# JACOB: A Dynamic Database for Computational Chemistry Benchmarking

Jack Yang and Mark P. Waller*

Theoretische Organische Chemie, Organisch-Chemisches Institut der Westfälische Wilhelms Universität Münster, Corrensstraße 40, 48149 Münster, Germany

S Supporting Information

ABSTRACT: JACOB (just a collection of benchmarks) is a database that contains four diverse benchmark studies, which in-turn included 72 data sets, with a total of 122 356 individual results. The database is constructed upon a dynamic web framework that allows users to retrieve data from the database via predefined categories. Additional flexibility is made available via user-defined text-based queries. Requested sets of results are then automatically presented as bar graphs, with parameters of the graphs being controllable via the URL. JACOB is currently available at www.wallerlab.org/jacob.
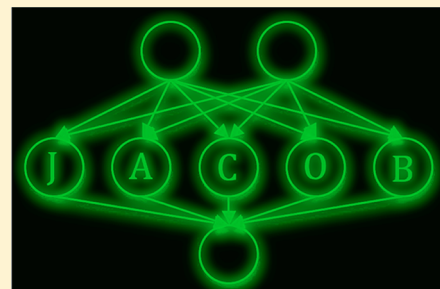
1. INTRODUCTION

With a rich repository of quantum chemical methods spread across the academic literature, researchers are often over-whelmed by a "cloud-of-acronyms" when choosing an appropriate method for solving their problems. Selecting the "right" method becomes an intellectual challenge for a nonspecialist in quantum chemistry. This is particularly the case in the domain of density functional theory (DFT), where the number of available functionals is frankly bewildering to the experimentalist, and many theoreticians alike. On the one hand, researchers select more "general" well performing functionals[1] such as B3LYP.[2−5] On the other hand, some researchers prefer to have a more specific functional for a given problem type,[6−10] which has the advantage of increased accuracy, perhaps at the expense of the range of applicability.

Unfortunately, in stark contrast to the situation with functional creation, there exist far fewer research endeavors that produce and/or collate large, accurate, and diverse data sets for density functional development, validation, testing, and benchmarking. Recently, the GMTKN30 benchmark proposed by Grimme et al.[11−13] contains a diverse superset of 30 chemically different subsets and provides a good general impression of the relative strengths and weaknesses of particular functionals. Interestingly, the "Jacob's Ladder" groupings are well recovered when one looks at the statistics for functionals on different rungs, which is reassuring.[14] Nevertheless, the differences between functionals on any given same rung of Jacob's ladder are harder to predict.

The idea of specialists deliberating over the relative strengths and weaknesses of a variety of functionals across carefully collated benchmarks is of limited practical use to experimental colleagues. Such information must be successfully transferred in a timely fashion and not buried away in a large number of overwhelming tables. The persistent popularity of certain DFT methods, may at least be partially due to this time-delay between method developers and users. In order to try and breach this gap, we now sketch out a (technical) solution that enables computational/quantum chemistry users (experimen-talists or theoreticians) to obtain up-to-date statistical data on which method performs best for a given problem domain.
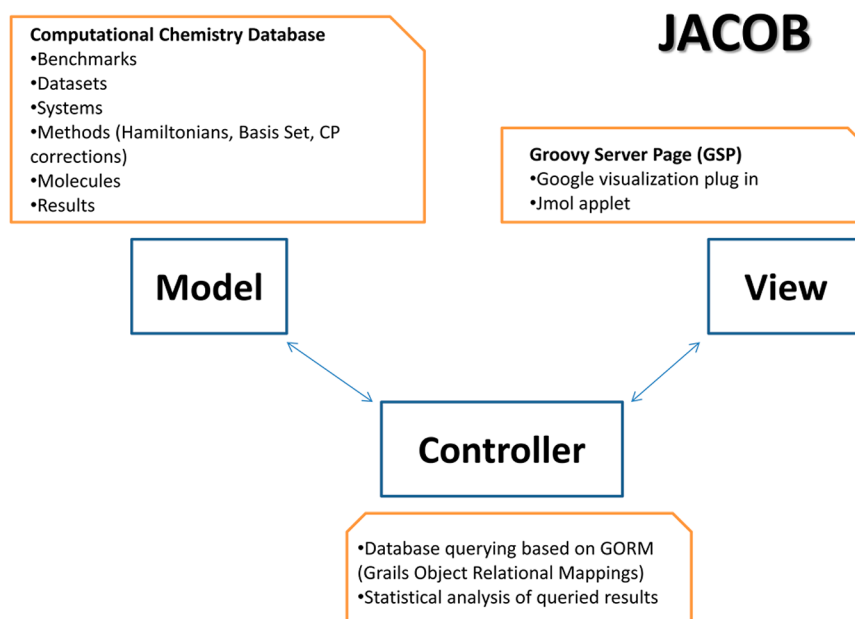
Instead of focusing on a way to systematically improve a specific quantum mechanical method, we present a pragmatic way of identifying the best currently available method for specific chemical systems based on published benchmark results. We present here preliminary findings using web 2.0 technology[15] for solving such a pertinent challenge.
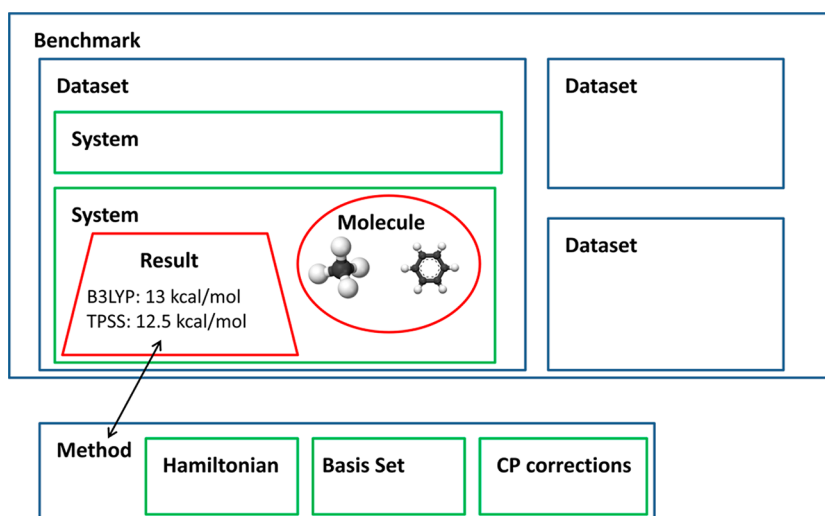
2. IMPLEMENTATION DETAILS

2.1. The Benchmarks. Currently JACOB has stored information from the following four benchmarks:

(a) Grimme's GMTKN30 (General Main-Group Thermo-chemistry and Kinetic Database)[11,12] which includes 30 data sets that can be divided into three subsets that deal with basic properties,[11,16−22] reaction energies,[14,21−29] and noncovalent interactions;[11,12,30−37]

(b) Hobza's BEGDB (Benchmark Energy and Geometry DataBase) database,[31,38−40] which collects results of highly accurate quantum mechanical calculations of molecular structures, energies, and properties;

(c) Transition metal complexes[41−43] which examines the metal−ligand bond lengths, and;

(d) Gigantic Non-Covalent Structure Benchmark (GNCSB) assembled by Schneebeli et al.[44] from 33 distinct data sets.[31,45−77]

JACOB

Computational Chemistry Database
•Benchmarks
•Datasets
•Systems
•Methods (Hamiltonians, Basis Set, CP corrections)
•Molecules
•Results

Groovy Server Page (GSP)
•Google visualization plug in
•Jmol applet

Model

View

Controller

•Database querying based on GORM (Grails Object Relational Mappings)
•Statistical analysis of queried results

Figure 1. Schematic illustration of the MVC framework that JACOB is constructed upon.

Benchmark

Dataset

System

System

Result

B3LYP: 13 kcal/mol
TPSS: 12.5 kcal/mol

Molecule

Dataset

Dataset

Method   Hamiltonian   Basis Set   CP corrections

Figure 2. Illustration of the domain models of JACOB. Each child object (inner shape) is fully contained within its parent object (outer shape). The result object is further linked to a method object, thus each result in the database can be uniquely identified by the benchmark, data set, and system that it belongs to, as well as the method used to obtain the numerical result.

2.2. Web Application Framework.

JACOB is developed upon the Grails[78] programming stack which includes a relational database (Postgres[79]/Hibernate[80]), Grails Object Relational Mapper (GORM), and Groovy[81] (dynamic version of java). The Grails framework utilizes the MVC (Model-View-Controller) design for dynamic web applications in which:

(a) The (domain) MODEL holds the data structures, which will be persisted into a relational database.

(b) The VIEW is the web page that will be presented to the user, which displays the result that the user has queried for.

(c) The CONTROLLER is the messenger between the view and the model. It can perform queries upon the domain models, based on the criteria in the URL address, and return the results as a set of parameters that can be rendered by the view.

Figure 1 is a schematic illustration of the key concepts behind the MVC framework upon which JACOB is implemented. The key implementations in each of these three components will be elaborated in the following subsections.

2.3. Domain Model.

The domain model holds a series of objects, which are linked together in a relational database. Our domain model contains a series of domain objects that are organized in a predefined hierarchy, which can be briefly outlined as follows:

(a) *Benchmark* is at the highest level of the database hierarchy.

(b) Every benchmark contains a number of *data sets*, each focuses on a set of *systems* that are typically related. As in GMTKN30, it consists of 30 data sets, while within each data set, such as the S22 data set, there are 22 systems that focus on noncovalent interaction energies in small molecular dimers.
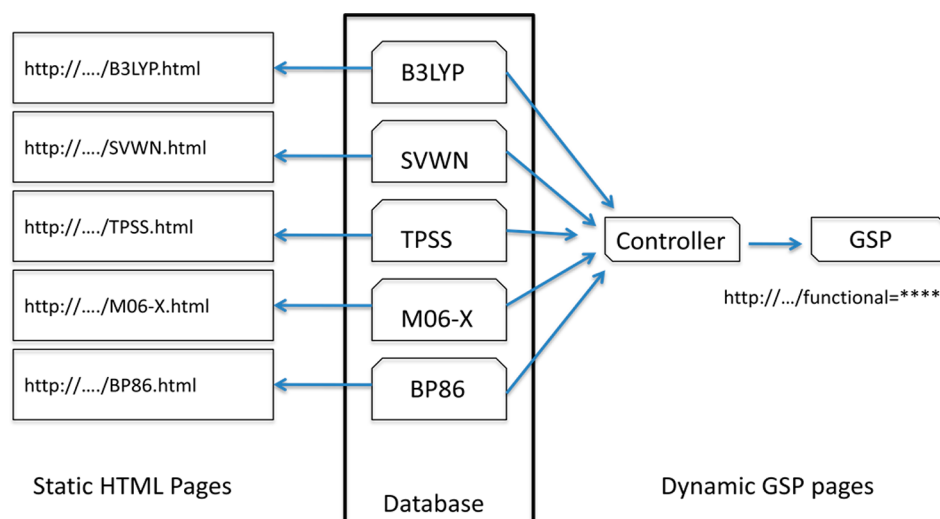
Figure 3. Comparison of static HTML and dynamic GSP view technology. For a database that consists of five functionals, under static HTML technology, five different HTML pages need to be constructed with the information about each functional typed in statically. In GSP technology, only one GSP page is set up which defines the general layout of the webpage. The data for a given functional is retrieved from the database and returned to the user on-the-fly based on the query.

(c) Each system can be linked to several *molecules*, which are stored in their respective domain classes. This allows users to query the database based on the molecules that they are interested in.

(d) Each system contains a series of *results* tested by a range of *methods*, as well as a unique reference value against which these results are compared to. Each result is thus associated with a unique method, which by itself is a separate domain object that contains a unique combination of *Hamiltonian*, *basis sets*, and *counterpoise (CP) correction*[82] (semiempirical methods will be identified with the name of the Hamiltonian only).

Figure 2 summarized the domain structure of JACOB. It can be seen that the data structure is highly nested. One advantage of adopting such a data structure is that it offers the possibility to more efficiently query the database, which will be discussed in the later session. A JACOB XML schema that clearly defines the data structure is deposited in the Supporting Information.

A searchable plugin[83] has been implemented which provides a text based search function for all domain models presented in Figure 2. For example, if a functional is searched for, a list of all results for that method is presented. The molecular systems stored in the JACOB database are also searchable via SMILES strings against the domain model. A SMILES string result will be directed to the corresponding information about the system and data set in which the structure can be found. A Jmol applet[84] is also implemented for structural visualization. The system ID names are consistent with the original publications, and the citation can be retrieved from the database by searching for the benchmark or data set name.

2.4. View. The view is a rendered Groovy Server Pages (GSP), which mixes the traditional HTML markup language with Groovy programming language to produce dynamic web content. One may consider the database as a collection of tables, each containing results tested for all of the systems for a given method. In a static view technology, one needs to include each numerical result for that given method into a HTML table. These pages can then be accessed by typing static addresses that link to the specific pages, which contain all the numerical

values in tabular form. In a dynamic framework, only one GSP page is required to be constructed for all the methods, where the page link contains an input parameter that allows users to request for the method of interest. This request parameter is then passed to the controller, which collects all the relevant information from the database and renders a corresponding view page to the end user. Figure 3 illustrates the concept behind the differences between HTML and GSP technologies.

One of the key features implemented in JACOB is the ability to visualize the ranked statistical results as bar charts generated on-the-fly. This is achieved by the application programming interface (API) provided by Google Visualization.[85] The Google visualization plugin[86] was installed. Figure 4 demon-
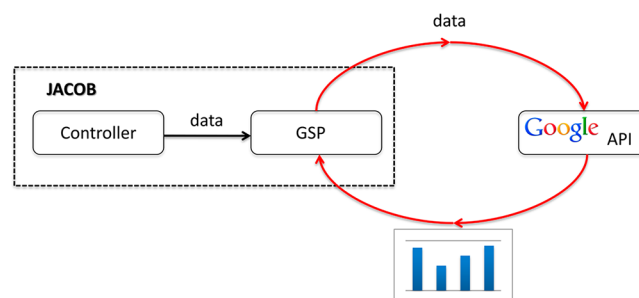
Figure 4. Schematic illustration of the service provided by Google visualization API.

strates the working principle behind the plugin. The Grails controller passes the data into a GSP page, which calls the plugin that intercepts the data and passes it further into the Google visualization API. The API is then responsible to convert the data into a corresponding bar chart and return it back to the GSP page for viewing. The advantage of such an API is that JACOB is not required to store charts as static objects; charts will always be generated dynamically.

2.5. Controller. The Grails controller is the message conveyer between domain model and GSP pages. In JACOB, a centralized data visualization controller has been implemented that performs the queries on the database, which then gathers

Scheme 1. Example of the Logic in the Grails Controller That Returns the MAPD for Each Data Set in GMTKN30 to Send to the Google Visualization API

```
find all Datasets in GMTKN30
for (each dataset in all DatasetsFound) {
        find all systems in this dataset
        for (each system in all SystemsFound) {
                get all absolute percentage deviations for this system with all
tested methods
        }
        perform statistical average on all absolute percentage deviations for
all systems in this dataset
}
gather a list of MAPDs for each dataset in GMTKN30
send the list to the view page for chart rendering
```

Table 1. Summary of Default Bar Charts Implemented in JACOB with Corresponding Accessing Method (Drop down Menu Link/View Page Link As Well As the URL Mapping) and the Contents of the Chart Returned in Terms of the Entries on the X- and Y-Axis, Respectively

| parameters	linked drop down menu	level of drop down menu	link on view page	URL mapping example	chart returned	
					X-axis	Y-axis
`benchmark`	YES	benchmark		/datavisualization/plot?benchmark=GMTKN30	all data sets in the benchmark	MAPD averaged over all systems and methods tested in each data set
`subset`	YES	Subset in GMTKN30		/datavisualization/plot?subset=NCI	all data sets in the subset	MAPD averaged over all systems and methods tested in each data set
`dataset`	YES	data set		/datavisualization/plot?dataset=PCONF	all systems in the data set	MAPD averaged over all methods tested for each system in the data set
`systemID`	NO		system show view	/datavisualization/plot?systemID=17	all methods tested for system with systemID	APD for each method tested for this system
`Hamiltonian`	YES	method		/datavisualization/plot?Hamiltonian=B3LYP	all systems tested with this Hamiltonian	APD for each system tested by this Hamiltonian
`Hamiltonian data setaverage`	NO		Hamiltonian show view	/datavisualization/plot?Hamiltonian=B3LYP-D3&data setaverage=true	all data sets that have systems tested by this Hamiltonian	MAPD over all systems tested by the Hamiltonian in each data set
`Hamiltonian data setName`	NO		Hamiltonian show view	/datavisualization/plot?Hamiltonian=B3LYP&data setName=ACChains	all systems in a particular data set that has been tested by this Hamiltonian	APD for each system tested by this Hamiltonian
`jacobLadder`	YES	method category		datavisualization/plot?jacobLadder=hybrid-GGA	all functionals in the database that belong to this level of Jacob's Ladder	MAPD over all systems in the database that have been tested by each functional as on the x-axis
`basisName`	YES	basis		/datavisualization/plot?basisName=aug-cc-pVDZ	all systems that have been tested by this basis set	MAPD over all methods with this basis set for each system

the corresponding results from the RESULT domain. A simple statistical analysis would be performed on the results, where the statistics are subsequently sent to the view page for plotting with Google visualization. For example, the following logic shown in Scheme 1 has been implemented to return a bar chart of the ranked mean absolute percentage deviations (MAPDs) for each data set in the GMTKN30 benchmark.

Since Grails queries the database with GORM, which supports the use of dynamic finders, the number of lines to implement a specific query can be greatly reduced compared to SQL. This allows different queries and statistical analysis to be easily implemented depending on the domain class (or database hierarchy) in which a user is interested. By default, a set of predefined queries on the JACOB database have been implemented, which return standard Google bar charts that display the best/worst performance for a specific level of JACOB. These statistics are always calculated based on the latest data in the database.

2.6. Data Visualization. The JACOB user interface has a series of drop down menus that redirect to the corresponding data visualization page. In general, three main categories have been defined, namely (a) Benchmark, which returns benchmark specific plots, (b) Method, which returns plots specific to different methods in the database, and (c) Basis, which returns basis set specific plots. Each category can be further subdivided, which returns specific statistical information as bar charts. A set of default queries has been implemented in the data visualization controller; these chart types are summarized in Table 1. The query parameters can also be sent to the controller via URL in the address bar, which provides a map to direct the web application to the desired data visualization pages. An example of a bar chart provided by JACOB for easy identification of best performing systems tested by B3LYP is presented in Figure 5.

As mentioned in section 2.1, currently there exist four different benchmarks in JACOB, while benchmarks a, b, and d focus on energy calculations with the numerical results given in
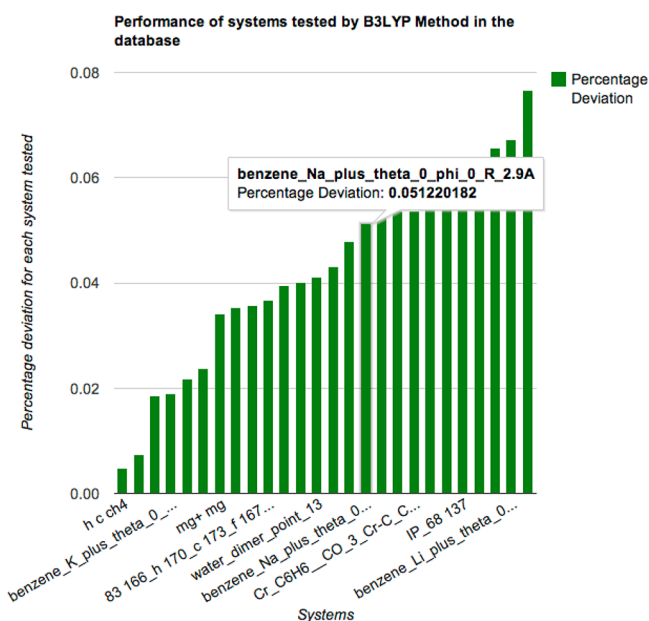
Figure 5. Example of Google chart for the performance of B3LYP in the JACOB database.

the units of kilocalories per mole, and the transition metal benchmark measures the bond distance in ångstrom. In order to implement a *unified controller* to generate statistical data for Google visualization, a dimensionless statistical merit based on the percentage deviation was implemented to allow different benchmarks to be compared on the same footing.

The absolute percentage deviation (APD) for a given system tested by a specific method is

$$\text{APD}(i, k) = |(\text{calculated result with method } i \text{ for system } k$$
$$- \text{ reference value for system } k)$$
$$/(\text{reference value for system } k)|$$

The mean absolute percentage deviation (MAPD) is calculated, for instance, for each data set in a given benchmark:

$$\text{MAPD}_{\text{dataset}} = \frac{1}{N} \sum_{k}^{N} \left(\frac{1}{p} \sum_{i}^{p} \text{APD}(i, k) \right)$$

where N is the number of systems in a given data set and P is the number of methods tested for a given system.

We acknowledge other types of statistical merits exist for some of the well-established benchmarks, such as the weighted total mean absolute deviation (WTMAD) as in GMTKN30.[12] Here, we are aiming at a more general merit that can be applied across all benchmarks in the database. The beta version has implemented only the APD and MAPD merits, but it is easily extendable to return other statistical merits.

Finally, queries based on either `Hamiltonian` could return a large set of results that cannot be all displayed in a single chart due to the large amount of data stored in the database. To further narrow down the search range, constraints can be passed to the controller as additional parameters in the URL address. These additional parameters are:

(a) `numberOfReturns` which defines the number of data points that will be displayed in the chart. By default in JACOB, this number is set to 50.

(b) `order` which defines the sorting order of the data returned, and only `ascending` or `descending` is allowed. This provides a way to examine the best/worst performing systems for a given database query.

(c) `lowBound` and `upBound` provides a plot such that only percentage deviations that are between these two bounds will be displayed. By default, `lowBound = 0` and `upBound = 1`.

3. CONCLUSIONS

To summarize, we have developed a dynamic database consisting of diverse and accurate benchmark data that may be useful when selecting quantum chemistry methods. We feel that such a real-time "evidence based" selection scheme would be of tremendous use to the traditional bench chemist wishing to perform (now routine) density functional theory calculations for structure optimizations. Furthermore, theoreticians can maintain up-to-date with density functional successes or failures. We anticipate that data mining may be of interest, for example, this may expose large inaccuracies/accuracies not typically observed across smaller scale studies. The JACOB database is under development within our laboratory, and the user-interface is currently accessible at www.wallerlab.org/jacob.

■ ASSOCIATED CONTENT

ⓢ Supporting Information
The structure of the JACOB XML and the corresponding schema for the JACOB version 0.1 are desposited in the Supporting Information. This information is available free of charge via the Internet at http://pubs.acs.org

■ AUTHOR INFORMATION

Corresponding Author
*E-mail: m.waller@uni-muenster.de.
Notes
The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

■ REFERENCES

(1) Sousa, S. F.; Fernandes, P. A.; Ramos, M. J. General Performance of Density Functionals. *J. Phys. Chem. A* **2007**, *111* (42), 10439−10452.
(2) Stephens, P. J.; Devlin, F. J.; Chabalowski, C. F.; Frisch, M. J. Ab Initio Calculation of Vibrational Absorption and Circular Dichroism Spectra Using Density Functional Force Fields. *J. Phys. Chem.* **1994**, *98* (45), 11623−11627.
(3) Becke, A. D. Density Functional Thermochemsitry 3: The Role of Exact Exchange. *J. Chem. Phys.* **1993**, *98* (7), 5648−5652.
(4) Lee, C.; Yang, W.; Parr, R. G. Development of the Colle-Salvetti Correlation-Energy Formula into A Functional of the Electron Density. *Phys. Rev. B* **1988**, *37* (2), 785−789.
(5) Vosko, S. H.; Wilk, L.; Nusair, M. Accurate Spin-Dependent Electron Liquid Correlation Energies for Local Spin-Density Calculations - A Critical Analysis. *Can. J. Phys.* **1980**, *58* (8), 1200−1211.
(6) Hohenstein, E. G.; Chill, S. T.; Sherrill, C. D. Assessment of the Performance of the M05-2X and M06-2X Exchange-Correlation Functionals for Noncovalent Interactions in Biomolecules. *J. Chem. Theory Comput.* **2008**, *4* (12), 1996−2000.

(7) Zhao, Y.; Truhlar, D. G. Density Functionals for Noncovalent Interaction Energies of Biological Importance. *J. Chem. Theory Comput.* **2007**, *3* (1), 289−300.

(8) Zhao, Y.; Truhlar, D. G. Applications and Validations of the Minnesota Density Functionals. *Chem. Phys. Lett.* **2011**, *502* (1−3), 1−13.

(9) Zhao, Y.; Truhlar, D. G. Density Functionals with Broad Applicability in Chemistry. *Acc. Chem. Res.* **2008**, *41* (2), 157−167.

(10) Jacquemin, D.; Perpete, E. A.; Ciofini, I.; Adamo, C.; Valero, R.; Zhao, Y.; Truhlar, D. G. On the Performances of the M06 Family of Density Functionals for Electronic Excitation Energies. *J. Chem. Theory Comput.* **2010**, *6* (7), 2071−2085.

(11) Goerigk, L.; Grimme, S. A General Database for Main Group Thermochemistry, Kinetics, and Noncovalent Interactions - Assessment of Common and Reparameterized (meta-)GGA Density Functionals. *J. Chem. Theory Comput.* **2010**, *6* (1), 107−126.

(12) Goerigk, L.; Grimme, S.; Thorough, A Benchmark of Density Functional Methods for General Main Group Thermochemistry, Kinetics, and Noncovalent Interactions. *Phys. Chem. Chem. Phys.* **2011**, *13* (14), 6670−6688.

(13) Goerigk, L.; Grimme, S. Efficient and Accurate Double-Hybrid-Meta-GGA Density Functionals-Evaluation with the Extended GMTKN30 Database for General Main Group Thermochemistry, Kinetics, and Noncovalent Interactions. *J. Chem. Theory Comput.* **2011**, *7* (2), 291−309.

(14) Krieg, H.; Grimme, S. Thermochemical Benchmarking of Hydrocarbon Bond Separation Reaction Energies: Jacob's Ladder Is Not Reversed! *Mol. Phys.* **2010**, *108* (19−20), 2655−2666.

(15) Oreilly, T. What is Web 2.0: Design Patterns and Business Models for the Next Generation of Software. *Comm. Strat.* **2007**, *1*, 17−37.

(16) Korth, M.; Grimme, S. "Mindless" DFT Benchmarking. *J. Chem. Theory Comput.* **2009**, *5* (4), 993−1003.

(17) Karton, A.; Tarnopolsky, A.; Lamere, J.-F.; Schatz, G. C.; Martin, J. M. L. Highly Accurate First-Principles Benchmark Data Sets for the Parametrization and Validation of Density Functional and Other Approximate Methods. Derivation of a Robust, Generally Applicable, Double-Hybrid Functional for Thermochemistry and Thermochemical Kinetics. *J. Phys. Chem. A* **2008**, *112* (50), 12868−12886.

(18) Curtiss, L. A.; Raghavachari, K.; Trucks, G. W.; Pople, J. A. GAUSSIAN-2 Theory for Molecular-Energies of 1st-Row and 2nd-Row Compounds. *J. Chem. Phys.* **1991**, *94* (11), 7221−7230.

(19) Parthiban, S.; Martin, J. M. L. Assessment of W1 and W2 Theories for The Computation of Electron Affinities, Ionization Potentials, Heats of Formation, and Proton Affinities. *J. Chem. Phys.* **2001**, *114* (14), 6014−6029.

(20) Zhao, Y.; Truhlar, D. G. Assessment of Density Functionals for pi Systems: Energy Differences between Cumulenes and Poly-ynes; Proton Affinities, Bond Length Alternation, and Torsional Potentials of Conjugated Polyenes; and Proton Affinities of Conjugated Shiff Bases. *J. Phys. Chem. A* **2006**, *110* (35), 10478−10486.

(21) Zhao, Y.; Lynch, B. J.; Truhlar, D. G. Development and Assessment of a New Hybrid Density Functional Model for Thermochemical Kinetics. *J. Phys. Chem. A* **2004**, *108* (14), 2715−2719.

(22) Zhao, Y.; Gonzalez-Garcia, N.; Truhlar, D. G. Benchmark Database of Barrier Heights for Heavy Atom Transfer, Nucleophilic Substitution, Association, and Unimolecular Reactions and Its Use to Test Theoretical Methods. *J. Phys. Chem. A* **2005**, *109* (9), 2012−2018.

(23) Neese, F.; Schwabe, T.; Kossmann, S.; Schirmer, B.; Grimme, S. Assessment of Orbital-Optimized, Spin-Component Scaled Second-Order Many-Body Perturbation Theory for Thermochemistry and Kinetics. *J. Chem. Theory Comput.* **2009**, *5* (11), 3060−3073.

(24) Zhao, Y.; Tishchenko, O.; Gour, J. R.; Li, W.; Lutz, J. J.; Piecuch, P.; Truhlar, D. G. Thermochemical Kinetics for Multireference Systems: Addition Reactions of Ozone. *J. Phys. Chem. A* **2009**, *113* (19), 5786−5799.

(25) Curtiss, L. A.; Raghavachari, K.; Redfern, P. C.; Pople, J. A. Assessment of Gaussian-2 and Density Functional Theories for The Computation of Enthalpies of Formation. *J. Chem. Phys.* **1997**, *106* (3), 1063−1079.

(26) Johnson, E. R.; Mori-Sanchez, P.; Cohen, A. J.; Yang, W. Delocalization Errors in Density Functionals and Implications for Main-Group Thermochemistry. *J. Chem. Phys.* **2008**, *129* (20), 204112−6.

(27) Grimme, S.; Kruse, H.; Goerigk, L.; Erker, G. The Mechanism of Dihydrogen Activation by Frustrated Lewis Pairs Revisited. *Angew. Chem., Int. Ed.* **2010**, *49* (8), 1402−1405.

(28) Grimme, S.; Steinmetz, M.; Korth, M. How to Compute Isomerization Energies of Organic Molecules with Quantum Chemical Methods. *J. Org. Chem.* **2007**, *72* (6), 2118−2126.

(29) Huenerbein, R.; Schirmer, B.; Moellmann, J.; Grimme, S. Effects of London Dispersion on the Isomerization Reactions of Large Organic Molecules: A Density Functional Benchmark Study. *Phys. Chem. Chem. Phys.* **2010**, *12* (26), 6940−6948.

(30) Bryantsev, V. S.; Diallo, M. S.; van Duin, A. C. T.; Goddard, W. A., III Evaluation of B3LYP, X3LYP, and M06-Class Density Functionals for Predicting the Binding Energies of Neutral, Protonated, and Deprotonated Water Clusters. *J. Chem. Theory Comput.* **2009**, *5* (4), 1016−1026.

(31) Jurečka, P.; Šponer, J.; Černý, J.; Hobza, P. Benchmark Database of Accurate (MP2 and CCSD(T) Complete Basis Set Limit) Interaction Energies of Small Model Complexes, DNA Base Pairs, and Amino Acid Pairs. *Phys. Chem. Chem. Phys.* **2006**, *8* (17), 1985−1993.

(32) Takatani, T.; Hohenstein, E. G.; Malagoli, M.; Marshall, M. S.; Sherrill, C. D. Basis Set Consistent Revision of the S22 Test Set of Noncovalent Interaction Energies. *J. Chem. Phys.* **2010**, *132*, 14.

(33) Grimme, S.; Antony, J.; Ehrlich, S.; Krieg, H. A Consistent and Accurate Ab Initio Parametrization of Density Functional Dispersion Correction (DFT-D) for the 94 Elements H-Pu. *J. Chem. Phys.* **2010**, *132*, 15.

(34) Csonka, G. I.; French, A. D.; Johnson, G. P.; Stortz, C. A. Evaluation of Density Functionals and Basis Sets for Carbohydrates. *J. Chem. Theory Comput.* **2009**, *5* (4), 679−692.

(35) Reha, D.; Valdes, H.; Vondrasek, J.; Hobza, P.; Abu-Riziq, A.; Crews, B.; de Vries, M. S. Structure and IR Spectrum of Phenylalanyl-Glycyl-Glycine Tripetide in the Gas-Phase: IR/UV Experiments, Ab Initio Quantum Chemical Calculations, and Molecular Dynamic Simulations. *Chem.—Eur. J.* **2005**, *11* (23), 6803−6817.

(36) Gruzman, D.; Karton, A.; Martin, J. M. L. Performance of Ab Initio and Density Functional Methods for Conformational Equilibria of C_nH_{2n+2} Alkane Isomers (n=4−8). *J. Phys. Chem. A* **2009**, *113* (43), 11974−11983.

(37) Wilke, J. J.; Lind, M. C.; Schaefer, H. F., III; Csaszar, A. G.; Allen, W. D. Conformers of Gaseous Cysteine. *J. Chem. Theory Comput.* **2009**, *5* (6), 1511−1523.

(38) Řezáč, J.; Hobza, P. Advanced Corrections of Hydrogen Bonding and Dispersion for Semiempirical Quantum Mechanical Methods. *J. Chem. Theory Comput.* **2011**, *8* (1), 141−151.

(39) Gráfová, L.; Pitoňák, M.; Řezáč, J.; Hobza, P. Comparative Study of Selected Wave Function and Density Functional Methods for Noncovalent Interaction Energy Calculations Using the Extended S22 Data Set. *J. Chem. Theory Comput.* **2010**, *6* (8), 2365−2376.

(40) Valdes, H.; Pluháčková, K.; Pitoňák, M.; Řezáč, J.; Hobza, P. Benchmark Database on Isolated Small Peptides Containing An Aromatic Side Chain: Comparison Between Wave Function and Density Functional Theory Methods And Empirical Force Field. *Phys. Chem. Chem. Phys.* **2008**, *10* (19), 2747−2757.

(41) Bühl, M.; Reimann, C.; Pantazis, D. A.; Bredow, T.; Neese, F. Geometries of Third-Row Transition-Metal Complexes from Density-Functional Theory. *J. Chem. Theory Comput.* **2008**, *4* (9), 1449−1459.

(42) Bühl, M.; Kabrede, H. Geometries of Transition-Metal Complexes from Density-Functional Theory. *J. Chem. Theory Comput.* **2006**, *2* (5), 1282−1290.

F

dx.doi.org/10.1021/ci300374g | *J. Chem. Inf. Model.* XXXX, XXX, XXX−XXX

(43) Waller, M. P.; Braun, H.; Hojdis, N.; Bühl, M. Geometries of Second-Row Transition-Metal Complexes from Density-Functional Theory. *J. Chem. Theory Comput.* **2007**, 3 (6), 2234−2242.

(44) Schneebeli, S. T.; Bochevarov, A. D.; Friesner, R. A. Parameterization of a B3LYP Specific Correction for Noncovalent Interactions and Basis Set Superposition Error on a Gigantic Data Set of CCSD(T) Quality Noncovalent Interaction Energies. *J. Chem. Theory Comput.* **2011**, 7 (3), 658−668.

(45) Berka, K.; Laskowski, R.; Riley, K. E.; Hobza, P.; Vondrášek, J. Representative Amino Acid Side Chain Interactions in Proteins. A Comparison of Highly Accurate Correlated ab Initio Quantum Chemical and Empirical Potential Procedures. *J. Chem. Theory Comput.* **2009**, 5 (4), 982−992.

(46) Riley, K. E.; Hobza, P. Investigations into the Nature of Halogen Bonding Including Symmetry Adapted Perturbation Theory Analyses. *J. Chem. Theory Comput.* **2008**, 4 (2), 232−242.

(47) Riley, K. E.; Hobza, P. Assessment of the MP2 Method, along with Several Basis Sets, for the Computation of Interaction Energies of Biologically Relevant Hydrogen Bonded and Dispersion Bound Complexes. *J. Phys. Chem. A* **2007**, 111 (33), 8257−8263.

(48) Šponer, J.; Zgarbová, M.; Jurečka, P.; Riley, K. E.; Šponer, J. E.; Hobza, P. Reference Quantum Chemical Calculations on RNA Base Pairs Directly Involving the 2′-OH Group of Ribose. *J. Chem. Theory Comput.* **2009**, 5 (4), 1166−1179.

(49) Rutledge, L. R.; Durst, H. F.; Wetmore, S. D. Evidence for Stabilization of DNA/RNA-Protein Complexes Arising from Nucleobase-Amino Acid Stacking and T-Shaped Interactions. *J. Chem. Theory Comput.* **2009**, 5 (5), 1400−1410.

(50) Churchill, C. D. M.; Wetmore, S. D. Noncovalent Interactions Involving Histidine: The Effect of Charge on π-π Stacking and T-Shaped Interactions with the DNA Nucleobases. *J. Phys. Chem. B* **2009**, 113 (49), 16046−16058.

(51) Wang, W.; Hobza, P. Theoretical Study on the Complexes of Benzene with Isoelectronic Nitrogen-Containing Heterocycles. *Chem. Phys. Chem.* **2008**, 9 (7), 1003−1009.

(52) Pluháčková, K.; Jurečka, P.; Hobza, P. Stabilisation Energy of C_6H_6..C_6X_6 (X = F, Cl, Br, I, CN) Complexes: Complete Basis Set Limit Calculations at MP2 and CCSD(T) Levels. *Phys. Chem. Chem. Phys.* **2007**, 9 (6), 755−760.

(53) Bates, D. M.; Anderson, J. A.; Oloyede, P.; Tschumper, G. S. Probing the Effects of Heterogeneity on Delocalized π-π Interaction Energies. *Phys. Chem. Chem. Phys.* **2008**, 10 (19), 2775−2779.

(54) Amicangelo, J. C.; Gung, B. W.; Irwin, D. G.; Romano, N. C. Ab Initio Study of Substituent Effects in the Interactions of Dimethyl Ether with Aromatic Rings. *Phys. Chem. Chem. Phys.* **2008**, 10 (19), 2695−2705.

(55) Molnar, L. F.; He, X.; Wang, B.; Merz, J. K. M. Further Analysis and Comparative Study of Intermolecular Interactions Using Dimers from the S22 Database. *J. Chem. Phys.* **2009**, 131 (6), 065102−16.

(56) Ran, J.; Hobza, P. On the Nature of Bonding in Lone Pair. π-Electron Complexes: CCSD(T)/Complete Basis Set Limit Calculations. *J. Chem. Theory Comput.* **2009**, 5 (4), 1180−1185.

(57) Kim, D. Y.; Singh, N. J.; Kim, K. S. Cyameluric Acid as Anion-π Type Receptor for ClO_4^- and NO_3^-: π-Stacked and Edge-to-Face Structures. *J. Chem. Theory Comput.* **2008**, 4 (8), 1401−1407.

(58) Morgado, C. A.; Jurečka, P.; Svozil, D.; Hobza, P.; Šponer, J. Balance of Attraction and Repulsion in Nucleic-Acid Base Stacking: CCSD(T)/Complete-Basis-Set-Limit Calculations on Uracil Dimer and a Comparison with the Force-Field Description. *J. Chem. Theory Comput.* **2009**, 5 (6), 1524−1544.

(59) Vogiatzis, K. D.; Mavrandonakis, A.; Klopper, W.; Froudakis, G. E. Ab Initio Study of The Interactions Between CO_2 and N-Containing Organic Heterocycles. *Chem. Phys. Chem.* **2009**, 10 (2), 374−383.

(60) Tsuzuki, S.; Honda, K.; Azumi, R. Model Chemistry Calculations of Thiophene Dimer Interactions: Origin of π-Stacking. *J. Am. Chem. Soc.* **2002**, 124 (41), 12200−12209.

(61) Copeland, K. L.; Anderson, J. A.; Farley, A. R.; Cox, J. R.; Tschumper, G. S. Probing Phenylalanine/Adenine π-Stacking

Interactions in Protein Complexes with Explicitly Correlated and CCSD(T) Computations. *J. Phys. Chem. B* **2008**, 112 (45), 14291−14295.

(62) Tsuzuki, S.; Mikami, M.; Yamada, S. Origin of Attraction, Magnitude, and Directionality of Interactions in Benzene Complexes with Pyridinium Cations. *J. Am. Chem. Soc.* **2007**, 129 (27), 8656−8662.

(63) Sinnokrot, M. O.; Sherrill, C. D. Highly Accurate Coupled Cluster Potential Energy Curves for the Benzene Dimer: Sandwich, T-Shaped, and Parallel-Displaced Configurations. *J. Phys. Chem. A* **2004**, 108 (46), 10200−10207.

(64) Hohenstein, E. G.; Sherrill, C. D. Effects of Heteroatoms on Aromatic π..π Interactions: Benzene, Pyridine and Pyridine Dimer. *J. Phys. Chem. A* **2009**, 113 (5), 878−886.

(65) Crittenden, D. L. A Systematic CCSD(T) Study of Long-Range and Noncovalent Interactions between Benzene and a Series of First- and Second-Row Hydrides and Rare Gas Atoms. *J. Phys. Chem. A* **2009**, 113 (8), 1663−1669.

(66) Sinnokrot, M. O.; Sherrill, C. D. Substituent Effects in π..π Interactions: Sandwich and T-Shaped Configurations. *J. Am. Chem. Soc.* **2004**, 126 (24), 7690−7697.

(67) Ringer, A. L.; Figgs, M. S.; Sinnokrot, M. O.; Sherrill, C. D. Aliphatic CH..π Interactions:, Methane-Benzene, Methane-Phenol, and Methane-Indole Complexes. *J. Phys. Chem. A* **2006**, 110 (37), 10822−10828.

(68) Tsuzuki, S.; Honda, K.; Uchimaru, T.; Mikami, M. High-Level Ab Initio Computations of Structures and Interaction Energies of Naphthalene Dimers: Origin of Attraction and Its Directionality. *J. Chem. Phys.* **2004**, 120 (2), 647−659.

(69) Tsuzuki, S.; Honda, K.; Uchimaru, T.; Mikami, M. Intermolecular Interactions of Nitrobenzene-Benzene Complex and Nitrobenzene Dimer: Significant Stabilization of Slipped-Parallel Orientation by Dispersion Interaction. *J. Chem. Phys.* **2006**, 125 (12), 124304−6.

(70) Tsuzuki, S.; Honda, K.; Uchimaru, T.; Mikami, M. Ab Initio Calculations of Structures and Interaction Energies of Toluene Dimers Including CCSD(T) Level Electron Correlation Correction. *J. Chem. Phys.* **2005**, 122 (14), 144323−8.

(71) Tsuzuki, S.; Honda, K.; Uchimaru, T.; Mikami, M.; Tanabe, K. The Magnitude of the CH/π Interaction between Benzene and Some Model Hydrocarbons. *J. Am. Chem. Soc.* **2000**, 122 (15), 3746−3753.

(72) Tsuzuki, S.; Honda, K.; Uchimaru, T.; Mikami, M.; Tanabe, K. The Interaction of Benzene with Chloro- and Fluoromethanes: Effects of Halogenation on CH/π Interaction. *J. Phys. Chem. A* **2002**, 106 (17), 4423−4428.

(73) Tsuzuki, S.; Honda, K.; Uchimaru, T.; Mikami, M.; Fujii, A. Magnitude and Directionality of the Interaction Energy of the Aliphatic CH/π Interaction: Significant Difference from Hydrogen Bond. *J. Phys. Chem. A* **2006**, 110 (33), 10163−10168.

(74) Tsuzuki, S.; Honda, K.; Uchimaru, T.; Mikami, M. Estimated MP2 and CCSD(T) Interaction Energies of n-Alkane Dimers at the Basis Set Limit: Comparison of the Methods of Helgaker et al. and Feller. *J. Chem. Phys.* **2006**, 124 (11), 114304−7.

(75) Tsuzuki, S.; Honda, K.; Fujii, A.; Uchimaru, T.; Mikami, M. CH/π Interactions in Methane Clusters with Polycyclic Aromatic Hydrocarbons. *Phys. Chem. Chem. Phys.* **2008**, 10 (19), 2860−2865.

(76) Marshall, M. S.; Steele, R. P.; Thanthiriwatte, K. S.; Sherrill, C. D. Potential Energy Curves for Cation..π Interactions: Off-Axis Configurations Are Also Attractive. *J. Phys. Chem. A* **2009**, 113 (48), 13628−13632.

(77) Rubesa, M.; Nachtigall, P.; Vondrasek, J.; Bludsky, O. Structure and Stability of the Water-Graphite Complexes. *J. Phys. Chem. C* **2009**, 113 (19), 8412−8419.

(78) *SpringSource GRAILS*. grails.org (accessed 1/10/2012).

(79) *PostgreSQL*. www.postgresql.org (accessed 15/9/2012).

(80) *Hibernate*. www.hibernate.org (accessed 1/10/2012).

(81) *SpringSource Groovy*. groovy.codehaus.org (accessed 1/9/2012).

(82) Boys, S. F.; Bernardi, F. The calculation of small molecular interactions by the differences of separate total energies. Some procedures with reduced errors. *Mol. Phys.* **1970**, *19*, 553.

(83) Nicholson, M. Searchable Plugin for Grails. http://grails.org/plugin/searchable (accessed 1/5/2012).

(84) Cass, M. E.; Rzepa, H. S.; Rzepa, D. R.; Williams, C. K. The Use of the Free, Open-Source Program Jmol To Generate an Interactive Web Site To Teach Molecular Symmetry. *J. Chem. Educ.* **2005**, *82* (11), 1736.

(85) *Google Visualization API.* https://developers.google.com/chart/interactive/docs/reference (accessed 1/10/2012).

(86) Muschko, B. Google Visualization Plugin for Grails. http://grails.org/plugin/google-visualization (accessed 1/10/2012).