JOURNAL OF CHEMICAL EDUCATION

Article

pubs.acs.org/jchemeduc

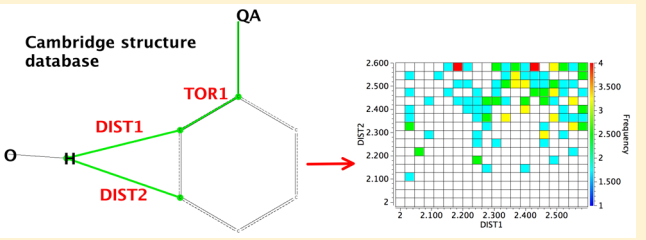
Discovering More Chemical Concepts from 3D Chemical Information Searches of Crystal Structure Databases

Henry S. Rzepa*

Department of Chemistry, Imperial College London, South Kensington Campus, London, SW7 2AZ, U.K.

Supporting Information

ABSTRACT: Three new examples are presented illustrating three-dimensional chemical information searches of the Cambridge structure database (CSD) from which basic core concepts in organic and inorganic chemistry emerge. These include connecting the regiochemistry of aromatic electrophilic substitution with the geometrical properties of hydrogen bonding interactions to the ring, the gauche conformational effect in 1,2-disubstituted ethanes, and square planar versus tetrahedral coordination in four-coordinate transition metals. The searches take only a short time and can be readily incorporated into a variety of teaching environments.



KEYWORDS: *Second-Year Undergraduate, Chemoinformatics, Organic Chemistry, Inorganic Chemistry, Aromatic Compounds, Conformational Analysis, Coordination Compounds, Computer-Based Learning, Inquiry-Based/Discovery Learning, Problem Solving/Decision Making*

Introducing students to chemical information prior to ~1980 largely involved familiarizing them with bound volumes such as Chemical Abstracts and Beilstein stacked in library shelves. Chemical information started to go online after this date, such that by 1996 at Imperial College, we were able to introduce a fully online-based course. This was structured around the metaphor of chemical dimensionality, with one-dimensional searches defined by metadata strings such as author, chemical, or property names. These were extended into two dimensions by the specification of molecular connection tables to allow structure and substructure searches of molecules, with only limited three-dimensional information such as stereochemistry added. The course ended fully in the third dimension by searching for structural attributes within a molecule such as bond lengths, angles, and torsions. This third category is the focus in this article, illustrating concepts inferred from the three-dimensional information obtained by searching the Cambridge crystal structure database (CSD).[1−3] The idea of using a statistical analysis of crystal structures to develop chemical principles was most famously introduced in 1974 in a paper[4] inferring the geometry of attack of a nucleophile on a trigonal unsaturated (sp^2) carbon in ketones and related carbonyls, a trajectory nowadays referred to as the Bürgi−Dunitz angle.

The three new examples presented here center around use of the commercial searching software ConQuest that is a component of the CSD.[5] Although a noncommercial open collection of crystal structures (COD) has more recently become available,[6] the accompanying software is currently less well suited for the types of search described here. The objective is to show how chemical principles taught in lecture courses taken by the students can be teased out from structural data. Each example was first disseminated to students via a blog[7] intended as support for lectures, taught workshops, laboratory courses, or tutorials.

Example 1: The Directing Influence of a Substituent on a Phenyl Ring toward Electrophilic Substitution Reactions[8]

This search is illustrated and explained in detail to show the procedures involved. It involves making a connection between chemical reactivity and kinetics as represented by the aromatic electrophilic substitution reaction of an aryl ring and the structural attributes of hydrogen bonding in a crystal structure.

The ConQuest software[5] includes a simple 2D sketching option for defining the atom connectivities in a chemical structure. Templates can be used to define a benzene ring, and a substituent is then attached to the ring using a variable atom QA (Figure 1), where the variable QA can be assigned to be any of the atoms N, O, F, Cl, or Br. Each of these atoms can

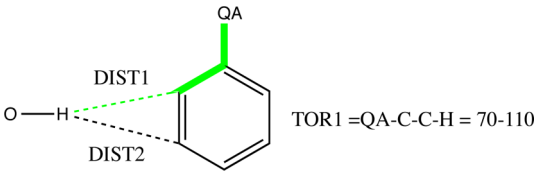


Figure 1. Search query for example 1 to illustrate how electron-donating substituents direct the location of a hydrogen bond on a phenyl ring.

Special Issue: Chemical Information

DOI: 10.1021/acs.jchemed.5b00346
J. Chem. Educ. XXXX, XXX, XXX−XXX

donate electrons into the phenyl ring via the π-mesomeric effect, as taught in introductory aromatic chemistry courses.

The electrophile will be modeled using an acidic proton, in this case attached to an oxygen as OH and drawn unconnected to the phenyl ring. Its 3D relationship with respect to the phenyl ring now needs definition (Figure 1). This allows the hydrogen atom of the OH group to have a constraint placed on it such that the distance between the hydrogen atom and a carbon atom ortho to the electron donating group QA is defined as being <0.3 Å than the sum of the van der Waals radii for H and C. This variable becomes known as DIST1. This one action introduces two important chemical concepts: that of typical values for interatomic distances and how they relate to van der Waals (VdW) values for the H and C atoms and the idea that atoms in either intra or intermolecular contact by 0.3 Å less than the VdW sum are considered to be interacting, in this case by a stabilizing hydrogen bond. DIST2 is similarly defined as the distance of the hydrogen to a meta-carbon of the ring. We are only interested in such interactions if they are above the plane of the ring via π-facial bonding rather than in the plane. This reflects that the π-electron density is indeed above and below the ring rather than in its plane. To achieve this constraint, a 3D torsion angle variable TOR1 is defined as constrained to the values 90 ± 20° in a similar manner to the distances. The search is further refined by specifying that only reasonably accurate crystallographic data be searched using the constraint that $R < 0.1$, that the structure is not disordered, that there be no crystallographic errors, and that hydrogen positions are normalized. These crystallographic terms can be explained to the students as part of a course or tutorial.

Around 126 hits might be expected for this definition (May 2015 data set) out of about 522,402 entries satisfying the quality criteria in the database (although a teaching subset of 733 compounds is available as a free resource,[1,2] it is not available for use with the ConQuest software and is not appropriate for the statistical searches described here). This is an example of data mining focusing the structures just to those compounds sustaining a hydrogen bond to a phenyl ring. The structural properties of these weak interactions are best presented using the free analysis program Mercury[9] as a *heat plot*, where the X and Y axes are defined by DIST1 and DIST2, and the color indicates the number of hits for each region. The red hotspots indicate the regions with the highest number of hits (Figure 2). The following conclusions may be drawn from the distribution. The diagonal line defines compounds with equal distance from the OH to the ortho and meta positions of a phenyl ring substituted with a (potentially) π-electron donating group such as OH, NH_2, halogen, etc. The highest number of hits clearly sits in the top left triangle of the square, a region where the distance of the H to the ortho position is shorter than to the meta position. There are two particular hot spots, one where the difference between DIST1 and DIST2 is ~0.17 Å and another where it is 0.4 Å. The curious student might wish to identify why there are two such regions. The students can also explore how the appearance of the heat plot varies according to the "number of bins" set to define the range of values or "resolution" encompassed by each square in the plot.

There are many variations possible on this basic search. The OH hydrogen bond donor might instead be replaced by another type of donor, for example, a nitrogen R_2NH or a halogen acid. The effects of the substituent on the oxygen atom of the OH bond (i.e., ROH) could be explored. The VdW

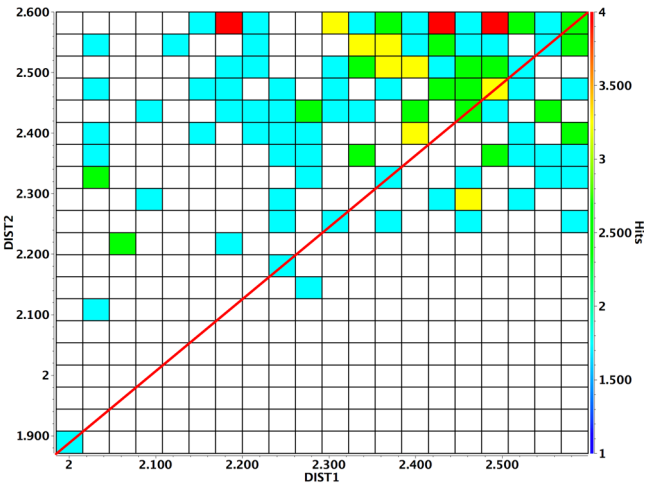


Figure 2. A heat plot display for example 1 showing the value of the two variables DIST1 and DIST2, with color indicating the number of examples for each pair of values (red = highest number) and a "bin" size of 20 corresponding to the number of squares shown along each axis.

contraction of 0.3 Å could be varied to see how this affects the distribution obtained, which allows the strength of the interactions to be assessed on the assumption that the shorter distances represent stronger hydrogen bonds. The 3D constraint TOR1 could be expanded or contracted to see how sensitive the effect is to the orientation of the axis of the hydrogen bond with respect to the plane of the aromatic ring. The quality of the crystal data searched could be increased by reducing R to <0.05 or constraining the data to reflect collection at, for example, <140 K. The substituent QA can be redefined as consisting only of electron withdrawing groups (NO_2, CN, SO_3H, R_3N^+, etc.) to see how the search responds. In fact, the number of hits decreases substantially, and the distances to the meta-carbon now outnumber the ortho-carbon and also lengthen. One may infer from this that electron donating groups produce stronger π-facial hydrogen bonding interactions than electron withdrawing groups and associate this with their respective activating and deactivating characteristics in aromatic substitution reactions. Further explorations could include studying phenyl rings with more than one substituent as QA and QB, exploring the distribution of hydrogen bonds between meta and para positions and investigating the properties of π-facial hydrogen bonds that focus instead on the centroid of the aromatic ring rather than the ortho and meta carbons. In doing so, students can also be encouraged to observe the crystal packing in examples they select for closer analysis and comment on the orientation of the molecules in relation to one another.

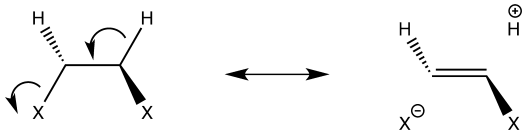
Since each search takes just a few minutes, it is possible to explore these variations quickly and for students to drive the experiment by using their own curiosity. The learning outcomes include achieving a connection between abstract concepts taught in lectures and the properties of real molecules, learning some of the basics of crystallography and acquiring a better feel for statistics by exploring the reasons for outliers. That is, those compounds containing electron donating substituents where the H-distance to the meta carbon appears to be shorter than to the ortho, or even bias toward particular structures in the database itself.

DOI: 10.1021/acs.jchemed.5b00346
J. Chem. Educ. XXXX, XXX, XXX−XXX

Example 2: The Gauche Effect in 1,2-Substituted Ethanes

In the area of confomational analysis, students are taught that the conformation of 1,2-difluoroethane is predominantly *gauche*, whereas that of 1,2-dimethylethane (butane) is predominantly *anti*. They may be shown[10] the orbital arguments that provide an theoretical explanation for this gauche effect. The same orbital arguments also predict that the C−C bond length will contract slightly compared to the *anti* conformation (Scheme 1).

Scheme 1



The procedure described here can be used to see if experimental support for the gauche effect can be found in crystal structures and whether there is any evidence for the bond length contraction when it occurs.[11] The search is defined as follows (Figure 3). The terminal carbon atoms chain of four

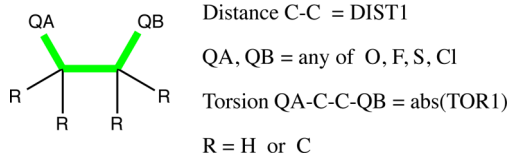


Figure 3. A search definition for example 2 suitable for probing the gauche effect in acyclic 1,2-disubstituted ethanes.

connected by single bonds are replaced by QA = QB where in this specific instance, either variable can be defined as being any one of O, F, S, or Cl. The central C−C bond is defined as being acyclic (so as not to restrict rotation), and two central carbons are defined with four further surrounding groups R that can be either H or C. The torsion angle about the central bond is defined as TOR1, and its absolute value is computed to avoid negative torsions. The C−C distance is defined as DIST1. The search has tighter crystallographic constraints than the previous example since we now need more accurate data to explore the difference in bond lengths. Hence $R < 0.05$, and the temperature at which the diffraction data were collected is specified as either <140 K or <90 K as part of a combined query illustrating the use of Boolean AND logic.

The search takes about 2−3 min and produces around 1271 hits at <140 K and 36 hits at <90 K (Figure 4). Two distinct clusters are observed, of which the larger is *gauche* (∼60°), and the smaller is *anti* (∼180°). The hot spot occurs at a distance of ∼1.5 A for each cluster for the <140 K data, but the lower temperature data now also reveal a small but significant contraction in the bond length for the *gauche* conformations. The students can be encouraged to find out the reason why there are so many apparent *anti* examples, to explore the effects of restricting QA = QB = F, or the properties of other halogens to see how the *gauche* effect changes with the nature of the halogen.

Example 3: Tetrahedral versus Square Planar Coordination in Transition Metal Complexes[12]

The definition for this search (Figure 5) includes the following: A central atom QA can be defined per column of the periodic

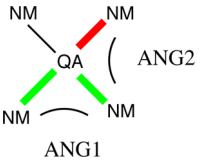


Figure 5. A search definition for example 3 suitable for probing the type of coordination of 4-coordinated transition metals.

table (here any element from column 7 (Mn−Re), 8 (Fe−Os), 9 (Co−Ir), or 10 (Ni−Pt)) or indeed specific elements. The coordination of QA is restricted to precisely 4. The four ligands are defined as nonmetals (NM) and connected to the central atom by acyclic bonds to avoid influencing the coordination. Two angles, ANG1 and ANG2, are defined at the central atom.

A heat plot (Figure 6) reveals three clusters at values of ANG1 and ANG2 of 90° and 180° (square planar coordination) and one cluster at values of 109° (tetrahedral) from around 8073 hits in total.

By repeating the search by stepping QA through each column of the transition metals individually, students can then discover for themselves how the position of an element in the

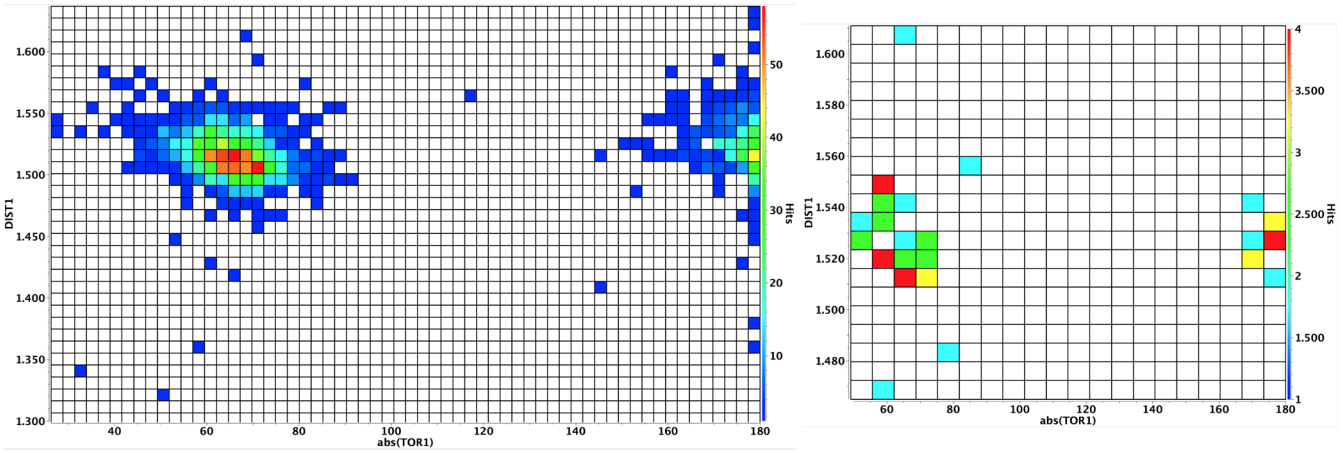


Figure 4. A heat plot for example 2 revealing the distribution of torsion angles in 1,2-disubstituted ethanes for (a) data recorded at <140 K and using a bin size of 60 and (b) < 90 K using a bin size of 20.

DOI: 10.1021/acs.jchemed.5b00346
J. Chem. Educ. XXXX, XXX, XXX−XXX

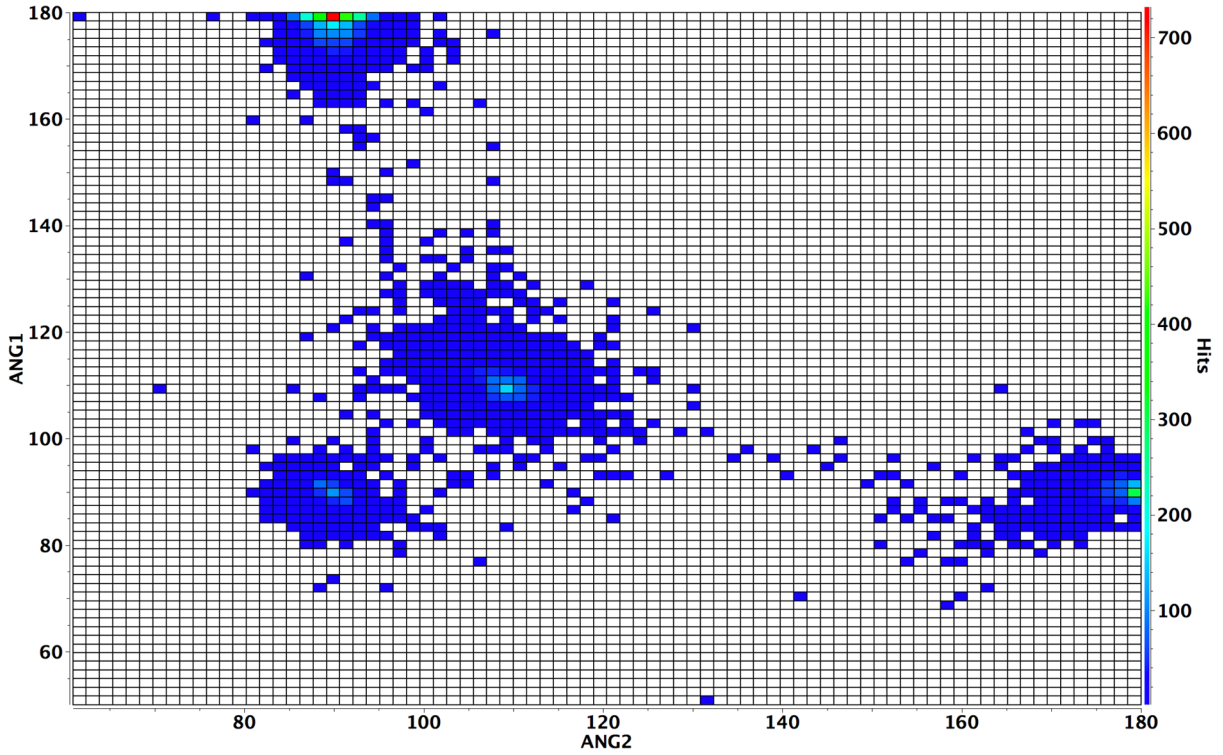


Figure 6. A heat plot for example 3 revealing square planar and tetrahedral geometries for four-coordinated transition metal complexes using a bin size of 80.

periodic table influences whether it is predominantly square planar or tetrahedrally coordinated.

The CSD contains coordinates for 713,006 crystal structures (May 2015) with a broad coverage across most molecule types. Many concepts in chemistry that are associated in some manner or other with molecular structure can be found by an appropriate search of this database; the three described here represent only a small proportion of those that could potentially be developed for teaching (see Supporting Information for further suggestions). Once familiarity with the ConQuest software is achieved, each of these searches takes less than 5 min to define and around the same time to perform and display. This allows the technique to be incorporated into a variety of teaching environments such as lectures, tutorials, workshops, and laboratory sessions as either an individual or as a group discovery activity. An experiment illustrating our use in a computational/synthesis laboratory was recently described.[15] It is also a visually stimulating tool since the objects are three-dimensional, colorful, and even lend themselves to tactile manipulation using, for example, 3D printed versions.[13−15] It is also suggested that posting to a blog (configured to Tweet any new entries) is an interesting way of alerting both students and staff to new searches and inviting commentary on the outcomes.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available on the ACS Publications website at DOI: 10.1021/acs.jchemed.5b00346.

Search query definitions for examples 1−3 for use with the ConQuest program and nine further search suggestions (ZIP)

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: rzepa@imperial.ac.uk.

Notes

The author declares no competing financial interest.

■ REFERENCES

(1) Battle, G. M.; Allen, F. H.; Ferrence, G. M. Teaching Three-Dimensional Structural Chemistry Using Crystal Structure Databases. 3. The Cambridge Structural Database: Information Content and Access Software in Educational Applications. *J. Chem. Educ.* **2011**, *88*, 886−890. Battle, G. M.; Allen, F. H.; Ferrence, G. M. Teaching Three-Dimensional Structural Chemistry Using Crystal Structure Databases. 4. Examples of Discovery-Based Learning Using the Complete Cambridge Structural Database. *J. Chem. Educ.* **2011**, *88*, 891−897.

(2) Henderson, S.; Battle, G. M.; Allen, F. H. Teaching chemistry in 3D using crystal structure data. *Education in Chemistry* **2011**, *48*, 175−178.

(3) Davis, T. V.; Zaveer, M. S.; Zimmer, M. Using the Cambridge Structural Database to introduce important inorganic concepts. *J. Chem. Educ.* **2002**, *79*, 1278−1280.

(4) Burgi, H.; Dunitz, J.; Lehn, J.; Wipff, G. Stereochemistry of reaction paths at carbonyl centres. *Tetrahedron* **1974**, *30*, 1563−1572.

(5) *ConQuest*. http://www.ccdc.cam.ac.uk/Solutions/CSDSystem/Pages/ConQuest.aspx (accessed May 8, 2015). A Web-based online version WebCSD does not have the full capability for performing the searches described here.

(6) Grazulis, S.; Daskevic, A.; Merkys, A.; Chateigner, D.; Lutterotti, L.; Quiros, M.; Serebryanaya, N. R.; Moeck, P.; Downs, R. T.; Le Bail, A. Crystallography Open Database (COD): an open-access collection of crystal structures and platform for world-wide collaboration. *Nucleic Acids Res.* **2012**, *40*, D420−D427.

(7) Rzepa, H. S. http://www.ch.imperial.ac.uk/rzepa/blog/?p=13962 (accessed May 8, 2015). Archived as Rzepa, H. S. A new way of exploring the directing influence of (electron donating) substituents

DOI: 10.1021/acs.jchemed.5b00346
J. Chem. Educ. XXXX, XXX, XXX−XXX

on benzene. *Winnower* **2015**, *2*, No. e143118.80307, DOI: 10.15200/winn.143118.80307.
(8) Armstrong, H. E. XXVIII. An explanation of the laws which govern substitution in the case of benzenoid compounds. *J. Chem. Soc., Trans.* **1887**, *51*, 258−268.
(9) *Mercury*. http://www.ccdc.cam.ac.uk/Solutions/FreeSoftware/Pages/FreeMercury.aspx (accessed May 8, 2015).
(10) Rzepa, H. S. The conformation of 1,2-difluoroethane. *Winnower* **2015**, *2*, No. e143125.51716, DOI: 10.15200/winn.143125.51716.
(11) Rzepa, H. S. The gauche effect: seeking evidence by a survey of crystal structures. *Winnower* **2015**, *2*, No. e142795.55871, DOI: 10.15200/winn.142795.55871.
(12) Rzepa, H. S. Tetrahedral or square planar? A ten minute exploration. *Winnower* **2015**, *2*, No. e143125.52611, DOI: 10.15200/winn.143125.52611.
(13) Moeck, P.; Stone-Sundberg, J.; SnyderT. J., Kaminsky W. Open Access Resources for Crystallography Education in Interdisciplinary College Courses: Crystallographic Databases and 3D Printed Models. *MRS Online Proc. Libr.* **2014**, *1716*, DOI: 10.1557/opl.2014.872.
(14) Chen, T.-H.; Lee, S.; Flood, A. H.; Miljanić, O. S. How to print a crystal structure model in 3D. *CrystEngComm* **2014**, *16*, 5488−5493.
(15) Hii, K. K. M.; Rzepa, H. S.; Smith, E. H. Asymmetric Epoxidation: A Twinned Laboratory and Molecular Modeling Experiment for Upper-Level Organic Chemistry Students. *J. Chem. Educ.* **2015**, *92*, 1385. Rzepa, H. S. Full-colour 3D printing of molecular models and orbitals (wavefunctions). *Winnower* **2015**, *2*, No. e143118.83281, DOI: 10.15200/winn.143118.83281.

DOI: 10.1021/acs.jchemed.5b00346
J. Chem. Educ. XXXX, XXX, XXX−XXX