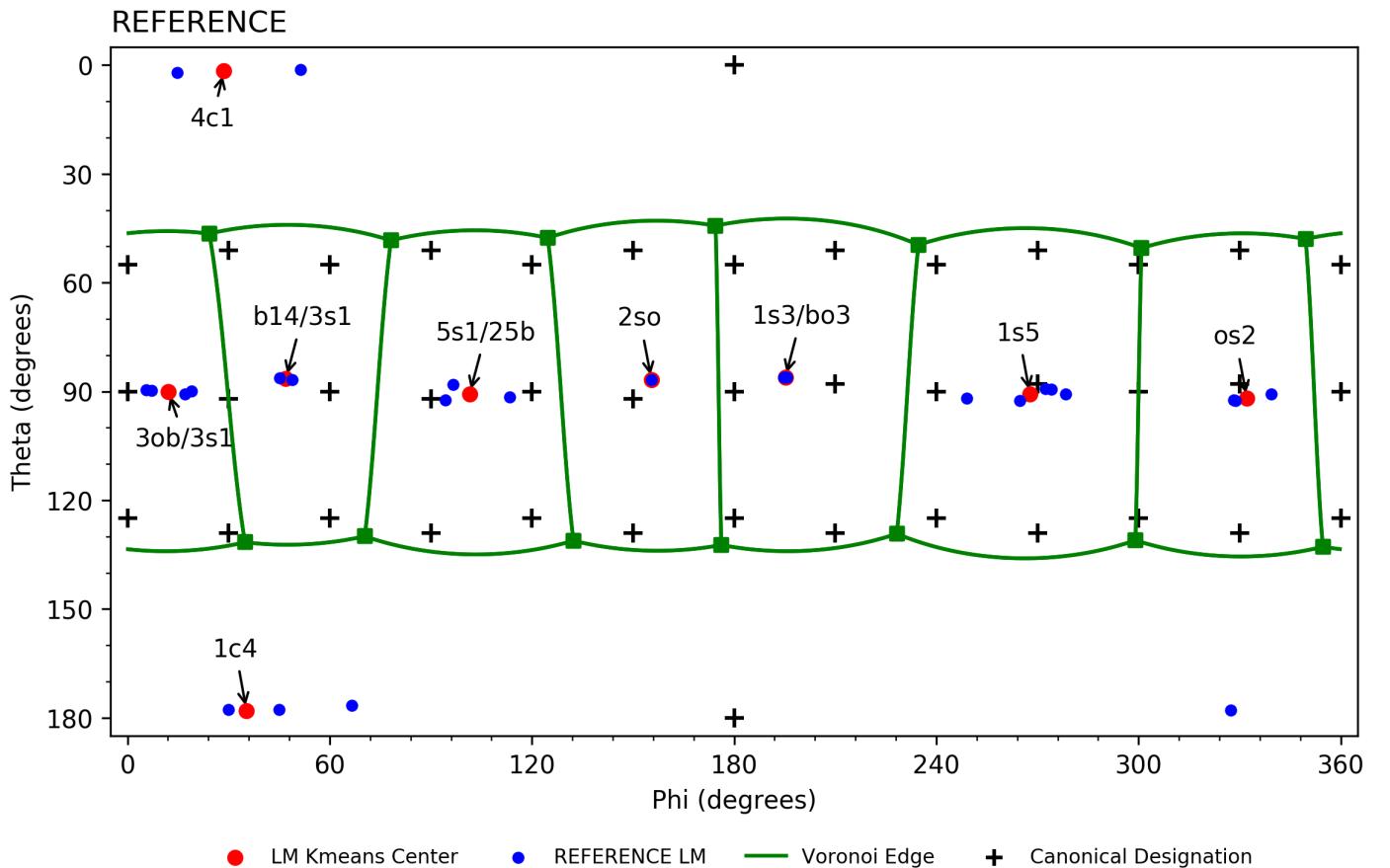


Figure 9: The LM reference landscape for β -xylose.

3.2.2 Transition State Reference Landscape

With the LM conformational landscape for β -xylose generated, the TS connecting each LM were then determined. Each TS found in Figures 10 and 11 represent a unique structure connecting two LMs. To aid in visualization, Northern

and Southern hemisphere plots were created to highlight the key pathways from 4c1 and 1c4 conformations. To distinguish the different pathways, each individual pathway was colored coded (Figure 11).

Figure 4 represents a unique structure that connects two LM. To aid in visualization, Northern and Southern hemisphere plots were created to highlight key pathways from 4c1 and 1c4 conformations.

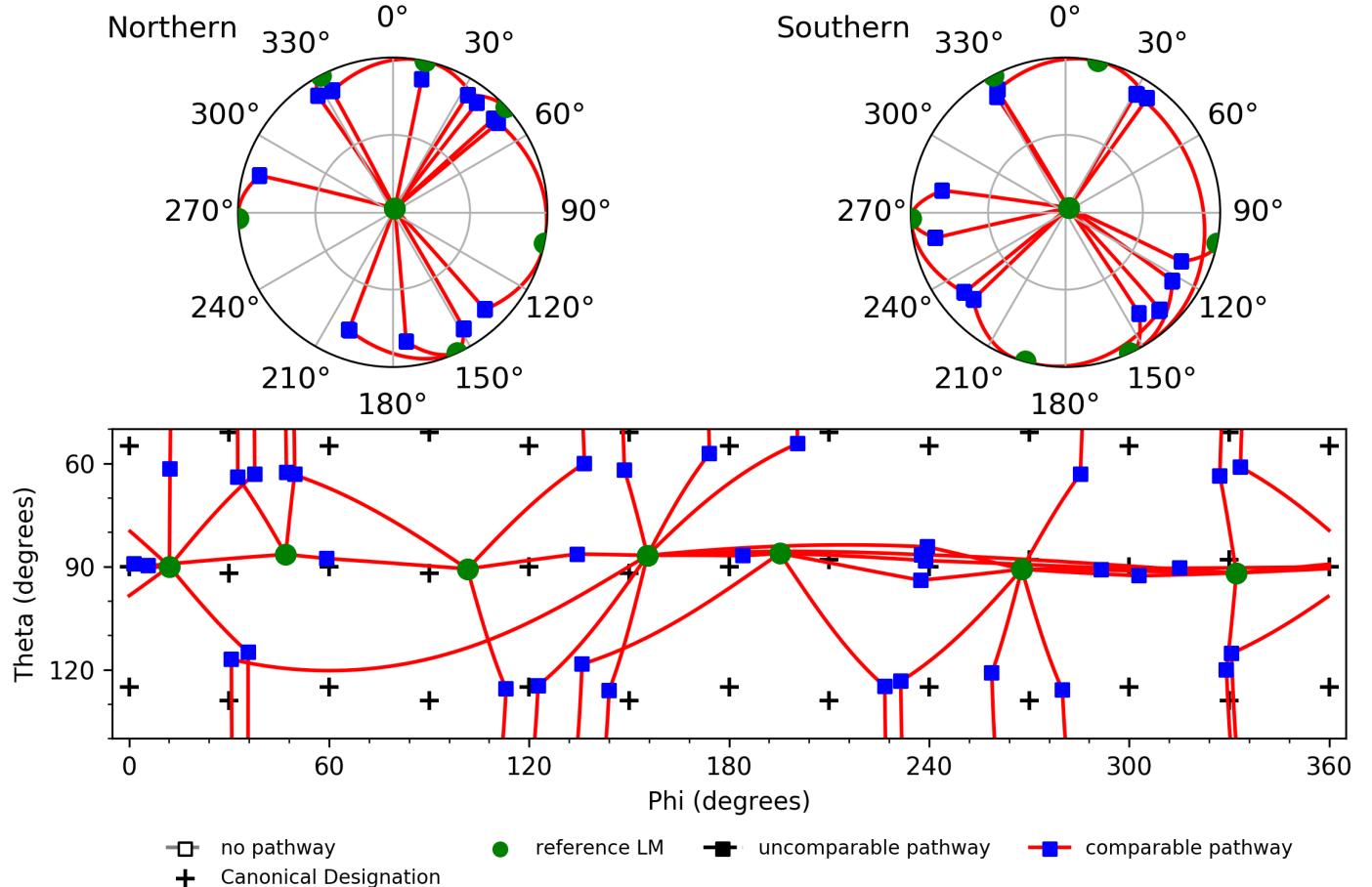


Figure 10: The TS reference landscape for β -xylose.

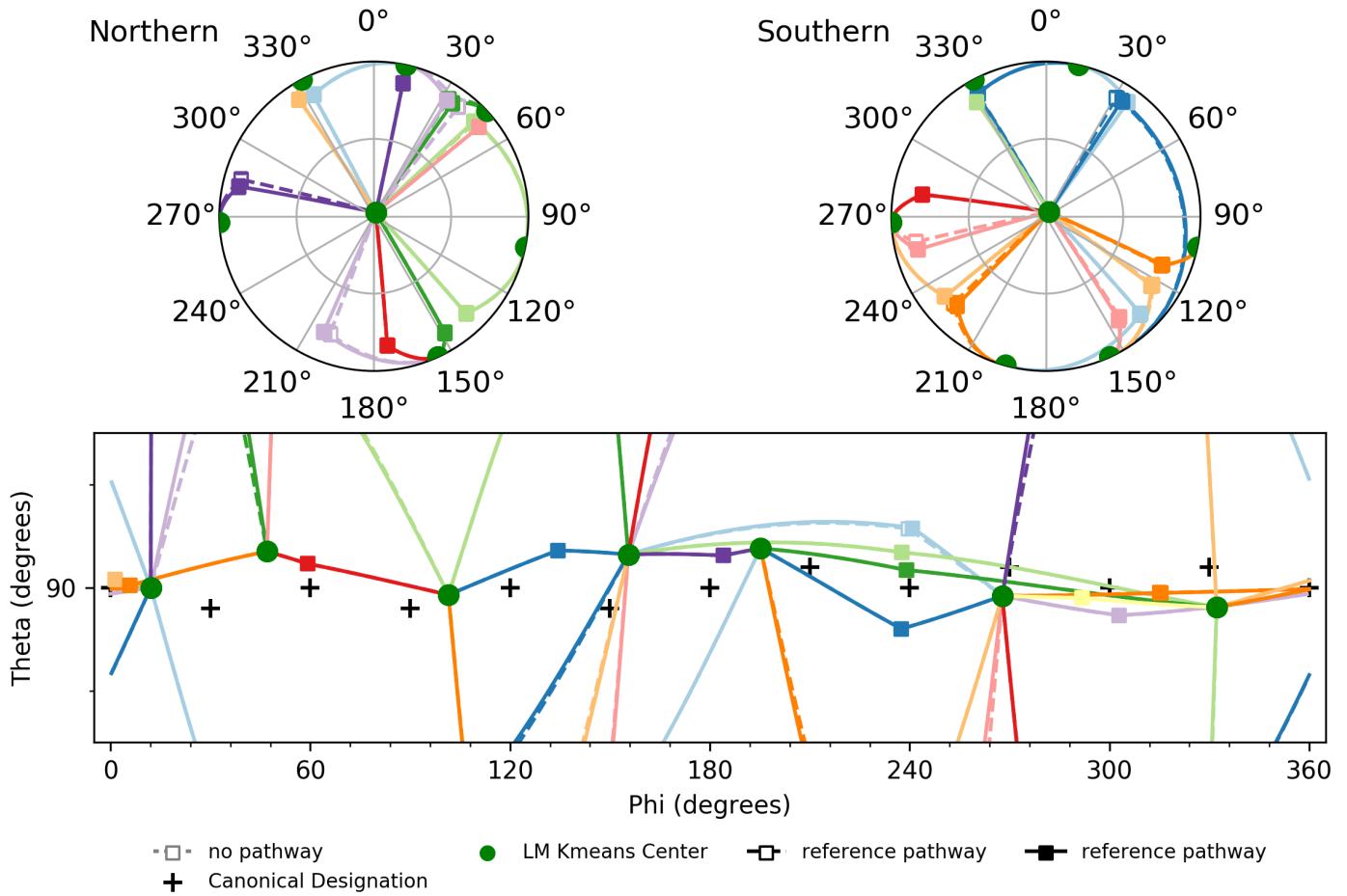
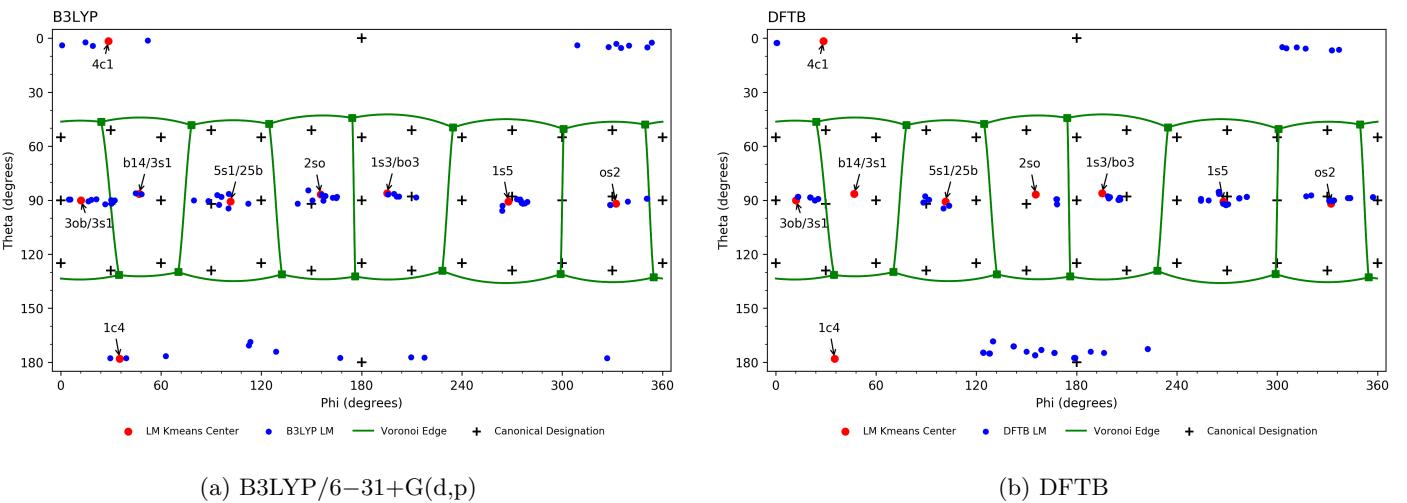
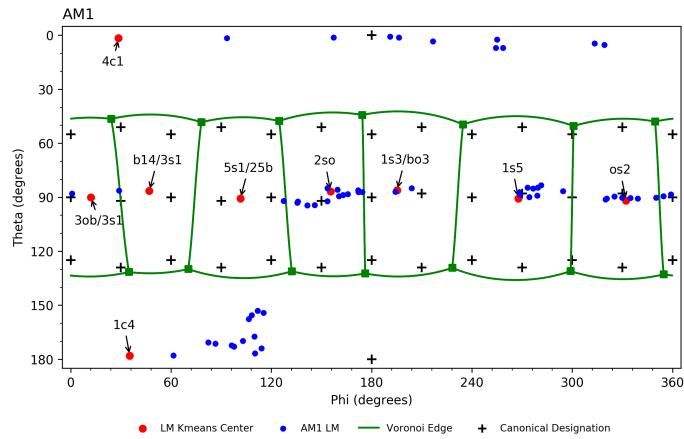


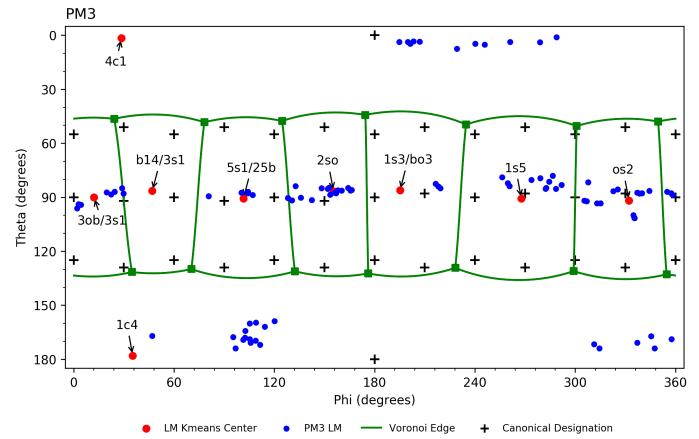
Figure 11: The TS reference landscape for β -xylose. Each pathway is colored individually to show the interconnectivity. Also, note the difference in theta values for the equatorial image compared to Figure 10

3.2.3 Local Minima Landscapes

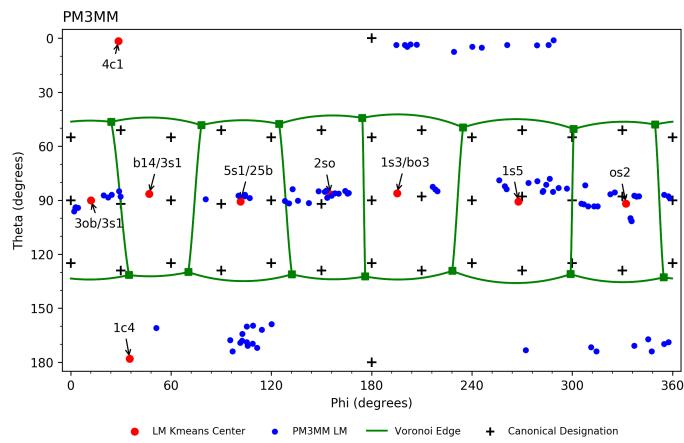




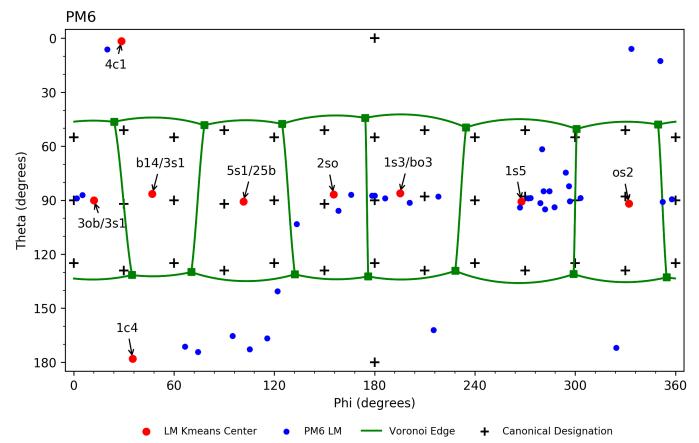
(c) AM1



(d) PM3



(e) PM3MM



(f) PM6

Figure 12: Insert Caption

3.3 β -Glucose

The last conformational landscape explored was β -glucose, which contains four hydroxyl groups and one hydroxymethyl group. The addition of the hydroxymethyl group increases the diversity in the conformational landscape.

3.3.1 Local Minima Reference Landscape

3.3.2 Transition State Reference Landscape

References

- [1] L. Misurcova, S. Skrovankova, D. Samek, J. Ambrozova, and L. Machu. Health Benefits of Algal Polysaccharides in Human Nutrition. *Advances in Food and Nutrition Research*, 66:75–145, 2012.
- [2] George W. Huber, Iborra Sara, and Avelino Corma. Synthesis of Transportation Fuels from Biomass. *Chem Rev.*, 2(106):4044–4098, 2006.
- [3] Ye Sun and Jiayang Cheng. Hydrolysis of lignocellulosic materials for ethanol production : a review q. *Bioresource technology*, 83(1):1–11, 2002.
- [4] R. Wolfenden, X. Lu, and G. Young. Spontaneous hydrolysis of glycosides. *Journal of the American Chemical Society*, 120(27):6814–6815, 1998.
- [5] D. Monti and S. Riva. Hydrolysis and Formation of Glycosidic Bonds. In K. Drauz, H. Groger, and O. May, editors, *Enzyme catalysis in organic synthesis Vol. II.*, chapter 10, pages 699–715. Wiley-VCH Verlag GmbH & Co. KGaA., 2002.
- [6] Gideon Davies and Bernard Henrissat. Structures and mechanisms of glycosyl hydrolases. *Structure*, 3(9):853–859, 1995.
- [7] Heather B Mayes, Michael W Nolte, Gregg T Beckham, Brent H Shanks, and Linda J Broadbelt. The Alpha Bet(a) of Glucose Pyrolysis: Computational and Experimental Investigations of 5 Hydroxymethylfurfural and Levoglucosan Formation Reveal Implications for Cellulose Pyrolysis. *ACS Sustainable Chemistry & Engineering*, 2014.
- [8] Gustavo De M Seabra, Ross C. Walker, Marcus Elstner, David A. Case, and Adrian E. Roitberg. Implementation of the SCC-DFTB method for hybrid QM/MM simulations within the Amber molecular dynamics package. *Journal of Physical Chemistry A*, 111(26):5655–5664, 2007.
- [9] Michael G. Medvedev, Ivan S. Bushmarinov, Jianwei Sun, John P. Perdew, and Konstantin A. Lyssenko. Density functional theory is Straying from the Path toward the Exact Functional. *Science*, 355(6320):49–52, 2017.
- [10] Mateusz Marianski, Adriana Supady, Teresa Ingram, Markus Schneider, and Carsten Baldauf. Assessing the accuracy of across-the-scale methods for predicting carbohydrate conformational energies on the example of glucose and α -maltose. *Journal of Chemical Theory and Computation*, 12(12):6157–6168, 2016.
- [11] Xevi Biarnes, Albert Ardevol, Antoni Planas, Carme Rovira, Alessandro Laio, and Michele Parrinello. The Conformational Free Energy Landscape of Beta-D-Glucopyranose . Implications for Substrate Preactivation in Beta-Glucoside Hydrolases. *Journal of the American Chemical Society*, 129(35):10686–10693, 2007.
- [12] Alessandro Laio and Michele Parrinello. Escaping Free-Energy Minima. *Proc. Natl. Acad. Sci. U. S. A.*, 99:12562, 2002.
- [13] Alessandro Laio, Antonio Rodriguez-Forte, Francesco Luigi Gervasio, Matteo Ceccarelli, and Michele Parrinello. Assessing the Accuracy of Metadynamics $\langle \sup \rangle \langle /sup \rangle$. *The Journal of Physical Chemistry B*, 109(14):6714–6721, 2005.

- [14] Andreia T. Pereira, António J. M. Ribeiro, Pedro A. Fernandes, and Maria J. Ramos. Benchmarking of density functionals for the kinetics and thermodynamics of the hydrolysis of glycosidic bonds catalyzed by glycosidases. *International Journal of Quantum Chemistry*, (February):e25409, 2017.
- [15] Heather B. Mayes, Linda J. Broadbelt, and Gregg T. Beckham. How sugars pucker: Electronic structure calculations map the kinetic landscape of fivebiologically paramount monosaccharides and their implications for enzymatic catalysis. *Journal of the American Chemical Society*, 136(3):1008–1022, 2014.
- [16] D Cremer and J A Pople. A General definition of ring puckering coordinates. *Journal of the American Chemical Society*, 97(6):1354–1358, 1975.
- [17] Anthony D. Hill and Peter J. Reilly. Puckering coordinates of monocyclic rings by triangular decomposition. *Journal of Chemical Information and Modeling*, 47(3):1031–1035, 2007.
- [18] Attila Bérces, Dennis M. Whitfield, and Tomoo Nukada. Quantitative description of six-membered ring conformations following the IUPAC conformational nomenclature. *Tetrahedron*, 57(3):477–491, 2001.
- [19] Pegah Khalili, Christopher B. Barnett, and Kevin J. Naidoo. Interpreting medium ring canonical conformers by a triangular plane tessellation of the macrocycle. *Journal of Chemical Physics*, 138(18), 2013.
- [20] Jason Laska. Clara-labs/spherecluster, 2016.
- [21] M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G. A. Petersson, H. Nakatsuji, M. Caricato, H. P. Li, X.; Hratchian, A. F. Izmaylov, J. Bloino, G. Zheng, J. L. Sonnenberg, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, Jr. Montgomery, J. A., J. E. Peralta, F. Ogliaro, M. Bearpark, J. J. Heyd, E. Brothers, K. N. Kudin, V. N. Staroverov, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, N. Rega, J. M. Millam, M. Klene, J. E. Knox, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, R. L. Martin, K. Morokuma, V. G. Zakrzewski, G. A. Voth, P. Salvador, J. J. Dannenberg, S. Dapprich, A.D. Daniels, O. Farkas, J. B. Foresman, J. V. Ortiz, J. Cioslowski, and D. J. Fox. Gaussian 09, revision D.01, 2009.
- [22] ChemCraft. ChemCraft.