

# Quantum Isostere Database: A Web-Based Tool Using Quantum Chemical Topology To Predict Bioisosteric Replacements for Drug Design

Mike Devereux,<sup>†,‡,§</sup> Paul L. A. Popelier,<sup>\*,‡,§</sup> and Iain M. McLay<sup>||</sup>

Manchester Interdisciplinary Biocentre (MIB), 131 Princess Street, Manchester M1 7DN, Great Britain,  
School of Chemistry, University of Manchester, Oxford Road, Manchester M13 9PL, Great Britain, and  
GlaxoSmithKline Medicines Research Centre, Computational and Structural Chemistry, Stevenage,  
Hertfordshire SG1 2NY, Great Britain

Received March 4, 2009

This paper introduces the ‘Quantum Isostere Database’ (QID), a Web-based tool designed to find bioisosteric fragment replacements for lead optimization using stored *ab initio* data. A wide range of original geometric, electronic, and calculated physical properties are stored for each fragment. Physical descriptors with clear meaning are chosen, such as distribution of electrostatic potential energy values across a fragment surface and geometric parameters to describe fragment conformation and shape from *ab initio* structures. Further fundamental physical properties are linked to broader chemical characteristics relevant to biological activity, such as H-bond donor and acceptor strengths. Additional properties with less easily interpretable links to biological activity are also stored to allow future development of QSAR/QSPR models for quantities such as  $pK_a$  and solubility. Conformational dependence of the *ab initio* descriptors is explicitly dealt with by storing properties for a variety of low-energy conformers of each fragment. Capping groups are used in *ab initio* calculations to represent different chemical environments, based on background research into transferability of electronic descriptors [J. Comput. Chem. 2009, 30, 1300–1318]. The resulting database has a Web interface that allows medicinal chemists to enter a query fragment, select important chemical features, and retrieve a list of suggested replacements with similar chemical characteristics. Examples of known bioisosteric replacements correctly identified by the QID tool are given.

## 1. INTRODUCTION

The concept of bioisosterism<sup>1–5</sup> continues to be widely applied in medicinal chemistry to make informed modifications to potential new drug compounds, while enhancing or maintaining the original biological activity. Such substitutions make it possible to alter key properties such as solubility or toxicity, without losing the potency of the original lead. While traditionally an intuitive concept, new insight can be gained by seeking the quantum chemical basis for bioisosterism and identifying the quantifiable physical characteristics that lead to similarity in biological activity. Here we introduce the Quantum Isostere Database (QID), a Web-based tool designed to store a range of novel quantum chemical descriptors for fragments of known drug compounds taken from the World Drug Index (WDI).<sup>6</sup> The descriptors stored were specifically designed to address the difficult problem of encapsulating biological activity directly from an isolated molecular wave function. They are combined in a user-friendly tool to help medicinal chemists to find bioisosteric fragment replacements in lead optimization projects.

Many of the chosen descriptors have their roots in Quantum Chemical Topology (QCT).<sup>7–9</sup> QCT extracts

precise physical properties for subsystems of molecular wave functions. It is possible to unambiguously assign regions of electron density to individual atoms or functional groups<sup>10,11</sup> and to obtain physical properties by means of a single volume integration<sup>12,13</sup> of corresponding property densities. In addition, bond properties are obtainable directly from the electron density at so-called ‘bond critical points’ (BCPs) lying between bonded nuclei. Surface properties can be defined by partitioning the molecular isodensity surface within the same QCT formalism. These fundamental theoretical quantities are then combined in the QID to form intuitive composite descriptors that encapsulate the chemistry of a fragment.

Successful prediction of biological activity and bioisosterism between fragments requires careful selection of chemical descriptors. Descriptors must encapsulate the propensity of a molecule to form interactions or undergo reactions within complex biomolecules, without explicitly modeling the process in question. They should be easily obtainable for as wide a range of compounds or compound fragments as possible, especially when applied to the large, diverse databases of many pharmaceutical companies. Many such descriptors have been developed and made widely available, especially for use in the field of QSAR. Holliday et al. recently demonstrated the ability of simple ‘R-group descriptors’,<sup>14</sup> to characterize biological activity of functional groups with a single attachment point. Significant relationships were found between descriptor similarity and bioisosteric relationships documented in the BIOSTER<sup>15</sup> database.

\* Corresponding author e-mail: pla@manchester.ac.uk.

<sup>†</sup> Current address: Department of Chemistry, University of Basel, Klingelbergstrasse 80, CH-4056 Basel, Switzerland.

<sup>‡</sup> Manchester Interdisciplinary Biocentre (MIB).

<sup>§</sup> University of Manchester.

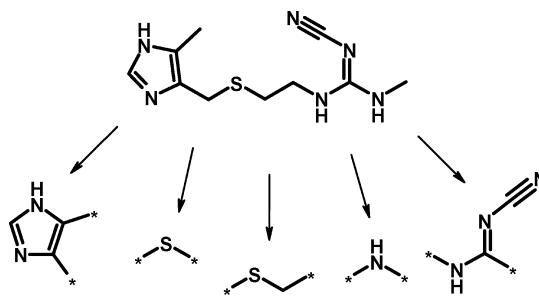
<sup>||</sup> GlaxoSmithKline Medicines Research Centre.

Large numbers of different descriptors are provided in software packages such as Dragon,<sup>16</sup> Sybyl-Selector,<sup>17</sup> TSAR,<sup>18</sup> and C<sup>2</sup>.Descriptor+.<sup>19</sup> Quantities such as estimates of molecular volume, components of the dipole moment vector, counts of methyl groups, and localized values of logP are commonly used. The descriptors included in such packages, and often used in industry,<sup>5,20,21</sup> are generally among the most rapid to obtain but not necessarily the most accurate or informative values available.

Geometric descriptors are also stored in the QID. Structural information is often used to describe molecular shape and distribution of physical features. Comparative Molecular Field Analysis (CoMFA), for example, incorporates 3D information into QSAR studies.<sup>22</sup> Pharmacophores are similarly used to compare the spatial distribution of molecular features, with efficient algorithms allowing rapid alignment into chemically significant orientations.<sup>23</sup> A further range of approaches have been introduced to describe a compound's physical, chemical, and biological characteristics. For example, preferences for formation of chemical interactions between different moieties have been identified by examining stored experimental X-ray structures<sup>24</sup> and by summarizing information from docking studies.<sup>25</sup> The Isostar<sup>26</sup> database also uses X-ray crystal structures combined with theoretical data to describe the preferred directionality of different intermolecular interactions such as H-bonds. Anzali et al.<sup>27</sup> described an approach to characterize molecular electrostatics, important to formation and stabilization of intermolecular interactions, using the electrostatic potential energy mapped onto a 2D tile with a Kohonen neural network, allowing rapid comparison of tiles to suggest bioisosteric replacements.

The task of finding bioisosteric fragment replacements has been addressed directly by the development of the BIOSTER database.<sup>15</sup> Rather than predicting bioisosteric relationships using similarity in molecular or fragment descriptors, the BIOSTER database acts as an extensive reference, storing details of many thousands of bioisosteric relationships documented in the literature. Such an approach cannot easily be used to predict new bioisosteric relationships, however, without understanding the origins of the preserved biological activity of stored compounds. It also difficult to predict whether a fragment pair that was found to be bioisosteric in one study against a given biological target will also be bioisosteric for a new class of compounds with a new target. A more recent, rational approach is the BROOD tool,<sup>28</sup> which combines 3D structures from force field calculations with a selection of readily available descriptors to suggest bioisosteric replacements for a query fragment. Shape comparisons can be made using a Gaussian overlay fingerprint approach,<sup>29</sup> and stored physical and chemical properties for a large number of fragments, with between 1 and 3 connection points, can be used to find suggested replacements.

The QID also uses a rational, descriptor-based approach to try to identify similarities in physical properties of a series of organic fragments that may lead to preserved biological activity of the parent compound. When selecting descriptors to encapsulate biological activity, a set of broad parameters, suggested in the early bioisosterism work of Thornber,<sup>3</sup> was taken as a starting point. Such quantities, including size, shape, and electronic distribution, can be interpreted and described in many different ways and at different levels of detail. Complementary *in silico* properties were therefore



**Figure 1.** Selected 1- and 2-atom linker fragments obtained after fragmentation of the drug cimetidine. Asterisks denote connection points.

selected for each descriptor category, allowing users to select the most appropriate descriptor for a given task. The computational expense required to generate the *ab initio* data used in the QID was carefully managed to provide an optimal balance between level of detail and reliability of information contained in the descriptors and the volume of data that could be generated with the computational resources available. While these descriptors are more computationally demanding than those of many other approaches, once they have been generated and stored they provide a wealth of chemical information that can be rapidly accessed via the QID Web interface using a standard Web browser.

## 2. METHODS

**2.1. Descriptor Calculation.** Daylight<sup>30</sup> substructure searching of the WDI provided unique fragments, termed 'linker groups', which are defined as fragments with two connection points to a parent compound. Figure 1 shows examples of linkers sampled from the drug cimetidine. In the present work, only linkers with either one or two atoms in the fragment backbone between the connection points are included. For example, in Figure 1 the one-atom linkers are \*-S-\* and \*-N[H]-\* where the atoms in brackets do not belong to the backbone. Examples of two-atom linkers are \*-S-C[H<sub>2</sub>]-\* and \*-N[H]-C[N]-\*. The stars represent either a carbon or nitrogen atom. Linkers were allowed to include rings as side chains or as part of their backbone, but no ring bonds were broken during fragmentation. In addition, no *sp* or *sp*<sup>2</sup> hybridized bonds were broken during fragmentation. Restriction of linkers to those with carbon or nitrogen neighbors at connection points allows the use of a minimal set of capping groups to generate *ab initio* fragment properties, as demonstrated in earlier work.<sup>31</sup> It is not yet clear whether the ethyl, phenyl, and pyrrole capping groups chosen could also be used to represent further common chemical environments, for example with sulfur or oxygen atoms at connection points.

So far, data have been generated for around 800 linkers (a number increased by associated low energy conformers for each) with ethyl capping groups and using the most abundant ionic form of each ionizable functional group at neutral pH. Conformers are generated for each linker group using a systematic conformational search of all rotatable internal degrees of freedom with the program Macromodel.<sup>32</sup> All unique conformers are then optimized in Gaussian98<sup>33</sup> at the AM1 semiempirical level of theory to obtain a set of low-energy structures, before further optimization at the RHF/6-31G\*\* level of theory. RHF/6-31G\*\* frequency calculations are performed to confirm the existence of a local

minimum, and the structures are optimized further at the B3LYP/6-311+G(2d,p) level. A final single point calculation is then performed at the MP2/6-311+G(2d,p)//B3LYP/6-311+G(2d,p) level to obtain a decent description of the molecular charge density, as demonstrated in the aforementioned transferability study.<sup>31</sup> The MP2 description of the molecular wave function is required to obtain reliable delocalization indices.

All QCT descriptors are calculated from the *ab initio* electron density using a combination of the AIMPAC suite of programs<sup>34</sup> and a modified, in-house version of the program MORPHY,<sup>35</sup> which locates<sup>36</sup> BCPs and performs atomic integrations.<sup>12,13</sup> The AIMPAC program was used to include atomic properties not currently available in the standard released version of MORPHY, such as the atomic overlap matrix that is needed to calculate delocalization indices. The delocalization indices were then calculated using an in-house FORTRAN program called ‘evaldili’. MORPHY was used to calculate surface properties across individual atomic grids of a given isodensity surface and to evaluate bond properties at BCPs. A BCP is a saddle point in the electron density, usually found between bonded nuclei, where  $\rho$  is a minimum in the direction running between the nuclei, and a maximum in the two remaining orthogonal directions. Grid points on atomic surface grids are evenly spaced, and the molecular isodensity surface is partitioned into atomic components using the same QCT formalism as to partition the molecular volume. The grid construction method is described in more detail elsewhere.<sup>37</sup> Grid points on the atomic surface are matched with corresponding points in a 3D grid of electrostatic potential energy values generated using the ‘CUBE’ option in Gaussian.

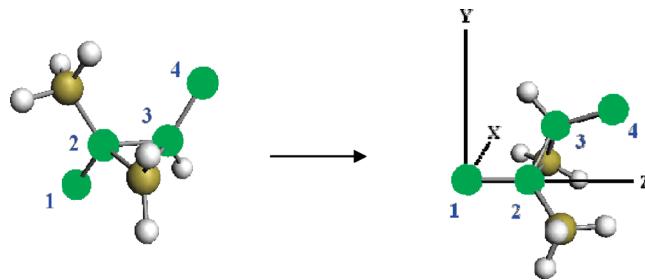
**2.2. Database.** Two versions of the fragment database were developed, the first using the commercial ORACLE package<sup>38</sup> and the second using ‘MySQL’.<sup>39</sup> Database queries were then made using the Perl DBI module to allow easy interfacing with the Web interface.

**2.3. Web Interface.** The Web interface exploits the CHIME PRO plug-in<sup>40</sup> to embed 3D representations of linkers. PERL CGI was used to dynamically create Web pages and to interface with the fragment database directly from the Web interface.

### 3. CHEMICAL DESCRIPTORS IN THE QID

This section introduces descriptors developed and stored in the QID. Section 4 then describes how these descriptors are incorporated into a Web tool and used to find bioisosteric replacements. Section 5 presents three case studies where the QID was used to find bioisosteric replacements or to search for linker replacements with similar physical or chemical characteristics.

**3.1. Shape Descriptors.** Size and shape are important in determining whether a ligand will physically fit into a target binding site. The stable conformations adopted by a linker will affect the geometry of the surrounding molecule and may favor or impede formation of a binding conformation. The detailed *ab initio* steric and conformational information stored in the QID was therefore exploited. A series of complementary descriptors was devised, including geometric measures to assess linker conformation, descriptions of linker



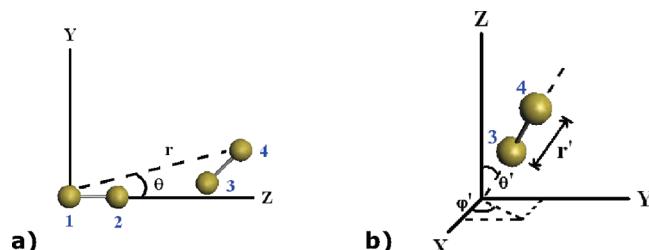
**Figure 2.** Standard orientation used to align linkers in the QID. The linker shown is a cyclopropyl linker with a methyl substituent. Atoms shown in green represent cleaved bonds connecting the linker to its original neighbors in the parent molecule, so atoms 1 and 4 correspond to the positions marked with asterisks shown in Figure 1.

volume and dimensions, and 3D fingerprints that allow fast comparison of volume overlap for prealigned structures.

**3.1.1. Conformation.** Fragment conformation strongly affects directional properties such as dipole moments, and the stable conformers a linker can adopt may influence whether a molecule is able to easily form a conformation that will bind to a given target. *Ab initio* structural and electronic properties were therefore calculated for a number of low-energy conformers of each linker. A diverse set of up to 20 stable conformers are stored in the QID for each flexible linker. Conformers chosen include all unique backbone conformers identified by a systematic search in Macromodel and the most diverse set of low-energy side chain conformations, where similarity is defined by average Euclidean distance between corresponding atoms. *Ab initio* optimizations are performed for each conformer, as outlined in the Methods section.

Similarity in linker conformation is most easily assessed when linkers are aligned. While meaningful alignment of molecules can be difficult, the linkers stored in the QID all contain two connection points. It is the relative orientation of these connection points that influences the global conformation of a molecule. Comparison of the relative orientation of linker group connection points in the QID is made by storing all linkers in a standard orientation. In the standard orientation, the first atom of the first linker bond is used to define the global origin, and the second atom is used to define the Z-axis (Figure 2). The tip of the second linker bond (atom 4) is then used to define the YZ plane. Atom 3 can lie outside the YZ plane, and the position and direction of bond 3–4 are used to compare the relative orientation of the two bonds. If a linker is asymmetric, the conformer can also be reversed so that bond 4–3 defines the global Z-axis, and atom 1 defines the YZ plane, producing a new unique conformer without the need for further *ab initio* calculations.

The conformation of the linker is described quantitatively using geometric measures of the separation and orientation of the second linker bond relative to the first. As the terminal atom of the first linker bond always lies at the origin, and the bond runs along the Z-axis, the direction of the other terminal bond is defined by a vector pointing from atom 3 to atom 4. The position of the second bond is defined using the angle and distance between atom 4 and the origin. This is made clear in Figure 3a. One angle is sufficient as atom 4 always lies within the ZY plane, by definition. Figure 3b shows how the direction of the second linker bond between atoms 3 and 4 is defined. A vector is described using

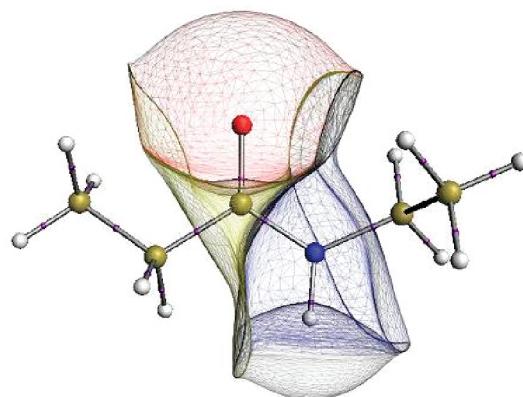


**Figure 3.** Geometric parameters describing a) position and b) direction of the second linker bond relative to the first.

spherical polar coordinates, where  $\theta'$  is the angle made with the global Z-axis,  $\varphi'$  is the angle between the global X-axis and the projection of the vector onto the XY plane, and  $r'$  is the bond length. The primes are used to distinguish these values from quantities shown in Figure 3a. The combined geometric parameters unambiguously describe the conformation of the linker backbone and can be weighted and quantitatively compared to describe similarity in linker conformation. Additionally, with linkers prealigned in a reference axis system, atomic Cartesian coordinates for all atoms and bonds inside different linkers can be compared directly, allowing searches in the QID for linkers with features at specified locations relative to linker connection points.

One potential problem in using this approach, however, is linker linearity, where the atoms of the second linker bond lie along or close to the Z-axis. For completely linear linkers, the YZ plane cannot be defined at all. For a linker that is close to being collinear, although the YZ plane is mathematically well-defined, the linker must be treated as if it were collinear. This is because the justification for using a standard axis system is that linkers with similar backbone geometries produce similar global conformations in a parent molecule. Similar coordinates relative to these connection points therefore mean that an atom will occupy a similar position relative to the remainder of the parent molecule. If a linker is completely linear, then the whole linker could be rotated about its connection point at one end while the second linker bond is still in the correct relative orientation to produce the same parent compound conformation. If the linker is only nonlinear to a small degree, then it could still be rotated about each connection point while producing a minimal effect on the global conformation of the parent compound. It is therefore not possible to exclude a fragment from search results on the basis that it lacks a feature at a specified set of Cartesian coordinates, as rotation might allow the feature to move while still approximately satisfying the global conformational constraints of the parent molecule. In both linear and nearly linear cases, then, Cartesian coordinates of noncollinear linker atoms are not compared directly between different linker groups. Instead, the distance and angle of each atom from the origin (analogous to Figure 3a) is used to identify atoms that could be superimposed by rotation about their respective Z-axes. Collinear and nearly collinear linker bonds are, however, uncommon among the linkers currently stored in the QID.

Finally, it should be noted that closely matching linker geometric parameters cannot always guarantee identical molecular geometries in a parent compound. Two linkers with identical backbone conformations may induce different preferred torsional rotations about their linker bonds. This

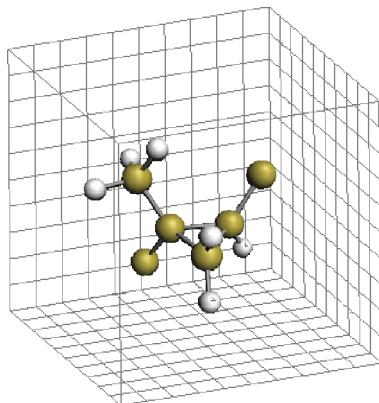


**Figure 4.** An amide linker group divided into atoms using QCT to partition the molecular charge density. A 0.001 au isodensity surface is used to cap the outer surface, and the electron density inside the volume associated with each atom is used to evaluate QCT properties such as charge and atomic volume.

is not easily predicted from analysis of the structures stored with their capping groups in the QID, as the favored torsional angle(s) of the capping group may not be representative of the favored torsional orientation of surrounding groups in a parent molecule. Linker geometric parameters should therefore be used primarily as a guide to suggest the likelihood of a similar impact on the global conformation of a parent compound.

**3.1.2. Size.** Linker size is compared using the spatial dimensions of each linker in its standard orientation, i.e. the maximum X, Y, and Z extent in the global axis system. This information is combined with the volume of the linker in order to distinguish between very large, bulky linkers and long, stretched linkers that may extend to large X-, Y-, and Z-coordinates without filling the ‘box’ that has these dimensions. Linker volume is defined as the sum of the QCT atomic volumes of all atoms inside the linker, capped by the 0.001 au isodensity surface (Figure 4).

**3.1.3. Shape Fingerprints.** Linker shape is another important consideration, affecting steric interactions with the rest of the parent molecule and the binding site of a target biomolecule. Where little is known about the steric interactions taking place inside a binding site, it may be safest to find linker isosteres with a similar shape to one another. One method to compare shape inside the QID has been based on the GaP method of Leach and co-workers.<sup>41</sup> In this approach, pharmacophores are rotated about a single connection point inside a 3D grid, and all grid elements that pharmacophoric points pass through are switched on in the corresponding bits of a ‘3D’ fingerprint. In the QID, the standard orientation of linkers allows them to be placed directly inside a 22 Å × 22 Å × 19 Å 3D grid without rotation (Figure 5), and any grid elements that lie within the Bondi radii<sup>42</sup> of any of the linker atoms are switched on in the corresponding bit of the fingerprint. As the atomic positions in the standard orientation are directly comparable, fingerprints can be rapidly compared and similarity-assessed using Tanimoto scoring<sup>43,44</sup> to gain an overall measure of the overlap of linker volumes. The large fingerprints produced, containing  $22 \times 22 \times 19 = 9196$  bits, are generally quite sparsely populated, but nonetheless provide a fast and relatively computationally efficient method to compare overlap of linker volumes.



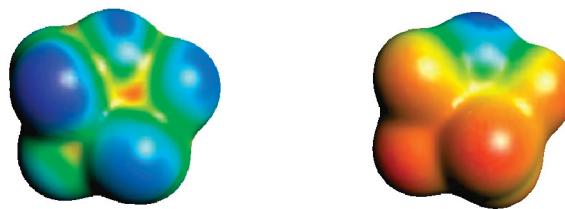
**Figure 5.** Linker group placed inside a 3D grid, representative of the technique used to obtain 3D structural fingerprints. Each grid element corresponds to a bit in the fingerprint, and elements occupied by part of the linker volume are switched on.

**3.1.4. Charge and Property Distribution Fingerprints.** Fingerprints are also used to compare the spatial distribution of individual physical or chemical features of stored linkers. Such fingerprints allow identification of linker replacements with, for example, similar spatial distribution of negatively charged atoms or H-bond basic atoms. These fingerprints are obtained in the same way as shape fingerprints, but only atoms with the requested physical properties are used to switch on the corresponding fingerprint bits. For example, the positions of H-bond acceptor atoms can be stored in a fingerprint for each linker and used to compare the spatial distribution of H-bond basic sites between linkers via fingerprint similarity scoring measures.

**3.2. Electronic Distribution.** Electronic distribution is another property considered by Thornber to be important to biological activity. The distribution of charge throughout a molecule will determine the electrostatic interaction with other moieties. Complementary electrostatics is important to ligand binding, both with the desired target, with solvent molecules, and with further biomolecules that may be significant to compound toxicity. Tight control of molecular electrostatics can be important in improving physical characteristics such as solubility, partition coefficients,<sup>45</sup> and H-bonding.<sup>46</sup> *Ab initio* calculations provide highly detailed descriptions of charge distribution, and QCT is used to describe the charge density belonging to atoms or functional groups.

**3.2.1. Atomic Multipole Moments.** Electron distribution within individual atoms is described by storing QCT atomic multipole moments up to the quadrupole moment. Directionality is included by rotating all multipole components into the global axis system of the linker's standard orientation. Multipole moments are described using spherical harmonics<sup>8,47,48</sup> and are rotated by first converting to their complex form, rotating, and then converting back to the real form, in accordance with a method outlined by Su et al.<sup>49</sup>

**3.2.2. Linker Multipole Moments.** Overall, the charge distribution within a linker group is described by summing the contributions of the component atoms, as described in the Appendix. Multipole moments up to the quadrupole moment are calculated and stored, again in the linker's standard orientation. Individual multipole moments are then directly compared, giving a quantitative and detailed measure of similarity in charge distribution within a linker. The reference origin used to define the linker multipole moments



**Figure 6.** Local ionization energy (left) and electrostatic potential (right) on 0.001 au isodensity surface of a THF molecule.

is the midpoint (geometric average, not the center-of-mass) between atoms 2 and 3 in the linker backbone (Figure 3a). This point was chosen as it is directly comparable between different linkers aligned in their standard orientation. Multipole moments then describe the distribution of charge relative to this origin, lying between the connection points of the parent molecule. This description is considered relevant when comparing the effect two linkers will have on the electrostatics of the parent molecule. Positions of other possible origins such as the center of mass may change significantly between linkers and do not allow direct comparison of the effect of substituting a linker on a parent molecule's charge distribution.

**3.2.3. Polarity.** It is not always necessary to compare the spatial distribution of charge. For example, the proportion of charged atoms in a linker or the proportion of a linker's isodensity surface that exhibits a large electrostatic potential energy may be more important to solubility than the precise way in which the total charge density is distributed. Linker polarity,  $\Pi$ , is therefore stored as an additional, nondirectional descriptor, defined as either the mean atomic charge in a linker (eq 1) or as an average of the mean value of the electrostatic potential across each atom's surface grid for all linker atoms (eq 2).

$$\Pi = \frac{1}{n} \sum_{i=1}^n |q_i(\Omega)| \quad (1)$$

$$\Pi = \frac{1}{n} \sum_{i=1}^n |\bar{V}(\mathbf{r}, \Omega_i)| \quad (2)$$

The sums either run over the atomic charge,  $q(\Omega)$ , or the mean electrostatic potential of the surface grid,  $\bar{V}$ , of each atom  $i$  of  $n$  atoms in a linker.

**3.3. Surface Properties.** The electrostatic potential,  $V(\mathbf{r})$ , and the local ionization energy,  $I(\mathbf{r})$ , are summarized and stored across atomic grids on the  $1 \times 10^{-3}$  and  $1 \times 10^{-5}$  au isodensity surfaces. Both properties are illustrated in Figure 6. The local ionization energy is defined<sup>50</sup> in eq 3

$$I(\mathbf{r}) = \sum_{i=1}^{HOMO} \frac{\rho_i(\mathbf{r})|\varepsilon_i|}{\rho(\mathbf{r})} \quad (3)$$

where  $\rho_i$  is the electron density associated with orbital  $i$ , and  $\varepsilon_i$  is this molecular orbital's energy. The local ionization energy provides a theoretical measure of the energy required, on average, to remove an electron from a point  $\mathbf{r}$ . The regions of lowest  $I(\mathbf{r})$  on a molecular surface have been found to correspond to regions most susceptible to charge transfer to an electrophile.<sup>51</sup> While the information contained in  $I(\mathbf{r})$  is known to be useful when describing the electron donor

properties of a molecule, it has been less successful in describing electron affinity and acceptor properties.<sup>52</sup> The information contained in  $I(\mathbf{r})$  is complementary to that contained in  $V(\mathbf{r})$ , with  $V(\mathbf{r})$  describing the propensity of a linker to attract another moiety with complementary electrostatics to a point  $\mathbf{r}$  on the linker's surface.  $I(\mathbf{r})$  describes the stabilizing effect obtained by withdrawing electron density from an atom to favor, for example, formation of an H-bond.

The distribution of  $I(\mathbf{r})$  and  $V(\mathbf{r})$  across each atomic surface is summarized by storing the minimum, maximum, mean, median, and standard deviation of the given property across grid points on each atomic surface. While the different statistical measures are clearly correlated, extra information can be gained by describing, for example, both the lowest value of the electrostatic potential (determining how strongly an oppositely charged species is attracted to a particular point) and the mean or median value, including information about what proportion of the remainder of the atomic surface carries a similar value of  $V(\mathbf{r})$ .

**3.4. H-Bonding.** H-Bonding is another of Thornber's suggested parameters. While H-bond donor and acceptor strengths of individual atoms are not easily calculated directly, it proved possible to develop accurate QSPR models using *ab initio* and QCT descriptors that could be readily generated and stored in the QID. The models, introduced elsewhere,<sup>46,53</sup> relate Abraham's H-bond donor and acceptor strengths ' $\alpha_2^H$ ' and ' $\beta_2^H$ ',<sup>54,55</sup> to the electrostatic potential energy at the nuclear position of an H-bond acidic proton and to a combination of the electrostatic potential and local ionization energy on the 0.001 au isodensity surface of H-bond acceptor atoms. The complete relations are shown below (eqs 4a–4c).

$$\alpha_2^H = 5.3664EP_{Nuc} + 5.7765 \quad (4a)$$

$$\beta_2^H = -2.536 \times 10^{-3}V_{\min} - 2.865 \times 10^{-3}V_{mean} - 0.163 \quad (4b)$$

$$\beta_2^H(N) = -2.243 \times 10^{-3}V_{\min} - 1.631 \times 10^{-3}V_{med} - 3.066I_{\min} + 1.5213 \quad (4c)$$

$EP_{Nuc}$  is the electrostatic potential at the proton's nuclear position,  $V_{\min}$ ,  $V_{mean}$ , and  $V_{med}$  are the minimum, mean, and median values of the electrostatic potential across the atomic 0.001 au isodensity surface of an H-bond acceptor atom, respectively.  $I_{\min}$  is the minimum value of the local ionization index on the atomic surface. Construction and definition of grids across the atomic surface has been outlined in Section 2 and in greater detail elsewhere.<sup>56</sup> The quantities  $\alpha_2^H$  and  $\beta_2^H$  are stored for all H-bond acidic protons and H-bond bases inside each linker. Searching for linkers with similar H-bonding characteristics is then possible, either by counting the number of donors and acceptors or by quantitatively comparing the positions and strengths of donor and acceptor atoms using Cartesian coordinates or shape fingerprint searches.

**3.5. Additional Electronic Descriptors.** Of Thornber's eight original parameters, descriptors for size, shape, charge distribution, and H-bonding have been introduced above and included in the QID. The remaining quantities (lipid and water solubility,  $pK_a$ , and reactivity) are not easily defined

**Table 1.** List of Atomic Properties and Formal Definitions Available in the AIM2000 Package and Stored in the QID

atomic property	definition
volume	$Vol(\Omega) = \int_{\Omega} d\tau$
charge	$q(\Omega) = Z_{\Omega} - \int_{\Omega} d\tau \rho(\mathbf{r})$
dipole moment	$Q_{10}(\Omega) = -\int_{\Omega} d\tau r_z \rho(\mathbf{r})$
quadrupole moment	$Q_{20}(\Omega) = -\int_{\Omega} d\tau (1)/(2)(3r_z^2 - r^2)\rho(\mathbf{r})$
Laplacian	$L(\Omega) = -\nabla^2 \rho(\mathbf{r})$
Lagrangian kinetic energy	$K(\Omega) = \int_{\Omega} d\tau (-1)/(4)Nf d\tau' [\psi^* \nabla^2 \psi + \psi \nabla^2 \psi^*]$
missing information	$I(\Omega) = \int_{\Omega} d\tau (-(\rho(\mathbf{r}))/N) \ln ((\rho(\mathbf{r}))/N))$
force exerted on nucleus by atomic density	$F_x = \int_{\Omega} d\tau (\rho(\mathbf{r})Z_{\Omega}(x)/(r^3))$
corrected attraction energy of atomic density to nucleus	$VC_{ne}^0(\Omega) = (-2(1 - \gamma))/(\gamma) \int_{\Omega} d\tau - (Z_{\Omega}\rho(\mathbf{r}))/(r)$
corrected electron - electron repulsion energy	$VC_{ee}(\Omega) = (-2(1 - \gamma))/(\gamma) \int_{\Omega} d\tau V_{ee}(\mathbf{r})$
total potential energy	$V_{tot}(\Omega) = 2 \int_{\Omega} d\tau (K(\mathbf{r})(1 - \gamma))$

within the framework of QCT and are not yet implemented. Progress has previously been made, however, in predicting  $pK_a$  values for sets of structurally related compounds using *ab initio* bond lengths and QCT descriptors at BCPs.<sup>57–59</sup> Platts et al. have also made substantial progress using QCT descriptors to predict solubility parameters<sup>45,53,60,61</sup> using Abraham's solvation equations.<sup>62,63</sup> Clearly no such general approach is likely to be successful in describing reactivity, as reactions involving drug compounds can take place in the body via many different pathways,<sup>64</sup> and it is not clear whether an individual linker can possess an inherent biological reactivity without considering the broader structure of the parent compound. QCT models have, however, shown some promise in predicting the propensity of a given class of compound to undergo specific biological reactions.<sup>59</sup> The basic approach common to the development of the QSPR and QSAR models for molecular solubility,  $pK_a$ , and reactivity was fitting of linear or nonlinear expressions to relate *ab initio* and QCT descriptors to the property of interest. A wide range of theoretical parameters has therefore been included in the QID, allowing future development of models such as these.

The additional theoretical properties, which will now be described, are also used to search for common features among sets of linkers that are known to act as bioisosteres. For example, if a series of bioisosteric linkers all possess an atom with a similar total potential energy at the same position, then it can be inferred that this feature is important to activity and used to find further possible bioisosteric replacements.

**3.5.1. Atom and Bond Properties.** The full range of QCT atomic properties currently stored, all available in the AIMPAc program,<sup>65</sup> is shown in Table 1, along with formal definitions. In the notation used in the table,  $\mathbf{r}$  is a position vector inside the volume of atom  $\Omega$  defined relative to the position of the nucleus,  $r = |\mathbf{r}|$ ,  $r_z$  is the z-component of  $\mathbf{r}$ ,  $Z_{\Omega}$  is the nuclear charge of atom  $\Omega$ , and  $\gamma = -V/T$  is the virial ratio of the molecule.<sup>7</sup> For true wave functions, this ratio of kinetic energy over potential energy must be exactly  $-2$ . Approximate wave functions can in principle be made to satisfy this ratio by multiplying each coordinate by a scale factor, which is not done here. More information on these atomic properties can be found in ref 66. A range of QCT bond properties is evaluated at BCPs and is included to describe bonding within each linker. The full list is given in Table 2.

**Table 2.** List of Bond Properties Available in MORPHY98 and Selected for Use in the QID

bond property (at BCP)	definition
density	$\rho(\mathbf{r})$
Laplacian	$L(\mathbf{r}) = -\nabla^2\rho(\mathbf{r})$
ellipticity	$\varepsilon(\mathbf{r}) = (\lambda_1/\lambda_2) - 1$
Lagrangian kinetic energy	$K(\mathbf{r}) = -(1)/(4)N_f d\tau [\psi^* \nabla^2 \psi + \psi \nabla^2 \psi^*]$
Hamiltonian kinetic energy	$G(\mathbf{r}) = (1)/(2)N_f d\tau \nabla \psi^* \cdot \nabla \psi$
eigenvalues of Hessian	$\lambda_1 + \lambda_2 + \lambda_3 = \nabla^2 \rho(\mathbf{r})$
bond length	$d_{AB} =  \mathbf{r}_A - \mathbf{r}_B $

**3.5.2. Electron Delocalization and Bond Order.** A final, more chemically intuitive bond property has also been included: the ‘delocalization index’, denoted as  $\delta(A,B)$ ,<sup>67–69</sup> where  $A$  and  $B$  are two topological atoms. A full discussion of the significance and interpretation of this quantity can be found in reference.<sup>70</sup> Derived from the electron pair density, the delocalization index of a pair of bonded atoms gives an intuitive measure of bond order, similar to that of Ángyán et al.<sup>71</sup> At RHF level,  $\delta(A,B)$  affords values for electron sharing that are in accord with the classical conceptual ‘Lewis model’, such as a bond order of approximately 3.0 for the triple bond of the  $N_2$  molecule.<sup>70</sup> At higher levels of theory, though, coulomb correlation leads to a decrease in  $\delta(A,B)$  from Lewis model predictions, but qualitative agreement is maintained. Evaluating  $\delta(A,B)$  between bonded atoms gives a quantitative and intuitive means of comparing the bond orders of different bonds, based on the degree of electron pair overlap between  $A$  and  $B$ . The delocalization index and a similar property called the ‘localization index’  $\lambda(A,A)$ , which describes the extent to which electrons are localized on a single atom, are defined as

$$\delta(A,B) = F(A,B) + F(B,A) \quad (5)$$

$$\lambda(A,A) + \sum_B F(A,B) = -N(A) \quad (6)$$

$$F(A,B) = - \sum_{lm} n_l^{1/2} n_m^{1/2} S_{lm}(A) S_{lm}(B) \quad (7)$$

Here  $n_l$  and  $n_m$  are occupation numbers of natural orbitals  $l$  and  $m$ ,  $S_{lm}(A)$  is element  $l,m$  of the atomic overlap matrix of atom  $A$ , and  $F(A,B)$  is the ‘Fermi correlation’. The atomic overlap matrix is evaluated via a volume integral of the product of each pair of orbitals (the so-called ‘pair density’) over the volume of atom  $A$ .

The second use of  $\delta(A,B)$  is quantifying charge delocalization throughout an entire structure. In this way, electron delocalization in aromatic ring systems, for example, can

be quantifiably compared using  $\delta(A,B)$  between atoms at corresponding ring positions.<sup>72–74</sup> As resonance stabilization affects fragment chemistry and aromatic ring systems form an important class of substructures in drug compounds,<sup>21</sup> characterization of aromaticity and electron delocalization is an important goal for the QID.

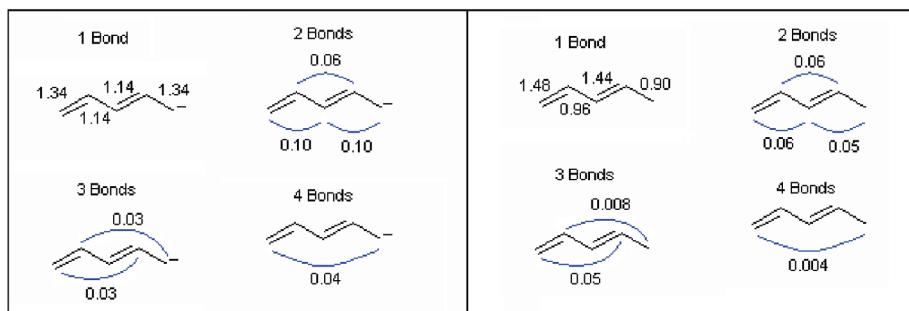
Figure 7 shows how electron delocalization can be quantitatively measured and illustrates, as noted previously,<sup>75</sup> that a conceptual distinction must be made between resonance and conjugation. Conjugated systems (right-hand side of Figure 7) show high degrees of charge delocalization across formal double bonds, but the alternating single bonds prevent electron delocalization throughout the remainder of the structure. Aromatic compounds and systems with formal resonant structures, however, show significant electron delocalization between even remote atom pairs (left-hand side of Figure 7).

$\delta(A,B)$  is currently only stored for bonded pairs of atoms as an estimate of bond order across each linker bond. This will be extended in the future, to describe electron delocalization in ring systems and resonant systems such as carboxyl groups that exist as anions or cations and exhibit resonance stabilization at neutral pH. The atomic overlap matrix of each linker atom is therefore stored for future use.

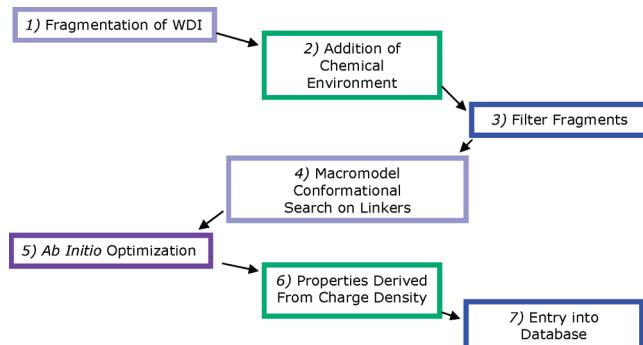
In summary, a diverse range of descriptors has been developed and implemented into the QID. The underlying philosophy when developing descriptors was to link fundamental, rigorous, and well-defined theoretical quantities from *ab initio* calculations to intuitive concepts widely used and understood in organic chemistry. Broad parameters suggested by Thornber were used to select descriptors with intuitive links to biological activity. Of his original eight suggested parameters, descriptors for size, shape, electronic distribution, and H-bonding have been included. Additional quantities such as bond order, that have clear chemical meaning, are also included for direct use. A wide range of other theoretical properties has also been stored to allow development of QSAR/QSPR models for Thornber’s remaining parameters and to help find trends in physical characteristics among series of known bioisosteres.

#### 4. ARCHITECTURE OF THE QID

**4.1. Linker Selection.** The linker groups chosen for inclusion in the QID were taken from the WDI, avoiding proprietary issues with GSK’s own, larger compound database. The WDI was considered suitable, as molecules contained therein are inherently ‘druglike’. Once fragmenta-



**Figure 7.** Pentadiene anion (left) with delocalization index values shown for bonded atoms and for pairs of atoms connected by blue lines. The neutral molecule (right) shows lower delocalization index values across corresponding bonds, despite conjugation, due to the lack of electron delocalization in this system.



**Figure 8.** Summary of process used to generate data for entry into the QID.

tion of the WDI had been completed, the resulting linkers had to be screened visually to filter out fragments such as peroxide bonds with known stability or reactivity issues. The used fragmentation approach implied that common linker groups were statistically much more likely to be found early on in the WDI compound fragmentation process. Methylenes linkers, for example, occur in almost every drug compound and are hence found after fragmenting only a few compounds. Uncommon linkers are not likely to be found until many compounds have been fragmented. Hence these linkers occur lower down the list, although some appear early on. Data were generated for linkers in the order they were identified during fragmentation, meaning that in the early stages of data generation the most common linkers are included alongside a few more unusual examples. This initial mixture of common and unusual linkers is advantageous for testing purposes because it allows identification of well-known bioisosteric replacements while also allowing a number of novel suggestions to be made.

**4.2. Data Storage.** Data were generated as outlined in the Methods section. A summary of the process, from fragmentation of the WDI to storage in the database, is shown in Figure 8. The data generated for each linker were stored in a relational database, with versions using both ORACLE and MySQL database software. A simple database architecture with 3 tables of properties was chosen, the first containing atomic properties, the second bond properties, and the third whole linker properties (Tables 3, 4 and 5, respectively). The data used to form the *primary key* of each table are marked with an asterisk. This key uniquely identifies each data entry and links atom, bond, and whole linker properties stored in the different tables.

**4.3. Web Interface.** The QID is accessed via a Web interface, allowing interactive construction of database search queries without the need to install new software. Users are guided through a series of Web pages to build each query. The linker of interest must first be entered, either by typing the fragment SMILES or building the structure using an ISIS Draw (Symyx) plug-in (Figure 9). The fragment SMILES is submitted to the database, and if data have been generated and stored, then the available conformers will be displayed using a CHIME plug-in. If prior knowledge of the most relevant conformation of a lead compound of interest exists, then the closest matching stored linker conformer can be selected. The full list of stored data for each conformer can also be viewed at this stage if requested.

The second step is to select search criteria that will be used to build a SQL query. The user is invited to select which

**Table 3.** Database Table Structure for Storage of Atom Data in the QID<sup>a</sup>

Name	Description	Data Type
SMILES *	SMILES representation of linker fragment	CHAR(130)
CONF_NO *	id number given to each conformer	CHAR(6)
ATOM_TYPE	Atom type	CHAR(4)
ATOM_NO *	Atom number from wfn file	INT
FRAG_NO	id number given to each linker	INT
VOL	Atomic volume, capped at 0.001 a.u. isosurface	FLOAT
X	Spatial x coordinate	FLOAT
Y	y coordinate	FLOAT
Z	z coordinate	FLOAT
CHARGE	Atomic Charge	FLOAT
DIP_X	x-component of atomic dipole moment	FLOAT
DIP_Y	y-component	FLOAT
DIP_Z	z-component	FLOAT
DIPOLE	Magnitude of atomic dipole moment	FLOAT
Q20	Q20-component of atomic quadrupole moment	FLOAT
Q21C	Q21c component	FLOAT
Q21S	Q21s component	FLOAT
Q22C	Q22c component	FLOAT
Q22S	Q22s component	FLOAT
QUADRUPOLE	Magnitude of atomic quadrupole moment	FLOAT
V_MAX_003	Max value of V(r) on 1d-3 au atomic surface	FLOAT
V_MIN_003	Min value of V(r) on 1d-3 au surface	FLOAT
MEAN_V_003	Mean value	FLOAT
V_DEV_003	Std deviation in values	FLOAT
V_MEDIAN_003	Median value	FLOAT
N PTS_003	No. of points in grid on 1d-3 au atomic surface	INT
V MAX_005	Max value of V(r) on 1d-5 au atomic surface	FLOAT
V MIN_005	Min value of V(r) on 1d-5 au surface	FLOAT
MEAN_V_005	Mean value	FLOAT
V DEV_005	Std deviation in values	FLOAT
V MEDIAN_005	Median value	FLOAT
N PTS_005	No. of points in grid on 1d-5 au atomic surface	INT
ERROR_L	Integration error measured via Laplacian	FLOAT
KE	Atomic kinetic energy	FLOAT
I	Missing information index	FLOAT
V_ATOM	Total atomic potential energy	FLOAT
V_NEO_COR	Corr. attraction energy of electron density to nucleus	FLOAT
V_EET_COR	Corr. electron - electron repulsion energy	FLOAT
FAA	Force exerted on nucleus by electron density	FLOAT
L INDEX	Localization index	FLOAT
IE_MIN	Min value of I(r) on 1d-3 au atomic surface	FLOAT
IE_MAX	Max value of I(r) on 1d-3 au atomic surface	FLOAT
IE_MEAN	Mean value of I(r) on 1d-3 au atomic surface	FLOAT
EPNUC	Electrostatic potential at nuclear position	FLOAT
PKHA	Estimate of acidic proton $\alpha_2^{\text{H}}$	FLOAT
PKHB	Estimate of basic atom $\beta_2^{\text{H}}$	FLOAT

\*Elements of Primary Key.

<sup>a</sup> Parameters are color-coded by type according to the following: orange = shape, blue = charge distribution, red = H-bonding, green = miscellaneous. Values written in black are used to identify a given atom. Asterisks indicate values form part of the primary key.

characteristics are considered important to the activity of the query linker, so that only linkers possessing these characteristics will be suggested as replacements (Figure 10). A list of intuitive descriptors is offered by default, including linker size, shape, conformation, charge distribution, polarity, and atomic H-bond donor/acceptor strength, defined by combinations of the descriptors introduced earlier. Advanced searches involving the full list of stored theoretical QCT values can be performed if preferred. The option to apply a weighting factor is also available, so that selected criteria will be given increased importance when assessing similarity scores. Results are summarized on a new page, ranked by similarity to the query linker (Figure 11). It should be noted that the specific search criteria and weightings required will depend on the role of the linker group in the parent lead compound. A given linker may perform a largely conformational role in one parent compound, while providing an H-bond donor or acceptor site in another. Search criteria should therefore be selected to maximally exploit any available chemical information describing the ligand–receptor interaction.

**Table 4.** Database Table Structure for Storage of Bond Data in the QID<sup>a</sup>

Name	Description	Data Type
SMILES *	SMILES representation of linker fragment	CHAR(130)
CONF_NO *	id number given to each conformer	CHAR(6)
ATOM_TYPE_1	Atom type of 1st bonded atom	CHAR(4)
ATOM_NO_1 *	Atom number of 1st bonded atom	INT
ATOM_TYPE_2	Atom type of 2nd bonded atom	CHAR(4)
ATOM_NO_2 *	Atom number of 2nd bonded atom	INT
X	Spatial x coordinate of bond critical point (BCP)	FLOAT
Y	y coordinate	FLOAT
Z	z coordinate	FLOAT
RHO	Electron density at BCP	FLOAT
LAPLACIAN	Laplacian at BCP	FLOAT
LAMBDA_1	1st eigenvalue of Hessian at BCP	FLOAT
LAMBDA_2	2nd eigenvalue of Hessian at BCP	FLOAT
LAMBDA_3	3rd eigenvalue of Hessian at BCP	FLOAT
ELLIPTICITY	Ellipticity of electron density at BCP	FLOAT
KE	Kinetic energy K(r) at BCP	FLOAT
G	Kinetic energy G(r) at BCP	FLOAT
LENGTH	Bond length	FLOAT
BOND_ORDER	Bond order (delocalization index)	FLOAT

\* Elements of Primary Key.

<sup>a</sup> Descriptors written in green are values that are used to characterize bonding. Values written in black are used to identify a given bond. Asterisks indicate that values form part of the primary key.

**4.4. Scoring.** Hits are scored using a weighted Euclidean distance from the query linker in property space (eq 8). Distances for each property are normalized using the user-defined search tolerance in the property value

$$D_{AB} = \sqrt{\sum_{i=1}^{nprops} \left( I_i \left( \frac{P_{i(A)} - P_{i(B)}}{P_{i(max)} - P_{i(min)}} \right) \right)^2} \quad (8)$$

Here ' $D_{AB}$ ' is the calculated distance between linkers A and B from selected properties indexed by  $i$ .  $P_{i(A)}$  ( $P_{i(B)}$ ) is the value of property ' $i$ ' for linker A (B), while  $P_{i(max)}$  and  $P_{i(min)}$  are the upper and lower search tolerances in property ' $i$ ', respectively.  $I_i$  is a multiplier representing the 'importance' of a search term, the optional user-specified weighting factor. An artifact of this definition of Euclidean distance is that increasing the tolerances in a particular property will affect normalization and decrease its contribution to  $D_{AB}$ , reducing the property's importance to overall linker similarity scores. There is clearly scope for testing the impact of different scoring functions as the QID reaches later stages of development.

## 5. CASE STUDIES

Three case studies are now presented, which help to validate the QID approach and to illustrate different applications of the tool.

**Example 1: Sample Search To Replace the Amide Linker.** In this study, bioisosteric replacements are suggested for two conformers of the amide linker. Results are presented in Figure 12. The amide linker was chosen as it has well-known bioisosteric replacements. The two conformers are shown in the top-left corners of Figure 12a,b. A search was performed to find replacements with similar overall shape and conformation to the query conformers and an H-bond acceptor at the position of the amide's carbonyl oxygen atom. The search demonstrates the ability of the QID to combine *ab initio* structural data with site-specific chemical characteristics, which can be difficult to achieve using simpler

approaches. Shape is described by linker volume and linker dimensions, geometry by angles and distances between the linker bonds, and H-bond acceptor strength by predicted  $\beta_2^H$  values.

Figure 12a contains five suggested replacements for the first conformer of the amide group. All suggested replacements display similar linker shape and conformation, as specified in the search criteria, and oxygen atoms that will act as H-bond acceptors of similar strength at the original amide oxygen position. The highest ranked replacements suggested by the QID substitute the NH group of the amide linker with either an oxygen atom to form the ester linker or by a methylene group (top right in Figure 12a). Both resulting linkers are known classical isosteres of the amide linker.<sup>76</sup> The remaining three suggested replacements also fit the search criteria supplied and may represent more novel bioisosteric replacements.

A search to replace the second conformer returned similar results (Figure 12b). The top 5 hits are again shown and include intuitive amide replacements. As well as the replacement of the amine group with a sulfur atom, another known classical isostere,<sup>77</sup> a ring system, is suggested. The electron rich nitrogen atom of the ring lies at the same geometric position as the carbonyl oxygen atom of the amide query linker, and overall linker conformation is preserved. The ring replacement is an example of nonclassical bioisosterism and could be used to reduce unwanted flexibility in a lead compound.

**Example 2: Identification of Common Bioisosteric Replacements.** The second, larger study uses the QID to find a variety of common fragment replacements published by Sheridan.<sup>4</sup> Sheridan searched the MDL Drug Data Report (MDDR) database to identify the most common replacements for fragments in drug molecules. Compounds were examined by therapeutic area and were clustered by overall structural similarity. Structurally related compounds were compared to identify those that differ by only one moiety. The portion of the molecule that differs represents a possible bioisosteric replacement, preserving the activity of the drug compound. Sheridan's method cannot, however, detect replacements of different topological length. R<sub>1</sub>-X-R<sub>2</sub> would therefore never be found as a replacement for R<sub>1</sub>-X-Y-R<sub>2</sub>, so such replacements are not examined here. Furthermore, Sheridan's approach does not show whether suggested replacements have some shared common biological functionality, or whether they are simply common organic replacements not involved in compound activity. Replacements studied may therefore not be true bioisosteres.

The fragment pairs selected from Sheridan's work are shown on the left-hand side of Table 6. For each pair of bioisosteric replacements, the fragment on the left was entered into the QID as a query in an attempt to retrieve the fragment on the right. Criteria were specified to characterize key features of the left linker, and a supervised search was performed to try to obtain Sheridan's replacement among the highest ranking hits (where a hit is any linker matching the specified search criteria). The results of each search performed in the QID are shown in the right-hand side of Table 6. The total number of hits retrieved by each search is given, using the search criteria specified in Table 6 and default property tolerances. The ranking of Sheridan's

**Table 5.** Database Table Structure for Storage of Linker Data in the QID<sup>a</sup>

Name	Description	Data Type
SMILES *	SMILES representation of linker fragment	CHAR(130)
CONF_NO *	id number given to each conformer	CHAR(6)
FRAG_NO	id number given to each linker	INT
N_ATOMS	Number of atoms in linker	INT
N_HEAVY	Number of heavy (non-hydrogen) atoms	INT
FRAG_VOL	Linker volume (capped at 1d-3au surface)	FLOAT
THETA_I	Conformational descriptor 'theta1'	FLOAT
R_I	Conformational descriptor 'r1'	FLOAT
THETA_II	Conformational descriptor 'theta2'	FLOAT
PHI_II	Conformational descriptor 'phi2'	FLOAT
R_II	Conformational descriptor 'r2'	FLOAT
R_III	Conformational descriptor 'r3'	FLOAT
X_MIN	Lowest x coordinate in standard orientation	FLOAT
X_MAX	Maximum x coordinate	FLOAT
Y_MIN	Lowest y coordinate	FLOAT
Y_MAX	Largest y coordinate	FLOAT
Z_MIN	Lowest z coordinate	FLOAT
Z_MAX	Highest z coordinate	FLOAT
FRAG_CHARGE	Total linker charge	FLOAT
FRAG_DIP_X	x component of total linker dipole moment	FLOAT
FRAG_DIP_Y	y component	FLOAT
FRAG_DIP_Z	z component	FLOAT
FRAG_DIPOLE	Magnitude of total linker dipole moment	FLOAT
FRAG_Q20	Q20 component of linker quadrupole moment	FLOAT
FRAG_Q21C	Q21c component	FLOAT
FRAG_Q21S	Q21s component	FLOAT
FRAG_Q22C	Q22c component	FLOAT
FRAG_Q22S	Q22s component	FLOAT
FRAG_QUAD	Magnitude of total linker quadrupole moment	FLOAT
POLAR_Q	Linker polarity, based on atomic charge	FLOAT
POLAR_V	Linker polarity, based on electrostatic potential	FLOAT
POLAR_WEIGHT	Weighted value of POLAR_V, based on size of atomic grids	FLOAT
SCF	SCF energy from <i>ab initio</i> calculation	DOUBLE
LINK_LAB1	Atom no. of 1st atom in capping group from wfn file	INT
LINK_LAB2	Atom no. of 1st atom in 2nd capping group from wfn file	INT
SUM_VMAX	Sum of atomic V <sub>max</sub> values	FLOAT
SUM_VMIN	Sum of atomic V <sub>min</sub> values	FLOAT
SUM_EPNUC	Sum of values of electrostatic potential at proton nucleus	FLOAT
SUM_PKHA **		
SUM_PKHB **		

\* Elements of Primary Key.

\*\*  $\sum \alpha_2^H$  and  $\sum \beta_2^H$ , values not currently stored.

<sup>a</sup> Parameters are color-coded by type according to the following: orange = shape, blue = charge distribution, red = H-bonding, green = miscellaneous. Values written in black are used to identify a given linker. Asterisks indicate that values form part of the primary key.

suggested replacement among the hits found by the QID is shown in the final column.

The prevalence of shape and conformation among the search criteria used clearly demonstrates the importance of

Step 1: Enter Fragment to be Replaced

Double click in the box below to invoke ISIS Draw. Enter a SINGLE structure only, e.g:

Alternatively, type in a SINGLE SMILES code,  
eg [1+]C(=O)N[2+]

**IMPORTANT:**

Fragments must be capped by '[1+]' and '[2+]' groups.  
[More Info](#)

Fragments should have skeletons no longer than 2 atoms long, should be joined to only carbon or nitrogen atoms in the parent molecule, and should not be connected to the parent molecule via a ring bond.  
[More Info](#)

**Figure 9.** First step: screenshot from the QID showing the first step in the process of performing a search: entering the query linker of interest.

Search for a Fragment with Overall:

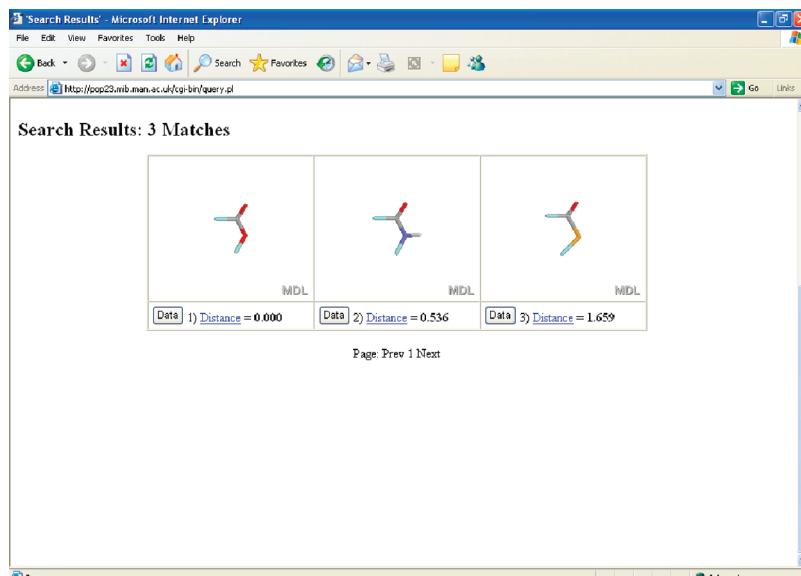
Property	Importance
<input checked="" type="checkbox"/> Conformation	low
<input type="checkbox"/> MEP Surface	low
<input checked="" type="checkbox"/> Charge Distribution	low
<input type="checkbox"/> Polarity	low
<input checked="" type="checkbox"/> Shape	low
<input type="checkbox"/> H-Bonding	low

Click atoms for Labels

Atom Label: O=1

Reset Form Submit Query Advanced Search

**Figure 10.** Step 2: After the desired linker conformer has been selected (not shown) characteristics are selected that must be present in a replacement linker. The search is then automatically converted to an SQL query, and linker groups matching the search specifications are returned. Linker connection points (asterisks in Figure 1) are marked in light blue here and in Figures 11 and 12.



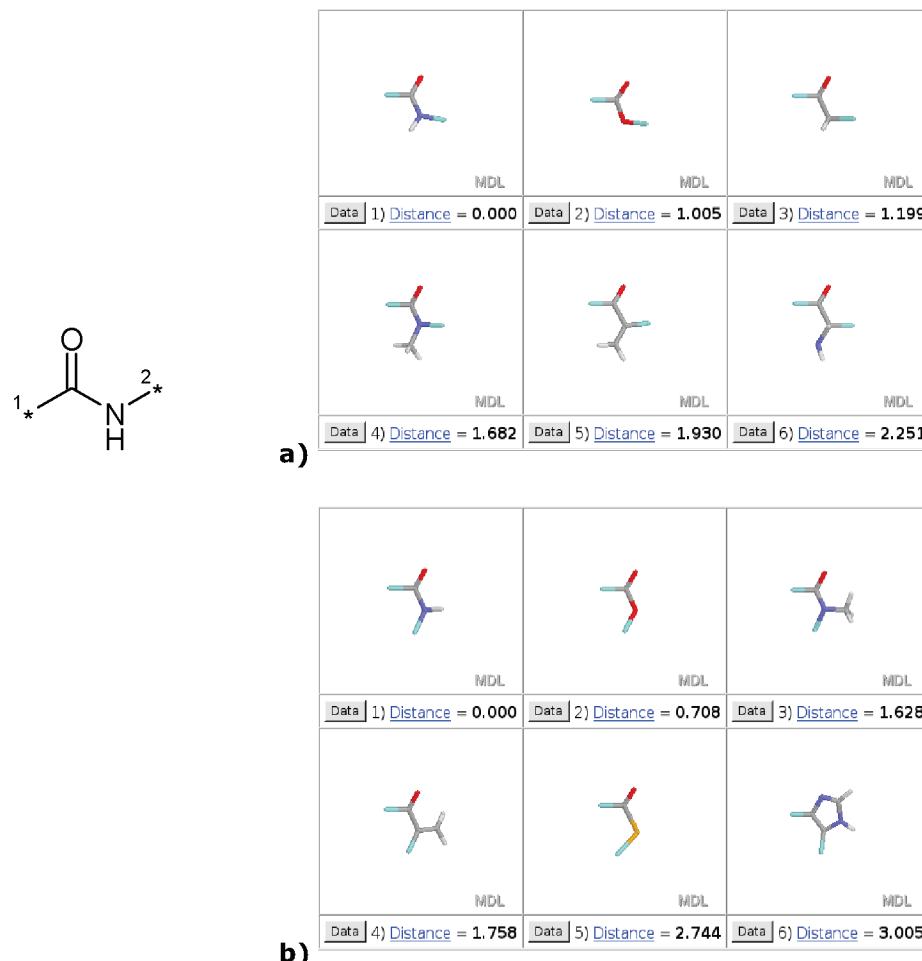
**Figure 11.** The final step is comparing similarity in the properties of interest and producing a similarity score. Results are then ranked by score and displayed for the user to inspect.

these factors in producing bioisosterism between Sheridan's fragments. Only the final pair of fragments (search 10) did not closely match in shape and conformation. This result suggests there is an advantage to storing detailed 3D structures over the simple connectivity data used in many tools, allowing more detailed comparison of linker size and shape. In addition to shape and conformation, fragment charge distribution was also important in identifying several of the bioisosteric pairs. The composite 'charge distribution' descriptor comprises overall linker charge, dipole, and quadrupole moments.

Searches 4 and 5 in Table 6 demonstrate how results can be refined by adding specific atom or bond information to complement general linker descriptors. Using shape and conformation alone, 40 hits were retrieved to replace the vinyl linker in search 4. However, by limiting results to those that share a similar bond order the number of hits is reduced to 5. Search 6 gives an example of the use of charge polarity

as a nondirectional descriptor of charge distribution. The reversed fragment maintains a similar shape, but directional charge distribution is clearly different so it cannot be the source of the observed bioisosterism. The mean absolute charge of the constituent atoms, however, remains identical. Search 10 provides the only example where the fragment and its replacement do not share particularly similar shape or conformation. They are similar, however, in several electronic and electrostatic characteristics, suggesting similarity in charge distribution may be the source of their similar biological effect.

**Example 3: Shape Searching Using Fingerprints.** The final example demonstrates the ability of 3D fingerprints to identify linkers with similar shape or charge distribution. Fingerprints offer an alternative to other shape descriptors stored in the QID, such as combinations of linker dimensions and QCT volume. The 3D fingerprint searches provide a fast, coarse means of performing full evaluation of volume



**Figure 12.** Results of a search to replace two conformers of the amide linker are displayed. Replacements were required to have similar shape and conformation to the amide and an H-bond acceptor at the carbonyl oxygen position. Results are ranked by similarity in these properties.

overlap, allowing quantitative comparison of shape similarity between very different structures using scoring functions such as the Tanimoto coefficient.

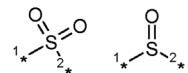
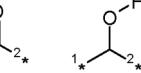
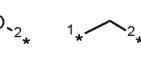
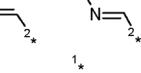
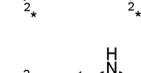
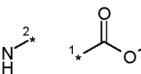
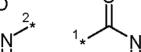
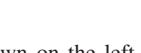
Three separate searches are presented (Figure 13a–c). The first identifies replacements with similar shape to the five-membered heterocycle shown in the top-left corner of Figure 13a. Shape fingerprints were used to evaluate overlap of the atomic volumes, defined using Bondi radii.<sup>42</sup> Results are shown in Figure 13a, sorted by the Tanimoto similarity score. It is encouraging to see that the fingerprint search returns five suggested replacements with very similar shapes. Ranking of the replacements, however, is difficult to rationalize, as the replacement with highest similarity score, while structurally similar to the query, appears more different than other high-ranking replacements in the list. It is likely that this ranking is a result of the coarse resolution of the grid elements ( $1 \text{ \AA}^3$ ) being insufficient to distinguish between very similar structures. Increasing resolution to  $0.5 \text{ \AA}^3$  grid elements may improve results but would increase the number of bits stored in each fingerprint from 9196 to 73568. As all of the top-ranking replacements are of very similar shape to the query linker, an increase in grid resolution seems unjustified.

The second search demonstrates the use of 3D fingerprints coupled with atomic charge information. The positions of all positively charged atoms ( $q(\Omega) > 0.6 \text{ au}$ ) or all negatively charged atoms ( $q(\Omega) < -0.6 \text{ au}$ ) are stored using their Bondi

radii in positive or negative charge 3D fingerprints. Use of fingerprints allows similarity scores that compare overlap of all polar atoms in both structures. Such fingerprint searches are particularly useful to perform flexible searches to compare global distribution of atoms with a given property between linkers that are not required to contain atoms with that property at all of the same positions as the original query. Alternative searches can be performed in the QID by selecting charged atoms in a query linker as individual search criteria. Users can then stipulate more rigidly that replacement linkers must possess atoms with quantitatively similar charges at the requested positions.

Figure 13b shows the results of a negative partial charge fingerprint search to find replacements for the heterocycle in the top-left corner. Search results are ranked by Tanimoto similarity score. All replacements suggested contain negatively charged atoms at the positions of the nitrogen atoms in the query linker. There is a significant drop in similarity score visible between the third and fourth replacements, falling from 0.946 to 0.722, as an extra nitrogen atom appears in the replacement. The final (5th) replacement has been rotated for clearer inspection. It is not a ring structure but contains two polar oxygen atoms at similar positions to the nitrogen atoms in the query. Charge fingerprints do, therefore, appear to provide a useful measure of overlap of polar atoms within linkers. Combination with shape or other descriptors improves search specificity.

**Table 6.** Results of the Study To Identify Common Fragment Replacements Found by Sheridan<sup>4</sup> from Analysis of MDDR<sup>a</sup>

Search	Replacements	Criteria Used	# Hits	Replacement Ranking
<b>1</b>	 	Shape, Conformation, Charge Distribution	3	1
<b>2</b>	 	Shape, Conformation, Charge Distribution	16	1
<b>3</b>	 	Shape, Conformation	7	1
<b>4</b>	 	Shape, Conformation	40	2
<b>5</b>	 	Shape, Conformation, Bond Order	5	1
<b>6</b>	 	Shape, Conformation, Charge Polarity	14	1
<b>7</b>	 	Shape, Conformation	7	6
<b>8</b>	 	Shape, Conformation	7	1
<b>9</b>	 	Shape, Conformation, Charge Distribution	23	1
<b>10</b>	 	Frag Dipole, Charge Polarity, Electrostatic Potential on Frag Surface	12	1

<sup>a</sup> Pairs of suitable replacements are shown on the left, criteria used to find the suggested replacement using QID are in the center, the total number of hits found in the QID search are on the right, and the ranking of the expected replacement in the list of search results is on the far right. For example, search 2 found 16 linkers matching the search criteria, and Sheridan's suggested replacement was the highest ranking hit.

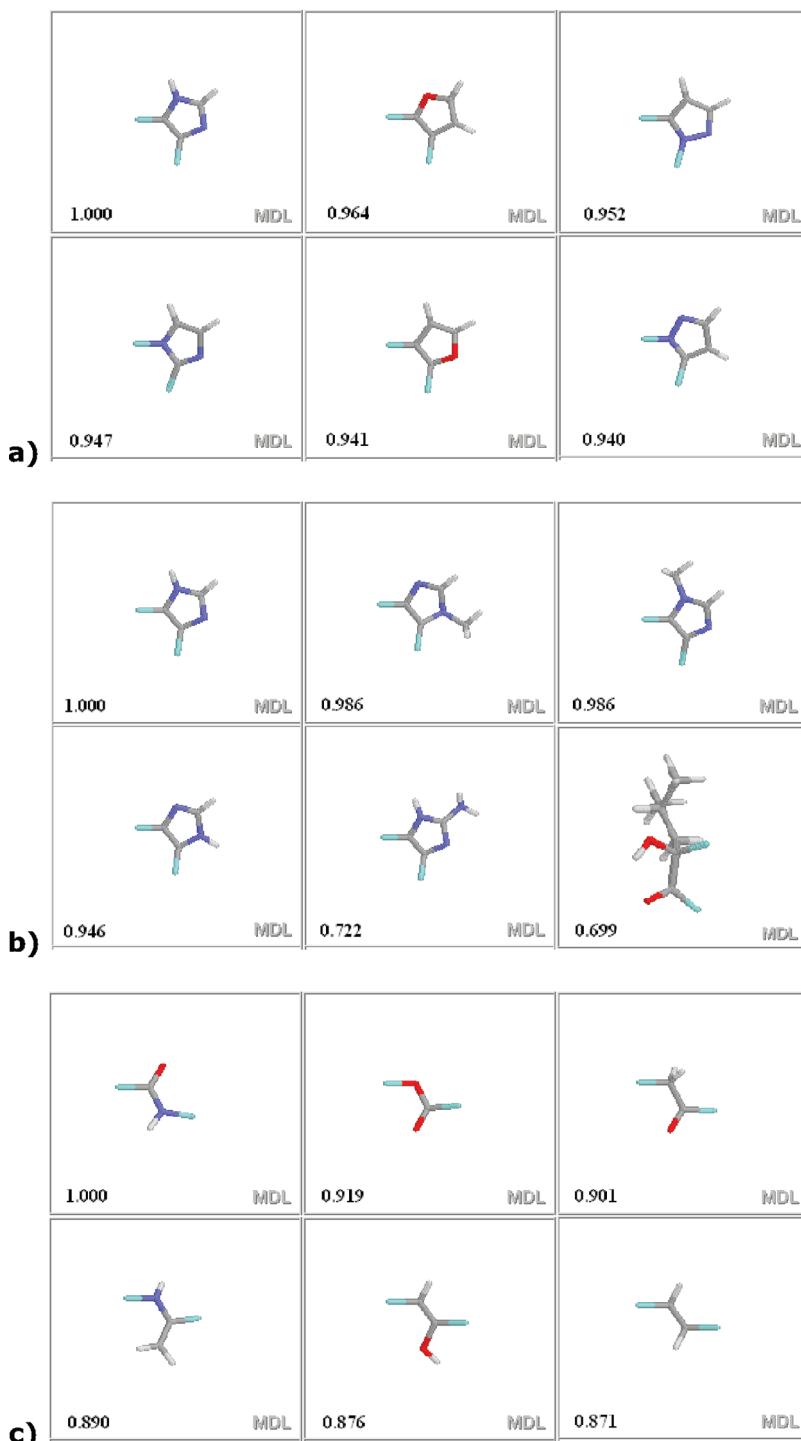
The final search again evaluates shape similarity, aiming to replace the amide linker shown in Figure 13c with a linker of similar shape. The amide linker was chosen to compare fingerprint searching results with those using linker volume and dimensions presented earlier. Results are again ranked by Tanimoto similarity score from the amide. While all search results are, indeed, of similar shape to the query, results are less precise than the quantitative results obtained in Figure 12a,b from linker dimensions, volume, and conformation. These 3D fingerprint searches can therefore be used as a general steric measure of shape similarity between structures, where exact geometry and small differences in atomic position are not important. Quantitative steric and geometric constraints, such as linker geometry and volume, can be used to constrain search results to produce lists of linkers with almost identical shape and geometric properties to the query.

The preceding examples provide an initial validation, demonstrating some of the basic functionality of the QID and providing comparison with other, simpler methodologies. Wider testing of the tool's predictive power has now begun, aimed at demonstrating the added value of the substantially

more detailed data stored in the QID than is commonly available using simpler methodologies.

## 6. CONCLUSIONS

A new tool, the Quantum Isostere Database (QID), has been introduced, using *ab initio* descriptors to identify bioisosteric replacements for 'linker group' fragments of active lead compounds. The theory of Quantum Chemical Topology (QCT) has been used extensively to obtain atom, bond and whole linker data from calculated charge densities. A diverse range of new descriptors has been developed, based on *ab initio* calculated properties, with intuitive links to biological activity. Additional, more theoretical quantities have been stored to allow future development of models to predict properties such as pK<sub>a</sub> and also to allow identification of new relationships between similarity in theoretical values and known bioisosterism between linkers. The data stored represent a significant investment of computational time, and contains a wealth of detailed chemical information. As well as allowing prediction of bioisosteric replacements from intuitive descriptors, there is much scope to use stored data



**Figure 13.** a) Results of a 3D fingerprint search to find linkers with similar shape to the query linker (top left). Pale blue bonds represent linker connection points. Tanimoto similarity scores are included beneath each structure. b) Results of 3D charge fingerprint search to find linkers with similar negatively charged atom distribution to the query linker (top-left). The bottom right structure has been rotated to allow closer inspection. c) Results of a 3D fingerprint search to find linkers with similar shape to the query linker (top left). The 5th linker (center-bottom) has a different conformation of the linker backbone.

with data analysis techniques to find subtle links between fragment physical properties and preserved biological activity in parent compounds.

Data have been stored in a relational database and can be accessed via a Web interface, allowing queries to be built interactively via a standard Web browser and avoiding the need for users to install specialist software. Scoring functions have been included to rank search results by similarity in specified characteristics.

Three short validation studies have been presented in the text, demonstrating the ability of the tool to suggest known isosteric replacements for common linker groups using a minimal set of intuitive descriptors. In addition, the ability of the tool to assist in rationalizing bioisosteric relationships between linkers using quantitative descriptions of their physical properties is clearly visible. It is anticipated that the full value of the QID will become apparent once additional data generation has been completed, and extensive

validation studies have been carried out. The added value of the *ab initio* descriptors used in the QID can then be fully assessed by direct comparison with other available methodologies.

#### ACKNOWLEDGMENT

We are grateful to the EPSRC for providing an EngD studentship and thank GlaxoSmithKline (GSK) for additional financial support.

#### APPENDIX

**Construction of Fragment Multipole Moments from Atomic Contributions.** Atomic multipole moments in the QID are stored using a standard axis system, described in the text, with the atomic nuclear coordinate  $\mathbf{R}$  as the local origin. Total fragment charge can be evaluated by simple summation of atomic charges of each atom  $n$

$$q_{tot} = \sum_n q(\Omega_n) \quad (\text{A1})$$

The total linker dipole moment, however, is typically origin dependent as fragments generally carry a net partial charge. The local origin used to describe linker multipole moments lies between the linker's connection points, again described in more detail in the text. The fragment dipole moment is constructed from atomic dipole moments  $\mu(\Omega)$  and atomic charges  $q(\Omega)$  as well as the nuclear coordinate  $\mathbf{R}$  (defined relative to the desired global origin and axis system) according to

$$\mu_{\alpha,tot} = \sum_n q(\Omega_n)R_{\alpha,n} + \mu_\alpha(\Omega_n) \quad (\text{A2})$$

where  $\alpha$  is the X, Y, or Z component.

The fragment quadrupole moment is also origin-dependent, and different definitions and formalisms exist. Buckingham defines the quadrupole moment operator as<sup>78</sup>

$$\hat{\theta}_{\alpha,\beta} = \hat{\theta}_{\beta,\alpha} = \frac{1}{2} \sum_i e_i (3r_{i\alpha}r_{i\beta} - r_i^2 \delta_{\alpha,\beta}) \quad (\text{A3})$$

Summation is over all  $i$  elements of charge, where  $\alpha$  and  $\beta$  are X, Y, or Z components,  $\delta_{\alpha,\beta}$  is Kronecker's delta, and  $r$  is the scalar distance of coordinate  $\mathbf{r}$  from the global origin. In ref 78, eq A3 is separated into a partial sum of  $I$  nuclear terms and a volume integral over electron density

$$\theta_{\alpha,\beta}^{(0)} = \frac{1}{2} \sum_I Z_I (3r_{I\alpha}r_{I\beta} - r^2 \delta_{\alpha,\beta}) - \frac{1}{2} \int d\tau \rho(r) (3r_\alpha r_\beta - r^2 \delta_{\alpha,\beta}) \quad (\text{A4})$$

Atomic quadrupole moments calculated in AIMPAC<sup>48,66</sup> are defined according to

$$Q_{\alpha,\beta}(\Omega) = - \int_\Omega d\tau \rho(r) (3r_{\alpha,\Omega}r_{\beta,\Omega} - r_\Omega^2 \delta_{\alpha,\beta}) \quad (\text{A5})$$

where  $r_\Omega$  is used to represent the quantity  $|\mathbf{r}|$ , and  $\mathbf{r}$  is a point defined relative to the atomic nucleus. Note that the volume integral is restricted to the atomic basin  $\Omega$ . The key difference between this and Buckingham's definition is the omission of the factor 1/2.

It should be noted that some *ab initio* programs<sup>33</sup> offer a different definition

$$\begin{aligned} Q_{20} &= \theta_{zz} & R_{20} &= \frac{1}{2}(3r_z^2 - r^2) \\ Q_{21c} &= \frac{2}{\sqrt{3}}\theta_{xz} & R_{21c} &= \sqrt{3}r_x r_z \\ Q_{21s} &= \frac{2}{\sqrt{3}}\theta_{yz} & R_{21s} &= \sqrt{3}r_y r_z \\ Q_{22c} &= \frac{1}{\sqrt{3}}(\theta_{xx} - \theta_{yy}) & R_{22c} &= \frac{1}{2}\sqrt{3}(r_x^2 - r_y^2) \\ Q_{22s} &= \frac{2}{\sqrt{3}}\theta_{xy} & R_{22s} &= \sqrt{3}r_x r_y \end{aligned}$$

**Figure 14.** (Left hand side) Conversion between Cartesian and spherical quadrupole moment components. (Right hand side) components of the second order spherical harmonics used to construct the quadrupole moment within the spherical formalism. Expressions taken from Appendix E of ref 47.

$$\hat{\theta}_{\alpha,\beta} = \sum_i e_i (r_{i\alpha}r_{i\beta}) \quad (\text{A6})$$

It is the original definition of Buckingham (eq A3) that has been used to define atomic and fragment quadrupole moments in the QID.

**The Quadrupole Moment Defined Using Spherical Harmonics.** A more compact alternative to Cartesian coordinates also exists, implemented in MORPHY98, that uses spherical harmonics to describe the components of the quadrupole moment. The 9 components of the Cartesian quadrupole moment (6 unique components, as  $\theta_{ij} = \theta_{ji}$ ) can then be reduced to 5 components, eliminating redundancy. The resulting quadrupole moments, still defined within Buckingham's definition but no longer using Cartesian coordinates, are shown in Figure 14.<sup>47</sup> The spherical quadrupole moment component  $Q_{20}$  is related to the Cartesian form,  $\theta_{zz}$ , with the second order spherical harmonics  $R_{20}$  used to construct the quadrupole moment component shown on the right.

To construct Buckingham's molecular quadrupole moment from atomic multipole moments defined within this spherical tensor formalism, the following series of equations were derived

$$\begin{aligned} Q_{20,tot} &= \sum_n Q_{20}(\Omega_n) + 2R_{z,n}\mu_{10}(\Omega_n) - R_{x,n}\mu_{11c}(\Omega_n) - \\ &\quad R_{y,n}\mu_{11s}(\Omega_n) + \frac{1}{2}(3R_{z,n}^2 - R_n^2)q(\Omega_n) \quad (\text{A7}) \end{aligned}$$

$$\begin{aligned} Q_{21c,tot} &= \sum_n Q_{21c}(\Omega_n) + \sqrt{3}R_{z,n}\mu_{11c}(\Omega_n) + \\ &\quad \sqrt{3}R_{x,n}\mu_{10}(\Omega_n) + \sqrt{3}R_{x,n}R_{z,n}q(\Omega_n) \quad (\text{A8}) \end{aligned}$$

$$\begin{aligned} Q_{21s,tot} &= \sum_n Q_{21s}(\Omega_n) + \sqrt{3}R_{z,n}\mu_{11s}(\Omega_n) + \\ &\quad \sqrt{3}R_{y,n}\mu_{10}(\Omega_n) + \sqrt{3}R_{y,n}R_{z,n}q(\Omega_n) \quad (\text{A9}) \end{aligned}$$

$$\begin{aligned} Q_{22c,tot} &= \sum_n Q_{22c}(\Omega_n) + \sqrt{3}R_{x,n}\mu_{11c}(\Omega_n) - \\ &\quad \sqrt{3}R_{y,n}\mu_{11s}(\Omega_n) + \frac{\sqrt{3}}{2}(R_{x,n}^2 - R_{y,n}^2)q(\Omega_n) \quad (\text{A10}) \end{aligned}$$

$$\begin{aligned} Q_{22s,tot} &= \sum_n Q_{22s}(\Omega_n) + \sqrt{3}R_{y,n}\mu_{11c}(\Omega_n) + \\ &\quad \sqrt{3}R_{x,n}\mu_{11s}(\Omega_n) + \sqrt{3}R_{x,n}R_{y,n}q(\Omega_n) \quad (\text{A11}) \end{aligned}$$

Total fragment quadrupole moment component  $Q_{2m,tot}$  is defined as a sum over all  $n$  linker atoms with atomic charge  $q(\Omega_n)$ , dipole  $\mu(\Omega_n)$ , and quadrupole moment  $Q(\Omega_n)$ .  $R_x$  is the

X component of the nuclear coordinate  $\mathbf{R}$  of atom  $n$ , defined relative to the desired global origin.  $R$  is again equal to  $|\mathbf{R}|$ .

## REFERENCES AND NOTES

- (1) Lima, L. M.; Barreiro, E. J. Bioisosterism: a useful strategy for molecular modification and drug design. *Curr. Med. Chem.* **2005**, *12*, 23–49.
- (2) Patani, G. A.; LaVoie, E. J. Bioisosterism: A rational approach in drug design. *Chem. Rev.* **1996**, *96* (8), 3147–3176.
- (3) Thornber, C. W. Isosterism and Molecular modification in Drug Design. *Chem. Soc. Rev.* **1979**, *8*, 563–580.
- (4) Sheridan, R. P. The most common chemical replacements in drug-like compounds. *J. Chem. Inf. Comput. Sci.* **2002**, *42*, 103–108.
- (5) Wagener, M.; Lommerse, J. P. M. The Quest for Bioisosteric Replacements. *J. Chem. Inf. Model.* **2006**, *46*, 677–685.
- (6) WDI, World Drug Index. *Derwent Information*; 1996.
- (7) Bader, R. F. W. *Atoms in Molecules. A Quantum Theory*; Oxford Univ. Press: Oxford, Great Britain, 1990.
- (8) Popelier, P. L. A. *Atoms in Molecules. An Introduction*; Pearson: London, Great Britain, 2000.
- (9) Popelier, P. L. A.; Aicken, F. M.; O'Brien, S. E. Atoms in Molecules. In *Chemical Modelling: Applications and Theory*; Royal Society of Chemistry Specialist Periodical Report; Hinchliffe, A., Ed.; 2000; Vol. 1, Chapter 3, pp 143–198.
- (10) Bader, R. F. W.; Popelier, P. L. A.; Keith, T. A. Theoretical Definition of a Functional-Group and the Molecular-Orbital Paradigm. *Angew. Chem., Int. Ed. Engl.* **1994**, *33* (6), 620–631.
- (11) Bader, R. F. W.; Bayles, D. Properties of atoms in molecules: Group additivity. *J. Phys. Chem. A* **2000**, *104* (23), 5579–5589.
- (12) Popelier, P. L. A. Integration of Atoms in Molecules - a Critical Examination. *Mol. Phys.* **1996**, *87* (5), 1169–1187.
- (13) Popelier, P. L. A. A method to integrate an atom in a molecule without explicit representation of the interatomic surface. *Comput. Phys. Commun.* **1998**, *108* (2–3), 180–190.
- (14) Holliday, J. D.; Jelfs, S. P.; Willett, P. Calculation of Intersubstituent Similarity Using R-Group Descriptors. *J. Chem. Inf. Comput. Sci.* **2003**, *43*, 406–411.
- (15) Ujvary, I. BIOSTER-A Database of Structurally Analogous Compounds. *Pestic. Sci.* **1997**, *51* (1), 92–95.
- (16) Talete-srl; Todeschini, R.; Consonni, V.; Pavan, V. *DRAGON, version 2.1*; Molecular Modeling Software: Milan, Italy, 2002.
- (17) Tripos-Associates. *SYBYL, Molecular Modeling Software, version 6.7*; St. Louis, MO, U.S.A.
- (18) Accelrys *TSAR, version 3.3*; San Diego, CA, U.S.A., 2005.
- (19) Accelrys *CERIUS2, version 4.7*; San Diego, CA U.S.A., 2005.
- (20) Ertl, P. World Wide Web-based system for the calculation of substituent parameters and substituent similarity searches. *J. Mol. Graphics Modell.* **1998**, *16* (1), 11–13.
- (21) Lewell, X. Q.; Jones, A. C.; Bruce, C. L.; Harper, G.; Jones, M. M.; McLay, I. M.; Bradshaw, J. Drug Rings Database with Web Interface. A tool for Identifying Alternative Chemical Rings in Lead Discovery Programs. *J. Med. Chem.* **2003**, *46*, 3257–3274.
- (22) Cramer, R. D., III; Patterson, D. E.; Bunce, J. D. Comparative molecular field analysis (CoMFA). 1. Effect of shape on binding of steroids to carrier proteins. *J. Am. Chem. Soc.* **1988**, *110*, 5959–5967.
- (23) Lemmen, C.; Lengauer, T.; Klebe, G. FLEXS: A method for fast flexible ligand superposition. *J. Med. Chem.* **1998**, *41* (23), 4502–4520.
- (24) Watson, P.; Willett, P.; Gillet, V. J.; Verdonk, M. L. Calculating the knowledge-based similarity of functional groups using crystallographic data. *J. Comput.-Aided Mol. Des.* **2001**, *15*, 835–857.
- (25) Weber, A.; Teckentrup, A.; Briem, H. Flexsim-R A virtual affinity fingerprint descriptor to calculate similarities of functional groups. *J. Comput.-Aided Mol. Des.* **2002**, *16* (12), 903–916.
- (26) Bruno, I. J.; Cole, J. C.; Lommerse, J. P. M.; Rowland, R. S.; Taylor, R.; Verdonk, M. L. IsoStar: A library of information about nonbonded interactions. *J. Comput.-Aided Mol. Des.* **1997**, *11* (6), 525–537.
- (27) Anzali, S.; Gerhard, B.; Krug, M.; Wagener, M.; Gasteiger, J. Kohonen Neural Network: A Novel Approach to Search for Bioisosteric Groups. In *Computer-Assisted Lead Finding and Optimization*; van de Waterbeemd, H., Testa, B., Folkers, G., Eds.; Helvetica Chimica Acta: Basel, 1997; pp 95–106.
- (28) BROOD; OpenEye Scientific Software Inc.: Santa Fe, NM, U.S.A., 2006.
- (29) Haigh, J. A.; Pickup, B. T.; Grant, A. J.; Nicholls, A. Small Molecule Shape-Fingerprints. *J. Chem. Inf. Model.* **2005**, *45*, 673–684.
- (30) Daylight; DaylightChemicalInformationSystemsInc.<http://www.daylight.com>.
- (31) Devereux, M.; Popelier, P. L. A.; McLay, I. M. Towards an ab initio Fragment Database for Bio-isosterism: Dependence of QCT Properties on Level of Theory, Conformation and Chemical Environment. *J. Comput. Chem.* **2009**, *30*, 1300–1319.
- (32) *Macromodel*; Schroedinger Inc.: New York, U.S.A., 2005.
- (33) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Zakrzewski, V. G.; Montgomery, J. A., Jr.; Stratmann, R. E.; Burant, J. C.; Dapprich, S.; Millam, J. M.; Daniels, A. D.; Kudin, K. N.; Strain, M. C.; Farkas, O.; Tomasi, J.; Barone, V.; Cossi, M.; Cammi, R.; Mennucci, B.; Pomelli, C.; Adamo, C.; Clifford, S.; Ochterski, J.; Petersson, G. A.; Ayala, P. Y.; Cui, Q.; Morokuma, K.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Cioslowski, J.; Ortiz, J. V.; Baboul, A. G.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Gomperts, R.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Gonzalez, C.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Andres, J. L.; Gonzalez, C.; Head-Gordon, M.; Replogle, E. S.; Pople, J. A. *GAUSSIAN98, Gaussian 98, Revision A.7*; Gaussian, Inc.: Pittsburgh, PA, U.S.A.; 1998.
- (34) Biegler-König, F. W.; Bader, R. F. W.; Tang, T. H. Calculation of the Average Properties of Atoms in Molecules. 2. *J. Comput. Chem.* **1982**, *3* (3), 317–328.
- (35) Popelier, P. L. A.; MORPHY. a Program For an Automated Atoms in Molecules Analysis. *Comput. Phys. Commun.* **1996**, *93* (2–3), 212–240.
- (36) Popelier, P. L. A. A Robust Algorithm to Locate Automatically All Types of Critical- Points in the Charge-Density and Its Laplacian. *Chem. Phys. Lett.* **1994**, *228* (1–3), 160–164.
- (37) Popelier, P. L. A.; Devereux, M.; Rafat, M. The quantum topological electrostatic potential as a probe for functional group transferability. *Acta Crystallogr., Sect. A: Found. Crystallogr.* **2004**, *A60*, 427–433.
- (38) Oracle; Oracle Corporation: Redwood Shores, CA, 2005.
- (39) MySQL; Sun Microsystems Inc.: U.S.A., 2008.
- (40) CHIME Pro; Symyx Technologies. <http://www.symyx.com>.
- (41) Leach, A. R.; Green, D. V. S.; Hann, M. M.; Judd, D. B.; Good, A. C. GaP method. *J. Chem. Inf. Comput. Sci.* **2000**, *40*, 1262–1269.
- (42) Bondi, A. Van der Waals volumes and radii. *J. Phys. Chem.* **1964**, *68*, 441–451.
- (43) Johnson, M. A. A review and examination of the mathematical spaces underlying molecular similarity analysis. *J. Math. Chem.* **1989**, *3*, 117–145.
- (44) Tanimoto, T. T. *IBM Internal Report 17th Nov.*; 1957.
- (45) Lamarche, O.; Platts, J. A.; Hersey, A. Theoretical prediction of the polarity/polarizability parameter  $\pi_2^H$ . *Phys. Chem. Chem. Phys.* **2001**, *3*, 2747–2753.
- (46) Devereux, M.; Popelier, P. L. A.; McLay, I. M. A refined model for prediction of hydrogen bond acidity and basicity parameters from quantum chemical molecular descriptors. *Phys. Chem. Chem. Phys.* **2009**, *11*, 1595–1603.
- (47) Stone, A. J. *The Theory of Intermolecular Forces*; Clarendon: Oxford, GB, 1996.
- (48) Bader, R. F. W. *Atom in Molecules. A Quantum Theory*; Oxford Univ. Press: GB, 1990.
- (49) Su, Z. W.; Coppens, P. Rotation of Real Spherical-Harmonics. *Acta Crystallogr., Sect. A: Found. Crystallogr.* **1994**, *A50*, 636–643.
- (50) Sjoberg, P.; Murray, J. S.; Brinck, T.; Politzer, P. Average Local Ionization Energies on the Molecular-Surfaces of Aromatic Systems as Guides to Chemical-Reactivity. *Can. J. Chem.* **1990**, *68* (8), 1440–1443.
- (51) Hagelin, H.; Murray, J. S.; Brinck, T.; Berthelot, M.; Politzer, P. Family-Independent Relationships between Computed Molecular-Surface Quantities and Solute Hydrogen-Bond Acidity Basicity and Solute-Induced Methanol O-H Infrared Frequency-Shifts. *Can. J. Chem.* **1995**, *73* (4), 483–488.
- (52) Ehresmann, B.; Martin, B.; Horn, A. H. C.; Clark, T. Local molecular properties and their use in predicting reactivity. *J. Mol. Model.* **2003**, *9* (5), 342–347.
- (53) Platts, J. A. Theoretical Prediction of Hydrogen Bond Donor Capacity. *Phys. Chem. Chem. Phys.* **2000**, *2*, 973–980.
- (54) Abraham, M. H.; Grellier, P. L.; Prior, D. V.; Duce, P. P.; Morris, J. J.; Taylor, P. J. Hydrogen-Bonding. 7. A Scale of Solute Hydrogen-Bond Acidity Based on Log K-Values for Complexation in Tetrachloromethane. *J. Chem. Soc., Perkin Trans. 2* **1989**, *(6)*, 699–711.
- (55) Abraham, M. H.; Grellier, P. L.; Prior, D. V.; Morris, J. J.; Taylor, P. J. Hydrogen-Bonding. 10. A Scale of Solute Hydrogen-Bond Basicity Using Log K Values for Complexation in Tetrachloromethane. *J. Chem. Soc., Perkin Trans. 2* **1990**, *(4)*, 521–529.
- (56) Rafat, M.; Devereux, M.; Popelier, P. L. A. Rendering of quantum topological atoms and bonds. *J. Mol. Graphics Modell.* **2005**, *24*, 111–120.
- (57) Chaudry, U. A.; Popelier, P. L. A. Estimation of pK<sub>a</sub> using Quantum Topological Molecular Similarity (QTMS) descriptors: Application

- to Carboxylic Acids, Anilines and Phenols. *J. Org. Chem.* **2004**, *69*, 233–241.
- (58) Popelier, P. L. A.; Chaudry, U. A.; Smith, P. J. Quantum Topological Molecular Similarity. Part 5. Further Development with an Application to the Toxicity of Polychlorinated Dibenz-p-dioxins. *J. Chem. Soc.* **2002**, *2*, 1231–1237.
- (59) Popelier, P. L. A.; Smith, P. J. QSAR models based on Quantum Topological Molecular Similarity. *Eur. J. Med. Chem.* **2006**, *41*, 862–873.
- (60) Platts, J. A. Theoretical Prediction of Hydrogen Bond Basicity. *Phys. Chem. Chem. Phys.* **2000**, *2*, 3115–3120.
- (61) Platts, J. A.; Butina, D.; Abraham, M. H.; Hersey, A. Estimation of Molecular Linear Free Energy Relation Descriptors Using a Group Contribution Approach. *J. Chem. Inf. Comput. Sci.* **2000**, *39*, 835–845.
- (62) Kamlet, M. J.; Doherty, R. M.; Abraham, M. H.; Taft, R. W. Solubility Properties in Polymers and Biological Media.4. Correlation of Octanol/Water Partition Coefficients with Solvatochromic Parameters. *J. Am. Chem. Soc.* **1984**, *106*, 464–466.
- (63) Abraham, M. H. Scales of Solute Hydrogen-Bonding- Their Construction and Application to Physicochemical and Biochemical Processes. *Chem. Soc. Rev.* **1993**, *22* (2), 73–83.
- (64) Kalgutkar, A. S.; Gardner, I.; Obach, R. S.; Shaffer, C. L.; Callegari, E.; Henne, K. R.; Mutlib, A. E.; Dalvie, D. K.; Lee, J. S.; Nakai, Y.; O'Donnell, J. P.; Boer, J.; Harriman, S. P. A comprehensive listing of bioactivation pathways of organic functional groups. *Curr. Drug Metab.* **2005**, *6* (3), 161–225.
- (65) Biegler-König, F. W.; Bader, R. F. W.; Tang, T. H. Calculation of the Average Properties of Atoms in Molecules. 2. *J. Comput. Chem.* **1982**, *3* (3), 317–328.
- (66) Biegler-König, F.; Schönbohm, J.; Bayles, D. Software news and updates - AIM2000 - A program to analyze and visualize atoms in molecules. *J. Comput. Chem.* **2001**, *22* (5), 545–559.
- (67) Fradera, X.; Austen, M. A.; Bader, R. F. W. The Lewis model and beyond. *J. Phys. Chem. A* **1999**, *103* (2), 304–314.
- (68) Poater, J.; Sola, M.; Duran, M.; Fradera, X. The calculation of electron localization and delocalization indices at the Hartree-Fock, density functional and post-Hartree-Fock levels of theory. *Theor. Chem. Acc.* **2002**, *107* (6), 362–371.
- (69) Wang, Y. G.; Werstiuk, N. H. A practical and efficient method to calculate AIM localization and delocalization indices at post-HF levels of theory. *J. Comput. Chem.* **2003**, *24* (3), 379–385.
- (70) Fradera, X.; Austen, M. A.; Bader, R. F. W. The Lewis Model and Beyond. *J. Phys. Chem. A* **1999**, *103*, 304–314.
- (71) Angyan, J. G.; Loos, M.; Mayer, I. Covalent Bond Orders and Atomic Valence Indexes in the Topological Theory of Atoms in Molecules. *J. Phys. Chem.* **1994**, *98* (20), 5244–5248.
- (72) Salzner, U.; Kiziltepe, T. Theoretical analysis of substituent effects on building blocks of conducting polymers: 3,4'-substituted bithiophenes. *J. Organomet. Chem.* **1999**, *64* (3), 764–769.
- (73) Krygowski, T. M.; Ejmont, K.; Stepien, B. T.; Cyranski, M. K.; Poater, J.; Sola, M. Relation between the substituent effect and aromaticity. *J. Organomet. Chem.* **2004**, *69* (20), 6634–6640.
- (74) Poater, J.; Sola, M.; Viglione, R. G.; Zanasi, R. Local aromaticity of the six-membered rings in pyracylene. A difficult case for the NICS indicator of aromaticity. *J. Organomet. Chem.* **2004**, *69* (22), 7537–7542.
- (75) Laidig, K. E.; Speers, P.; Streitwieser, A. Origin of depressed dipole moments in five-membered, unsaturated heterocycles. *Can. J. Chem.* **1996**, *74* (6), 1215–1220.
- (76) Grimm, H. G. Structure and Size of the Non-metallic Hydrides. *Z. Electrochem.* **1925**, *31*, 474–480.
- (77) Erlenmeyer, H.; Leo, M. On Pseudoatoms. *Helv. Chim. Acta* **1932**, *15*, 1171–1186.
- (78) Höfinger, S.; Wendland, M. Method/Basis Set Dependence of the Traceless Quadrupole Moment Calculation for N<sub>2</sub>, CO<sub>2</sub>, SO<sub>2</sub>, HCl, CO, NH<sub>3</sub>, PH<sub>3</sub>, HF and H<sub>2</sub>O. *Int. J. Quantum Chem.* **2002**, *86*, 199–217.

CI900085D