Polarization of Electron Density Databases of Transferable Multipolar Atoms

Théo Leduc, Emmanuel Aubert, Enrique Espinosa, Christian Jelsch, Cristian Iordache, and Benoît Guillot*

Université de Lorraine, CNRS, CRM2, F-54000 Nancy, France

S *Supporting Information*

ABSTRACT: Polarizability is a key molecular property involved in either macroscopic (i.e., dielectric constant) and microscopic properties (i.e., interaction energies). In rigid molecules, this property only depends on the ability of the electron density (ED) to acquire electrostatic moments in response to applied electric fields. Databases of transferable electron density fragments are a cheap and efficient way to access molecular EDs. This approach is rooted in the relative conservation of the atomic ED between different molecules, termed transferability principle. The present work discusses the application of this transferability principle to the polarizability, an electron density-derived property, partitioned in atomic contributions using the Quantum Theory of Atoms In Molecules topology. The energetic consequences of accounting for in situ deformation (polarization) of database multipolar atoms are investigated in detail by using a high-quality quantum chemical benchmark.
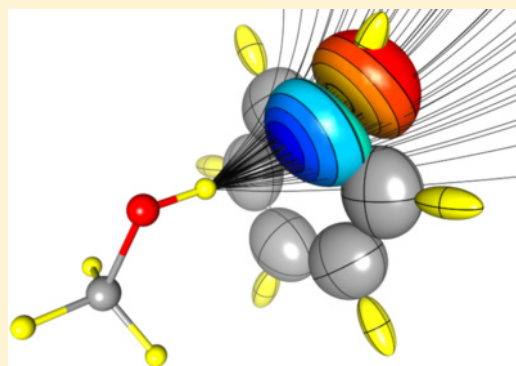
1. INTRODUCTION

Electrostatic interactions play a major role in noncovalent intermolecular interactions occurring in crystal packings and between ligands and biological macromolecules. These interactions are often a main driving force in intermolecular binding. At equilibrium distances, the usual order of magnitude of electrostatic interactions makes them an important contribution to the total interaction energy.[1] Moreover, binding sites in proteins usually show electrostatic complementarity with their ligand. Molecular dynamics force fields in biomolecular modeling include calculations of electrostatic interaction energies with, for instance, the use of partial atomic charges to describe the Coulombic term. However, molecular electrostatic properties originate from the total charge distribution of the molecule under examination, that is, from the atom nuclei and from the electron distribution in the molecule. The latter property, the molecular electron density (ED), can actually be obtained by first-principle calculations or be measured by ultrahigh-resolution X-ray diffraction experiments. The molecular ED can be represented as a summation of appropriately modeled atomic ED. Such approach is commonly followed in charge density research, which recently evolved into a new field termed "Quantum Crystallography".[2] ED modeling relies, in a vast majority of experimental cases, on the multipolar atom model of Hansen and Coppens (HCMM).[3,4] This model provides not only reliable atomic coordinates and thermal displacement parameters but also allows for accounting most of the aspherical features of atomic EDs. HCMM describes the molecular ED using multipolar parameters (see Methods section) that can be fitted against subatomic resolution X-ray diffraction data. HCMM allows to describe and quantify fine details in the molecular ED, such as bonding and lone pair electrons. It also allows the modeling of ED in intermolecular regions, giving access to the quantitative characterization of intermolecular interactions such as hydrogen bonds. Globally, all ED deformation features, such as the electron redistribution upon covalent bond formation in a molecule or intermolecular polarization effects when a molecular ED is under the influence of neighboring molecules in a crystal, are within the scope of the HCMM modeling.

Consequently, HCMM modeling gives access to a wealth of ED descriptors [source function,[5] quantum theory of atoms in molecules (QTAIM) topology.[6]..] used to characterize intra- and intermolecular interactions. Finally, an experimental ED, once modeled in the HCMM approach, allows derivation of important molecular properties such as electrostatic potential (ESP), electric moments, or atomic charges.

The literature describing various applications of charge density research is extremely vast and goes to organic small molecules with biological activities to inorganic or organometallic compounds, with applications toward crystal engineering,[7,8] medicinal chemistry,[9,10] catalysis,[11] etc. The experimental approach of charge density modeling, however, strongly relies on the accuracy and resolution of the diffraction data used for the refinement of the ED model. Indeed, in the case of an experimental determination of a molecular ED, it is

commonly admitted within the charge density science community that an X-ray data resolution of about 0.5 Å is compulsory to obtain reliable results. This obviously hinders the application of a charge density refinement to biological macromolecules whose crystals, besides very few exceptional cases,[12,13] cannot give access to such high quality and high-resolution diffraction data. However, scientists of the charge density community have observed that multipolar parameters of atoms characterized by similar chemical environments in different crystals are very similar and could reliably be transferred from a molecule to another.[14] This, termed as the *transferability principle*, has led to the construction of libraries of multipolar pseudoatoms, (i.e., sets of atomic multipolar parameters), allowing quasi-instantaneous reconstruction of ED distributions of very large molecular systems such as biological macromolecules, without the need of performing a full multipole refinement against experimental data. The seminal paper for the creation of pseudoatom databanks dates back to 1995 when Pichon-Pesme et al.[15] have reconstructed the charge distribution of a peptide backbone using multipolar parameters previously obtained from the experimental ED refinements of small peptides and isolated amino acids. This has led the Nancy group to the construction of the experimental library of multipolar atom model (ELMAM), which contains all the possible chemically unique pseudoatoms (i.e., atoms as defined in the multipolar formalism) found in the 20 proteinogenic natural amino acids.[16] The database has been improved afterward (ELMAM2) so that it covers now most of the common functional groups found in biological molecules.[17,18] ELMAM2 is implemented in the MoProSuite software,[19] a package dedicated to the structural and charge density refinement using the HCMM formalism. The strategy to build these databanks consisted in averaging the values of multipole parameters of chemically equivalent atoms, which had been previously obtained from high-quality experimental charge density refinements of small compounds (organic small molecules, isolated amino acids, di- or tripeptides). Theoretical libraries (UBDB,[20−22] invariom[23]) have also been constructed from first-principle calculations of molecular EDs. These libraries of ED parameters have been successfully used in various applications, notably the computation of electrostatic properties of several biomolecules.[24,25] Moreover, they have been validated by comparing molecular EDs obtained by the pseudoatom approach to ED fragments (extremely localized molecular orbitals) computed by quantum chemical methods.[26−28]

Our general aim is to compute accurate electrostatic properties, including electrostatic interaction energies, derived from ED transferred on macromolecular complexes. To reach such an objective, it is obvious that the quality of the ED model is a crucial factor. Hence, we discuss in this work the development of a model, based on the HCMM formalism, able to account for polarization effects that were previously overlooked in the transferable database approximation. Indeed, as mentioned before, the transferable pseudoatoms listed in the ELMAM2 library bear averaged parameters associated with atom types,[18] defined as similar atoms issued from different charge density experiments (several molecules and crystal environments). The determination of these atom types is based on rules to identify atoms with similar chemical characteristics (chemical nature, hybridization state) and presenting the same covalent neighborhood (up to the second

degree) but neglecting the noncovalent environment, such as crystal field effect, ionic bridges, van der Waals contacts, or hydrogen bonds. Hence, the averaging process on which the construction of the ELMAM2 library relies includes similar atomic multipolar EDs extracted from various noncovalent environments. In addition to the averaging procedure, the spherical harmonic functions used in the multipolar expansion of ELMAM2 atoms are constrained to follow the local symmetry of the atom. This leads to a reduction in the number of multipole populations needed to model a given atom type.[29] As a consequence, the transferable ED fragments lack noncovalent context-sensitive features such as intermolecular polarization effects, limiting possible interpretations of fine, local, electrostatic details that could be induced by a molecular environment. This averaging process is believed to cause some partial loss of atomic polarization that this work aims to reconstruct.

It must be noted that some molecular mechanics force fields already include explicit computation of inducible electrostatics, such as AMOEBA[30] or SIBFA.[31] In these force fields, iterative algorithms apply polarization on each atom, until convergence, using a tensor formalism to symbolize the electrostatic induction up to the quadrupolar level. In the field of charge density research, noteworthy efforts to obtain polarization energies based on the ED encompass up to now the PIXEL method from Gavezzotti[32,33] and the CrystalExplorer software from Spackman.[1,34]

The present article, however, describes and tests a new method designed to account for permanent (zero-frequency) atomic dipole induction in database-transferred multipolar EDs. This will allow improving the accuracy of the derived properties such as ESP and, most importantly, intermolecular electrostatic interaction energies. By using ad hoc computed atomic polarizabilities, this method explicitly modifies the transferred multipolar ED of interacting molecules through inclusion of induced atomic dipolar functions. It is expected that the transferred EDs, once modified by induced atomic dipole contributions, yield more reliable computed intermolecular interaction energies.

To summarize, this work discusses the following points concerning polarizabilities and multipolar ED polarization. First, atomic polarizability values are needed to assess a polarization model. Theoretical molecular and atomic polarizabilities have been determined for a benchmark database of interacting dimers (S66 dataset[35,36]). These atomic polarizabilities (fully available in the Supporting Information) are analyzed in detail. Notably, their modifications upon dimer formation are described, and their transferability, once clustered according to the atom types found in the ELMAM2 database, is discussed. Next, the ED polarization method is presented, and its implementation in the MoProSuite software is briefly described. Finally, the polarization algorithm is applied on the interacting dimers of the S66 and S66x8 datasets using the theoretical atomic polarizabilities. Electrostatic and polarization energies, respectively based on the transferred and polarized ELMAM2 ED, are then compared to related energy terms provided by the symmetry adapted perturbation theory (SAPT) decomposition method.[37] The ELMAM2 electrostatic energy is directly compared to the electrostatic part of the first-order SAPT perturbation. Similarly, polarized ELMAM2 energies are assessed against the SAPT total induction energy in order to quantify to what

extent the method proposed here can account for the induction contribution.

2. METHODS

2.1. Hansen and Coppens Multipolar Model. The HCMM[3,4] used in ELMAM2 and in the MoProSuite software splits the atomic ED in a sum of spherical EDs representing core and valence shells. A third deformation term, designed to account for deviation from the spherical approximation of the atomic ED, is written as atom-centered spherical harmonics, modulated by Slater-type radial functions (R_{nl} in eq 1).

$$\rho(\mathbf{r}) = \rho_{core}(r) + P_v \kappa^3 \rho_{val}(\kappa r)$$
$$+ \sum_{l=0}^{l\,max} \kappa'^3 R_{nl}(\kappa' r) \sum_{m=0}^{+l} P_{l\pm m} y_{l\pm m}(\theta, \varphi) \tag{1}$$

In this expression, ρ_{core} and ρ_{val} are spherically symmetric EDs. Both are computed from the orbital functions determined by Roetti and Clementi.[38] Whereas ρ_{core} is kept frozen, ρ_{val} is normalized to one electron and scaled by the atomic valence population P_v, a parameter which can be refined against diffraction intensities (experimental ones or theoretically computed) to give access to an estimate of the atomic charge. $y_{l\pm m}$ are real spherical harmonics weighted by the electron populations $P_{l\pm m}$, which are also adjustable parameters. κ and κ' are expansion−contraction parameters for the radial parts of the spherical valence and multipolar EDs, respectively.

The ELMAM2 ED database contains experimental parameters that allow the reconstruction of molecular EDs of organic molecules using HCMM. These parameters came from least squares ED refinements of small organic molecules against high-resolution X-ray diffraction data. The parameters stored in the ELMAM2 library were obtained based on the transferability principle. To build one entry (atom type) in ELMAM2, chemically similar atoms, which are supposed to have closely related ED parameters in a different crystalline environment according to the transferability principle, have been averaged and symmetrized.[15,18]

2.2. Polarization and Moments in HCMM. The valence population P_v may be used to describe the atomic charge (zeroth-order moment), while deformation populations (with $l > 0$) only affect the aspherical repartition of electrons around nuclei and thus are related only to higher-order electrostatic moments. Consequently, the atomic electrostatic moments are simple functions of the parameters of the multipole formalism. Moreover, only the harmonic functions ($y_{l\pm m}$) with a number $l = 1$ present a dipole moment.[4] Similarly, considering the traceless definition of higher-order moments, only the lth-order multipolar functions contribute to the lth-order moments. Hence, in this work, atomic dipoles induced in the polarization process are represented in terms of dipolar ED functions, making use of the corresponding dipole populations ($P_{1,-1}$, $P_{1,0}$, and $P_{1,+1}$) of the multipolar expansion.

The present work focuses on explicit polarization of the ELMAM2 ED. Indeed, in this approach, distributed multipoles are used to compute the electrostatic interaction energy between two entities (atoms or molecules). Until now, transferred ELMAM2 parameters for a given atom type are always the same regardless of the specific noncovalent context of the target atom. However, the polarization phenomenon corresponds to an induction of electrostatic moments in the ED in response to an external electric field. The computation

of the ESP from the multipolar ED gives access, through numerical differentiation, to the electric field vector

$$\mathbf{E}(\mathbf{r}) = -\nabla\varphi(\mathbf{r}) \tag{2}$$

HCMM only uses nuclei-centered spherical harmonics to model the ED. During computation of the dipole induction using eq 3, the electric field $\mathbf{E}(\mathbf{r})$ is computed at the coordinates of atomic nuclei. The following linear relationship links atomic-induced dipole moments $\boldsymbol{\mu}_i$ and the applied electric field $\mathbf{E}_i(\mathbf{r})$

$$\boldsymbol{\mu}_i = \boldsymbol{\alpha}_i \cdot \mathbf{E}_i(\mathbf{r}) \tag{3}$$

The atomic polarizability $\boldsymbol{\alpha}_i$ has the dimension of a volume and is expressed in Å^3. $\boldsymbol{\alpha}_i$ may be a scalar or a tensor with 3^2 elements. In case the scalar (isotropic) polarizability is needed, the usual $\boldsymbol{\alpha}_{iso} = 1/3\mathrm{trace}(\boldsymbol{\alpha}_i)$ relationship is used. Partition of the molecular polarizability $\boldsymbol{\alpha}_{mol}$ in atomic contributions $\boldsymbol{\alpha}_i$ is discussed in the Computational Details section, along with the method used to modify the transferred ED according to $\boldsymbol{\mu}_i$.

2.3. Reference Database and Energies. The S66 database and its extensions[35,36] are chosen to test and validate the present polarization procedure. This ensemble contains theoretical structures of 14 simple neutral organic molecules ($n_{atoms} \leq 17$) that form 66 molecular dimers. The original authors selected the dimers to mimic interactions relevant for biomolecular simulations. The energies and molecular geometries for monomers and dimers (at equilibrium distance or in a transition geometry, e.g., T-stack of aromatic rings) were optimized using the MP2/cc-pVTZ CP level of theory.[35] The S66x8 extension of the dataset uses the S66 geometries to mimic a discrete dissociation curve. While retaining molecular geometries and relative orientation, the distance between centers of masses is multiplied by a factor ranging from 0.9 to 2, obtaining eight complexes per dimer. The interaction energy was computed at the same level of theory at all distances.

To be able to handle all molecules in the S66 set with the ELMAM2 library, four database entries had to be added. The hydrogen and carbon from ethyne have been added and constrained to a cylindrical symmetry. These were obtained from a theoretical calculation[39] with software CRYSTAL-09. Hydrogen and carbon from ethene were added by extension of existing types of comparable hybridization (H101 and C301). Overall, the ED reconstruction used 34 different ELMAM2 atomic entries to cover the atom types present in the S66 database. A comprehensive list is available in the Supporting Information (Table S1).

The MP2/cc-pVTZ CP level of theory, used to optimize S66 geometries, outputs total interaction energies, like most theoretical methods. Dispersion, exchange, and correlation contributions cannot be derived from the experimental crystallographic ED, so they are not extractable from a transferred ED. For comparison, theoretical reference energy is needed, and this reference should feature energetic contributions that are partitioned in a physically meaningful way.

This is the case for SAPT[37] that splits interaction energies in permanent electrostatic, induction, dispersion, and exchange−correlation energetic contributions (eq 4). SAPT computes induction contributions using the perturbation theory. In this framework, the perturbation represents the interaction between otherwise isolated monomers as a power series. The first-order perturbation is E_{ELEC}, the Coulombic interaction

between charge densities. E_{IND}, the SAPT induction energy, regroups the second- and third-order perturbations.

$$E_{\text{SAPT}} = E_{\text{ELEC}} + E_{\text{IND}} + E_{\text{DISP}} + E_{\text{EX,CORR}} \quad (4)$$

Li et al.[40] benchmarked different versions of the SAPT theory against the S66x8 and other datasets from the BEGDB,[41] along with DFT and PMx theories. The present work uses the values published by Li et al. for the S66x8 dataset as reference SAPT energies and focuses on the electrostatic and induction part of the SAPT decomposition. Both are accessible through the ED: the electrostatic (E_{ELEC}) contribution from SAPT can be compared to the electrostatic interaction energies computed from the transferred ELMAM2 ED (eq S2). The induction energy from SAPT decomposition (E_{IND}), which includes dipolar induction effects, is compared to the ELMAM2-derived polarization energy (eq 5).

3. COMPUTATIONAL DETAILS

3.1. Permanent Electrostatic and Polarization Energy Computations with EP/MM. For electrostatic interaction energies, the EP/MM scheme (exact potential/multipolar model[42,43]) is implemented in the MoProSuite program. EP/MM has already found a number of applications briefly summarized along with ESP computations in the Supporting Information (eqs S1 and S2). Bojarowski et al. used the EP/MM scheme to assess their fast energy computation method (aug-PROmol).[44,45] Another example of the use of an explicit ED model can be found in Vandenbrande et al. (2017), where a force field is built using a functional describing the penetration term (to account for overlap between ED in intermolecular regions) and has been parameterized using theoretical explicit ED and EP/MM energy computation method.[46] It has to be noted that their force field does not contain explicit polarization term; polarization energy is handled by an empirical function of interatomic distances.

The polarization method proposed here leads to modifications of the atomic parameters used to represent the multipolar ED of interacting molecules. Consequently, applying this procedure modifies the total electrostatic interaction energy between the molecules undergoing mutual induction. We can therefore define $E_{\text{POL,DB}}$, the polarization energy based on the polarized ELMAM2 database ED as

$$E_{\text{POL,DB}} = E_{\text{ELEC,POLARIZED}} - E_{\text{ELEC,DB}} \quad (5)$$

where $E_{\text{ELEC,DB}}$ represents the EP/MM electrostatic interaction energy between the database-transferred charge densities of the monomers and $E_{\text{ELEC,POLARIZED}}$ represents the corrected electrostatic interaction energy between the charge densities obtained after the application of polarization. By definition, the polarization energy $E_{\text{POL,DB}}$ is always negative. Considering this application to the dataset of S66 dimers, it should be stressed that $E_{\text{POL,DB}}$ is here the polarization energy of the dimer system. As the polarizability is a positive definite tensor, the dipole moments $\boldsymbol{\mu}_i$ are always in an energetically favorable orientation compared to the electric field $\mathbf{E}(\mathbf{r})$. The scalar product $\mathbf{E}^t(\mathbf{r}) \cdot \boldsymbol{\mu}_i$ is always positive,[47] as seen in eq 3. Both $E_{\text{ELEC,DB}}$ and $E_{\text{ELEC,POLARIZED}}$ energy values are computed from eq S2 using slightly different EDs. $E_{\text{ELEC,DB}}$ is obtained using $\rho_{\text{DB}}(\mathbf{r})$, the unperturbed charge density, while $E_{\text{ELEC,POLARIZED}}$ is computed using $\rho_{\text{total}}(\mathbf{r})$, the polarized (perturbed) charge density. The latter can be written as

$$\rho_{\text{total}}(\mathbf{r}) = \rho_{\text{DB}}(\mathbf{r}) + \rho_{\text{pol}}(\mathbf{r}) \quad (6)$$

where $\rho_{\text{pol}}(\mathbf{r})$ is the induced deformation density, built from dipolar components, associated with in situ polarization of the multipolar ED. $\rho_{\text{pol}}(\mathbf{r})$ can be fully represented using the last term from eq 1, with R_{nl} functions and κ' parameter taken from the ELMAM2 atom (only $P_{1,\pm m}$ are adjusted, see below).

3.2. Calculation of Molecular and Atomic Polarizabilities. Calculations are performed on the isolated monomers and molecular complexes of the S66 dataset[35,36] at their equilibrium geometries. The present partitioning of molecular polarizability in atomic contributions has a lot in common with the method developed by Krawczuk, Pérez, and Macchi in the PolaBer software.[48] For each molecule or complex, single point density functional theory calculations have been performed at the B3LYP/6-311++G(3df,3pd) level of theory using the Gaussian 09 software.[49] Static electric fields are applied in X, Y, and Z directions with increasing amplitudes ranging from -0.005 to 0.005 a.u. (± 0.0014 e/Å^2) with 0.001 a.u. increment (2.8×10^{-4} e/Å^2). The PolaBer method usually employs two electric field values per direction of space. The resulting EDs are then analyzed with the AIMAll software,[50] performing the integration of the dipolar charge density within the Bader atomic basins, resulting in integrated atomic dipole moments. A 0.002 e/a.u.^{3} isosurface closes the open ρ-basins. For a neutral atomic basin, the dipole moment $\boldsymbol{\mu}_\Omega$ is independent of the origin and is given by

$$\boldsymbol{\mu}_\Omega = \int_\Omega \rho_{\text{total}}(\mathbf{r})(\mathbf{r} - \mathbf{r}_\Omega) d\mathbf{r} \quad (7)$$

where ρ_{total} represents the total charge density (electrons and nucleus). The integration runs over the volume of the atomic ρ-basin. $\mathbf{r}_\Omega$ is the position vector of the (point) nucleus, which is associated with the basin (all ρ-basins have one). ρ-basin charges (q_Ω) are determined in a consistent way (eq S5). In our work, the dipole moment is the atomic polarization term $\boldsymbol{\mu}_p(\Omega)$ as defined in Krawczuk, Pérez, and Macchi,[48] but in contrast to their work, the charge translation ($\boldsymbol{\mu}_c(\Omega)$) term is neglected here. The very weak effect of this omission is discussed in Section 5.2 and in the Supporting Information.

Atomic polarizabilities were then derived from the linear dependencies of these atomic dipole moments with respect to the applied external electric field. Summation of all atomic polarizabilities $\boldsymbol{\alpha}_i$ led to molecular polarizabilities $\boldsymbol{\alpha}_{\text{mol}}$ in close agreement with the values obtained directly from the DFT calculation (eq 8, the average and maximum absolute deviations calculated over all molecular $\boldsymbol{\alpha}_{\text{mol},i,j}$ components are 0.011 and 0.015 Å^3, respectively).

$$\boldsymbol{\alpha}_{\text{mol}} = \sum_i \boldsymbol{\alpha}_i \quad (8)$$

Nine over the 66 dimers of the dataset led to computational problems and were discarded from the results (e.g., some atomic polarizabilities were nonpositively defined and/or dipole moments did not scale linearly with respect to the applied electric field).

Such topologically defined atomic volumes, charges, and polarizabilities were computed for all 14 monomers and all 66 dimers of the S66 dataset, yielding 139 and 1103 atomic polarizabilities and topological charges from the monomer and dimer EDs, respectively. Comparing topological charges between monomers and dimers allows defining topological

Scheme 1. Workflow of the Polarization Algorithm Implemented in MoProViewer
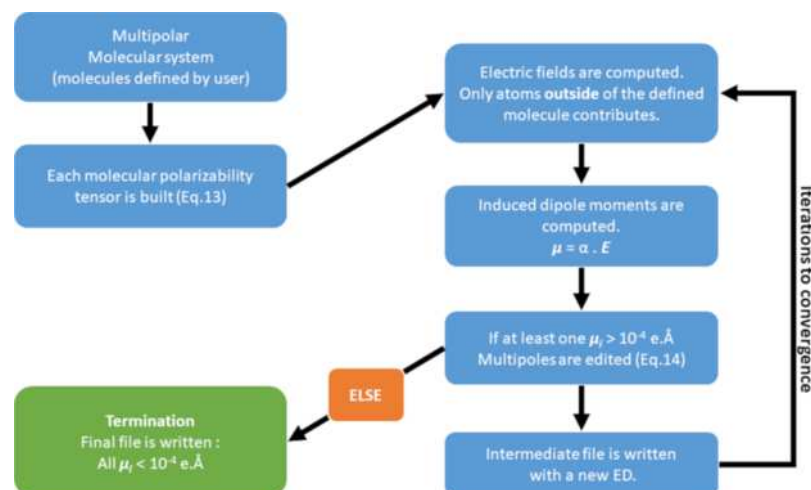
charge transfers, which are also discussed in Section 5.2 and in the Supporting Information.

Variations of the topological properties upon dimerization were tabulated in order to compare each atom in each dimer to its homologue in the monomer. Comparison criteria include isotropic and oriented anisotropic polarizabilities (in tensor form) along with charge and volume of the atomic ρ-basin. To assess finer variations of the topological atomic polarizabilities, polarizability ellipticity ε, from eq S3, and the longitudinal-to-transverse ratio η, from eq S4, are examined.[51]

3.3. Polarization Algorithm and Implementation. The previous theoretical polarizabilities are then used to polarize ELMAM2 multipolar atoms. The polarization algorithm has been implemented in the software MoProViewer,[52] the molecular viewer of MoProSuite, in C++ programing language using Qt5 and the Armadillo linear algebra library.[53] The chosen implementation allows the application of point dipole induction on a HCMM ED. It accounts for the future possibility of adapting the Thole modified point dipole interaction model[47,54] to the present ED polarization method (more details on this well-known[55] tensor formalism can be found in eq S11).

The polarization module uses an iterative formalism to compute induction of dipole moments, described in Scheme 1. The electric field is computed by numerical differentiation (using fourth-order Taylor series expansion) of the HCMM exact ESP at each nuclear position. At each polarization step, the software keeps track of the sequence of electric fields that convey induction on each atom. At each step, if the magnitude of at least one induced atomic dipole is larger than a small threshold, dipoles are converted into the $P_{l,m}$ populations corresponding to the computed dipole moment magnitude (order $l = 1$, see eq 9). Only multipolar populations with $l = 1$ contribute to the atomic dipole moment. If no such modification occurs, convergence is considered to be reached and the computation stops. A termination threshold of 10^{-4} e· Å (0.48×10^{-3} D) is chosen for several reasons. First, most force-field approaches use a 10^{-3} D threshold in their iterative procedures.[30,56] Second, the MoPro file system stores $P_{l,m}$ population with 10^{-3} e precision. With usual κ' and ζ values (expansion−contraction coefficients), this 10^{-4} e·Å threshold ensures that the last computed dipoles are smaller than the precision on $P_{l,m}$. Each multipolar expansion is expressed in an

atom-centered local axes system. In this reference, eq 9, where κ' and ζ are radial parameters of the considered pseudoatom, modifies the dipolar ($l = 1$) parameters according to the computed dipole induction.[4]

$$P_{1,+1} = u \cdot \boldsymbol{\mu}_x \qquad P_{1,-1} = u \cdot \boldsymbol{\mu}_y \qquad P_{1,0} = u \cdot \boldsymbol{\mu}_z$$

$$\text{with } u = -\frac{3\kappa'\zeta}{20} \tag{9}$$

4. RESULTS: POLARIZABILITIES FOR THE S66 DATASET

4.1. Theoretical Molecular Polarizabilities. Molecular polarizabilities were computed, along with the partition of this property in atomic contributions. Table 1 contains eigenvalues of the diagonalized molecular polarizability, computed as the sum of topologically defined atomic polarizabilities, as in eq 8. The result of these determinations is in good agreement with experimental values from the literature [an article by Gussoni

Table 1. Theoretical Polarizabilities (Principal Axes), Computed in Vacuum for the 14 Molecules in the S66 Dataset[a]

molecule	α_{xx} (Å³)	α_{yy} (Å³)	α_{zz} (Å³)	α, iso. (Å³)	α, exp. (Å³)	rel. error (%)
acetamide	6.75	6.37	4.46	5.86	5.67	3.34
acetic acid	5.89	5.52	3.91	5.11	5.15	−0.81
benzene	12.07	12.03	6.62	10.24	9.96	2.83
cyclopentane	9.25	9.25	7.99	8.83	9.17	−3.73
ethene	5.29	3.66	3.35	4.10	4.09	0.20
ethyne	4.75	2.79	2.79	3.44	3.40	1.17
methylamine	4.19	3.66	3.55	3.80	3.97	−4.30
methanol	3.42	3.00	2.93	3.12	3.21	−2.88
neopentane	9.70	9.69	9.69	9.69	10.24	−5.34
pentane	11.45	9.19	8.56	9.73	9.88	−1.47
peptide	9.37	7.74	5.88	7.66	7.82	−2.03
pyridine	11.48	10.82	6.03	9.44	9.17	2.96
uracil	13.86	11.14	6.04	10.35	10.09	2.58
water	1.44	1.36	1.33	1.38	1.45	−5.17

[a]For comparison, isotropic theoretical and experimental electronic polarizability values from Gussoni (1998)[57] are also given, along with their relative discrepancy.

(1998)[57] compiled electronic molecular polarizabilities measured in the gas phase]. Comparing the isotropic molecular polarizability of the 14 monomers to their experimental values leads to a root-mean-square deviation (rmsd, eq S7) of 0.23 Å^3 and a mean absolute relative error (MARE, eq S8) of only 2.8%. This error compares well to historical references [3.5% from Thole (1981)[47,54]] and to a current work on the topic [6% from Dos Santos, Krawczuk, and Macchi (2015)[58]]. Linear regression between experimental and theoretical isotropic molecular polarizabilities yields a slope of 0.995, with a determination coefficient R^2 of 0.994. The least agreement is obtained for neopentane and cyclopentane, for which molecular polarizabilities are underestimated by, respectively, 0.55 Å^3 (−5.3%) and 0.34 Å^3 (−3.7%) when compared to the Gussoni experimental values. Conversely, theoretical polarizabilities for ethene and ethyne are in best agreement, overestimating experimental values by, respectively, 0.01 Å^3 (+0.2%) and 0.04 Å^3 (+1.1%). Acetic acid theoretical polarizability is also very close to the experimental value, with an underestimation of 0.08 Å^3 (−0.8%). Complete tables of topological atomic polarizability tensors can be found in the Supporting Information for both monomers (Table S4) and dimers (Table S5), given to ease their reusability in the same Cartesian coordinate system than the one used in the BEGDB web server (www.begdb.com).[35,36,41]

Figure 1 displays average isotropic polarizability values for each of the four chemical elements found in the S66 dataset.
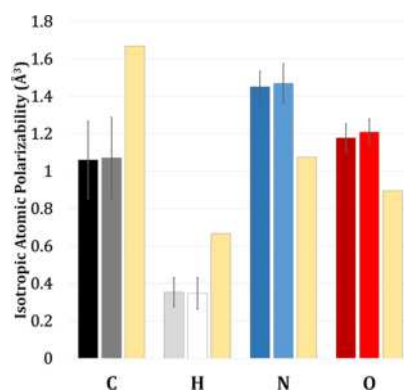

Figure 1. Average values of theoretical isotropic atomic polarizabilities for all four chemical elements in the S66 and S66x8 datasets in Å^3. Error bars display the sample standard deviation (SSD) of each group. Darker hues: isolated molecules, lighter hues: dimers. Light yellow bars: theoretical values for isolated atoms (carbon,[59] hydrogen,[60] nitrogen,[61] oxygen[62]).

The theoretical free atom values[59−62] are given for scale purposes. The O and N atoms are the most electronegative. Therefore, their polarizability appears to increase in molecules compared to free atoms because of their increased electronic charge when embedded in a molecular environment. The trend is reversed for the two relatively electropositive atoms (H, C). There is little variation of average atomic isotropic polarizabilities between the computation in monomers and dimers. A slightly more detailed picture of the computed polarizabilities, in the light of Bader partition, is proposed in the Supporting Information (Table S2, Figures S1 and S2).

4.2. Polarizability Eigenvalues. The polarizability shapes and magnitudes are now examined using quantities based on the eigenvalues of their representation tensors. To appreciate the global shape of these theoretical anisotropic polarizabilities,

one may compare their ellipticity[51] ε (eq S3) and the longitudinal-to-transverse ratio[51] η (S4). Figure 2 regroups
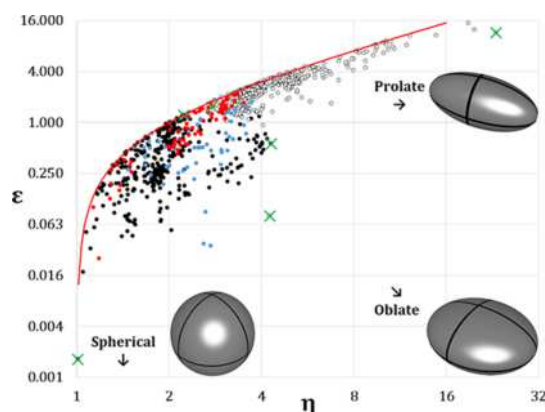

Figure 2. Topological parameters of all atomic polarizabilities. The ellipticity ε is described by eq S3, and the transverse-to-longitudinal ratio η is found in eq S4. Both axes use a base-two logarithmic scale to spread the points cloud. The red line ($\varepsilon = \eta - 1$) displays the limit of the two plotted parameters. C: black, H: white, N: blue, O: red. Green crosses represent peculiar atoms found in Table S3.

ε and η for all computed atomic polarizabilities of the S66 dataset (notice the log axes). The red line ($\varepsilon = \eta - 1$, hence $\lambda_2 = \lambda_3$) displays the limit of the two plotted parameters. Polarizabilities along this line are cylindrically symmetric around their first eigenvector ranging from spherical (bottom, left) to prolate (top, right). Considering a constant η value, oblate ellipsoids have lower ε values compared to the cylindrically symmetric limit. Thereby, oblate ellipsoids are found deviating from the limit toward the bottom-right corner in Figure 2. With the exception of one outlying hydrogen atom, oblate polarizabilities are mostly associated with aromatic carbon and aromatic nitrogen atoms. All the highly prolate polarizability tensors correspond to hydrogen atoms (with the direction associated with their greatest eigenvalue nearly always oriented along their single covalent bond). Oxygen and carbon atoms occupy the central region of the graph, where polarizability ellipsoids display prolate to spherical characters. The water oxygen and the neopentane central carbon show the most spherical polarizabilities found in the S66 dataset. Table S3 describes a few noticeable atomic polarizabilities (also highlighted in Figure 2 by green crosses) with their shape parameters to help the reader grasp the magnitude of variations presented by carbon and hydrogen atoms.

To assess transferability properties of anisotropic polarizabilities, it is necessary to look for their similarities among atoms sharing common chemical characteristics. The very first level of such classification concerns the chemical nature of the available atoms in the S66 dataset. Figure S3 presents eigenvalues of atomic polarizability tensors in the S66 dimers and monomers, averaged over H, C, N, and O atoms, along with their associated SSDs (eq S9). Using this atom clustering, it appears that SSD values are rather high ranging between 11.1% and 37.7% of the corresponding average eigenvalue. This indicates significant variations of eigenvalue magnitudes among atoms of identical chemical specie in the S66 dataset. However, polarizability ellipsoid shapes are more variable for C than for N, O, and H atoms. The systematically prolate character of hydrogen atom polarizabilities appears as λ_1 has a
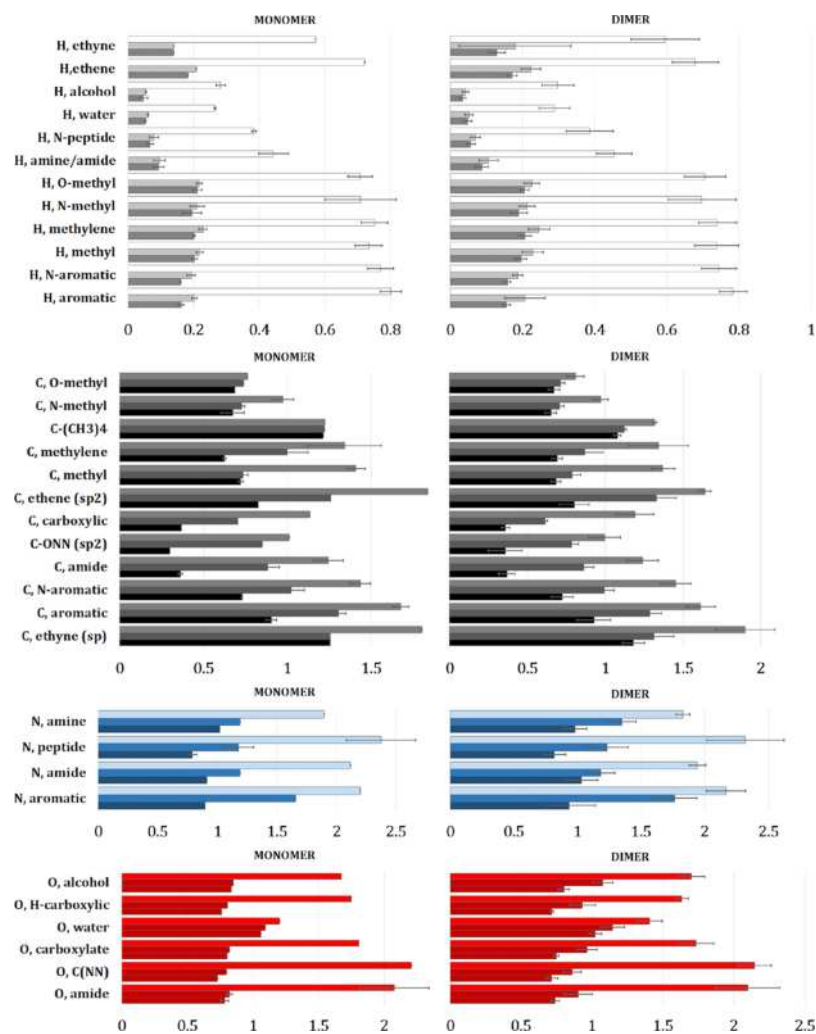
Figure 3. Eigenvalues of atomic polarizability ellipsoids found in the S66 dataset, averaged over atoms sharing an ELMAM2 atom type, computed in isolated molecules (on the left) and computed in the S66 dimers (on the right). Error bars display the SSD of each sample (atom type). If the atom type is only present once, the error bars are set to zero. Polarizability tensors used to build this figure can be found in Tables S4 and S5. First panel: hydrogen, second panel: carbon, third panel: nitrogen, fourth panel: oxygen.

significantly larger value compared to λ_2 and λ_3. The average λ_2 and λ_3 values are very similar and are associated with clearly large and overlapping SSD values. The case of carbon atom polarizabilities is especially noteworthy here. Their three averaged eigenvalues are indeed associated with large overlapping SSD. Therefore, a wide variety of polarizability ellipsoids, with variable sizes and shapes, can be found among carbon atoms of the S66 dataset. Consequently, anisotropic atomic polarizabilities averaged per chemical species could not be considered as transferable entities, especially for carbon atoms.

To capture the diversity of anisotropic polarizabilities and to cluster those in a way that would allow enhancing their transferability, it is necessary to investigate further trends associated with finer descriptions of atomic similarities. For this purpose, ELMAM2 atom types have been considered because, by construction, they share similar chemical characteristics, one of which being the same covalent bonding. Averaged eigenvalues of the polarizability tensors, computed in monomers, can be found in the left panel of Figure 3, along with their associated SSD. The averaging is carried out on all occurrences of an ELMAM2 atom type in the S66 molecules (Table S4). The right panel of Figure 3 shows the same

average of eigenvalues over atoms sharing an ELMAM2 atom type, but this time computed in dimers (Table S5).

In both cases, the pattern noticed in Figure 2 is still visible: hydrogen atoms are the least polarizable and display similar λ_2 and λ_3, which are small compared to λ_1. All nitrogen atoms have a slightly oblate polarizability ($\lambda_2 > \lambda_3$). Oxygen polarizabilities (with the exception of water) have a more prolate character, but the relative difference between λ_1 and the other eigenvalues is smaller. Using the clustering based on ELMAM2 atom types, oxygen atom polarizability now appears distinct when, for instance, water oxygen is compared to oxygen atoms in hydroxyl groups. Again, the case of carbon atoms is noteworthy. They have the widest variations of polarizability sizes and shapes and may either have a spherical, prolate, or oblate polarizability. However, this property now appears clearly in Figure 3 depending on the ELMAM2 atom type. Furthermore, for all atoms, considering averages over atom types (Figure 3, regardless of the monomer/dimer system), all SSDs are smaller than the corresponding SSD from average values over chemical elements (see Figure S3). For example, the relative SSD on polarizability of atoms of the methyl carbon type is 1.8%, while the same value for all carbons is 19.4%. For methyl hydrogen, the relative SSD is

4.7% compared to 22.0% for all hydrogen atoms. Excluding the methylene carbon in monomers (discussed below), the SSD intervals for λ_1 do not overlap with the SSD intervals for the other two eigenvalues. This is also a sign of increased specificity of shape and magnitude of an atom-type polarizability compared to the raw chemical element average presented in Figure S3. This way to define transferable properties is also used by the authors of the UBDB databank.[20] In their approach, parameters of a given atom type (occurring in the small compounds used to build the library) must present small variances, and these variances must be sufficiently small so that the considered atom type can be distinguished from other related atom types. This is exactly what is achieved here, considering polarizability tensor eigenvalues. Consequently, all these observations suggest that atomic polarizability tensors, associated with a given ELMAM2 atom type (among those represented in the S66 dataset), present similarities, which can be exploited in a transferability approach.

The case of the methylene carbon atom type, mentioned above, is an example of averaging polarizabilities over a heterogeneous sample. Indeed, we included in this averaging the methylene carbon atoms from pentane (relatively prolate) along with those from cyclopentane (relatively oblate). The cyclic character of the molecule appears as a chemical difference for the carbon atom, which is reflected in terms of polarizability. The two carbon atom types display the same covalent neighborhood. Therefore, only one atom type covers both cases in the ELMAM2 database, despite different average C−C−C angles (pentane: 112.8° ± 0.4, cyclopentane: 103.8° ± 1.3). This suggests that in this case, a differentiation of the ELMAM2 atom type would increase the consistency of the pseudoatom and polarizability transferability hypothesis. Two other atoms have notably high SSD on λ_1 (see Figure 3): the amide oxygen (O102) and the peptide nitrogen (N302). These atom types are found in uracil and N-methyl-acetamide (labeled "peptide" in the S66 dataset). To account for these findings, future versions of the ELMAM2 library might feature separated atom types for the methylene carbon and peptide nitrogen, according to their involvement in a cyclic molecule, in order to increase their transferability. The case of the amide oxygen atom is less clear (low SSD for λ_2 and λ_3), so this atom type shall not need a similar duplication. The Supporting Information contains more insights about atomic polarizability transferability (see Figure S4).

Automated transfer and comparison of monomer polarizabilities allowed treating the complete S66x8 dataset consistently. Hence, to measure the global variations of atomic polarizabilities upon dimer formation, polarizabilities in the dimers in the complete S66x8 dataset were compared to the ones computed in the monomers. Computation of average atom-wise variations of eigenvalues ($\alpha_{dimer} - \alpha_{monomer}$) and their SSD is shown in Figure S5. For 20 out of the 34 atom types, the SSD encompasses 0 for the variations of every eigenvalue. More details about these small variations of atomic polarizabilities (regarding ellipsoid eigenvalues and orientations) can be found in Figures S5 and S6.

5. RESULTS: POLARIZATION OF THE S66 DATASET

5.1. Permanent Electrostatic Interaction. As stated earlier, the SAPT decomposition of the total interaction energy between two molecular charge densities gives access to physically interpretable contributions. Among those, the permanent electrostatics and the induction terms can be

compared to corresponding quantities computed from an experimental charge density model. At first, SAPT E_{ELEC} term as of eq 4 is compared to the electrostatic interaction energy ($E_{ELEC,DB}$) obtained from the transferred ELMAM2 ED on the S66 and S66x8 dimers (Figure 4). It must be recalled here that
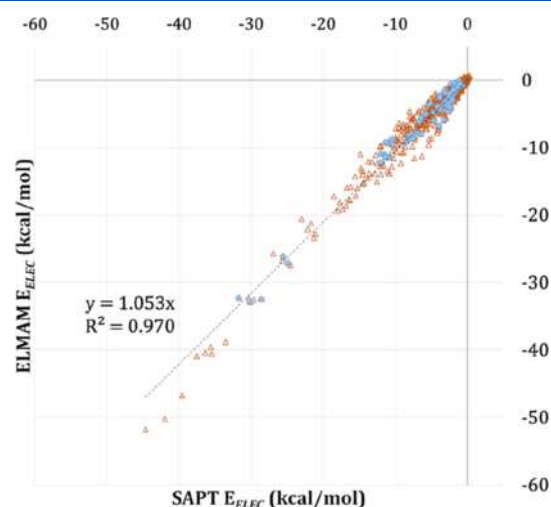

Figure 4. Correlation between SAPT E_{ELEC} and ELMAM $E_{ELEC,DB}$ on the S66 (blue diamonds, equilibrium distances) and on the S66x8 dataset (orange dots, all distances). The linear regression carried out with equilibrium distances only yields a determination coefficient of 0.973 and a slope of 1.038. This linear regression was constrained to have a zero intercept.

these two energies are computed through two very different means: E_{ELEC} is a purely theoretical result, while $E_{ELEC,DB}$ is obtained using ED parameters of experimental origin. The only common roots between the two approaches are the geometries of the considered dimers and the use, in the multipolar formalism, of Hartree−Fock spherically averaged core and valence EDs (eq 1). Nevertheless, a linear regression between E_{ELEC} and $E_{ELEC,DB}$ values, based on the S66 dimers at equilibrium distance, yields an excellent agreement with a determination coefficient $R^2 = 97.3\%$ (correlation coefficient R of 98.6%) and a slope reasonably close to unity with $E_{ELEC} = 1.04 \times E_{ELEC,DB}$ (Table 2). A similar regression was done on the complete S66x8 dataset, leading to a slightly worsened agreement with $R^2 = 97.0\%$ and a slope of 1.05. An increased error of E_{ELEC} estimation for the shortest interaction distances explains this behavior (Figure 5). Indeed, below 110% of equilibrium distance, it appears that the $E_{ELEC,DB}$ energy term overestimates the strength of the interaction (with respect to the SAPT reference), with an average discrepancy (rmsd) reaching 2.6 kcal·mol⁻¹ for the shortest distance (90%, Figure 5). Considering all interaction distances, the rmsd between the ELMAM2 and SAPT energies is 1.4 kcal·mol⁻¹, which can be considered as a good result given the differences of the methods and the absence of any ad hoc parameterization. Considering all interaction groups, the rmsd at equilibrium distance is 1.3 kcal·mol⁻¹. The rmsd falls below the "symbolic" 1.0 kcal·mol⁻¹ accuracy above 110% of equilibrium distance.

Li et al. 2014[40] proposed a classification of the interacting dimers described in the S66 dataset, based on which contribution is the most important drive to the interaction (electrostatic, dispersive, or a mix of both), defining this way the electrostatic, dispersion, and mixed groups of dimers. The dispersion group presents the best agreement in absolute

Table 2. Comparison of Linear Regressions between the SAPT Reference and Electrostatic Energies Computed from ELMAM2[a]

	SAPT electrostatics (E_{ELEC})	SAPT electrostatic + SAPT induction ($E_{ELEC} + E_{IND}$)
ELMAM2 electrostatics ($E_{ELEC,DB}$)	R^2 = 0.973 slope = 1.038 rmsd = 1.3 kcal·mol^{-1}	R^2 = 0.974 slope = 0.721 rmsd = 4.0 kcal·mol^{-1}
ELMAM2 electrostatics + induction ($E_{ELEC,DB} + E_{POL,DB}$)	R^2 = 0.950 slope = 1.455 rmsd = 5.0 kcal·mol^{-1}	R^2 = 0.960 slope = 1.014 rmsd = 2.2 kcal·mol^{-1}

[a]S66 dimers at equilibrium distances were used with EP/MM energy and monomer polarizabilities. The SAPT energy is linearly fitted to the ELMAM2 energy, so that $E_{ELMAM} = $ slope $\times E_{SAPT}$.

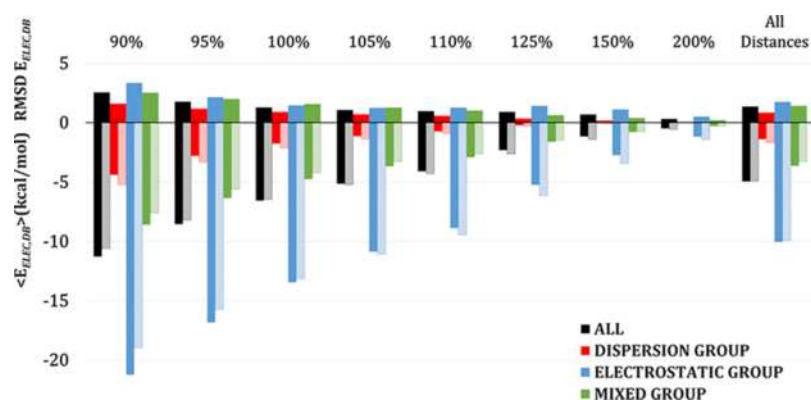**Figure 5.** Average electrostatic interaction energies, $\langle E_{ELEC,DB}\rangle$ and SAPT $\langle E_{ELEC}\rangle$, and rmsd between these values. The dimers are sorted by relative distance to equilibrium and by dimer groups. rmsd values are above the zero line; all $\langle E_{ELEC,DB}\rangle$ and $\langle E_{ELEC}\rangle$ are negative and displayed below this line. Dark-colored bars indicate the present $\langle E_{ELEC,DB}\rangle$ value. Light-colored bars represent the reference SAPT $\langle E_{ELEC}\rangle$ values.

values between ELMAM2 and SAPT energies (Figure 5), with an rmsd value of 1.6 kcal·mol^{-1} for the shortest interaction distances but only 0.9 kcal·mol^{-1} at equilibrium geometry (similar to 0.9 kcal·mol^{-1} for all distances). Conversely, the electrostatic group presents a lower agreement with rmsd at the shortest distance of 3.4 kcal·mol^{-1}, 1.5 kcal·mol^{-1} at equilibrium distance, and 1.8 kcal·mol^{-1} globally. As expected, the mixed group appears to behave in an intermediate fashion. In terms of relative errors however, the electrostatic group provides globally stronger interaction energies ($\langle E_{ELEC}\rangle$ = −13.5 kcal·mol^{-1}) than the dispersion group ($\langle E_{ELEC}\rangle$ = −1.8 kcal·mol^{-1}) and the mixed group ($\langle E_{ELEC}\rangle$ = −4.8 kcal·mol^{-1}). The electrostatic group is the one presenting lowest MARE values, with 10% at equilibrium distance. The dispersion and mixed groups of interacting dimers present at equilibrium distance MARE values that are significantly larger, reaching 33% in both cases.

It is worth mentioning that the same approach has been followed by Kumar et al.,[63] who compared the first-order SAPT electrostatic interaction energy to E_{ELEC} obtained using the EP/MM method and the UBDB transferable pseudoatoms. Compared to our results, they obtained for the S66 dataset (equilibrium distances) slightly better overall statistics (rmsd = 1.1 kcal·mol^{-1}, R^2 = 98%, and slope = 1.03), but comparing theoretical electrostatic interaction energies to E_{ELEC} values derived from theoretical structure factors. Indeed, UBDB library parameters have been obtained by DFT single point calculations (6-31G** basis set + polarization) on small molecules extracted from the Cambridge Structural Database.[64] Theoretical high-resolution structure factors have then been obtained by Fourier transform of the resulting DFT EDs arranged in pseudocubic crystals of 30 Å unit cell edges.

Finally, multipolar refinements against these theoretical structure factors yield averaged multipole parameters needed to build the UBDB library. This procedure leads to transferable ED parameters build from actually pseudo-isolated molecules. The electrostatic interaction energies computed from ELMAM2 parameters strongly correlate (correlation coefficient R = 98.8%) with similar energies provided by UBDB. Again, this fact suggests that the average ELMAM2 pseudoatoms are mostly devoid of the environment effects occurring in the crystals from where their parameters are originally extracted. This is also seen when $E_{ELEC,DB}$ is compared to the total SAPT electrostatic plus induction energy, leading to an agreement which is not as good as the comparison covering only the first-order SAPT term (Table 2). Reproduction of SAPT induction is discussed in Section 5.3.

5.2. Charge Transfer and Other Induction Contributions. It is obvious that the total induction term from the SAPT method (used as reference energy) covers more contributions, such as charge transfer effects or induction of higher-order electrostatic moments, than the point dipole induction considered here. It has been recently shown, for instance, that intermolecular charge transfer and polarization energy contributions can be modeled separately from the SAPT induction energy,[65] albeit at shorter interaction ranges than the ones considered here (down to 70% of equilibrium distances). Evidencing and quantifying electronic charge flow within a molecule (intramolecular charge transfers) depend on the way atomic charges are actually defined. Here, it has been observed that some topological charges can vary significantly when charges from the monomer ED are compared to the dimer ED. For instance, oxygen atoms in dimers have an average topological charge 0.026 e lower than their charges in

the isolated geometry (using theoretical ED). The average effect on other chemical species is significantly less visible. Moving ED from one atomic basin to another will induce electrostatic moments, so that this effect actually contributes to the total induced molecular dipole. The transferred ED polarization model proposed here focuses only on atomic dipole induction. It appears that this dipole induction in multipolar pseudoatoms causes ED zero-flux surfaces to shift, leading to the possibility to account partially for topological charge-transfer effects, in the polarized ELMAM2 ED. QTAIM partition of the ELMAM2 ED was obtained using the BADER software.[66−68] Figure S7 shows more details about topological charge transfer.

It is obvious that an HCMM density fitted on interacting dimers would reflect polarization effects both in terms of modified atomic valence populations (compared to monomeric state) and in terms of induced higher-order electric moments, including dipoles. However, it must be noted that the magnitudes of the topological charge transfers observed here are low and actually lower than the statistical uncertainties obtained in a recent study on both topological and valence population-derived charges.[69] This suggests that even if intramolecular charge-transfer effects were explicitly modeled, the required perturbation through alteration of the transferred atomic valence populations could actually end up to be poorly significant when compared to uncertainties on the corresponding parameter. Nevertheless, this does not mean that their impact in the total induction energy would be negligible. Hence, we do not expect our dipole induction model in the HCMM model to reproduce totally the SAPT induction energy. This is why these theoretical energies are used for comparison and not in the context of a fitting procedure where some parameters of the polarization model would be adjusted to reproduce them as reference values. Explicitly accounting for polarization effect due to intra- or intermolecular charge transfer through modification of atomic valence populations of HCMM pseudoatoms is not within the scope of this study. Likewise, Krawczuk, Macchi, and collaborators[7,48,58] propose a method to evaluate charge translation that was not employed here because it involves the evaluation of finer—and nonatomic—topological features (bond critical points of the ED). The ELMAM2 ED has the ability to reproduce such features.[27] Still, this topic deserves a dedicated investigation.

5.3. Induction Energies. *5.3.1. Determination of the Best Computation Method for Polarizabilities.* Using the computation scheme described here, eight polarization cycles allowed to reach convergence on all considered dimers of the S66x8 dataset. The ELMAM2 polarization energy values as defined in eq 5 are negative for all dimers. The polarization energy from ELMAM2 parameters is compared to the induction energy E_{IND} from SAPT in Figures 6 and S8. The energetic effect of polarizabilities computed in dimers and the correlation between $E_{POL,DB}$ and the SAPT E_{IND} reference is shown—for dimers at equilibrium distances—only in Figure S8. Nine dimers from S66 were excluded from this analysis because of missing dimer polarizabilities (see above). Using dimer polarizabilities, the $E_{POL,DB}$ energy overestimates the reference SAPT E_{IND}, with a regression slope of 1.132. The determination coefficients in Figure S8 indicate similar performance for dimer polarizabilities ($R^2 = 0.859$) and monomer polarizabilities ($R^2 = 0.864$). The slope for the regression between the reference SAPT E_{IND} and $E_{POL,DB}$ using monomer polarizabilities is 0.953 and the rmsd is 1.6 kcal·

mol^{-1} on the 57 considered dimers. For the $E_{POL,DB}$ energies using dimer polarizabilities, the rmsd to SAPT E_{IND} is slightly higher at 1.8 kcal·mol^{-1}. Usage of dimer polarizabilities yields $E_{POL,DB}$ that are on average 19% more intense compared to energies obtained with polarizabilities computed on isolated molecules. This overestimation of E_{IND} appears large, considering the low differences in polarizability eigenvalues between monomer and dimer states (Figure S5). The exaggerated induction energies yielded by polarizabilities computed in dimers hint at their relationship to an ED that is already perturbed by the other monomer. Hence, the present polarization energy calculations use monomer polarizabilities. More information on the induced dipole moments in dimers, computed from monomer polarizabilities, is available in the Supporting Information.

5.3.2. Computations of E_{POL}/E_{IND} on S66 and S66x8 Datasets. The best estimation of E_{IND} by $E_{POL,DB}$ is found when polarizing the ELMAM2 ED using monomer polarizabilities and EP/MM energy computation, yielding a slope of 0.955 and $R^2 = 0.865$ (complete S66 dataset). Thus, in the rest of this work, monomer polarizabilities are used. Using this method on the 66 dimers, the rmsd between E_{IND} and $E_{POL,DB}$ is also 1.6 kcal·mol^{-1}. The mean ELMAM2 polarization energy is −2.6 kcal·mol^{-1} at equilibrium distances compared to an average of −2.9 kcal·mol^{-1} for E_{IND} from the SAPT method. This underlines a slight underestimation of E_{IND} by $E_{POL,DB}$. Still, the presented polarization model accounts for a significant part of the induction contributions included in the E_{IND} SAPT term. The largest disagreement, in magnitude, is observed for the acetic acid homodimer, with a 6.7 kcal·mol^{-1} underestimation of E_{IND} (−38.3%). The N-methylacetamide-ethene dimer yields the best prediction, underestimating E_{IND} by a mere 0.017 kcal·mol^{-1} (−2.3%).

Figure 6 summarizes the polarization energies at equilibrium distance (blue diamonds) using the best computation parameters previously determined (monomer polarizabilities). Considering all distances in the S66x8,[36] $E_{POL,DB}$ from ELMAM2 and the reference SAPT E_{IND} are correlated with
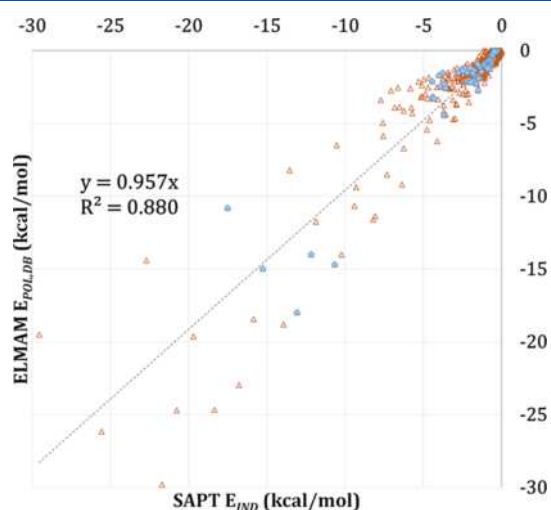

Figure 6. Correlation between SAPT E_{IND} and $E_{POL,DB}$ on the S66 (equilibrium distances, blue diamonds) and on the S66x8 dataset (full range of documented distances, orange dots) using monomer polarizabilities. The linear regression was carried out on the S66x8 dataset using the equation $E_{POL,DB}$ = slope × E_{IND}. For equilibrium distances, the determination coefficient is 0.865 and the slope is 0.955.
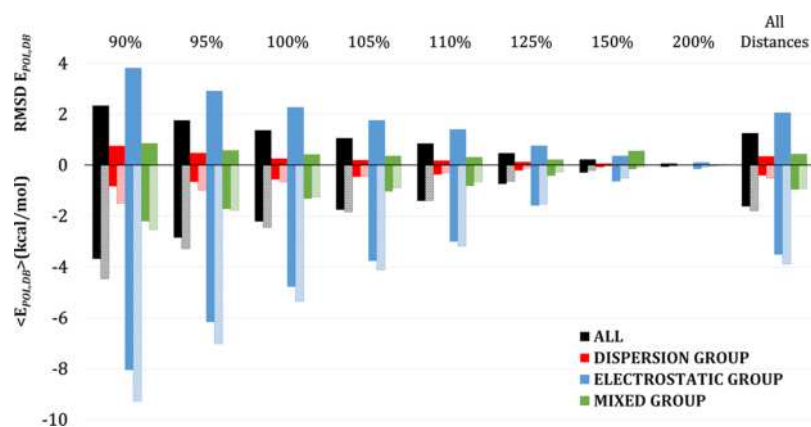
Figure 7. Average polarization energies, $\langle E_{POL,DB} \rangle$ and SAPT $\langle E_{IND} \rangle$, and rmsd between those values, sorted by relative distance to equilibrium and by dimer groups. rmsd values are above the zero line; all $\langle E_{POL,DB} \rangle$ and $\langle E_{IND} \rangle$ are negative and displayed below this line. Dark-colored bars indicate the present $\langle E_{POL,DB} \rangle$ values. Light-colored bars represent the reference SAPT $\langle E_{IND} \rangle$ values.

a determination coefficient $R^2 = 0.880$ and a slope of 0.957. The rmsd on all distances is 1.3 kcal·mol^{-1} for an average $E_{POL,DB}$ of -1.6 kcal·mol^{-1}. The maximal absolute error is once again computed in the acetic acid homodimer, but this time for the dimer interacting at 90% of the equilibrium distance. E_{IND} is underestimated there by 10.1 kcal·mol^{-1} (-34.0%). The best absolute errors are found in dimers with minimal induction (e.g., cyclopentane with neopentane at 200% of the equilibrium distance). For such molecules at long distances, all induced dipoles are inferior to the convergence threshold ($|\mu_i| < 10^{-4}$ e·Å). Hence, no significant ED modification occurs; the polarized ELMAM2 ED is strictly the same as the unpolarized ELMAM2 ED. The lowest relative error interestingly occurs in the acetic acid−uracil dimer, interacting at 95% of the equilibrium distance. $E_{POL,DB}$ in this case is 0.068 kcal·mol^{-1} below E_{IND} (-0.35%).

The performance of this polarization model varies among the group of dimers in the S66 dataset proposed in Li et al.[40] Figure 7 displays the average polarization energies (ELMAM2 and SAPT) for each class of molecules according to this classification and for each relative interacting distance. At equilibrium distances, the average $\langle E_{POL} \rangle$ varies between groups. The weakest polarization energy was computed for the dispersion group, with $\langle E_{POL} \rangle = -0.56$ kcal·mol^{-1}. Mixed and electrostatic groups display stronger $\langle E_{POL} \rangle$, respectively, -1.3 and -4.8 kcal·mol^{-1}. The proposed model reproduces well the distance dependence of the polarization energy, even if it tends to underestimate E_{IND} at short distances. The regression slope at equilibrium distances is 0.955, and comparable values are obtained at 90 and 95% of the equilibrium distance (respectively, 0.951 and 0.950). This confirms the slight underestimation of E_{IND} at short distance. At longer distances, the magnitude of the interaction energies diminishes drastically; the reference $\langle E_{IND,200\%} \rangle$ is 0.035 kcal·mol^{-1}. Indeed, the polarization energy[47] decreases with the interatomic distance as r^{-3}. At 200% of the equilibrium distances, the regression slope rises up to 1.729 for the fit $E_{POL,DB} = \text{slope} \times E_{IND}$.

Assessment of the method performance for various kinds of interacting dimers is of interest. Figure 7 regroups rmsd values between ELMAM2 $E_{POL,DB}$ and SAPT E_{IND}, following the same interaction classes and the same distance sampling. The magnitude of rmsd follows the same trend as polarization energies. High induction groups as electrostatically bound

dimers at 90% distance have high rmsd, while groups containing mixed or dispersion-bound dimers show an rmsd below 1.0 kcal·mol^{-1} at all distances. Figure 7 also puts forward that the highest absolute discrepancies between $E_{POL,DB}$ and E_{IND} are observed for dimers of the electrostatic group, which display the largest polarization energies. The relative error is not the largest for the electrostatic group. At equilibrium distances, the MARE for the electrostatic group is 32.6% (on par with the MARE of whole S66, 34.5%). The mixed group scores best on relative errors (at equilibrium distances, the MARE is 22.2%). Conversely, at equilibrium distances, the least good MARE (43.5%) is obtained for the dispersion group (where the induction energies are weakest of all groups). Figure S10 displays the size of error groups, by slices of 1.0 kcal·mol^{-1}, along Gaussian reference curves of the same mean and standard deviation. Errors on such small samples follow the Gaussian envelope only approximately. Errors in the range $[-1;0]$ and $[0;1]$ are more populated than the Gaussian approximation for all of the considered energies, suggesting that our model has some far outliers that exaggerate the standard deviation of error distributions. Hence, while most of factors determining the induction energy are well captured, the authors cannot put forward that the proposed model is totally devoid of systematic errors. Discrepancies may originate from the different source of the results: experimental for the database ED and theoretical for SAPT.

5.3.3. Sum of Electrostatic and Induction Energies. The goal is, eventually, to propose a method giving access to the total electrostatic interaction energy (accounting for permanent and induced contributions) from a transferred electron distribution. Hence, it is fair to compare $E_{ELEC} + E_{IND}$ from the SAPT method to the equivalent quantity obtained using the polarized ELMAM2 ED. Figure 8 shows this comparison using monomer polarizabilities for the polarization energy. At equilibrium distances, the average $\langle E_{ELEC,DB} + E_{POL,DB} \rangle$ is -8.8 kcal·mol^{-1}, and the rmsd is 2.2 kcal·mol^{-1}, as shown in Table 2 along with regression parameters. The best absolute error is obtained for the uracil pentane dimer, with an electrostatic plus induction energy discrepancy of 0.019 kcal·mol^{-1} with the SAPT computation. Considering the same quantity at equilibrium distances, the largest error is obtained for the acetamide uracil dimer, with an energy 8.8 kcal·mol^{-1} below the SAPT reference.
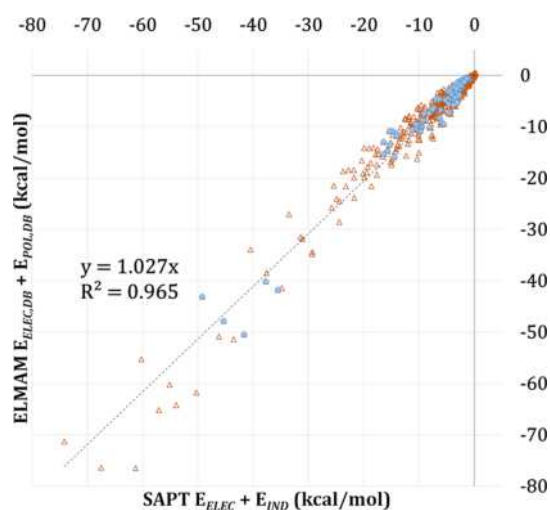
Figure 8. Correlation between SAPT $E_{ELEC} + E_{IND}$ and ELMAM2 $E_{ELEC,DB} + E_{POL,DB}$ on the S66 (equilibrium distances, blue diamonds) and S66x8 dataset (all distances, orange dots) using monomer polarizabilities. Linear regression is carried out on all distances; the determination coefficient for equilibrium distance data only is found in Table 2.

Considering all interaction distances, the largest error is again observed for the acetamide uracil dimer; this time at 90% of equilibrium distance, the ELMAM2 energy is 15.2 kcal·mol^{-1} below the SAPT reference (-67.5 kcal·mol^{-1}). The overall slope on the complete S66x8 dataset is close to unity (1.027) and the determination coefficient is 0.965, which is comparable to $E_{ELEC,DB}$ alone and sensibly better than $E_{POL,DB}$ alone, as shown in Figure 8 and Table 2. Importantly, the agreement between $E_{ELEC,DB} + E_{POL,DB}$ and its theoretical counterpart is better than the one obtained when the transferred ED is used as it is to reproduce the total theoretical electrostatic energy ($E_{ELEC,DB}$ = slope × ($E_{ELEC} + E_{IND}$) yields similar $R^2 = 0.974$ but dissimilar slope = 0.721 and rmsd = 4.0 kcal·mol^{-1}). This indicates that the polarization procedure we propose here is actually needed as a complement to the energy provided by the unperturbed transferred ELMAM2 density to model faithfully all electrostatic effects arising from interactions between molecular charge densities. Indeed, at equilibrium distances, $\langle E_{POL} \rangle$ represents 33.6% of $\langle E_{ELEC} \rangle$. This proportion is consistent between all dimer groups from Li et al.[40] (dispersion: $\langle E_{POL} \rangle / \langle E_{ELEC} \rangle = 31.4\%$; electrostatic: $\langle E_{POL} \rangle / \langle E_{ELEC} \rangle = 35.5\%$; mixed: $\langle E_{POL} \rangle / \langle E_{ELEC} \rangle = 27.6\%$). It must be noted that ELMAM2 parameters, such as the ones of other transferable multipolar pseudoatoms, have been used previously to estimate electrostatic interaction energies for instance in protein−ligand complexes.[12,25,70,71] In these studies, the permanent electrostatic contribution to the total interaction energy was only considered, neglecting the intermolecular polarization contribution, which can be now estimated using the method proposed here.

Using the monomer polarization method yields $E_{POL,DB}$, which slightly underestimates SAPT E_{IND}, while $E_{ELEC,DB}$ displays a weak overestimation of SAPT E_{ELEC}. The linear regression slopes shown in Figure 8 and Table 2 suggest that these errors may compensate in $E_{ELEC,DB} + E_{POL,DB}$. This could be explained by a reminiscent polarization in the transferred ELMAM2 ED. The induction of dipoles on ELMAM2 ED occurs on top of preexisting moments. This could explain the slight overestimation of the permanent electrostatic contribu-

tion (ELMAM2 ED retains part of the electrostatic moments induced by the crystal field in the original charge density studies). This effect remains to be confirmed and is too weak to be objectivized when analyzing the errors in the presented computations.

The least good predictions in the dataset concern acetamide, uracil, and acetic acid. These molecules involve polar chemical functionalities with double bonds. Still, not all dimers involving these molecules are associated with important errors. These atoms feature relatively high polarizabilities, so improvement of the model is expected with implementation and parameterization of dampened intramolecular dipole interaction.[47]

6. CONCLUSIONS

This article reports theoretical atomic and molecular polarizabilities computed at the B3LYP/6-311++G(3df,3pd) level of theory for the S66 monomers and for 57 of the 66 dimers of the S66 datasets. QTAIM partition of this property in atomic contributions is made fully available in the Supporting Information. The computed polarizabilities are analyzed in detail. Topological atomic polarizability tensors display large amplitude of variation, in both shape and intensity, among atoms of the same chemical specie. This is especially true for carbon atoms, which accommodate a wide range of chemical functionalities. Clustering this property by ELMAM2 atom types reduces the variability in each group. This finding hints toward a certain degree of transferability of atomic polarizabilities. The use of polarizability library is a step beyond the ED databases transferability approximation, which has already many applications even when ignoring this effect. Indeed, transferability is an approximation that is only valid up to some precision tolerance.

The present article also demonstrates the direct application of this property to a model of transferable multipolar pseudoatoms through a new method implemented in the MoProViewer module of the MoProSuite program package. This method is designed to give access to intermolecular dipolar induction using the database transferred multipolar ED. Dipolar functions are iteratively added to the transferred pseudoatoms to account for mutual induction phenomenon. At convergence of this procedure, the transferred ED is complemented by induced dipolar terms on each atom of the interacting molecules, giving access to an estimate of the dipolar polarization component of the total electrostatic interaction energy.

The S66 and S66x8 datasets are used as the benchmarking system for the interaction energies computed from the permanent and polarized ELMAM2 multipolar EDs. It should be stressed that the SAPT energy reference is independent of both the ELMAM2 ED and the DFT method used to compute polarizabilities. Evaluated energetic contributions encompass permanent electrostatics and polarization energies arising from atomic dipole induction. Electrostatic interaction energies computed from ELMAM2 are in close agreement with those obtained from SAPT. No adjustable parameter is used in the proposed model, and the slope from the linear regression is reasonably close to unity. This is a new and important step to demonstrate transferability of $\rho(\mathbf{r})$ from ELMAM2 pseudoatoms.

Application of isolated molecule polarizability on ELMAM2 ED yields encouraging induction energies predictions, suggesting that a significant part of the total induction energy is accounted for by the polarization model. The agreement

with the SAPT reference is actually surprisingly good considering the simple and straightforward approach presented in this work. The best result to reconstruct the intermolecular induction contribution is obtained using atomic polarizabilities derived from the isolated monomers.

Better energy predictions are obtained when computing $E_{\text{ELEC,DB}} + E_{\text{POL,DB}}$. Once again, only the slope of the linear regression was adjusted to compute the determination coefficient between the SAPT reference and $E_{\text{ELEC,DB}} + E_{\text{POL,DB}}$, which has no free parameter. This slope is closer to unity than the ones obtained for $E_{\text{ELEC,DB}}$ alone (compared to SAPT E_{ELEC}) or for $E_{\text{POL,DB}}$ alone (compared to SAPT E_{IND}), suggesting a degree of error compensation between these two quantities. Such error compensation is expected because ELMAM2 ED is obtained from the solid state and might contain an average in situ intermolecular polarization of the pseudoatoms.

It has been found that using dimer polarizabilities with the ELMAM2 ED yields energy predictions that are slightly overestimated. These dimer polarizabilities are computed from the QTAIM partition of the total dimer theoretical ED. The theoretical ED of each dimer is already perturbed (polarized) by the other monomer. The goal is to reproduce this perturbation on the ELMAM2 ED. Polarization energy computations showed that polarizabilities computed from monomers are more suitable for this purpose, compared to dimer polarizabilities.

The aim of this article is double. First, polarizabilities, an important molecular property, have been computed for a benchmark database of interacting molecules. Second, a tool to exploit this property in a specialized multipolar modeling software is introduced. Improvement of the polarization model for HCMM ED will involve answering several questions. The relationship between atomic ED transferability and atomic polarizability has to be explored more systematically. The S66 dataset only samples a subset of ELMAM2 atom types (34/154) that do not cover all proteinogenic amino acids. Therefore, application to protein–ligand complexes calls for more benchmark studies. The partition of polarizability in supra atomic groups (synthons or promolecules) has also been overlooked in this study. Dos Santos et al. (2015)[58] and Krawczuk et al. (2009)[48] have observed that the polarizability tensors are more transferable for atom groups (methyl, phenyl, COOH, etc.) than for individual atoms. Partitioning polarizability by groups of atoms needs to be explored in the future, along with bond properties (through critical points and charge translations). This implies a deeper look at the difference between HCMM and QTAIM partitioning of molecular properties from the ED. As stated before, this alluring topic deserves its own study and the present one outlines its feasibility. Application of local symmetry constraints from ELMAM2 to atomic polarizability and the effect of intramolecular dipole coupling needs to be evaluated extensively. Intramolecular dipole coupling is expected to improve induction energies and also to yield a different decomposition of molecular polarizabilities in atomic contributions, compared to the QTAIM partition. Moreover, intramolecular dipole coupling must be attenuated by some dampening function. Such function will also modify the partition of polarizability in atomic contributions.[47,54,55] This may be of peculiar relevance to define polarizability and polarization of synthons or promolecules in the context of the HCMM. The last four axes—namely, completing the database, looking for supra-

atomic transferability, applying symmetry constraints, and coupling dipole induction—constitute an integrated strategy to lift the limitations of the model presented here. This strategy may lead to some fruitful insights to apply the presented model in large molecular edifices such as protein–ligand complexes.

■ ASSOCIATED CONTENT

ⓢ Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.jpca.9b05051.

List of ELMAM2 atom types, the EP/MM scheme, topological and statistical equations, average polarizabilities and their variations, details about the tensor formalism, details about charge/volume/polarizability relationships, details about the polarizabilities/dipole uncertainties relationship, discussions about charge transfer, comparisons of polarization energy computations, average of induced dipoles and error repartition on energy computations, and all atomic polarizabilities for each dataset studied in the present work (PDF)

Polarization tool for the MoProViewer polarization module, displaying the settings used for the polarization of transferred ED (TXT)

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: benoit.guillot@univ-lorraine.fr.

ORCID ⓘ

Enrique Espinosa: 0000-0002-8911-1887
Benoît Guillot: 0000-0002-9992-4427

Author Contributions

The manuscript was written through contributions of all authors. All authors have given approval to the final version of the manuscript.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

■ ABBREVIATIONS

BP, base pair; DB, database; DFT, density functional theory; ED, electron density; Exp., experimental; HCMM, Hansen and Coppens multipolar model; ELMAM, experimental library of multipolar atom model; EP/MM, exact potential/multipolar model; ESP, electrostatic potential; Iso., isotropic; MARE, mean absolute relative error; QTAIM, quantum theory of atoms in molecules; SSD, sample standard deviation; rmsd, root-mean-square deviation; SAPT, symmetry-adapted perturbation theory; TS, T-stack; UBDB, University of Buffalo database

■ REFERENCES

(1) Mackenzie, C. F.; Spackman, P. R.; Jayatilaka, D.; Spackman, M. A. CrystalExplorer Model Energies and Energy Frameworks: Extension to Metal Coordination Compounds, Organic Salts, Solvates and Open-Shell Systems. *IUCrJ* **2017**, *4*, 575−587.

(2) Grabowsky, S.; Genoni, A.; Bürgi, H. B. Quantum Crystallography. *Chem. Sci.* **2017**, *8*, 4159−4176.

(3) Hansen, N. K.; Coppens, P. Testing Aspherical Atom Refinements on Small-Molecule Data Sets. *Acta Crystallogr., Sect. A: Cryst. Phys., Diffr., Theor. Gen. Crystallogr.* **1978**, *34*, 909−921.

(4) *X-ray Charge Densities and Chemical Bonding*; Coppens, P., Ed.; Oxford Science Publications: Oxford, 1997.

(5) Gatti, C.; Cargnoni, F.; Bertini, L. Chemical Information from the Source Function. *J. Comput. Chem.* **2003**, *24*, 422−36.

(6) Bader, R. F. W.; Carroll, M. T.; Cheeseman, J. R.; Chang, C. Properties of Atoms in Molecules: Atomic Volumes. *J. Am. Chem. Soc.* **1987**, *109*, 7968−7979.

(7) Krawczuk, A.; Macchi, P. Charge Density Analysis for Crystal Engineering. *Chem. Cent. J.* **2014**, *8*, 68.

(8) Dos Santos, L. H. R.; Macchi, P. The Role of Hydrogen Bond in Designing Molecular Optical Materials. *Crystals* **2016**, *6*, 43.

(9) Matta, C. F.; Arabi, A. A. Electron-Density Descriptors as Predictors in Quantitative Structure−Activity/Property Relationships and Drug Design. *Future Med. Chem.* **2011**, *3*, 969−994.

(10) Dittrich, B.; Matta, C. F. Contributions of Charge-Density Research to Medicinal Chemistry. *IUCrJ* **2014**, *1*, 457−469.

(11) Farrugia, L. J.; Evans, C.; Tegel, M. Chemical Bonds without "Chemical Bonding"? A Combined Experimental and Theoretical Charge Density Study on an Iron Trimethylenemethane Complex. *J. Phys. Chem. A* **2006**, *110*, 7952−7961.

(12) Guillot, B.; Jelsch, C.; Podjarny, A.; Lecomte, C. Charge-Density Analysis of a Protein Structure at Subatomic Resolution: the Human Aldose Reductase Case. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2008**, *64*, 567−588.

(13) Howard, E. I.; Guillot, B.; Blakeley, M. P.; Haertlein, M.; Moulin, M.; Mitschler, A.; Cousido-Siah, A.; Fadel, F.; Valsecchi, W. M.; Tomizaki, T.; et al. High-Resolution Neutron and X-Ray Diffraction Room-Temperature Studies of an H-FABP−Oleic Acid Complex: Study of the Internal Water Cluster and Ligand Binding by a Transferred Multipolar Electron-Density Distribution. *IUCrJ* **2016**, *3*, 115−126.

(14) Brock, C. P.; Dunitz, J. D.; Hirshfeld, F. L. Transferability of Deformation Densities among Related Molecules: Atomic Multipole Parameters from Perylene for Improved Estimation of Molecular Vibrations in Naphthalene and Anthracene. *Acta Crystallogr., Sect. B: Struct. Sci.* **1991**, *47*, 789−797.

(15) Pichon-Pesme, V.; Lecomte, C.; Lachekar, H. On Building a Data Bank of Transferable Experimental Electron Density Parameters: Application to Polypeptides. *J. Phys. Chem.* **1995**, *99*, 6242−6250.

(16) Zarychta, B.; Pichon-Pesme, V.; Guillot, B.; Lecomte, C.; Jelsch, C. On the Application of an Experimental Multipolar Pseudo-Atom Library for Accurate Refinement of Small-Molecule and Protein Crystal Structures. *Acta Crystallogr., Sect. A: Found. Crystallogr.* **2007**, *63*, 108−125.

(17) Domagała, S.; Munshi, P.; Ahmed, M.; Guillot, B.; Jelsch, C. Structural Analysis and Multipole Modelling of Quercetin Monohydrate − a Quantitative and Comparative Study. *Acta Crystallogr., Sect. B: Struct. Sci.* **2011**, *67*, 63−78.

(18) Domagała, S.; Fournier, B.; Liebschner, D.; Guillot, B.; Jelsch, C. An Improved Experimental Databank of Transferable Multipolar Atom Models − ELMAM2. Construction Details and Applications. *Acta Crystallogr., Sect. A: Found. Crystallogr.* **2012**, *68*, 337−351.

(19) Jelsch, C.; Guillot, B.; Lagoutte, A.; Lecomte, C. Advances in Protein and Small-Molecule Charge Density Refinement Methods Using *MoPro*. *J. Appl. Crystallogr.* **2005**, *38*, 38−54.

(20) Volkov, A.; Li, X.; Koritsanszky, T.; Coppens, P. *Ab Initio* Quality Electrostatic Atomic and Molecular Properties Including Intermolecular Energies from a Transferable Theoretical Pseudoatom Databank. *J. Phys. Chem. A* **2004**, *108*, 4283−4300.

(21) Volkov, A.; Messerschmidt, M.; Coppens, P. Improving the Scattering-Factor Formalism in Protein Refinement: Application of the University at Buffalo Aspherical-Atom Databank to Polypeptide Structures. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2007**, *63*, 160−170.

(22) Jarzembska, K. N.; Dominiak, P. M. New Version of the Theoretical Databank of Transferable Aspherical Pseudoatoms, UBDB2011 − towards Nucleic Acid Modelling. *Acta Crystallogr., Sect. A: Found. Crystallogr.* **2012**, *68*, 139−147.

(23) Dittrich, B.; Hübschle, C. B.; Pröpper, K.; Dietrich, F.; Stolper, T.; Holstein, J. J. The Generalized Invariom Database (GID). *Acta Crystallogr., Sect. B: Struct. Sci., Cryst. Eng. Mater.* **2013**, *69*, 91−104.

(24) Dominiak, P. M.; Volkov, A.; Li, X.; Messerschmidt, M.; Coppens, P. A Theoretical Databank of Transferable Aspherical Atoms and Its Application to Electrostatic Interaction Energy Calculations of Macromolecules. *J. Chem. Theory. Comput.* **2007**, *3*, 232−247.

(25) Fournier, B.; Bendeif, E.-E.; Guillot, B.; Podjarny, A.; Lecomte, C.; Jelsch, C. Charge Density and Electrostatic Interactions of Fidarestat, an Inhibitor of Human Aldose Reductase. *J. Am. Chem. Soc.* **2009**, *131*, 10929−10941.

(26) Meyer, B.; Guillot, B.; Ruiz-Lopez, M. F.; Genoni, A. Libraries of Extremely Localized Molecular Orbitals. 1. Model Molecules Approximation and Molecular Orbitals Transferability. *J. Chem. Theory Comput.* **2016**, *12*, 1052−1067.

(27) Meyer, B.; Guillot, B.; Ruiz-Lopez, M. F.; Jelsch, C.; Genoni, A. Libraries of Extremely Localized Molecular Orbitals. 2. Comparison with the Pseudoatoms Transferability. *J. Chem. Theory Comput.* **2016**, *12*, 1068−1081.

(28) Meyer, B.; Genoni, A. Libraries of Extremely Localized Molecular Orbitals. 3. Construction and Preliminary Assessment of the New Databanks. *J. Phys. Chem. A* **2018**, *122*, 8965−8981.

(29) Domagała, S.; Jelsch, C. Optimal Local Axes and Symmetry Assignment for Charge-Density Refinement. *J. Appl. Crystallogr.* **2008**, *41*, 1140−1149.

(30) Shi, Y.; Xia, S.; Zhang, J.; Best, R.; Wu, C.; Ponder, J. W.; Ren, P. Polarizable Atomic Multipole-Based AMOEBA Force Field for Proteins. *J. Chem. Theory Comput.* **2013**, *9*, 4046−4063.

(31) Gresh, N.; Andrès Cisneros, G.; Darden, T. A.; Piquemal, J.-P. Anisotropic, Polarizable Molecular Mechanics Studies of Inter- and Intramolecular Interactions and Ligand-Macromolecule Complexes. A Bottom-Up Strategy. *J. Chem. Theory Comput.* **2007**, *3*, 1960−1986.

(32) Gavezzotti, A. Calculation of Intermolecular Interaction Energies by Direct Numerical Integration over Electron Densities. I. Electrostatic and Polarization Energies in Molecular Crystals. *J. Phys. Chem. B* **2002**, *106*, 4145−4154.

(33) Gavezzotti, A. Calculation of Intermolecular Interaction Energies by Direct Numerical Integration over Electron Densities. 2. An Improved Polarization Model and the Evaluation of Dispersion and Repulsion Energies. *J. Phys. Chem. B* **2003**, *107*, 2344−2353.

(34) Spackman, M. A. Towards the Use of Experimental Electron Densities to Estimate Reliable Lattice Energies. *CrystEngComm* **2018**, *20*, 5340−5347.

(35) Řezáč, J.; Riley, K. E.; Hobza, P. S66: A Well-Balanced Database of Benchmark Interaction Energies Relevant to Biomolecular Structures. *J. Chem. Theory Comput.* **2011**, *7*, 2427−2438.

(36) Řezáč, J.; Riley, K. E.; Hobza, P. Extensions of the S66 Data Set: More Accurate Interaction Energies and Angular-Displaced Nonequilibrium Geometries. *J. Chem. Theory Comput.* **2011**, *7*, 3466−3470.

(37) Jeziorski, B.; Moszynski, R.; Szalewicz, K. Perturbation Theory Approach to Intermolecular Potential Energy Surfaces of van der Waals Complexes. *Chem. Rev.* **1994**, *94*, 1887−1930.

(38) Roetti, C.; Clementi, E. Simple Basis Sets for Molecular Wavefunctions Containing Atoms from Z = 2 to Z = 54. *J. Chem. Phys.* **1974**, *60*, 4725−4729.

(39) Nassour, A.; Domagala, S.; Guillot, B.; Leduc, T.; Lecomte, C.; Jelsch, C. A Theoretical-Electron-Density Databank Using a Model of Real and Virtual Spherical Atoms. *Acta Crystallogr., Sect. B: Struct. Sci., Cryst. Eng. Mater.* **2017**, *73*, 610−625.

(40) Li, A.; Muddana, H. S.; Gilson, M. K. Quantum Mechanical Calculation of Noncovalent Interactions: A Large-Scale Evaluation of PMx, DFT, and SAPT Approaches. *J. Chem. Theory Comput.* **2014**, *10*, 1563−1575.

(41) Řezáč, J.; Jurečka, P.; Riley, K. E.; Černý, J.; Valdes, H.; Pluháčková, K.; Berka, K.; Řezáč, T.; Pitoňák, M.; Vondrášek, J.; et al. Quantum Chemical Benchmark Energy and Geometry Database for Molecular Clusters and Complex Molecular Systems (www.begdb.com): A Users' Manual and Examples. *Collect. Czech. Chem. Commun.* **2008**, *73*, 1261−1270.

(42) Volkov, A.; Coppens, P. Calculation of Electrostatic Interaction Energies in Molecular Dimers from Atomic Multipole Moments Obtained by Different Methods of Electron Density Partitioning. *J. Comput. Chem.* **2004**, *25*, 921−934.

(43) Bouhmaida, N.; Bonhomme, F.; Guillot, B.; Jelsch, C.; Ghermani, N. E. Charge Density and Electrostatic Potential Analyses in Paracetamol. *Acta Crystallogr., Sect. B: Struct. Sci.* **2009**, *65*, 363−374.

(44) Bojarowski, S. A.; Kumar, P.; Dominiak, P. M. A Universal and Straightforward Approach to Include Penetration Effects in Electrostatic Interaction Energy Estimation. *ChemPhysChem* **2016**, *17*, 2455−2460.

(45) Bojarowski, S. A.; Kumar, P.; Wandtke, C. M.; Dittrich, B.; Dominiak, P. M. Universal Method for Electrostatic Interaction Energies Estimation with Charge Penetration and Easily Attainable Point Charges. *J. Chem. Theory Comput.* **2018**, *14*, 6336−6345.

(46) Vandenbrande, S.; Waroquier, M.; Speybroeck, V. V.; Verstraelen, T. The Monomer Electron Density Force Field (MEDFF): A Physically Inspired Model for Noncovalent Interactions. *J. Chem. Theory Comput.* **2017**, *13*, 161−179.

(47) Thole, B. T. Molecular Polarizabilities Calculated with a Modified Dipole Interaction. *Chem. Phys.* **1981**, *59*, 341−350.

(48) Krawczuk, A.; Pérez, D.; Macchi, P. *Polaber*: a Program to Calculate and Visualize Distributed Atomic Polarizabilities Based on Electron Density Partitioning. *J. Appl. Crystallogr.* **2014**, *47*, 1452−1458.

(49) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Petersson, G. A.; Nakatsuji, H.; et al. *Gaussian 09*, Revision D.01; Gaussian, Inc.: Wallingford CT, 2016.

(50) Keith, T. A. *AIMAll*, version 17.11.14; TK Gristmill Software: Overland Park KS, 2017.

(51) Souhassou, M.; Blessing, R. H. Topological Analysis of Experimental Electron Densities. *J. Appl. Crystallogr.* **1991**, *32*, 210−217.

(52) Guillot, B.; Espinosa, E.; Huder, L.; Jelsch, C. MoProViewer: a tool to study proteins from a charge density science perspective. *Acta Crystallogr., Sect. A: Found. Adv.* **2014**, *70*, C279.

(53) Sanderson, C.; Curtin, R. Armadillo: a Template-Based C++ Library for Linear Algebra. *J. Open Source Softw.* **2016**, *1*, 26.

(54) van Duijnen, P. T.; Swart, M. Molecular and Atomic Polarizabilities: Thole's Model Revisited. *J. Phys. Chem. A* **1998**, *102*, 2399−2407.

(55) Applequist, J.; Carl, J. R.; Fung, K.-K. An Atom Dipole Interaction Model for Molecular Polarizability. Application to Polyatomic Molecules and Determination of Atom Polarizabilities. *J. Am. Chem. Soc.* **1972**, *94*, 2952−2960.

(56) Cieplak, P.; Dupradeau, F.-Y.; Duan, Y.; Wang, J. Polarization Effects in Molecular Mechanical Force Fields. *J. Phys.: Condens. Matter* **2009**, *21*, 3331021.

(57) Gussoni, M.; Rui, M.; Zerbi, G. Electronic and Relaxation Contribution to Linear Molecular Polarizability. An Analysis of the Experimental Values. *J. Mol. Struct.* **1998**, *447*, 163−215.

(58) Dos Santos, L. H. R.; Krawczuk, A.; Macchi, P. Distributed Atomic Polarizabilities of Amino Acids and their Hydrogen-Bonded Aggregates. *J. Phys. Chem. A* **2015**, *119*, 3285−3298.

(59) Thierfelder, C.; Assadollahzadeh, B.; Schwerdtfeger, P.; Schäfer, S.; Schäfer, R. Relativistic and Electron Correlation Effects in Static Dipole Polarizabilities for the Group 14 Elements from Carbon to Element 114. *Phys. Rev. A: At., Mol., Opt. Phys.* **2008**, *78*, 052506.

(60) Goldman, S. P. Gauge-Invariance Method for Accurate Atomic-Physics Calculations: Application to Relativistic Polarizabilities. *Phys. Rev. A: At., Mol., Opt. Phys.* **1989**, *39*, 976−980.

(61) Das, A. K.; Thakkar, A. J. Static Response Properties of Second-Period Atoms: Coupled Cluster Calculations. *J. Phys. B: At., Mol. Opt. Phys.* **1998**, *31*, 2215−2223.

(62) Werner, H.-J.; Meyer, W. Finite Perturbation Calculations for the Static Dipole Polarizabilities of the First-Row Atoms. *Phys. Rev. A: At., Mol., Opt. Phys.* **1976**, *13*, 13−16.

(63) Kumar, P.; Bojarowski, S. A.; Jarzembska, K. N.; Domagała, S.; Vanommeslaeghe, K.; MacKerell, A. D., Jr.; Dominiak, P. M. A Comparative Study of Transferable Aspherical Pseudoatom Databank and Classical Force Fields for Predicting Electrostatic Interactions in Molecular Dimers. *J. Chem. Theory Comput.* **2014**, *10*, 1652−1664.

(64) Groom, C. R.; Bruno, I. J.; Lightfoot, M. P.; Ward, S. C. The Cambridge Structural Database. *Acta Crystallogr., Sect. B: Struct. Sci., Cryst. Eng. Mater.* **2016**, *72*, 171−179.

(65) Deng, S.; Wang, Q.; Ren, P. Estimating and Modeling Charge Transfer from the SAPT Induction Energy. *J. Comput. Chem.* **2017**, *38*, 2222−2231.

(66) Henkelman, G.; Arnaldsson, A.; Jónsson, H. A Fast and Robust Algorithm for Bader Decomposition of Charge Density. *Comput. Mater. Sci.* **2006**, *36*, 354−360.

(67) Sanville, E.; Kenny, S. D.; Smith, R.; Henkelman, G. Improved Grid-Based Algorithm for Bader Charge Allocation. *J. Comput. Chem.* **2007**, *28*, 899−908.

(68) Tang, W.; Sanville, E.; Henkelman, G. A Grid-Based Bader Analysis Algorithm without Lattice Bias. *J. Phys.: Condens. Matter* **2009**, *21*, 084204−084211.

(69) Fournier, B.; Guillot, B.; Lecomte, C.; Escudero-Adán, E. C.; Jelsch, C. A Method to Estimate Statistical Errors of Properties Derived from Charge-Density Modelling. *Acta Crystallogr., Sect. A: Found. Crystallogr.* **2018**, *74*, 170−183.

(70) Dominiak, P. M.; Volkov, A.; Dominiak, A. P.; Jarzembska, K. N.; Coppens, P. Combining Crystallographic Information and an Aspherical-Atom Data Bank in the Evaluation of the Electrostatic Interaction Energy in an Enzyme−Substrate Complex: Influenza Neuraminidase Inhibition. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2009**, *65*, 485−499.

(71) Liebschner, D.; Elias, M.; Moniot, S.; Fournier, B.; Scott, K.; Jelsch, C.; Guillot, B.; Lecomte, C.; Chabrière, E. Elucidation of the Phosphate Binding Mode of DING Proteins Revealed by Sub-angstrom X-Ray Crystallography. *J. Am. Chem. Soc.* **2009**, *131*, 7879−7886.