

RESEARCH ARTICLE OPEN ACCESS

Non-Covalent Molecular Interaction Rules to Define Internal Dimer Coordinates for Quantum Mechanical Potential Energy Scans

Suliman Sharif | Anmol Kumar | Alexander D. MacKerell Jr.

Department of Pharmaceutical Sciences, School of Pharmacy, University of Maryland, Baltimore, Maryland, USA

Correspondence: Alexander D. MacKerell Jr. (alex@outerbanks.umaryland.edu)**Received:** 15 January 2025 | **Revised:** 22 March 2025 | **Accepted:** 6 May 2025**Funding:** This work was supported by National Institute of General Medical Sciences (Grant R35 GM131710).**Keywords:** ab initio | database | interaction energy | Lennard-Jones | non-covalent interactions | van der Waals

ABSTRACT

Non-covalent interactions (NCI) dominate the properties of condensed phase systems. Towards a detailed understanding of NCI, quantum mechanical (QM) methods allow for accurate estimates of interaction energies and geometries, allowing for the contributions of different types of NCI to condensed phase properties to be understood. In addition, such information can be used for the optimization of empirical force fields, including the specific contribution of electrostatic versus van der Waals interactions. However, to date, the relative orientation of monomers defining molecular interactions of dimers is often based on full geometry optimizations of all degrees of freedom or extracted from known experimental structures of biological molecules. In such cases, the spatial relationship of the monomers often leads to multiple atoms in each monomer making significant contributions to the interactions occurring in the dimer, confounding understanding of the contributions of specific atoms or functional groups. To overcome this, a workflow is presented that allows for systematic control of the interaction orientation between monomers to be performed through the use of molecular interaction rules (MIR) in an extendable tool that can be applied to a broad range of chemical space. Using the “MIR workflow” allows a user to perform automation of the determination of well-defined monomer interaction orientations in dimers using Z-matrices, allowing for potential energy scans (PES) to be performed on combinatorial pairs of the monomers. In addition, compiled monomer and dimer geometries and PES data are stored in an extendable database. Illustration of the utility of the workflow is performed based on a collection of 89 monomers encompassing a variety of functional group classes from which 10,616 interaction dimers can be automatically generated. PES between all dimers were calculated at the QM HF/6-31G*, MP2/6-31G*, and ω b97x-d3/6-31G* model chemistries. In addition, analysis of the benzene dimer in three interaction orientations, a hydrogen bond interaction between azetidinone and *N*-methylacetamide, and the interaction of pyridine with acetone in the Burgi–Dunitz orientation are presented including results with the aug-cc-pVDZ basis set. Results show the impact of different QM model chemistries on minimum interaction energies and distances over a large ensemble of intermolecular interactions with emphasis on the contributions of dispersion.

1 | Introduction

Non-covalent interactions (NCI) are essential for the function of biological as well as all condensed phase systems [1]. For

example, the central dogma is the transfer of sequential information from nucleic acids through the process of transcription and translation whose fidelity involves hydrogen bond interactions to ultimately produce proteins for which the three-dimensional

This is an open access article under the terms of the [Creative Commons Attribution-NonCommercial-NoDerivs](https://creativecommons.org/licenses/by-nc-nd/4.0/) License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

© 2025 The Author(s). *Journal of Computational Chemistry* published by Wiley Periodicals LLC.

structure and stability are dependent on multiple types of NCI [2, 3]. While NCI take a wide variety of forms, from ionic to van der Waals (vdW) interactions, the energy may be broken down into five subcomponents: electrostatic, exchange-repulsion, polarization, charge transfer, and dispersion [4]. In commonly used additive empirical force fields (FF) NCI are treated using a combination of electrostatics, modeled using Coulomb's law, and vdW interactions, often treated using the 12-6 Lennard-Jones (LJ) potential [5]. In addition, explicit treatment of polarization has now been included in force fields such as AMOEBA [6, 7] and the classical Drude oscillator model [8–12]. When parametrizing a FF, the electrostatic and polarization terms can be directly extracted from quantum mechanical (QM) calculations using a variety of methods [13]. Treatment of the vdW interactions is more challenging as missing terms not explicitly treated in the FF as well as approximations in the treatment of the Coulombic and polarization terms that contribute to the NCI are often implicitly included in the LJ term. For example, it has long been known that appropriate optimization of the dispersion term based on the LJ formalism is central to accurately modeling the condensed phase properties of small polar molecules [14]. This makes optimization of the dispersion term particularly challenging. This has led to their optimization being performed targeting a combination of experimental condensed phase data and QM data [15–19].

Towards facilitating this task as well as more general optimization of FFs and for investigating NCI in general, a collection of databases based on a variety of QM methods has been developed over the years. Open-source molecular interaction databases include A24 [20], NENCI [21], D1200 [22], GMTKN24 [23], GMTKN55 [24], D442 [22], SH250 [25], R160 [26], R739 [26], HBC6 [27], HB300SPX [27], HB375 [28], HSG [27], IHB100 [28], S22 [29], S66 [30], DES15K [31], DES370K [31], X40 [32], ACCDB [33], PPS5/05 [34], RGC10 [35], L7 [36], S12L [37], XB18 [38], NBC10 [39], NCIR [40], and many others [41–43]. A modern example is the Biofragment Database (BFDB) used to explore NCI motifs based on model compounds representative of fragments from biomolecules [44]. All these databases include dimer interaction geometries that users can access and acquire interaction energies in conjunction with their choice of QM or molecular modeling (MM) model chemistry. However, the dimer interaction orientations are often poorly defined, being based on gas phase optimization of manually aligned monomers, from macro-molecular structures, or extracted from condensed phase simulations. These issues, as well as the structures often being stored as Cartesian coordinates, make it difficult to perform systematic interaction potential energy scans (PES) to capture specific, well-defined interactions between molecules. Toward this, FFParam is a parameter optimization tool developed in our lab that also allows automated generation of NCI of any molecule with water in well-defined orientations [45]. As our long-term goal is the optimization of LJ parameters associated with specific atom type pairs, it is necessary to automatically generate dimer structures that have well-defined interactions between any two molecules that may be investigated through PES of specific intermolecular degrees of freedom. To generalize dimer interaction definition and circumvent the conformation problem with published databases, we employ an internal coordinate system that is more intuitive and may be systematically manipulated to study specific monomer interaction orientations and generate dimer structures

required for PES [45–47]. The internal coordinates, commonly termed Z-matrices, can be converted to Cartesian coordinates to define the interaction orientation between two molecules, from which the interaction energy between those molecules may be calculated using either QM or MM model chemistries. Notably, such coordinates serve as base interoperability that is available to any computational chemist to use for their specific models. In this study, presented is a workflow that generates combinatorial pairs of small molecule dimers in well-defined interaction orientations. This involves a set of general interactions with a programmable decision tree that can be extended to provide avenues for additional model compounds and interactions of interest where the interaction energy for PES can be obtained for any level of QM or MM. Once the QM PES are acquired for a comprehensive collection of model compound dimers, that data is stored for subsequent use.

The workflow is based on molecular interaction rules (MIR) that are used to generate the dimer PES, which are subsequently stored. In addition, the monomer Z-matrices and the MIR are stored in the database and can be used to form combinations of monomers of domain-specific chemicals that will greatly help investigations of NCI. The “MIR workflow” is applied to 89 model compounds, from which 10,616 dimer interactions are generated and used in QM calculations with selected model chemistries. Results from the calculations offer a picture of how the MIR workflow can be used to systematically study the impact of model chemistry on a large number of NCI of various classes. Presently, the emphasis is on the impact of selected methods to more rigorously account for dispersion. Specific analyses of the benzene dimer in different interaction orientations and interactions of azetidinone with *N*-methylacetamide and pyridine with acetone are presented.

2 | Methodology

2.1 | Workflow Development

The MIR workflow was developed to automate the process of molecule selection through generation of the PES and database creation. The full workflow begins with molecule registration along with the associated SMILES [48, 49], transformation into the internal coordinate representation, formation of the dimers, and acquisition of interaction PES that are then stored. The overall workflow is shown in Figure 1. It consists of a number of automated steps in combination with manual user input. The initial step is registering the molecule in the database, with the class being the name of the molecule, “Benzene.” Its two meta data are its SMILES and identifiers of the interactions of interest involving that monomer.

In Step 2 of the MIR workflow, SMILES are passed into RDKit's [50] structure utility from which initial geometries in Cartesian coordinates are generated, with resulting structures then minimized with the Merck Molecular Force Field (MMFF) [51, 52]. At this stage, one or more Z-matrices for the molecule are created depending on the desired interaction orientations to be studied. A Z-matrix defines a molecular structure in terms of internal coordinates involving bond distances (BD), valence angles (VA), and dihedral angles (DA) defining the location of each

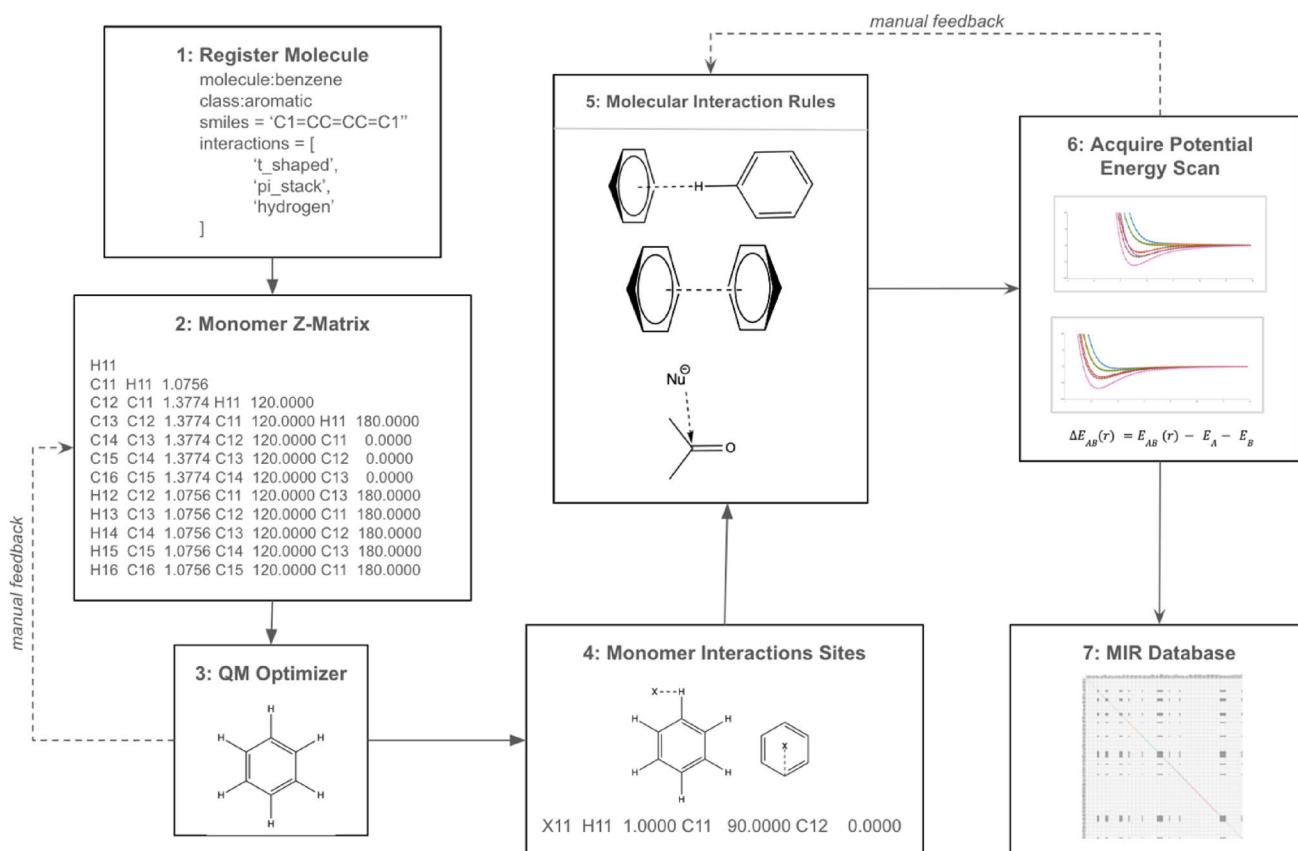


FIGURE 1 | Molecular interaction rule workflow.

atom relative to the remainder of the molecule. Importantly, Z-matrices are not unique, allowing for different atom ordering that can facilitate the design of different interaction orientations of each model compound through application of the proper MIR as described below. This represents the first significant manual input by the user in the workflow. For example, in the case of acetone, the SMILES can be written as CC(=O)C such that the rank order of the atoms in the Z-matrix is determined based on how the atoms are sequentially written in the SMILES reading from left to right. Typically, non-hydrogen atoms, and preferably heteroatoms, are selected as the first atom when creating the Z-matrix. This selection is based on the atom of the molecule that will define the interaction with the second monomer in a dimer, although this may be altered in step 4 below. The code then generates the subsequent non-hydrogen atoms based on the SMILES string, with the hydrogens subsequently being built. In certain cases, as in preparation for hydrogen-bond donor dimer interactions, such as those involving alcohols, hydrogens can be used as the initial atom, with the Z-matrix typically constructed manually. Once the Z-matrix is constructed, the internal geometry, including the BD and VA, is obtained from the MMFF optimized geometries. The DA required for the Z-matrix is calculated with an in-house implementation of the Rosetta NeRF algorithm [53].

Once an initial Z-matrix of a monomer is created, the compound is subjected to a QM optimization in step 3. If problems are encountered during the optimization, manual correction is undertaken as indicated by the feedback between steps 2 and 3 in the workflow. The resulting optimized geometric information is

then stored in the form of internal coordinates. The QM monomer geometries were used in all subsequent steps. This included the calculation of the PES in which the two monomer geometries are fixed at those optimized in the gas phase. If a user wants to register a new molecule in the system, the user must first identify its functional group classification, which determines how it will interact with other molecules during potential energy scans. After determining the appropriate classification, a new class object is created representing the molecule. This object requires several essential properties, including the SMILES notation that defines its chemical structure. The key aspect of molecule registration involves adding specific methods to the class object, with each method representing a distinct orientation of the molecule in space. These orientations correspond to keywords like “t_shaped,” “pi_stack,” and “hydrogen,” which define how the molecule can interact with others during computational analyses. Once defined, these orientations undergo a rigorous review process to verify their geometric accuracy and interaction potential. After validation, the orientations are sequentially added to the database. This systematic approach to molecule registration and orientation validation ensures the expanding molecular library maintains consistent quality and interaction parameters across the computational framework.

Step 4 involves verification and, if needed, modification of the initial Z-matrices of each molecule that define the monomer interaction sites, which act as the basis of the dimer interaction orientations based on the MIR described below. This is a manual intervention in which the Z-matrix is modified to define the site that will be directly involved in the dimer

calculations. These are performed in the context of a monomer acting as the “receptor,” which we refer to as monomer A, or as the “ligand,” which we refer to as monomer B. The sites can be individual atoms or virtual “dummy” atoms that, for example, define the centroid of a ring. Additional dummy atoms may be included to allow for the orientation of Monomer A to Monomer B to be defined and to facilitate the building of the rest of the molecule. In the case of monomer B, the Z-matrix is also designed with the initial site being involved in the interaction defining the PES with monomer A, with that site being either a real or dummy atom. In addition, the monomer B Z-matrix explicitly includes sites defined as variables that refer to specific sites in monomer A. Those variables are simply numbers, :1, :2, and so on that refer to the first, second, and so on sites in monomer A, with the colon signifying a variable defining the site number. Thus, the monomer B Z-matrix contains the information to build its structure in a specific orientation relative to monomer A. It is the combination of the monomer A and B Z-matrices built based on one of the MIR that define the PES. Therefore, for each model compound, one or more monomer A type and one or more monomer B type Z-matrices are created. This allows for a monomer A type Z-matrix to be combined with monomer B type Z-matrices of all the model compounds, including itself, from which the Z-matrices of specific, well-defined monomer interaction orientations in a dimer are created. This capability defines the PES, which is the core capability of allowing for all possible model compound dimer Z-matrices to be automatically generated, the corresponding Cartesian coordinates generated, and the dimer subjected to QM and MM PES calculations. In the present study, the monomer A and B Z-matrices for 86 compounds were initially created from the SMILES strings and manually corrected based on the MIR, a significant task. For users to add additional model compounds to the workflow, that process must be performed. However, given the wide range of monomer A and B type Z-matrices created in this study, the user will typically be able to find examples that include the type of monomer interaction sites they require.

The MIR encoded in the monomer A and monomer B Z-matrices are then combined to create the dimer Z-matrix, which is used for the calculation of the PES represented by step 5 in Figure 1. Included in Table 1 are the MIRs based on monomer A and B Z-matrices that are combined to create the dimer Z-matrix for selected interaction pairs. Seven general MIRs are defined that can be applied to create specific types of interaction orientations. In addition, a MIR for a specific interaction of interest, the Burgi–Dunitz angle [54], was created. Example Z-matrices of each molecule with their names, SMILES, atom name rank order, and skeletal diagram are presented in Table S1. Real and dummy atom names include numerical values of 1 or 2 indicating monomer A or B, respectively, followed by sequential numbers with individual numbering for non-hydrogen, hydrogen, and dummy atoms. Multiple Z-matrices for each molecule associated with the monomer A and B designations for one or more molecular interaction rules are accessible in the workflow through the Github repository listed below.

The MIR are presented in Table 1 with rule 1 described in detail. Rule 1 defines the interaction of the centroid of a ring in monomer A with an atom in monomer B that is bonded to only one

atom in the molecule. The latter will be referred to as “donor” atom. An example of such interaction is the T-shaped benzene dimer interaction shown as MIR 1 in Table 1. For this rule, the Z-matrix of monomer A is designed to define the centroid of the benzene ring as the interaction site using the dummy atom X11. X11 is defined with respect to three ring carbons with the BD, VA, and DA values obtained from the QM optimized monomer structure. Monomer B is defined with hydrogen as the donor interaction site. The monomer B Z-matrix is then constructed using the variables :1, :2, and :3 to define the orientation of B with respect to monomer A. This involves X11 (:1), defining the interaction of H21 of monomer B with the centroid of the monomer A ring, along with a VA to C11 (:2) of 90° and a DA to C12 (:3) of 90°, defining the perpendicular spatial relationship of the benzene rings. To allow for the remainder of monomer B to be built requires the use of a dummy atom, X21, with BD assigned a value of 1.0 Å and the remaining two sites defining the orientation relative to monomer A from the monomer A Z-matrix. The third site in monomer B is then defined in an analogous fashion, with the remainder of the monomer B Z-matrix then defined based on the first three sites. Note that the use of the variables DISTANCE, ANGLE, and DIHEDRAL in the monomer B Z-matrix is to allow for PES to be performed. The combination of the monomer A and B Z-matrices built with respect to a MIR yields the full Z-matrix for the dimer in the defined orientation. That Z-matrix can then be used to perform a linear PES between the interaction sites by simply varying the value of the variable DISTANCE built into monomer B. We note that the dimer Z-matrix also allows PES of the angle between the monomers through varying the ANGLE variable and the dihedral between the monomers through varying the DIHEDRAL variable. In the present study, only linear distance PES were performed.

While the above example involved a hydrogen atom of benzene interacting with the ring centroid, the same MIR may be used for a carbonyl oxygen interacting with the ring centroid. This could be the oxygen of acetone having a direct interaction with the benzene ring centroid. Thus, each rule can be applied to a range of diverse interacting pairs. In addition, the same monomer Z-matrix may be used with multiple MIR. For example, the monomer A Z-matrix is used in MIR 1, 2, 4, 5, and 6. Accordingly, the rules in Table 1 are named in general terms rather than specific interaction orientations, with the exception of the Burgi–Dunitz angle.

In step 6 of the MIR workflow, the PES are calculated using the dimer Z-matrix generated in step 5. This is performed by inputting the dimer Z-matrix into the QM program that converts the internal coordinates into Cartesian coordinates. Distance PES were performed using 72 single point energy scans from a distance of 8.0–1.0 Å in steps of −0.1 Å. Initiating the scans at longer distances allows for capturing and retaining PES data in cases when the monomers approach small distances. When a PES fails, for example, from a convergence error due to atomic overlap, the PES is captured up to the distance of the point of failure and moved into the database for subsequent use.

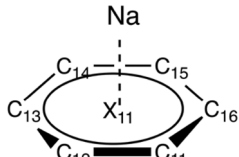
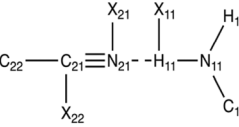
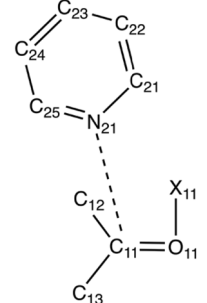
Error checking is an essential aspect of the workflow to ensure that the interaction orientations for the individual dimers are correct. This error check of the PES is performed through calculation of the first derivative of the potential energy with

TABLE 1 | Molecular interaction rules (MIR) defining dimer interaction orientations including the components of the Z-matrices used to define the MIR of the interaction type.

Monomer interaction rules	Type	Example	Z-matrix components of sites defining the intermolecular interactions
1. Ring Centroid to Donor atom	T-shaped (C-H- π)	<p>T-Shaped Benzene Dimer T-Shaped Benzene Dimer</p>	X11 C11 X11 1.3940 C12 C11 1.3774 X11 60.0000 C13 C12 1.3774 C11 120.0000 X11 0.0000 — H21 :1 DISTANCE :2 ANGLE :3 90.0000 X21 H21 1.0000 :1 90.0000 :2 180.0000 C21 H21 1.0756 X21 90.0000 :1 90.0000 C22 C21 1.3774 H21 120.0000 X21 DIHEDRAL
2. Ring Centroid to Ring Centroid	π -stacking	<p>Pi-Stack Benzene Dimer Pi-Stack Benzene Dimer</p>	X11 C11 X11 1.3940 C12 C11 1.3774 X11 60.0000 C13 C12 1.3774 C11 120.0000 X11 0.0000 — X21 :1 DISTANCE :2 ANGLE :3 90.00000 C21 X21 1.3940 :1 90.0000 :2 180.0000 C22 C21 1.3774 X21 60.0000 :1 90.0000 C23 C22 1.3774 C21 120.0000 X21 DIHEDRAL
3. Donor to Donor atom	H-H repulsion	<p>Cyclobutane Dimer hydrogen-hydrogen repulsion Cyclobutane dimer hydrogen-hydrogen repulsion</p>	H11 X11 H11 1.0000 C11 H11 1.1001 X11 90.0 C12 C11 1.5573 H11 118.5122 X11 180.0 C13 C12 1.5573 C11 87.8052 H11 -143.6662 — H21: 1 DISTANCE: 2 ANGLE: 3180.0 X21 H21 1.0000: 1 90.0000: 2 DIHEDRAL C21 H21 1.1013 X21 90.0000: 2180.0000 C22 C21 1.5345 H21 111.6112: 1180.0000
4. Ring Centroid to Acceptor	Lone Pair- π	<p>Benzene to Furan Benzene to Furan</p>	X11 C11 X11 1.3940 C12 C11 1.3774 X11 60.0000 C13 C12 1.3774 C11 120.0000 X11 0.0000 — O21 :1 DISTANCE :2 ANGLE :3 DIHEDRAL C21 O21 1.3723 :1 126.5480 :2 90.0000 C22 O21 1.3723 :1 126.5480 :2 -90.0000
5. Ring Centroid to sp3 Acceptor	Lone Pair- π	<p>Benzene to Methanol</p>	X11 C11 X11 1.3940 C12 C11 1.3774 X11 60.0000 C13 C12 1.3774 C11 120.0000 X11 0.0000 — O21 :1 DISTANCE :2 ANGLE :3 DIHEDRAL H21 O21 0.9657 :1 236.0000 :2 0.0000 X21 H21 1.0000 O21 90.0000 :1 0.0000 C21 O21 1.4349 H21 111.8699 X21 180.0000

(Continues)

TABLE 1 | (Continued)

Monomer interaction rules	Type	Example	Z-matrix components of sites defining the intermolecular interactions
6. Ring Centroid to monoatomic species	Cation- π	 <p>Benzene pi-cloud to Na⁺ Benzene pi-cloud to Na⁺</p>	X11 C11 X11 1.3940 C12 C11 1.3774 X11 60.0000 C13 C12 1.3774 C11 120.0000 X11 0.0000 — NA21 :1 DISTANCE :2 ANGLE :3 DIHEDRAL
7. Triple Bond to Donor atom	H-bond	 <p>Cyano N to Amine H Cyano N to Amine H</p>	H11 X11 H11 1.0000 N11 H11 1.0199 X11 90.0000 C11 N11 1.4728 H11 110.0093 X11 180.000 H12 N11 1.1057 H11 115.0533 X11 0.000 — N21 :1 DISTANCE :2 ANGLE :3 180.0000 X21 N21 1.0000 :1 90.0000 :2 DIHEDRAL C21 N21 1.1854 X21 90.0000 :1 180.0000 X22 C21 1.0000 N21 90.0000 X21 0.00000 C22 C21 1.4710 X22 90.0000 N21 180.0000
8. Nucleophile to Trigonal unsaturated center (Special Rule)	Burgi–Dunitz Angle	 <p>Pyridine to Acetone Pyridine to Acetone</p>	O11 C11 O11 1.2307 X11 O11 1.0000 C11 90.0000 C12 C11 1.5180 O11 121.7139 X11 90.0000 C13 C11 1.5180 O11 121.7139 X11 -90.0000 — N21 C11 DISTANCE O11 ANGLE X11 0.0000 C21 N21 1.3208 C11 111.1750 O11 DIHEDRAL C22 C21 1.3208 N21 116.0000 C11 180.0000

Note: The atom numbers (:1, :2, :3) in the second Z-matrix for each dimer correspond to the sites in the first Z-matrix. A ring centroid (RC) is the center of an aromatic ring, donors are functional groups that interact non-covalently with other molecules through an atom bonded to only one atom in the rest of the molecule, acceptors are the functional groups that generally form non-covalent interactions via an atom bonded to two or more atoms in rest of the molecule, triple-bond refers to functional group like cyano group (–CN) that exhibit linear bond angle within the molecule, and monoatomic species are chemical entities that do not have any bond description requirement. Note that the terms donor and acceptor do not necessarily correspond to the definitions commonly associated with hydrogen bonds. MIR 4 and 5 differ by the need to introduce an additional dummy atom, X21, in rule 5 to define the orientation of the second monomer (e.g., atoms C21, H21, H22, and H23) relative to the first monomer.

distance. The most common correct PES with a single favorable minimum, going from the shortest distance, will have a negative continuously decreasing derivative as it approaches the minimum, a zero derivative at the minimum, and then followed by a gradually decreasing positive derivative until the PES asymptotes. Correct repulsive surfaces will have continuously decreasing negative derivatives throughout the PES. To identify correct PES, a series of checks are made as defined in Equation S1. If alternative behavior is identified, then the PES is flagged for manual intervention. This is represented as the feedback loop between steps 5 and 6 in the workflow in Figure 1. Examples of “bad” PES are shown in Figure S1 and a repulsive interaction in Figure 5D (see below).

Following completion of the PES, the energetic data along with the Cartesian coordinates as a function of distance are stored as shown in step 7 of the workflow. This involves a knowledge graph associated with a functional group (FG) category with

the parents and molecules belonging to that group as child nodes. Each child node contains atom names, internal coordinates, and force field atom types. The individual FG nodes are conjoined and stored for analysis. The database has a respective application programming interface to browse through data, render status matrix plots, and build input files for QM and MM software.

To obtain PES target data on a mass scale, the workflow was parallelized across the UMB CADD Center cluster on a server with 64 cores per node, with one thread assigned to each core for a total of 24 nodes and 1536 CPUs. Each node is assigned one FG with either itself or another FG set, with the FG type for each molecule included in the database. This allows not only for selective capture of FGs as well as for organized data analysis. A key item to note is the flexibility of the QM or MM engine used in the workflow. This allows for tunable QM theory/basis set or MM force field based on user interest.

2.2 | Computational Methods

QM calculations were performed using the program Psi4 [55]. Monomer geometry optimizations were initiated from the MMFF [51, 52] optimized geometries calculated with RDKit. Monomer optimizations were performed at the MP2/aug-cc-pVDZ model chemistry [56–58] to the “Gau” defined convergence criteria. Monomers that failed initially were manually checked and optimized at the HF/6-31G* model chemistry [59, 60] and re-optimized at MP2/aug-cc-pVDZ. The MP2/aug-cc-pVDZ optimized geometries were then used to create the monomer Z-matrices. These were then stored in the database for subsequent PES calculations using dimer Z-matrices created using the MIR described in Table 1. Dimer PES was performed at the QM model chemistries reported below. Counterpoise correction for orbital overlap is applied as a default parameter to account for the basis set superposition error (BSSE) which is known to be pronounced in Gaussian double basis set expansion [61].

3 | Results

Presented in Figure 1 is the MIR workflow developed to perform PES between molecules in well-defined interaction orientations. It encompasses initial model compound and interaction orientation selection, creation of monomer geometries through a SMILES interpreter followed by QM geometry optimization. Z-matrices for the model compounds are then generated and populated based on optimized geometries. Manual manipulation of the Z-matrices is performed, implementing specific MIR that defines the interaction orientations in which the model compound will participate. This procedure required the creation of both monomer A (receptor) and monomer B (ligand) specific Z-matrices based on the desired MIR described in Table 1. While the MIR in Table 1 are general, they may be altered by altering the definition of the ANGLE or DIHEDRAL, based on user requirements. These Z-matrices may then be combined in a pairwise fashion to generate all possible dimer Z-matrices over the selected model compounds. The dimer Z-matrices may then be input into a number of programs allowing for creation of the Cartesian coordinates for QM or MM model chemistry calculations.

The MIR establishes a framework for calculating different types of molecular interactions in PES. Rule 1 handles interactions between a donor atom (e.g., hydrogen in benzene or oxygen in O=C) and a ring centroid. Rule 2 addresses stacking interactions with any two ring types based on user-defined centroids. The flexible Z-matrix approach enables the calculation of stacking interactions across various orientations, which has proven valuable for analyzing nucleic acid base stacking [62]. Rule 3 serves as a general approach for various interactions, particularly hydrogen bonds, such as between N–H donors and O=C acceptors in the *N*-methylacetamide dimer. Rule 4 specifically focuses on interactions between acceptors in heteroaromatic rings and ring centroids. Rule 5 deals with sp³ hybridized atoms, including hydroxyl oxygen or sulfhydryl sulfur atoms, acting as acceptors with ring centroids, using dummy atoms in the Z-matrix to define the remainder of the structure, which is also valuable for water as a hydrogen bond acceptor. The rule may also be used for interactions of the oxygen or sulfur atom with

a hydrogen bond donor, for example, an N–H. Rule 6 is a variation of Rule 1 for interactions involving monatomic species like ions or rare gases, particularly useful for optimizing Lennard–Jones parameters [63]. Rule 7 is useful for managing interactions between donor atoms and functional groups with linear bond angles, such as the cyano (–CN) group. Rule 8 is a special rule that allows the interaction of a nucleophile towards a trigonal unsaturated center such as C in the C=O group of acetone, also referred to as the Burgi–Dunitz angle [54].

Selected PES generated using the workflow are shown in Figure 2. Panels A, B, and C show the benzene dimer as calculated using Rules 1, 2, and 3, respectively. As is evident, the use of the Z-matrices and associated MIR allows for PES along 3 well-defined interaction orientations to be calculated. Figure 2D shows a hydrogen bond interaction between azetidinone and NMA also calculated using Rule 3, emphasizing the generality of the MIR for different types of NCI. The utility of the special rule 8 associated with the Burgi–Dunitz angle is shown in panel E of Figure 2. The ability to create this specific interaction orientation emphasizes the power of the use of Z-matrices to define intermolecular interactions. In this case, the Burgi–Dunitz [54] angle is associated with the approach of nucleophilic nitrogen of pyridine and the electrophilic carbon of acetone. Notably, the special MIR would allow the energetics of the approach as a function of intermolecular angle and dihedral to readily be studied. Clearly, the use of Z-matrices combined with the MIR allows for well-defined interactions between molecules to be defined and studied using a variety of model chemistries. The range of accessible interactions can be expanded through additional MIR, which will often be small variations of those already available from the workflow.

A particular area of interest to investigations of NCI is the treatment of dispersion interactions as they are important for the accurate modeling of condensed phase interactions using empirical force fields. The workflow in conjunction with the specific interactions in Figure 2 demonstrates the power of the MIR to facilitate the application of different basis sets and QM levels of theory. To better include dispersion effects, U_{disp} , in QM, theories have been built on the early perturbation methods described by Eischenschitz and London [64]. From these efforts it was shown that the dispersion energy, at a large distance R between atoms i and j , may be modeled as a pair-wise specific r^6 interaction term as described in Equation (1).

$$U_{\text{disp}}^{(i,j)} \cong -C_6^{(i,j)} R^{-6} \quad (1)$$

In Equation (1), C_6 defines the magnitude of the dispersion interaction between the two atoms. This model is the basis of improved treatment of dispersion in density functional theory (DFT) through the use of dispersion (DFT-D) corrections [65]. Grimme et al., as well as a number of other labs, have proposed a variety of dispersion corrections to DFT [66] leading to the widely used DFT-D method of Becke and Johnson with GGA functionals to achieve near CCSD(T) accuracy [67]. Following this approach, Gordon et al. demonstrated that the dispersion correction could be used with Hartree–Fock and “ ω B97x-D” long-range corrected hybrid density functional worked well for the treatment of NCI [68, 69]. A further

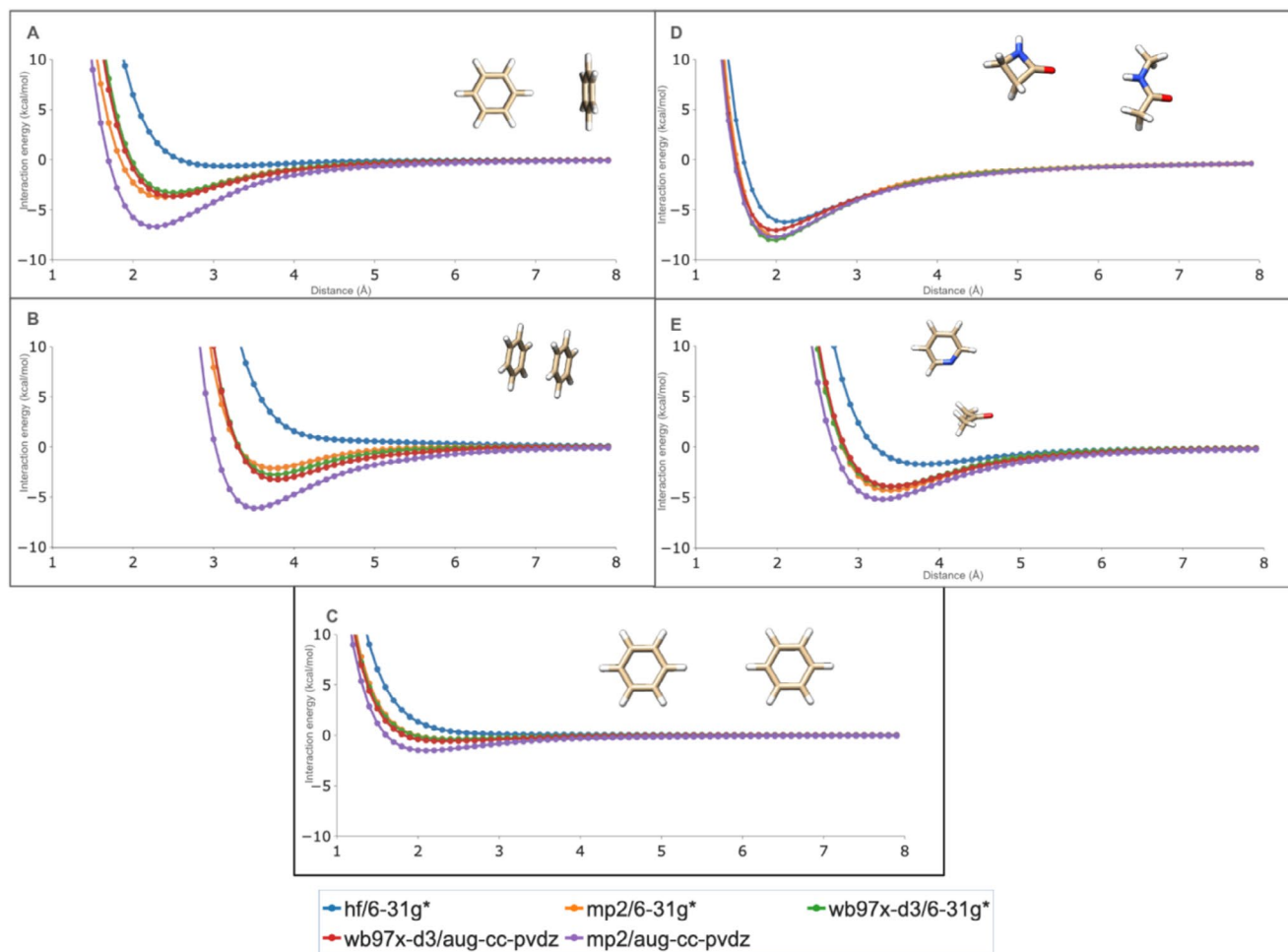


FIGURE 2 | PES of the benzene dimer in the (A) T-shaped (MIR 1), (B) π - π stack (MIR 2) and (C) H-H planar orientations (MIR 3), of the (D) azetidinone-methylacetamide dimer (MIR 3), and of the (E) Burgi-Dunitz angle of pyridine-acetone dimer (MIR 8). PES with QM model chemistries of HF/6-31G* (blue), MP2/6-31G* (orange), ω B97x-D3/6-31G* (green), ω B97x-D3/aug-cc-pVDZ (red), and MP2/aug-cc-pVDZ (purple).

development, ω B97x-D3, was able to demonstrate accuracy for non-bonded interactions over other methods for a wide range of interactions [69].

Building on the PES performed for the dimers in Figure 2 QM PES calculations using different model chemistries were undertaken. This included ab initio and dispersion-corrected DFT model chemistries including HF/6-31G*, MP2/6-31G*, ω B97x-D3/6-31G*, MP2/aug-cc-pVDZ, and ω B97x-D3/aug-cc-pVDZ. MP2 calculations offer a limited treatment electron correlation; couple-clustered methods such as CCSD(T), are typically used to “fully” account for dispersion interactions. However, CCSD(T) calculations with larger basis sets, such as aug-cc-pVDZ, are computationally challenging, especially with the size of many of the dimers being studied presently as well as the number of dimers accessible to the workflow [70]. Accordingly, we considered the use of the ω B97x functional with the D3 dispersion correction as an alternative to model electron correlation effects. The impact of the use of methods to better include dispersion is evident in Figure 2. In all cases, the least favorable interaction energies are obtained with HF/6-31G*. Use of MP2 with the 6-31G* basis sets accounts for a significant amount of dispersion contributions. A similar amount of dispersion contributions is obtained with DFT with the D3 dispersion correction, ω B97x-D3, with the 6-31G*

basis set. Moving to the larger basis set with MP2, aug-cc-pVDZ, yields the most favorable interaction energies due to the more flexible basis set. However, the ω B97x functional with D3 correction with the aug-cc-pVDZ basis set does not offer additional dispersion correction over 6-31G*, as expected given the “fixed” atom-based dispersion correction associated with that method. Notable is the variation in the differences in the impact of the model chemistry for the different dimers in Figure 2. The largest differences occur with the T-shaped (Figure 2A) and stacked (Figure 2B) benzene dimers followed by the Burgi-Dunitz pyridine-acetone interaction (Figure 2E). This is due to the larger orbital overlap in those dimers as compared with the hydrogen bond (Figure 2D) and benzene dimer H-H planar (Figure 2C) interactions. These differences indicate the utility of using well defined monomer interaction orientations when performing detailed analysis of dimer interactions and the impact on different model chemistries on those interactions.

To show the full power of the workflow, a larger number of model compounds for dimer calculations were selected. The selection criteria were such that the model compounds contained functional groups associated with the NCI of interest and had minimal complexity. For example, to study hydrogen bonding associated with the oxygen of an ether functional group, dimethyl

ether is preferred over diethyl ether or methyl acetate due to its greater simplicity and presence of a single type of heteroatom in the former. This reduces confounding contributions from other atoms when calculating the interaction energies. Such considerations will facilitate the planned optimization of the off-diagonal atom-pair specific LJ parameters in the CHARMM additive and Drude FFs [9, 71, 72]. The chemical names of these compounds were then used to generate the associated SMILES. The molecules were then organized in an object-oriented architecture based on their FG class. Each object is a molecule (e.g., Benzene) associated with an object defining the different interaction orientations for that molecule. The full list of FG classes included in the present study and their associated molecules are presented in Table 2.

Table 2 includes two columns: Functional Group Class and Molecules. When a molecule is added to the workflow, it is placed into an appropriate generalized class determined by its dominant FG. This allows for a systematic way, specific to the user's domain chemical space, to create the desired types of interaction dimers when investigating NCI. The collection of molecules used in the present study includes a variety of aromatic, aliphatic, and polar species, allowing for investigation of a broad range of NCI. There are a large number of molecules in the Aromatic class because of general interest in investigations of π - π stacking, cation- π , anion- π , and other related interactions.

Beyond the interactions involving aromatic groups, the interactions also include direct hydrogen to hydrogen interactions, which are typically repulsive in nature, emphasizing the ability of the workflow to investigate the full range of molecular interactions. The specific interactions studied in this work were based on MIR 1 to 5 (Table 1). A total of 89 monomer pairs were used to generate 10,616 dimer Z-matrices and were subject to QM calculations. The PES were successfully generated per level of theory/basis set: 6355 (HF/6-31G*), 6123 (MP2/6-31G*), and 4735 (ω B97x-D3/6-31G*). The number of successful PES was based on several factors. The majority of failures were due to geometric errors in the Z-matrices. Efforts to correct these are ongoing.

The number of successful PES being lower than the total possible PES emphasizes the importance of error checking. A valid PES must follow four essential rules. Starting from the smallest distance between the monomers, all first derivatives of the energy must be negative and continuously decrease until reaching a minimum where the derivative equals zero. The minimum cannot occur at the start or end of the sequence. After reaching the minimum, all subsequent values must be positive, continuously increase, and then gradually decrease without encountering a maximum upon going to distances in the PES longer than the minimum. In addition, throughout the entire PES, no two consecutive points can have the same value, ensuring there are no plateaus. This assures that a smooth transition from a negative decreasing slope to a positive increasing slope through a clear minimum point is present in the PES. The equations defining the correct PES are presented in Equation S1.

Initial tests were run with the HF/6-31G* model chemistry. Examples of incorrect PES that did not pass error checking are shown in Figure S1. The discontinuous PES is due to the incorrect interaction orientations of the monomers. When

such errors are encountered, manual intervention is required to correct the problem. Such problems are typically associated with an incorrect Z-matrix of a specific monomer, such that correction of the problematic Z-matrix leads to the PES containing that model compound being corrected. Correction of such errors typically leads to the correction of a number of PES, given that each monomer is involved in a minimum of at least 89 PES. Additional failures with MP2 and ω B97x-D3 are primarily associated with memory errors and lack of SCF convergence.

In Figure 3, the full range of successfully calculated PES is displayed at different levels of theory (HF, ω B97x-D3, and MP2) with the 6-31G* basis set and interaction charge type (neutral/neutral, neutral/ion, ion/ion). In Row A, as higher levels of theory are applied to the neutral to neutral interactions the presence of the correlation energy is evident with more minimum energies extending from -10 kcal/mol in panel A.1 up to -15 kcal/mol in A.3. The dispersion correction from ω B97x-D3 A.2 leads to more well-defined minima for selected interactions and the magnitude of the effect similar with the MP2 data in A.3, though this will be shown in more detail below. The interactions impacted most by the improved treatment of dispersion include—stacking of ring centroids of azulene to phenoxazine and methylene oxindole to phenoxazine.

In Row B in Figure 3, the correlation energy for the neutral to ion PES does not have as holistic an impact as for the neutral to neutral interactions, as expected given the greater role of electrostatic versus dispersion with ionic species. In B.2 and B.3 there is a greater density of minimum energies of less than -30 kcal/mol for certain interactions that contain 4-pyridinone, pyridinone, pyridinium, imidazole, and imidazolium following Rules 1, 2, and 5. Thus, dispersion contributions associated with either ω B97x-D3 or MP2 do contribute to this class of interactions though to a far lesser extent than with the neutral to neutral interactions.

For the ion-to-ion interactions shown in Row C in Figure 3 only PES for oppositely charged monomers are shown. The contribution of electron correlation via MP2 or through the D3 dispersion correction is the smallest versus the interactions involving neutral species, as expected given the dominance of pure electrostatics that are reasonably treated by HF. Also, as expected, the interactions are still quite favorable at the end of the PES, with interaction energies of approximately -40 kcal/mol at 8 Å associated with the interaction of oppositely charged species, as defined by Coulomb's law. These interactions include amidinium, pyridinium, imidazolium, pyridinium, charged prolineamide, ethylammonium, trimethylammonium, and tetramethylammonium with acetate and methylthiolate, following Rules 1, 3, and 5. PES with minima that are approximately -80 to -60 kcal/mol include methylthiolate sulfur atom interactions with ethyl/tri/tetra-methylammonium and the pyridinium ring centroid, following Rules 1 and 2. Interactions that exceed -100 kcal/mol consist of the same aforementioned positively charged compounds with acetate. PES with longer minimum interaction energy distances are associated with, for example, interactions involving the methyl group of a charged species. Again, this emphasizes the range of NCI that may be explored with the MIR workflow.

TABLE 2 | Functional group classes and specific molecules included in this study.

Functional group class	Molecules
Aromatic	Azulene, Benzene, Bipyrrole, Bromobenzene, Chlorobenzene, Fluorobenzene, Four Pyridinone, Furan, Imidazole, Imidazolium, Indole, Indolizine, Isoxazole, Methylene Oxindole, Nitrobenzene, 1 Phenyl-4-Pyridinone, Phenol, Phenoxazine, Pyridine, Pyridinium, Pyrroline, Pyrrolidine, Thiophene, 3-Aminopyridine, 2-H-Pyran, Uracil
Alcohols	Methanol
Alkanes	Cyclobutane, Cyclohexane, Cyclopropane, Neopentane, Propane
Alkenes	Cyclohexene, Cyclopentene, Methoxyethene, 1,3-dibutene, Propene, 2-Pyrroline
Alkynes	Propyne
Amides	Acetamide, Amidinium, Azetidinone, DimethylFormamide, Methylacetamide, 2-pyrrolidinone, Prolinamide, Prolinamide Charged
Amines	Ammonia, Dimethylamine, Ethyl Ammonium, Hydrazine, Methylamine, Piperidine, (Z)-N-methylethanimine, Tetramethylammonium, Trimethylamine, Triethylammonium
Carbonyls	Acetaldehyde, Acetate, Acetic Acid, Acetone, Carbon Dioxide, Formaldehyde, Methylacetate, Urea
Ethers	Dimethyl ether, Epoxide, Oxetane, Tetrahydrofuran, Tetrahydropyran
Imines	Ethenamine
Halogens	Bromoethane, Chloroethane, Dibromoethane, Dichloroethane, Fluoroethane, Difluoroethane, Tribromoethane, Trichloroethane, Trifluoroethane
Nitriles	Acetonitrile
Organophosphorus	Methyl Phosphate, Dimethyl Phosphate
Organosulfur	Dimethyl sulfone, Dimethyl Sulfoxide, Dimethyl trithiocarbonate, Dimethyl Disulfide, Methanethiol, Methylthiolate

To more closely analyze the impact of investigating large numbers of interaction orientations on the minimum interaction energies and distances, their probability distributions are presented in Figure 4. Results are again shown for neutral to neutral, neutral to ion, and ion to ion interactions. Results from only those PES with a local minimum are included in the distance analysis, with interaction energies at 4–5 Å included for the repulsive interactions in the energy analysis. Concerning the minimum interaction distances, for all three combinations of compounds, there are a range of minimum distances. This is associated with different types of interactions, from idealized hydrogen bond or ionic interactions to interactions involving ring centroids, as well as those involving interactions between nonpolar groups. Also, the interaction distances are defined based on the rules used to define the dimers, such that atoms defining the distances could be two hydrogens, hydrogens and heteroatoms, two heteroatoms, two carbons, and so on, as well as real atoms with dummy atoms as used for interactions with ring centroids. Example interactions are shown as inserts in the three upper panels in Figure 4.

The large variety of interaction orientations leads to the wide distribution of minimum interaction energies seen in Figure 4D–F. For all three classes of interactions, the majority of energies are favorable. However, unfavorable energies are obtained with all three classes associated with “repulsive” interactions. Examples of dimer PES for monomer pairs that involve direct interactions between atoms that would be expected to yield repulsive interactions are shown in Figure 5. For example, direct interactions are present between hydrogen bond acceptors (Figure 5A), two hydrogens (Figure 5B,D), and two halogens (Figure 5C). However, the interactions occurring in Figure 5A–C have local minima, showing that in certain regions of the PES, overall favorable interactions between the two monomers overcome the repulsion between the directly interacting atoms. Only in the case of panel D is the interaction repulsive over the entire PES. These results further emphasize the utility of using well-defined interaction orientations on PES to systematically investigate the full range of NCI, especially in the context of force field development.

Results in Figure 4 also show the impact of the QM level of theory on the populations of the minimum interaction distances and energies of the dimers. For the neutral to neutral interactions shown in Figure 4A the peak of the probability density of the minimum interaction distances between dimers is reduced from ~3 to 2 Å upon going from the HF to the MP2 and D3 corrected levels of theory. This shows that the explicit treatment of electron correlation or the use of a dispersion correction has a significant impact on a large number of intermolecular interactions for the neutral species. This is also evident in the energies in panel D. Similar differences are present with the neutral to ion interactions in panels B and E, though to a smaller extent. Interestingly, the overlap between the MP2 and ω B97x-D3 distributions in panels A, B, D, and E shows the energetic impact of the explicit treatment of electron correlation or the use of the D3 dispersion correction with the 6-31G* basis set yields very similar dispersion contributions to the neutral–neutral and neutral–ion interactions for the large majority of dimers. However, as noted above for the benzene dimer, the use of a larger, more flexible basis set allows for

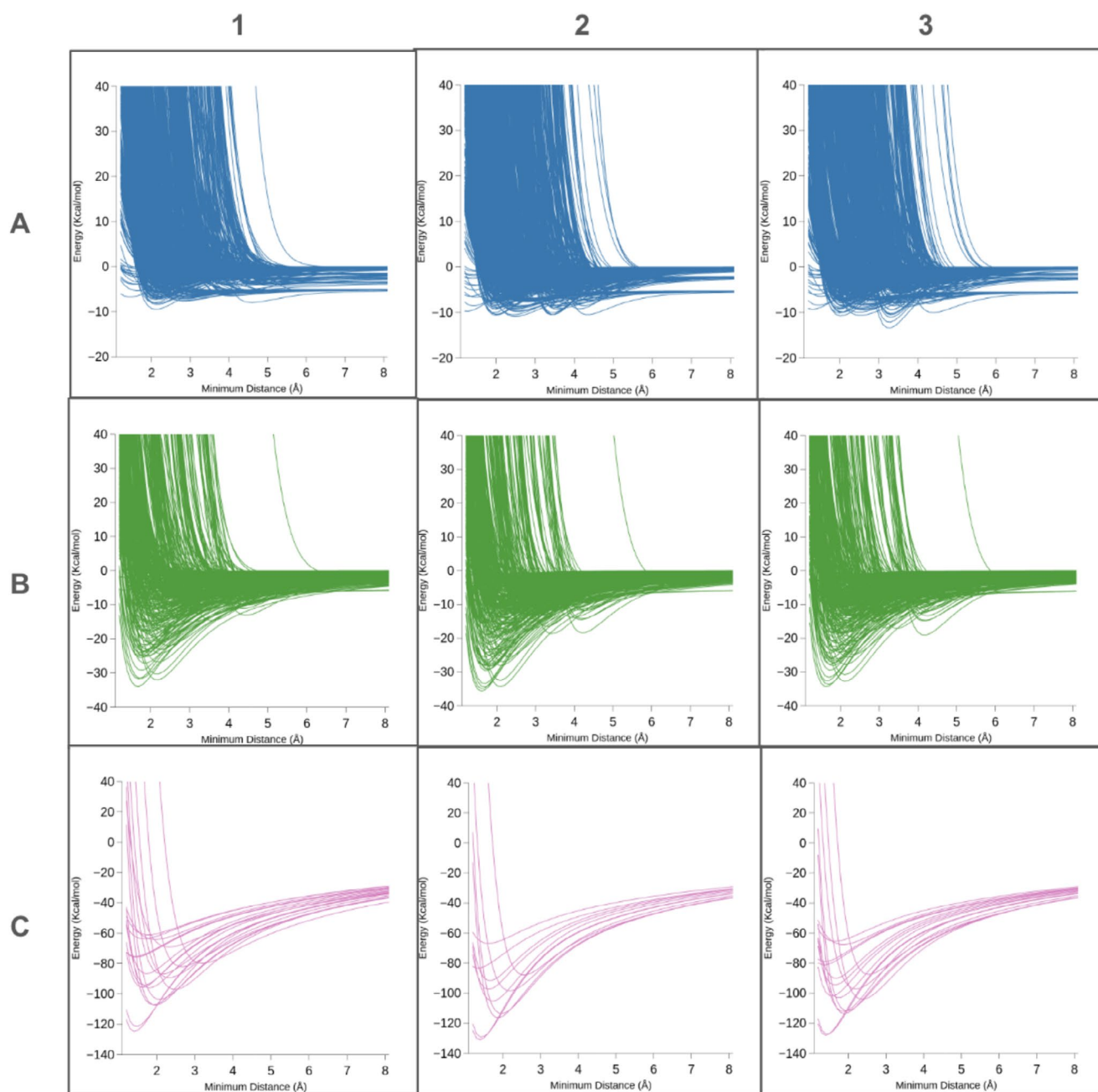


FIGURE 3 | Potential energy scans of analyzed non-covalent interactions (NCI). Columns by model chemistry: (1) HF/6-31G*, (2) ω b97x-d3/6-31G*, and (3) MP2/6-31G* and rows by NCI type: (A) neutral to neutral (blue), (B) neutral to ion (green), and (C) ion to ion (pink). For the ion to ion interactions only attractive dimers were studied.

more dispersion to be accounted for with the explicit treatment of electron correlation via MP2 (Figure 2). This difference is indicative of the known overestimation of electron correlation by MP2 for stacking interactions [73, 74], though such overestimation may potentially be desirable in the context of force field development for condensed phase studies where many-body contributions to dispersion may need to be implicitly included in the vdW parameters. Future studies are needed to address this issue.

In Figure 4C,F, the impact of improved treatment of dispersion versus HF is shown to be less with the charged interactions. The peak of the probability density of the minimum interaction distances is similar (Figure 4C). However, there is a shift in the

distributions upon going from HF to the improved treatment of dispersion to more favorable energies indicating that the use of the higher levels of theory still plays a role. The ω B97x-D3 probability density is downshifted more than MP2 indicating that the D3 dispersion corrections may be somewhat overestimated for the ion-to-ion interactions.

4 | Summary

Presented is a MIR workflow for the generation of PES between arbitrary monomers in well-defined interaction orientations. Input requires identification of the model compound of interest, the associated SMILES strings, and developing a Z-matrix of the

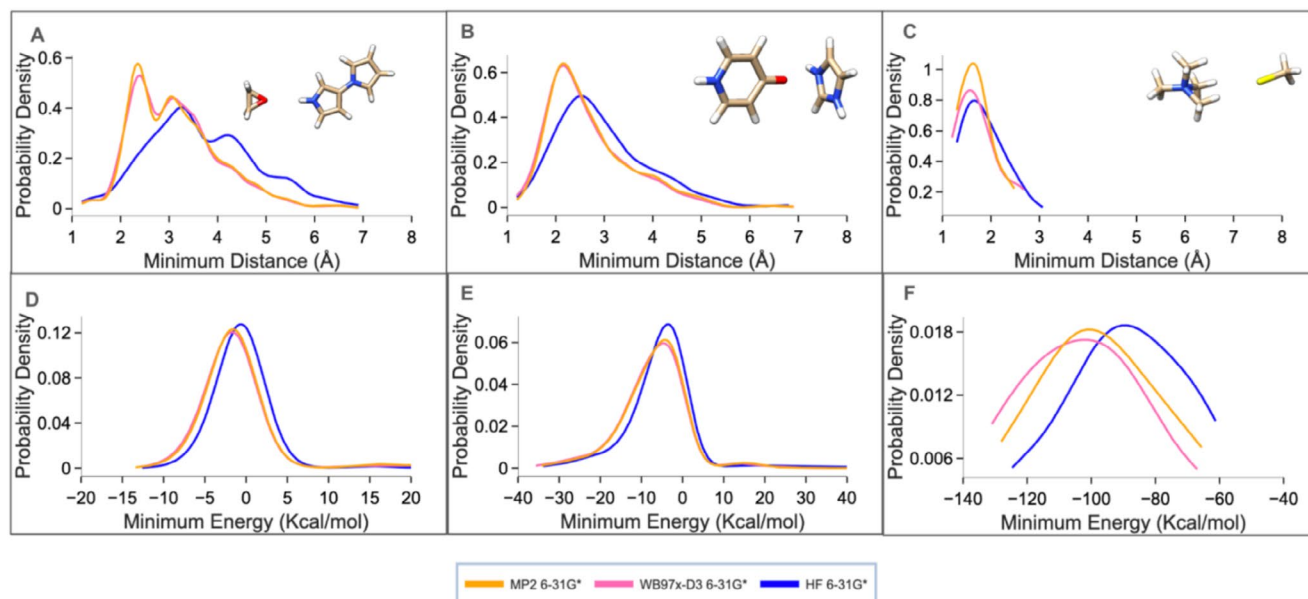


FIGURE 4 | Probability density plots of the minimum interaction (A–C) distances and (D–F) energies normalized with a bin size of 0.5 Å or 0.5 kcal/mol, respectively, for the HF/6-31G* (Blue), ωB97x-D3/6-31G* (Pink), and MP2/6-31G* (orange) model chemistries with a representative dimer model in each upper panel. Panels A/D correspond to neutral-to-neutral interactions with epoxide to 1,3-bipyrrole shown, Panels B/E correspond to neutral to ion Interactions with four-pyridinone to imidazolium shown, and Panel C/F correspond to ion-to-ion interactions with tetramethylammonium to methylthiolate shown. Repulsive interaction minimum energies are included in the energy distributions for distances of 4–5 Å but are not included in the minimum distance distributions. For the ion-to-ion interactions only attractive dimers were studied.

molecule. Following QM optimization of the molecule, the Z-matrix is registered and manually modified following the MIR to allow for specific interaction orientations of the molecule with other monomers to be generated. This requires creation of Z-matrices as either monomer A (receptor) or monomer B (ligand) or both. Once the Z-matrices are created, target PES with the other monomers in the database can be acquired at any QM model chemistry.

Successful PES require careful visual inspection and manipulation of the Z-matrices modified based on the MIR to create the dimers. A critical element is the proper placement of the dummy atom, particularly in relation to monomer B's orientation to monomer A. The dummy atom typically forms part of the first three atoms in the Z-matrix, which collectively determine the overall molecular orientation. When encountering problematic orientations, a systematic approach proves valuable; building monomer B incrementally, atom by atom, helps pinpoint where the orientation deviates from the intended configuration. If achieving the desired orientation remains challenging, repositioning the virtual atom within the first three atom positions often resolves geometrical inconsistencies. For future automation efforts, the development of a library containing correctly configured Z-matrices would be invaluable. From these reference structures, systematic rules could be established to characterize optimal Z-matrix connections. As new molecular structures are incorporated into computational workflows, these rules would enable automated flagging of potentially problematic geometries before running resource-intensive PES calculations. This approach would streamline workflow efficiency by identifying geometry issues before computation begins,

reducing wasted computational resources on invalid configurations and ensuring more reliable PES results.

To show the utility of the MIR workflow, it was used to investigate the treatment of electron correlation and basis set on a wide set of FGs and dimer interaction PES. While only a few model chemistries are applied in the present study, the known importance of using post-HF methods in capturing accurate NCI is evident. Notable were the similarities and differences in the use of ωB97x-D3 versus MP2 to account for dispersion interactions. The functional with D3 corrections, ωB97x-D3, versus MP2 is attractive due to the $O(n^3)$ versus $O(n^5)$ computational demand, where n is the number of basis-functions. In the context of the small 6-31G* basis set ωB97x-D3 is clearly similar to MP2 as we show for a large number of intermolecular interactions. However, MP2 used with a larger, correlation consistent basis set, aug-cc-pVDZ, has a significant energy difference versus ωB97x-D3 as observed with the PES for the benzene dimers, the azetidinone-methylacetamide dimer and the Burgi–Dunitz angle of pyridine-acetone shown in Figure 2. These differences represent the significant challenge in adequately treating dispersion using QM methods for large numbers of intermolecular pairs that the presented MIR workflow allows to be calculated. In addition, analysis of selected dimers that involve direct interactions between atoms that are repulsive in nature reveal the presence of favorable, local minima. Such a result shows how studies of a wide range of monomer pairs in well-defined interaction orientations can yield unexpected insights into NCI.

Taking advantage of this capability, in future studies we plan to use the PES at the model chemistry MP2/aug-cc-pVDZ to

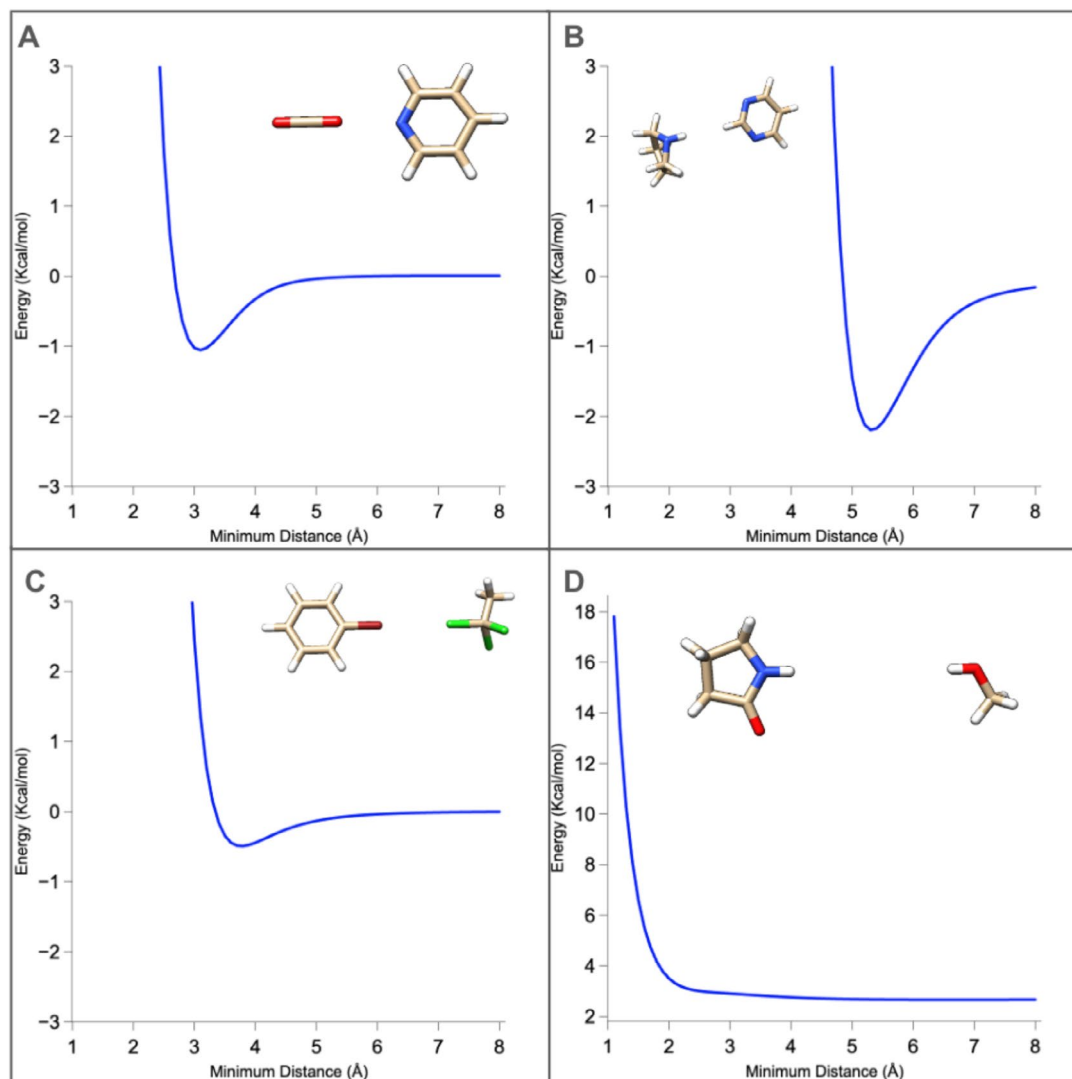


FIGURE 5 | Examples of potential energy scans of locally repulsive interactions. Panel A shows directional interactions between the oxygen of carbon dioxide and the pyridine nitrogen. Panel B shows a direct hydrogen-to-hydrogen interaction between pyrrolidine and pyrimidine. Panel C shows a direct interaction between the bromine of bromobenzene and the chlorine atom of trichloroethane. Panel D shows an interaction of the hydrogen of 2-pyrrolidinone and the hydroxyl hydrogen of the methanol. Potential energy scans at the MP2/6-31G* model chemistry.

facilitate the optimization of LJ parameters. Fitting will target PES for small molecule dimers, with those parameters then applied to mixed condensed phase simulations. This will include expansion of our software to the CHARMM General FF and Drude FFs potential energy functions to acquire MM PES [72].

The MIR workflow, molecules, Z-matrices based on the molecular interaction rules, and Cartesian coordinates can be installed as a python package. The package is available at <https://github.com/mackereil-lab/Molecular-Interaction-Rules>. The package includes an example QM PES calculated using the Psi4.

Acknowledgments

Financial support from the NIH (R35 GM131710) and computational support from the University of Maryland Computer-Aided Drug Design Center are acknowledged. The authors acknowledge helpful discussions and support from Tobias Huefner, Bettina Lier, and Ronald Kasl.

Conflicts of Interest

A.D.M. is cofounder and CSO of SilcsBio LLC.

Data Availability Statement

The data that support the findings of this study are openly available in Molecular-Interaction-Rules at <https://github.com/mackereil-lab/Molecular-Interaction-Rules>.

References

1. V. A. Adhav and K. Saikrishnan, "The Realm of Unconventional Non-covalent Interactions in Proteins: Their Significance in Structure and Function," *ACS Omega* 8 (2023): 22268–22284.
2. K. Müller-Dethlefs and P. Hobza, "Noncovalent Interactions: A Challenge for Experiment and Theory," *Chemical Reviews* 100, no. 1 (1999): 143–168.
3. A. S. Mahadevi and G. N. Sastry, "Cooperativity in Noncovalent Interactions," *Chemical Reviews* 116 (2016): 2775–2825.

4. P. Kollman, "Bacterial Cell Wall," in *New Comprehensive Biochemistry*, vol. 6, ed. M. I. Page (Elsevier, 1984), 55–71.
5. A. D. Mackerell, "Empirical Force Fields for Biological Macromolecules: Overview and Issues," *Journal of Computational Chemistry* 25 (2004): 1584–1604.
6. P. Ren and J. W. Ponder, "Consistent Treatment of Inter- and Intramolecular Polarization in Molecular Mechanics Calculations," *Journal of Computational Chemistry* 23 (2002): 1497–1506.
7. J. W. Ponder, C. Wu, P. Ren, et al., "Current Status of the Amoeba Polarizable Force Field," *Journal of Physical Chemistry B* 114 (2010): 2549–2564.
8. G. Lamoureux and B. Roux, "Modeling Induced Polarization With Classical Drude Oscillators: Theory and Molecular Dynamics Simulation Algorithm," *Journal of Chemical Physics* 119 (2003): 3025–3039.
9. J. A. Lemkul, J. Huang, B. Roux, and A. D. MacKerell, "An Empirical Polarizable Force Field Based on the Classical Drude Oscillator Model: Development History and Recent Applications," *Chemical Reviews* 116, no. 9 (2016): 4983–5013.
10. A.-P. E. Kunz and W. F. van Gunsteren, "Development of a Nonlinear Classical Polarization Model for Liquid Water and Aqueous Solutions: COS/D," *Journal of Physical Chemistry A* 113 (2009): 11570–11579.
11. S. W. Rick, "A Polarizable, Charge Transfer Model of Water Using the Drude Oscillator," *Journal of Computational Chemistry* 37 (2016): 2060–2066.
12. G. Lamoureux, E. Harder, I. V. Vorobyov, B. Roux, and A. D. MacKerell, "A Polarizable Model of Water for Molecular Dynamics Simulations of Biomolecules," *Chemical Physics Letters* 418 (2006): 245–249.
13. E. Heid, M. Fleck, P. Chatterjee, C. Schröder, and A. D. MacKerell, "Toward Prediction of Electrostatic Parameters for Force Fields That Explicitly Treat Electronic Polarization," *Journal of Chemical Theory and Computation* 15, no. 4 (2019): 2460–2469.
14. A. D. MacKerell and M. Karplus, "Importance of Attractive Van der Waals Contribution in Empirical Energy Function Models for the Heat of Vaporization of Polar Liquids," *Journal of Physical Chemistry* 95 (1991): 10559–10560.
15. C. R. Rupakheti, A. D. MacKerell, and B. Roux, "Global Optimization of the Lennard-Jones Parameters for the Drude Polarizable Force Field," *Journal of Chemical Theory and Computation* 17 (2021): 7085–7095.
16. W. L. Jorgensen, "Optimized Intermolecular Potential Functions for Liquid Alcohols," *Journal of Physical Chemistry* 90 (1986): 1276–1284.
17. J. D. Madura and W. L. Jorgensen, "Ab Initio and Monte Carlo Calculations for a Nucleophilic Addition Reaction in the Gas Phase and in Aqueous Solution," *Journal of the American Chemical Society* 108 (1986): 2517–2527.
18. W. L. Jorgensen and D. L. Severance, "Aromatic-Aromatic Interactions: Free Energy Profiles for the Benzene Dimer in Water, Chloroform, and Liquid Benzene," *Journal of the American Chemical Society* 112 (1990): 4768–4774.
19. W. L. Jorgensen, "Intermolecular Potential Functions and Monte Carlo Simulations for Liquid Sulfur Compounds," *Journal of Physical Chemistry* 90 (1986): 6379–6388.
20. J. Řezáč, M. Dubecký, P. Jurečka, and P. Hobza, "Extensions and Applications of the A24 Data Set of Accurate Interaction Energies," *Physical Chemistry Chemical Physics* 17 (2015): 19268–19277.
21. Z. M. Sparrow, B. G. Ernst, P. T. Joo, K. U. Lao, and R. A. DiStasio, "NENCI-2021. I. A Large Benchmark Database of Non-Equilibrium Non-Covalent Interactions Emphasizing Close Intermolecular Contacts," *Journal of Chemical Physics* 155, no. 18 (2021): 184303.
22. J. Řezáč, "Non-Covalent Interactions Atlas Benchmark Data Sets 5," *ChemRxiv* (2022), <https://doi.org/10.26434/chemrxiv-2022-pl3r8>.
23. L. Goerigk and S. Grimme, "A General Database for Main Group Thermochemistry, Kinetics, and Noncovalent Interactions – Assessment of Common and Reparameterized (Meta-)GGA Density Functionals," *Journal of Chemical Theory and Computation* 6 (2010): 107–126.
24. L. Goerigk, A. Hansen, C. Bauer, S. Ehrlich, A. Najibi, and S. Grimme, "A Look at the Density Functional Theory Zoo With the Advanced GMTKN55 Database for General Main Group Thermochemistry, Kinetics and Noncovalent Interactions," *Physical Chemistry Chemical Physics* 19 (2017): 32184–32215.
25. K. Kříž and J. Řezáč, "Non-Covalent Interactions Atlas Benchmark Data Sets 4: σ -Hole Interactions," *Physical Chemistry Chemical Physics* 24 (2022): 14794–14804.
26. K. Kříž, M. Nováček, and J. Řezáč, "Non-Covalent Interactions Atlas Benchmark Data Sets 3: Repulsive Contacts," *Journal of Chemical Theory and Computation* 17 (2021): 1548–1561.
27. J. Řezáč, "Non-Covalent Interactions Atlas Benchmark Data Sets 2: Hydrogen Bonding in an Extended Chemical Space," *Journal of Chemical Theory and Computation* 16 (2020): 6305–6316.
28. J. Řezáč, "Non-Covalent Interactions Atlas Benchmark Data Sets: Hydrogen Bonding," *Journal of Chemical Theory and Computation* 16 (2020): 2355–2368.
29. P. Jurecka, J. Sponer, J. Cerný, and P. Hobza, "Benchmark Database of Accurate (MP2 and CCSD(T) Complete Basis Set Limit) Interaction Energies of Small Model Complexes, DNA Base Pairs, and Amino Acid Pairs," *Physical Chemistry Chemical Physics* 8 (2006): 1985–1993.
30. J. Řezáč, K. E. Riley, and P. Hobza, "S66: A Well-Balanced Database of Benchmark Interaction Energies Relevant to Biomolecular Structures," *Journal of Chemical Theory and Computation* 7 (2011): 2427–2438.
31. A. G. Donchev, A. G. Taube, E. Decolvenaere, et al., "Quantum Chemical Benchmark Databases of Gold-Standard Dimer Interaction Energies," *Scientific Data* 8 (2021): 55.
32. J. Řezáč, K. E. Riley, and P. Hobza, "Benchmark Calculations of Noncovalent Interactions of Halogenated Molecules," *Journal of Chemical Theory and Computation* 8 (2012): 4285–4292.
33. P. Morgante and R. Peverati, "ACCDDB: A Collection of Chemistry Databases for Broad Computational Purposes," *Journal of Computational Chemistry* 40 (2019): 839–848.
34. Y. Zhao and D. G. Truhlar, "Assessment of Model Chemistries for Noncovalent Interactions," *Journal of Chemical Theory and Computation* 2 (2006): 1009–1018.
35. K. T. Tang and J. P. Toennies, "The Van der Waals Potentials Between All the Rare Gas Atoms From He to Rn," *Journal of Chemical Physics* 118 (2003): 4976–4983.
36. F. Ballesteros, S. Dunivan, and K. U. Lao, "Coupled Cluster Benchmarks of Large Noncovalent Complexes: The L7 Dataset as Well as DNA–Ellipticine and Buckycatcher–Fullerene," *Journal of Chemical Physics* 154 (2021): 154104.
37. S. Grimme, "Supramolecular Binding Thermodynamics by Dispersion-Corrected Density Functional Theory," *Chemistry* 18 (2012): 9955–9964.
38. S. Kozuch and J. M. L. Martin, "Halogen Bonds: Benchmarks and Theoretical Analysis," *Journal of Chemical Theory and Computation* 9 (2013): 1918–1931.
39. C. D. Sherrill, T. Takatani, and E. G. Hohenstein, "An Assessment of Theoretical Methods for Nonbonded Interactions: Comparison to Complete Basis Set Limit Coupled-Cluster Potential Energy Curves for the Benzene Dimer, the Methane Dimer, Benzene–Methane, and Benzene–H₂S," *Journal of Physical Chemistry A* 113 (2009): 10146–10159.
40. U. Nagaswamy, M. Larios-Sanz, J. Hury, et al., "NCIR: A Database of Non-Canonical Interactions in Known RNA Structures," *Nucleic Acids Research* 30 (2002): 395–397.

41. V. Bakken and T. Helgaker, "The Efficient Optimization of Molecular Geometries Using Redundant Internal Coordinates," *Journal of Chemical Physics* 117 (2002): 9160–9174.
42. E. Soydaş and U. Bozkaya, "Assessment of Orbital-Optimized Third-Order Møller–Plesset Perturbation Theory and Its Spin-Component and Spin-Opposite Scaled Variants for Thermochemistry and Kinetics," *Journal of Chemical Theory and Computation* 9 (2013): 1452–1460.
43. Y. Zhao, N. González-García, and D. G. Truhlar, "Benchmark Database of Barrier Heights for Heavy Atom Transfer, Nucleophilic Substitution, Association, and Unimolecular Reactions and Its Use to Test Theoretical Methods," *Journal of Physical Chemistry A* 109 (2005): 2012–2018.
44. L. A. Burns, J. C. Faver, Z. Zheng, et al., "The BioFragment Database (BFD): An Open-Data Platform for Computational Chemistry Analysis of Noncovalent Interactions," *Journal of Chemical Physics* 147, no. 16 (2017): 161727.
45. A. Kumar, O. Yoluk, and A. D. MacKerell, "FFParam: Standalone Package for CHARMM Additive and Drude Polarizable Force Field Parametrization of Small Molecules," *Journal of Computational Chemistry* 41 (2020): 958–970.
46. J. Baker, "Techniques for Geometry Optimization: A Comparison of Cartesian and Natural Internal Coordinates," *Journal of Computational Chemistry* 14 (1993): 1085–1100.
47. J. Parsons, J. B. Holmes, J. M. Rojas, J. Tsai, and C. E. M. Strauss, "Practical Conversion From Torsion Space to Cartesian Space For Siliconprotein Synthesis," *Journal of Computational Chemistry* 26 (2005): 1063–1068.
48. D. Weininger, "SMILES, a Chemical Language and Information System. 1. Introduction to Methodology and Encoding Rules," *Journal of Chemical Information and Computer Sciences* 28 (1988): 31–36.
49. D. Weininger, A. Weininger, and J. L. Weininger, "SMILES. 2. Algorithm for Generation of Unique SMILES Notation," *Journal of Chemical Information and Computer Sciences* 29 (1989): 97–101.
50. G. Landrum, *RDKit: Open-Source Cheminformatics Software* (NIBR IT, 2024).
51. T. A. Halgren, "Merck Molecular Force Field. I. Basis, Form, Scope, Parameterization, and Performance of MMFF94," *Journal of Computational Chemistry* 17 (1996): 490–519.
52. T. A. Halgren, "Merck Molecular Force Field. V. Extension of MMFF94 Using Experimental Data, Additional Computational Data, and Empirical Rules," *Journal of Computational Chemistry* 17 (1996): 616–641.
53. C. A. Rohl, C. E. M. Strauss, K. M. S. Misura, and D. Baker, "Protein Structure Prediction Using Rosetta," *Methods in Enzymology* 383 (2004): 66–93.
54. H. B. Bürgi, J. D. Dunitz, J. M. Lehn, and G. Wipff, "Stereochemistry of Reaction Paths at Carbonyl Centres," *Tetrahedron* 30 (1974): 1563–1572.
55. R. M. Parrish, L. A. Burns, D. G. A. Smith, et al., "Psi4 1.1: An Open-Source Electronic Structure Program Emphasizing Automation, Advanced Libraries, and Interoperability," *Journal of Chemical Theory and Computation* 13 (2017): 3185–3197.
56. C. Møller and M. S. Plesset, "Note on an Approximation Treatment for Many-Electron Systems," *Physics Review* 46 (1934): 618–622.
57. S. M. Cybulski and M. L. Lytle, "The Origin of Deficiency of the Supermolecule Second-Order Møller–Plesset Approach for Evaluating Interaction Energies," *Journal of Chemical Physics* 127 (2007): 141102.
58. F. Jensen, "Polarization Consistent Basis Sets: Principles," *Journal of Chemical Physics* 115 (2001): 9113–9125.
59. R. Ditchfield, W. J. Hehre, and J. A. Pople, "Self-Consistent Molecular-Orbital Methods. IX. An Extended Gaussian-Type Basis for Molecular-Orbital Studies of Organic Molecules," *Journal of Chemical Physics* 54 (1971): 724–728.
60. T. H. Dunning, "Gaussian Basis Sets for Use in Correlated Molecular Calculations. I. The Atoms Boron Through Neon and Hydrogen," *Journal of Chemical Physics* 90, no. 2 (1989): 1007–1023.
61. J. G. Brandenburg, M. Alessio, B. Civalieri, M. F. Peintinger, T. Bredow, and S. Grimme, "Geometrical Correction for the Inter- and Intramolecular Basis Set Superposition Error in Periodic Density Functional Theory Calculations," *Journal of Physical Chemistry A* 117 (2013): 9282–9292.
62. A. R. McDonald, E. J. Denning, and A. D. MacKerell, "Impact of Geometry Optimization on Base-Base Stacking Interaction Energies in the Canonical A- and B-Forms of DNA," *Journal of Physical Chemistry A* 117, no. 7 (2013): 1560–1568.
63. I. J. Chen, D. Yin, and A. D. MacKerell, "Combined Ab Initio/Empirical Approach for Optimization of Lennard–Jones Parameters for Polar-Neutral Compounds," *Journal of Computational Chemistry* 23, no. 2 (2002): 199–213.
64. R. Eisenschitz and F. London, "Über das Verhältnis der van der Waalsschen Kräfte zu den homöopolaren Bindungskräften," *Zeitschrift für Physik* 60 (1930): 491–527.
65. J. S. Cohen and R. T. Pack, "Modified Statistical Method for Intermolecular Potentials. Combining Rules for Higher Van der Waals Coefficients," *Journal of Chemical Physics* 61 (1974): 2372–2382.
66. S. Grimme, S. Ehrlich, and L. Goerigk, "Effect of the Damping Function in Dispersion Corrected Density Functional Theory," *Journal of Computational Chemistry* 32 (2011): 1456–1465.
67. E. R. Johnson and A. D. Becke, "A Post-Hartree-Fock Model of Intermolecular Interactions: Inclusion of Higher-Order Corrections," *Journal of Chemical Physics* 124 (2006): 174104.
68. J. A. Conrad and M. S. Gordon, "Modeling Systems With π – π Interactions Using the Hartree–Fock Method With an Empirical Dispersion Correction," *Journal of Physical Chemistry A* 119 (2015): 5377–5385.
69. J.-D. Chai and M. Head-Gordon, "Long-Range Corrected Hybrid Density Functionals With Damped Atom–Atom Dispersion Corrections," *Physical Chemistry Chemical Physics* 10 (2008): 6615–6620.
70. L. F. Molnar, X. He, B. Wang, and K. M. Merz, "Further Analysis and Comparative Study of Intermolecular Interactions Using Dimers From the S22 Database," *Journal of Chemical Physics* 131, no. 6 (2009): 065102.
71. P. Chatterjee, M. Y. Sengul, A. Kumar, and A. D. MacKerell, "Harnessing Deep Learning for Optimization of Lennard–Jones Parameters for the Polarizable Classical Drude Oscillator Force Field," *Journal of Chemical Theory and Computation* 18, no. 4 (2022): 2388–2407.
72. K. Vanommeslaeghe, E. Hatcher, C. Acharya, et al., "CHARMM General Force Field: A Force Field for Drug-Like Molecules Compatible With the CHARMM All-Atom Additive Biological Force Fields," *Journal of Computational Chemistry* 31 (2010): 671–690.
73. P. Hobza and J. Sponer, "Toward True DNA Base-Stacking Energies: MP2, CCSD(T), and Complete Basis Set Calculations," *Journal of the American Chemical Society* 124 (2002): 11802–11808.
74. M. O. Sinnokrot, E. F. Valeev, and C. D. Sherrill, "Estimates of the Ab Initio Limit for π – π Interactions: The Benzene Dimer," *Journal of the American Chemical Society* 124 (2002): 10887–10893.

Supporting Information

Additional supporting information can be found online in the Supporting Information section.